ASTES

Advances in Science, Technology & Engineering Systems Journal

VOLUME 7-ISSUE 3|MAY-JUN 2022

www.astesj.com
ISSN: 2415-6698

Editorial

We are pleased to present this issue showcasing 20 accepted research articles across a diverse array of topics in science and engineering.

The issue opens with El Kadiri et al.'s study on performance measurement practices in the automotive industry [1]. Through a survey of Moroccan firms, they identified barriers to effective overall performance management and recommended key indicators for sustainability. Their findings can guide improved decision-making.

Sentiment analysis is the focus of Laaz and Mazroui's paper proposing an Arabic text classification framework [2]. Combining lexicon-based approaches with machine learning achieved promising accuracy. Advancing natural language processing for under-resourced languages provides broad societal benefits.

Shifting focus to systems design, Wohlrab et al. put forth a model-based, variant-oriented methodology for developing systems of systems [3]. Identifying commonalities early on facilitates standardized interfaces and loose coupling. This work provides an efficient engineering approach for complex aggregations.

In the artificial intelligence domain, Ouahmane et al. empirically investigate embedding chaos in neural network models [4]. For both substance classification and breast cancer detection tasks, chaotic bidirectional associative memories yielded perfect accuracy, demonstrating the power of brain-inspired computing.

Gamification of instructional media is examined by Pratama et al. who designed science lessons incorporating the Genially platform [5]. Evaluations showed improved critical thinking and creativity in primary students. This highlights the potential of engaging tools to strengthen foundational skills.

Predictive modeling is spotlighted in Wulandari and Amin's comparison of algorithms for credit card default risk [6]. On real-world data, generalized linear models achieved high interpretability and accuracy. Practical insights are provided for financial risk assessment.

Multi-agent reinforcement learning is explored by Laaziz who proposes knowledge sharing to accelerate decentralized learning [7]. Simulations verified faster convergence through communicating learned state values. Such techniques may enable more nimble autonomous systems.

In the machine learning systems domain, Garg et al. tailored a CNN architecture called High Performance SqueezeNext for efficient deployment on embedded hardware [8]. Optimizations yielded high accuracy given memory constraints. This facilitates expanding AI to the edge.

Semiconductor physics is the focus of Almutairi and Wang's work modeling hole-confined optical phonon interactions in quantum well structures [9]. Their analysis revealed key insights into scattering mechanisms. Advancing physics-based simulations guides device optimization.

Integrated circuit design is spotlighted in Huang et al.'s paper on a CMOS process-voltage-temperature-leakage monitoring system [10]. A novel circuit topology achieved highly accurate real-time detection capabilities. Such self-awareness allows mitigating variability and enhancing robustness.

Secure wireless sensor networks for the Internet of Things are examined by Razaque et al. [11]. They propose an ant colony optimization routing protocol combined with trust management. Simulations demonstrated improved resilience to malicious nodes and energy efficiency. Hardening IoT systems against evolving threats remains crucial.

Supply chain provenance through blockchain technology is explored by Tran et al. who put forth aggregated verification schemes [12]. A multi-layer architecture minimized blockchain storage costs for tracking agricultural production. Integrity improvements have far-reaching potential.

Alternative energy sources are evaluated by Olajire et al. who characterized coal-biomass briquette blends [13]. Mixing agricultural wastes with coal boosted energy density while lowering emissions. This provides a more sustainable fuel option.

Wireless power transfer is the focus of Collot's paper on maximizing efficiency for near-field charging systems [14]. Optimizing coil parameters and alignment can reach up to 80% efficiency. Enabling robust contactless power has many applications from consumer electronics to EVs.

Indoor positioning is targeted by Kawasaki et al. through deep learning on BLE beacon fingerprints [15]. Temporal interpolation of RSSI sequences combined with a DNN model achieved 0.33m accuracy. Precise location services augment many contexts from navigation to asset tracking.

Automation for analytical chemistry is demonstrated by Stelzle et al. who developed a robotic platform for heavy metal quantification in dust samples [16]. The automated workflow provided faster, reliable measurements. Intelligent systems have potential to accelerate and expand environmental monitoring.

Information security is revisited by Flowerday and Blundell who apply facial expression analysis to evaluate reactions to policies [17]. Deep learning based sentiment classification provided promising results. Novel affective computing techniques can complement surveys.

Optoelectronic devices are explored by Abd-Ellah et al. who fabricated a perovskite-TiO2 heterojunction for broad-spectrum photodetection [18]. Incorporating NiO quantum dots yielded high responsivity and detectivity. Further advances in hybrid materials can lead to next-generation solar technologies and imaging systems.

Remote sensing for cartography is the theme of Abdelhafiz's work on segmenting buildings in aerial images by leveraging shadow information [19]. The proposed technique demonstrated over 97% precision on complex scenes. Reliable feature extraction facilitates geospatial applications.

Robotics control is targeted by Lu and Liu who developed a two-layer scheme for redundant manipulators [20]. Simulations verified improved performance in navigating constraints. Intelligent algorithms continue enabling more nimble automation across domains.

In summary, the diverse excellence of the research compiled in this special issue exemplifies leading-edge innovations pushing the boundaries of science and technology. We hope these contributions will stimulate researchers across disciplines to consider synergistic approaches leading to impactful discoveries. We thank the authors and reviewers for their diligent efforts in creating this issue.

References:

[1] A. Lamjahdi, H. Bouloiz, M. Gallab, “Heuristic Analysis of Overall Performance Measurement Perception and Management in Automotive Industry,” Advances in Science, Technology and Engineering Systems Journal, **7**(3), 1–11, 2022, doi:10.25046/aj070301.

[2] B. Marieme, Z. El Houssaine, “Analysis Methods and Classification Algorithms with a Novel Sentiment Classification for Arabic Text using the Lexicon-Based Approach,” Advances in Science, Technology and Engineering Systems Journal, **7**(3), 12–18, 2022, doi:10.25046/aj070302.

[3] S. Melzer, S. Thiemann, H. Peukert, R. Möller, “Towards a Model-based and Variant-oriented Development of a System of Systems,” Advances in Science, Technology and Engineering Systems Journal, **7**(3), 19–31, 2022, doi:10.25046/aj070303.

[4] H. Naoum, S.M. Benslimane, M. Boukadoum, “Encompassing Chaos in Brain-inspired Neural Network Models for Substance Identification and Breast Cancer Detection,” Advances in Science, Technology and Engineering Systems Journal, **7**(3), 32–43, 2022, doi:10.25046/aj070304.

[5] N. Hermita, R. Vebrianto, Z.H. Putra, J.A. Alim, T.T. Wijaya, U. Sulistiyo, “Effectiveness of Gamified Instructional Media to Improve Critical and Creative Thinking Skills in Science Class,” Advances in Science, Technology and Engineering Systems Journal, **7**(3), 44–50, 2022, doi:10.25046/aj070305.

[6] L. Xiong, S. Duncan-Williams, “Generalized Linear Model for Predicting the Credit Card Default Payment Risk,” Advances in Science, Technology and Engineering Systems Journal, **7**(3), 51–61, 2022, doi:10.25046/aj070306.

[7] N.E.A. Amrani, E.F. Ezzahra, M. Youssfi, S.M. Snineh, O. Bouattane, “A New Technique to Accelerate the Learning Process in Agents based on Reinforcement Learning,” Advances in Science, Technology and Engineering Systems Journal, **7**(3), 62–69, 2022, doi:10.25046/aj070307.

[8] J.K. Duggal, M. El-Sharkawy, “High Performance SqueezeNext: Real time deployment on Bluebox 2.0 by NXP,” Advances in Science, Technology and Engineering Systems Journal, **7**(3), 70–81, 2022, doi:10.25046/aj070308.

[9] M. Boumaza, Y. Boumaza, “Hole-Confined Polar Optical Phonon Interaction in Al0.35Ga0.65As/GaAs/Al0.25Ga0.75As Quantum Wells,” Advances in Science, Technology and Engineering Systems Journal, **7**(3), 82–86, 2022, doi:10.25046/aj070309.

[10] P.-Y. Lou, Y.-X. Chen, C.-C. Wang, W.-C. Chang, “A CMOS On-Chip High-Precision PVTL Detector,” Advances in Science, Technology and Engineering Systems Journal, **7**(3), 87–94, 2022, doi:10.25046/aj070310.

[11] A. Sharmin, F. Anwar, S.M.A. Motakabber, A.H.A. Hashim, “A Secure Trust Aware ACO-Based WSN Routing Protocol for IoT,” Advances in Science, Technology and Engineering Systems Journal, **7**(3), 95–105, 2022, doi:10.25046/aj070311.

[12] M.F. Sie, J. Wu, S.A. Harding, C.-L. Lin, S.-T. Wang, S. Liao, “Secured Multi-Layer Blockchain Framework for IoT Aggregate Verification,” Advances in Science, Technology and Engineering Systems Journal, **7**(3), 106–115, 2022, doi:10.25046/aj070312.

[13] C.N. Anyanwu, C.J. Animoke, B.U. Agu, I.F. Okafor, N.J. Ogbuagu, S. Bentson, O. Ojike, “Physical and Emission Properties of Blended Bio-Coal Briquettes Derived from Agro-Wastes in Nigeria,” Advances in Science, Technology and Engineering Systems Journal, **7**(3), 116–122, 2022, doi:10.25046/aj070313.

[14] J. Quignon, A. Tornambe, T. Deleruyelle, P. Pannier, “Antenna System Design To Increase Power Transfer Efficiency with NFC Wireless Charging Technology,” Advances in Science, Technology and Engineering Systems Journal, **7**(3), 123–128, 2022, doi:10.25046/aj070314.

[15] K. Echizenya, K. Kondo, “Indoor Position and Movement Direction Estimation System Using DNN on BLE Beacon RSSI Fingerprints,” Advances in Science, Technology and Engineering Systems Journal, **7**(3), 129–138, 2022, doi:10.25046/aj070315.

[16] H. Fleischer, S. Statkevych, J. Widmer, R. Stoll, T. Roddelkopf, K. Thurow, “Automated Robotic System for Sample Preparation and Measurement of Heavy Metals in Indoor Dust Using Inductively Coupled Plasma Mass Spectrometry (ICP-MS),” Advances in Science, Technology and Engineering Systems Journal, **7**(3), 139–151, 2022, doi:10.25046/aj070316.

[17] T. du Toit, H. Kruger, L. Drevin, N. Maree, “Deep Learning Affective Computing to Elicit Sentiment Towards Information Security Policies,” Advances in Science, Technology and Engineering Systems Journal, **7**(3), 152–160, 2022, doi:10.25046/aj070317.

[18] Y.A. Alwadei, M.S. Alshatwi, N.M. Alwadai, M.A. AlWehaibi, M.F. Alotaibi, M.M. Lashin, M.H. Alotaibi, “NiO Quantum dots Doped Triple Cation Perovskite CsMAFAPbI2Br2 Heterojunction Photodetector with High Responsivity,” Advances in Science, Technology and Engineering Systems Journal, **7**(3), 161–166, 2022, doi:10.25046/aj070318.

[19] A. Benchabana, M.-K. Kholladi, R. Bensaci, B. Khaldi, “A Supervised Building Detection Based on Shadow using Segmentation and Texture in High-Resolution Images,” Advances in Science, Technology and Engineering Systems Journal, **7**(3), 167–174, 2022, doi:10.25046/aj070319.

[20] D.M. Hung, D.X. Ba, “A Constrained Intelligent Nonlinear Control Method for Redundant Robotic Manipulators,” Advances in Science, Technology and Engineering Systems Journal, **7**(3), 175–182, 2022, doi:10.25046/aj070320.

Editor-in-chief

Prof. Passerini Kazmersk

ADVANCES IN SCIENCE, TECHNOLOGY AND ENGINEERING SYSTEMS JOURNAL

Volume 7 Issue 3 **May-June 2022**

CONTENTS

ASTES

Advances in Science, Technology and Engineering Systems Journal Vol. 7, No. 3, 1-11 (2022)

www.astesj.com

ASTESJ

ISSN: 2415-6698

Heuristic Analysis of Overall Performance Measurement Perception and Management in Automotive Industry

Aicha Lamjahdi*,[1], Hafida Bouloiz[1], Maryam Gallab[2]

[1]*National School of Applied Sciences (ENSA), Systems Engineering and Decision Support Laboratory (LISAD), IBN ZOHR University, Agadir, BP 80000, Morocco*

[2]*MINES-RABAT School (ENSMR), Computer Science Department, Rabat, BP 753, Morocco*

ARTICLE INFO

Article history:
Received: 07 February, 2022
Accepted: 21 April, 2022
Online: 16 May, 2022

Keywords:
Key performance indicators (KPIs)
Overall Performance management
Automotive industry
Overall performance measurement
Sustainable manufacturing
Decision making

ABSTRACT

Overall Performance (OP) measurement is an essential instrument in sustainable manufacturing implementation and management. The effective use of key performance indicators (KPIs) can potentially contribute to identify the firm's overall performance, provide the crucial gaps between desired results and current actions, and thus facilitate the implementation and execution of improvement strategies. This study attempts to give an insight into the perception of OP measurement and management in the automotive industry and explore the KPIs pertinent to this sector. This research is conducted in automobile organizations based in Morocco. We first carried out a literature review to determine the commonly used indicators of sustainability management in manufacturing. We then conducted a survey on a sample of firms to investigate the OP system management and how the proposed initial set of KPIs is perceived and used in the field. Findings reveal some management problems of the OP measurement system in the automobile sector. It was found that only a minority of companies use dedicated applications to manage their set of indicators. There is a lack of well-defined and standardized KPIs, which generally affect data quality. Moreover, most companies have a minimum percentage of decisions based on KPIs use, and only a few are satisfied with their overall performance measurement systems. However, analysis indicates no substantial differences in the perception of KPIs' importance among various respondents. Results showed that the most used KPIs are perceived as the most important. Consequently, sixteen indicators under the three dimensions of sustainability were presented as KPIs for OP measurement in the automotive industry. These indicators will hopefully serve the development of an OP management tool to support sustainability in this sector. The study is evenly valuable for other developing countries as Morocco in sustainability implementation in the automobile field.

1. Introduction

The automotive industry plays a significant role in the world economy's progress and evolution. It is a highly capitalistic sector, a catalyst of innovation, and an important generator of investments and employment around the world. In Morocco, the industry recorded the highest job creation between 2014 and 2019, with over 29% of the industrial sector's total job creations, reflecting the strong growth dynamic in which it is part [1]. Its performance is particularly remarkable in exports, ensuring a third of exports for a value of 7 billion dollars, which make it the country's leading export sector [2]. Thus, Morocco has become the leading car manufacturer in North Africa, with a market share of 38% in 2018 against 5% in 2003 [3].

However, the automotive sector significantly impacts the environment and society [4]. Governments, consumers, and investors are pushing manufacturers to adopt sustainability in the way they work, their culture, and their products. It is a strategic priority and long-term implication that can help companies gain several benefits such as improving product quality, enhancing the company's image, and increasing market share [5–7]. Similarly, the employees' interactions and skills improve by adopting the sustainable culture, which is reflected in their communication,

*Corresponding Author: Aicha Lamjahdi, aichalamjahdi@gmail.com

motivation, and satisfaction [8]. In [7], the author suggested that cost reduction and better attraction of funding sources are among the essential advantages of sustainability. Thus, eliminating waste sources and adopting sustainable practices results in cost minimization, motivates investors, and helps the company profit from the funding opportunities that are increasingly dedicated to encouraging this strategy in the firms.

To support the sustainability implementation in the manufacturing sector, managers need to monitor the firm's overall performance [9]. This approach combines the assessment of the economic, social, and environmental performances [10,11]. To fulfill this purpose, manufacturing organizations use performance measurement systems to ensure that they achieve their desired objectives and control possible deviations. Performance measures play an essential role within these systems. They provide detailed information on the current situation of the manufacturing process [12], support realization of improvement actions, and allow to evaluate decisions effectiveness [13]. A large number of manufacturing firms tend to surcharge themselves with performance indicators, which will affect decision-making and defocus fundamental issues that require consistent control and monitoring [14]. Thus, rigorous identification of Key Performance Indicators (KPIs), practically needed, should be a priority for organizations that aim to implement sustainability and assess OP [13,15].

Many researchers have proposed different types of indicators for overall performance measurement. However, a well-defined set of KPIs for OP management in the manufacturing industry has not yet been established to date [16–18]. In [16], the author stressed that this might be attributed to the difference between manufacturing organizations' characteristics, as each business industry has its special features and work environment. Moreover, in [18], the author confirmed that identifying the appropriate KPIs is affected by managers' perception of these indicators, which varies depending on different factors such as level of overall experience [19]. Further, many managers still find it challenging to identify the interrelationships between performance indicators. These barriers are added to the main challenges that can hinder the implementation of OP measurement system in the firm due to lack of use of structured tool and platform to assess the OP [20], the implementation complexity, and cost [8, 11] and the lack of qualification and training [16].

This paper aims to explore the perception of OP measurement in the context of automotive companies in Morocco and investigate the substantial KPIs in this field. Therefore, a literature review was carried out to collect the important key performance indicators demonstrated to be used in manufacturing practices. A survey was then conducted to investigate how OP is managed in the automotive firms, the challenges faced and the general adequacy of the OP measurement system used. The effective use of the established KPIs was examined, and the perception of these KPIs by various managers was also captured. Finally, a set of KPIs, perceived as the most relevant for OP measurement and management, was proposed, and the correlation between these indicators was evenly established. This research will be useful for automotive companies to get an overview of OP management and the appropriate KPIs for a meaningful implementation and monitoring of sustainability.

2. Methodology

2.1. Identification of key performance measures: Literature review

OP measurement is considered an indispensable approach to evaluate an organization's strategy and situate its results compared to its goals and targets. It provides essential information needed and supports decision building to achieve organization strategic objectives [12, 20]. Thus, this approach is essentially based on identifying, evaluating, and monitoring appropriate indicators and metrics that enable a detailed description of performance progress in the firm[15]. Different studies in the literature have suggested various sets of metrics for overall performance measurement. Thus, in [12] the author proposed eleven dimensions with 106 metrics for overall performance management in different manufacturing organizations in Pakistan. In [21], the researchers adopted the Triple Bottom Line (TBL) to define 18 indicators for lean manufacturing integration in the Brazilian industry. Similarly, in [22] the authors identified 41 indicators based on the TBL to develop an AHP model for sustainability assessment in the automotive sector. In [23], the authors defined a pool of 43 indicators to get an overview of sustainable scheduling in manufacturing industries.

From our review of literature [24], we have proposed an initial catalog of KPIs commonly considered by authors as necessary for OP management. We have adopted the TBL of sustainable management consisting of three dimensions- economic, environmental, and social. The economic dimension aims to ensure financial performance through an enhanced long-term income that exceeds the organization‘s charges and expenditures [25]. Thus, from the literature, the most popular indicators of the economic dimension measure quality of products and services, cost reduction, net profit, one time delivery, and investments in innovation and research and development (R&D). Quality of products and services refers to a firm's capacity to meet customer specifications and expectations [12, 26, 27]. Cost reduction relates to manufacturing and organization costs decreasing, including material cost, labor cost, and operational and capital costs [23, 28]. Net profit indicator measures the profit gained through the firm's activities [28]. One time delivery refers to the rate of parts delivered on time compared to the total number of units ordered in a predefined period [12]. It is considered a major factor of firm performance to increase customer service level and strengthen its competitive position in the market [12,29]. Similarly, investments in innovation and research and development measure is considered an essential aspect of firm's economic health. It ensures the organization's commitment to innovation that grantees its sustainability and growth [11, 28].

The environmental dimension addresses the rational use of natural resources within an organization's activities [21]. According to several authors, it covers indicators related to resource consumption, emissions impacts, waste management, and environmental certification. Resources consumption indicators seek to optimize energy and water during manufacturing operations [30, 31]. Emissions impacts criteria measure the impact of polluting emissions produced on the environment, involving the amount of greenhouse gases generated [32]. Waste management aims to minimize the solid and liquid waste produced and monitor

its treatment during the production process [23]. Thus, environmental certification is considered an important driver of environmental performance improvement by the firm and a relevant way to prove its commitment to environmental policy [29].

The social dimension is related to employees' social welfare and the surrounding community in which operate the organization [21,32]. Authors in the literature categorized this dimension through various measures, including injury rate, occupational health, safety, ratio of training hours, level of employee satisfaction, and investments in community development activities. Injury rate and occupational health and safety reflect mainly the work conditions in the firm, they measure the effect of the manufacturing process on worker health and safety and provide an overview of the social performance level in the workplace [23,33,34]. The training hours ratio is associated with employees' career development and reflects the firm's strategy in its staff's professional evolution [11,32]. Thus, employee satisfaction is considered an important factor to build and strengthen the firm's overall performance given the strong relationship between workers and the firm's efficiency and productivity [12,32]. It ensures employee well-being within the organization and supports decision-making for working environment improvement [32,33]. In [32], the authors confirmed that investments in community development activities criteria help assess the firm's impact on the surrounding community and support its involvement in community advancement projects.

Table1 outlines the literature on the relevant indicators for overall performance management.

2.2. Automotive industry survey

In an attempt to explore the established catalog of KPIs in the automotive sector, an industrial survey in Morocco-based automotive companies was carried out. A questionnaire made up of four parts was then used for this purpose. The first part consists of background information, such as the firm's size, product type, and respondents' experience in the automotive industry. The second part aims to provide an insight into the measurement systems management in terms of adequacy, challenges that affect the process, type of calculation and pilotage system used, and the percentage of decisions made as a result of KPIs use. The third part investigates the KPIs used in practice in the automotive industry. The last part focuses on the perception of each proposed KPI's importance, from respondents' point of view, based on their experiences and their firms' practices.

The questionnaire was submitted to a group of experts in the sustainable management field for a preliminary test. Ten copies were distributed to three researchers and seven practitioners in other automotive companies that were not part of the sample. This test allowed verification and validation of the questionnaire through the improvement of the questions and the statements' clarity and accuracy to avoid possible misinterpretations.

This survey population was selected in January 2020 from the Ministry of Industry, Trade, and Green and Digital Economy (MITGDE) automotive directory. The ministry directory was used due to its diverse and updated automotive companies' database, which provides a reliable and well-represented sampling reference. Thus, the study was conducted among project managers, deputy project managers, and other senior experts in the surveyed firms. They are high-ranking informants that most consider the KPIs management approach to control and manage their projects.

The survey was sent to respondents by email that explained its objective and ensured their anonymity. Some undelivered questionnaires by email, were issued to the respondents through the business online platform Linkedin, which provided more interactivity with the respondents and contributed significantly to raising the response rate.

3. Results and discussions

3.1. Background of respondents

The questionnaire was delivered to 127 automotive firms in Morocco. Thus, 51 replies were received, which resulted in a response rate of 40%. This rate seems to be acceptable. According to the authors in [35] a lot of this type of study, in the literature, obtain less than 30% response rate. The table 2 presents the background of respondents.

Of the study sample, over half of the respondents (53%) were project managers, 29% were deputy project managers, 18% were directors and senior professionals and the remaining 10% of 'others' consisted of production managers, quality control managers and process engineers. For the years of respondents' experience in the automotive sector, over 66% had from 6 to 10 years' experience in the field, then 21.6% with more than 10 years and finally 11.8% with 1 to 5 years in the sector.

The majority of respondents (58.8%) were from large companies categorized in this research as having an annual turnover equal to or greater than $ 20 million[36]. The other 41.2% were from Small and Medium-sized Enterprises (SMEs) having an annual turnover between 1 and 20 million dollars [36]**.** Almost 23% of the companies surveyed manufactured plastic and rubber parts, 21.6 % produced automotive wiring equipments, 17.6 % manufactured metal equipments, 11.8% manufactured mechanical equipments, 5.9% manufactured electrical and electronic equipments and 19,6% manufactured other automotive equipments. This diversity in product type of the sample firms showed that the automotive sector population is well represented in this study. Furthermore, 86% of the companies were certified to IATF 16949 that provides specifications for quality system management for the automobile industry.

Besides, 44% are certified to ISO 14001, the environmental management standard, 28% were certified to ISO 45001 related to health and safety management, 13% were certified to ISO 9001, and 2% were certified to ISO 26000 related to corporate social responsibility implementation. This implies the companies rely on essential standards in the management of their overall performance.

3.2. Perception of KPIs management in automotive industry

The objective of this section is to check how the KPIs are managed and how they support performance management in automotive companies. For this aim, the respondents were asked which system they use to manage their set of KPIs, which challenges are faced in their KPIs monitoring process, the

Table 1: Relevant indicators for performance management in literature

Overall Performance dimensions	Indicators	Authors
Economic	Quality of products and services	[21] ; [28] ; [12] ;[37];[26];[29]; [38];[30];[27]
	Material cost	[23] ; [21]; [28]; [31]; [12]; [37]; [26]; [39]; [40]; [38]; [41]
	Labour cost	[23] ; [28]; [31]; [12]; [37]; [39]; [40]; [41]
	Operational and capital costs	[23] ; [28]; [31]; [12]; [37]; [26]; [39]; [29]; [38]; [30]; [27]
	Net profit	[23] ; [21]; [12]; [39]; [30]; [27]; [42]; [41]
	Rate of one time delivery	[12]; [37]; [26]; [29]; [38]; [30]; [27]
	Investments in innovation and research and development	[21]; [12]; [37]; [26]; [39]; [30]; [27]
Environmental	Total water consumption	[23] ; [21];[28]; [31]; [4]; [39]; [40]; [29]; [33]; [30]
	Total energy consumption	[23] ; [21]; [28]; [31]; [4]; [26]; [39]; [40]; [29]; [33]; [30]
	Amount of greenhouse gases generated	[23] ; [28]; [31]; [4]; [26]; [39]; [40]; [29] ; [33]; [30]
	Solid and liquid waste produced	[23] ; [31]; [37]; [39] ; [29]; [33]
	Disposal of waste	[23] ; [21]; [28]; [31]; [26]; [29]; [27]
	Environmental certification	[12]; [26]; [29]; [38]; [30] ; [27]
Social	Injury rate	[23] ; [31]; [12]; [26]; [40]; [29]; [33]; [27]
	Occupational health and safety	[23] ; [21]; [28]; [31]; [12]; [37]; [26]; [39]; [40]; [29]; [33]; [38] ; [30] ; [27]
	Ratio of training hours	[23] ; [21]; [31]; [37]; [40]; [33]
	Level of employee satisfaction	[21]; [28]; [31]; [12]; [37]; [39]; [33]
	Investments in community development activities	[21]; [31]; [12]; [26]; [39]; [29]; [33] ; [30]

percentage of changes and decisions based on KPIs use, and the adequacy of their system of performance measurement.

As shown in Figure 1, over half of the respondents' companies (56.9 %) use spreadsheets or manual processes to control and monitor their KPIs, 29.4% use on-premise financial software modules, and only 13.7 % use advanced applications. This indicates a significant lack of management tools that help save time, which could be reinvested in value-adding activities. Moreover, the majority of respondents (56,9%), as shown in Figure 2, stated that “Too many indicators” is the challenge that most affects them in the process of monitoring KPIs, followed by “too many departments involved” challenge with 41.2%, then “Not enough time” challenge with 21.6%. This finding confirms the OP management complexity advocated in the literature [8,11,20] due to the multiplicity of metrics and departments involved, which leads to enormous difficulties for managers.

For the percentage of decisions and changes based on KPIs use (Figure 3), the greatest percentage of respondents (39.2%) have less than 25%. Only 17.6% have over 75% of changes and decisions established based on KPIs deployment. This unexpected result contrasts with the role of decision support and change orientation in which KPIs use is supposed to help managers, as confirmed in the literature [13].

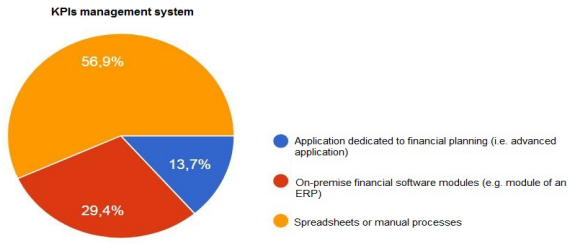


Figure 1: KPIs management system

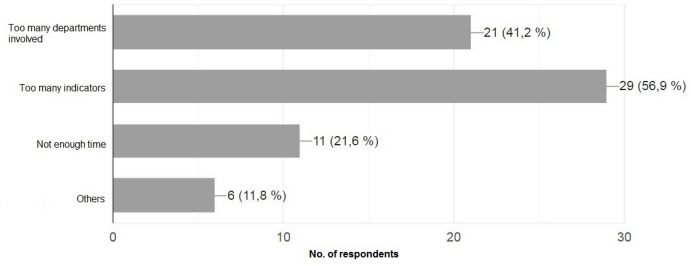


Figure 2: Challenges faced within KPIs management process

Furthermore, as illustrated in Figure 4, 60.8% of those surveyed indicated that their KPIs measurement system needs improvement, while the remaining 39.2% said their measurement system is adequate. This reveals that the majority of managers are not satisfied with the KPIs measurement process in their companies.

Table 2: Background of respondents

Background of respondents	Frequency	Percentage
Respondents job position		
Director/General manager	9	18
Project manager	27	53
Deputy project manager	10	19
Others	5	10
Respondents experience in automotive sector		
1-5 years	6	11.8
6-10 years	34	66.7
>10 years	11	21.6
Company turnover		
$ 1-20m	21	41.2
>= $ 20m	30	58.8
Product type		
Automotive wiring	11	21.6
Metal equipments	9	17.6
Mechanical equipments	6	11.8
Electrical and electronic equipments	3	5.9
Plastic and rubber parts	12	23.5
Others	10	19.6
Certification		
IATF 16949	43	86
ISO 14001	22	44
ISO 45001	14	28
ISO 26000	1	2
ISO 9001	7	13
Others	1	2

These notable findings reveal a significant problem of effectiveness of KPIs management process in automotive companies' practice. The lack of in-depth and precise definition of metrics and data that should be used coupled with the non-implementation of the appropriate tools of performance management affect the adequacy of system measurement and impact its ability to support decision making in the companies.

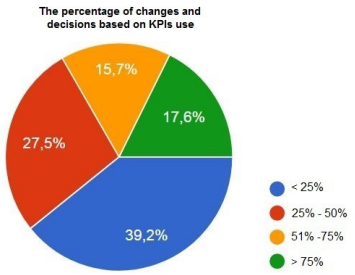


Figure 3: The percentage of changes and decisions based on KPIs use

This ineffectiveness can widen the gap between the firms' actual needs and the measured indicators and eventually hampered the good governance and the improvement of companies' overall performance.

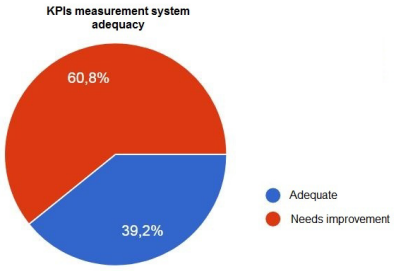


Figure 4: KPIs measurement system adequacy

3.3. Perception of KPIs use in automotive industry practice

In order to examine the use of the proposed set of KPIs in the automotive sector practices, the participants were asked to indicate the metrics they are using to manage their companies' overall performance. Table 3 presents the results.

As shown in table 3, the "Quality of products and services" is the most used indicator (98%) in the surveyed automotive companies. This is not unexpected since the quality ensures that the firms provide products and services that respect clients' requirements and expectations [43]. Further, most of the companies are certified to IATF 16949, and they should consider product quality measurement in their priorities while monitoring OP.

The injury rate is the second most used indicator with a rate of 92%. There is also occupational health & safety in the fifth rank with a rate of 75%. This result is evident, given the criticality of risks and dangers within the automotive workplace industry [44] . Economic metrics are used evenly with important rates: one time delivery with 88%, labor cost with 78%, material cost with 71%, and net profit with 57%. Besides, environmental metrics are high ranked between the 18 KPIs proposed, with a usage rate of 69% for total energy consumption and 51% for both total water consumption and solid and liquid waste produced. This may be attributed to the legislation requirements to control the automotive industry's environmental impact as an important resource consumption source**.** For the social metrics, as well as injury rate and occupational health & safety, the ratio of training hours is also commonly used with a rate of 61%. Overall, it can be seen, from the results, that companies in the automotive sector, as confirmed previously, focus on the different OP dimensions metrics (social, economic, and environmental) in their practices. However,

environmental certification and investments in community development activities are the least used metrics, with 12% and 8%, respectively. This may be due to the complexity of implementation and the high cost of such investments as advocated in the literature [8].

Table 3: Usage of KPIs in automotive industry practices

Performance indicator	Percentage	Rank
Quality of products and services	98	1
Injury rate	92	2
Rate of one time delivery	88	3
Labour cost	78	4
Occupational health and safety	75	5
Material cost	71	6
Total energy consumption	69	7
Ratio of training hours	61	8
Net profit	57	9
Total water consumption	51	10
Solid and liquid waste produced	51	11
Operational and capital costs	49	12
Level of employee satisfaction	49	13
Investments in innovation and research and development	39	14
Disposal of waste	29	15
Amount of greenhouse gases generated	20	16
Environmental certification	12	17
Investments in community development activities	08	18

3.4. Perception of KPIs importance in automotive industry

- **Analysis of variance ANOVA and T-test results**

To investigate automotive companies' perspective of the to investigate automotive companies' perspective of the proposed KPIs' importance for overall performance management, the respondents were asked to rate the importance level of each KPI. A five-point Likert-type scale ranging from 1 (not important at all) to 5 (very important) was used. The analysis of variance (ANOVA) and the t-test were performed to detect if the perceived importance of KPIs is affected by differences between respondent groups. For this purpose, the respondents were categorized under three categories according to experience in the automotive sector, the type of product manufactured, and the company size.

Analysis of variance (ANOVA) and t-test are decision-making techniques for detecting any statistical differences among various means of independent groups [45,46]. The ANOVA is used to compare means of three or more independent groups, whereas the t-test is used when the difference between the means of only two groups are to be studied [46]. Therefore, the analysis of variance (ANOVA) was carried out to investigate the important differences among the "Type of product" groups on one side and the "years' experience in the automotive sector" groups on the other side. Afterward, the t-test was performed to compare the means between large companies and SMEs, the two groups of "company size" category, surveyed in this study. The table 4 shows the results of these tests.

The null hypothesis for these statistical tests is that the means for all the groups in the same category are supposed to be identical to each other. The null hypothesis is rejected when the p value is less than alpha. The p-value represents the probability of obtaining the observed results, assuming the null hypothesis is true. Alpha is the risk that a null hypothesis will be rejected when it is actually true. When p- value is less than alpha, this means that there is a statistically significant difference in the groups' means and they may represent separate categories. By convention, alpha is typically set to 0.05 [46]. The test statistic for ANOVA and t-test are respectively the F score and the T score.

As shown in table 4, the ANOVA results suggest that there is no significant difference in respondents' perception rating of KPIs for "Type of product" groups. Similarly, for "years' experience in the automotive sector", respondents' groups show a general agreement in importance perception rating of KPIs, except for "training hours". They have a clear significant difference in means at a 95% confidence level ($p < .05$). Tukey's post hoc test was then carried out to explore the pairwise comparison differences between "years' experience" groups for "training hours" indicator. Results show that managers with 6 to 10 years' experience and those with more than ten years experience give more importance rating to "training hours" in comparison with those with less than five years experience in the field ($p < .05$).

This is in line with Cox findings [19], which suggests that managers with less experience tend to place more importance on indicators related to the field level, while more experienced managers have a wider company vision where other important performance issues are considered when choosing KPIs for OP assessment. For the t-test, as shown in table 5, results indicate that respondents had a complete agreement about KPIs importance perception when company size was regarded. Taken together, these results suggest that there are no substantial differences in the perceived importance of KPIs among various groups of each category of respondents, namely 'years' experience in the automotive sector', "Type of product manufactured" and "Company size". Thus, respondents in the automotive sector generally tend to have similar views about the proposed KPIs importance to manage and assess overall performance.

Table 4: ANOVA for different categories of respondents groups

Key performance indicator	Type of product		Experience in automotive sector	
	F score	Signification	F score	Signification
Quality of products and services	0.347	0.908	1.285	0.286
Material cost	0.954	0.467	0.688	0.507
Labour cost	0.638	0.700	1.698	0.194

Operational and capital costs	1.207	0.321	2.661	0.080
Net profit	0.285	0.941	2.949	0.062
Rate of one time delivery	0.451	0.841	0.435	0.650
Investments in innovation and research and development	1.486	0.205	0.233	0.793
Total water consumption	0.804	0.573	1.380	0.261
Total energy consumption	1.898	0,102	1.673	0.198
Amount of greenhouse gases generated	0.298	0.935	1.750	0.185
Solid and liquid waste produced	1.368	0.248	0.424	0.657
Disposal of waste	1.426	0.226	2.516	0.091
Environmental certification	0.796	0.578	1.912	0.159
Injury rate	0.863	0.530	0.224	0.800
Occupational health and safety	0.497	0.807	0.860	0.429
Level of employee satisfaction	0.735	0.624	0.221	0.802
Investments in community development activities	0.773	0.595	2.143	0.128
Ratio of training hours	0.978	0.452	21.303	0.001*

* $p < .05$

- **KPIs importance ranking**

The mean importance of KPIs was calculated and ranked, as presented in table 6, for the overall population since no significant differences in perceived KPIs were captured among various respondents groups.

In general, the mean importance values range from 3.37 to 4.82, which implies that all the proposed KPIs are relevant for OP management in the automotive industry. Among the 18 KPIs, quality, one-time delivery, occupational health & safety, labor cost, and material cost are highly ranked. As customers are very demanding on the quality of products and services, companies consider ensuring high quality at minimum costs, a primary objective to improve their competitiveness [12].

Table 5: T-test for company size categories

Key performance indicator	Mean importance		T-test	
	Large entrep-rises	Small and medium-sized enterprises	T score	Significat ion
Quality of products and services	4.77	4.90	-1.160	0.252
Material cost	4.40	4.45	-0.267	0.790
Labour cost	4.33	4.50	-0.859	0.394
Operational and capital costs	4.00	3.40	1.755	0.090
Net profit	4.33	4.35	-0.058	0.954
Rate of one time delivery	4.67	4.70	-0,184	0.855
Investments in innovation and research and development	3.97	3.80	0.543	0.589
Total water consumption	3.77	3.85	-0.317	0.753
Total energy consumption	3.63	3.95	-0.951	0.347
Amount of greenhouse gases generated	3.87	3.95	-0.262	0.795
Solid and liquid waste produced	3.83	4.05	-0.758	0.452
Disposal of waste	3.87	3.45	1.356	0.182
Environmental certification	3.27	3.50	-0.836	0.407
Injury rate	4.17	4.05	0.392	0.697
Occupational health and safety	4.47	4.60	-0.547	0.587
Level of employee satisfaction	4.17	4.40	-0.875	0.386
Investments in community development activities	3.70	3.55	0.535	0.595
Ratio of training hours	4.27	4.25	0.074	0.942

Likewise, time is considered an important factor of a firm's commitment and performance [47]. Interest in safety is also understandable due to its critical effect on cost, production, and workplace components [13]. These indicators are followed by environmental dimension measures where solid and liquid waste produced is the high ranked. However, investments in community development activities and environmental certification are considered the least important indicators and were ranked at the bottom of the KPIs list in terms of importance perception.

Table 6: T-test for company size categories

Performance indicator	Mean	Rank
Quality of products and services	4.82	1
Rate of one time delivery	4.69	2
Occupational health & safety	4.53	3
Labour cost	4.41	4
Material cost	4.41	5
Net profit	4.35	6
Level of employee satisfaction	4.27	7
Ratio of training hours	4.27	8
Injury rate	4.12	9
Solid and liquid waste produced	3.94	10
Investments in innovation and and research and development	3.90	11
Amount of greenhouse gases generated	3.90	12
Operational and capital costs	3.78	13
Total energy consumption	3.78	14
Total water consumption	3.78	15
Disposal of waste	3.71	16
Investments in community development activities	3.65	17
Environmental certification	3.37	18

In comparison with the usage perception results reported in section 3.3, it can be seen that practically almost all the indicators that were most used in automotive practices are perceived as the most important in OP assessment and management in the sector. Quality, time, safety, and cost are the highly perceived indicators in terms of usage and importance for OP management. Investments in community development activities and environmental certification are evenly the least perceived. It is suggested then to remove these two indicators from the initial proposed catalog of KPIs for OP management in the automotive industry. As shown in Figure 5, a total of sixteen measures with three dimensions of economic, social, and environmental have been proposed as the KPIs for OP assessment and management in the automotive industry.

So as to ensure the adequacy and the compliance of the proposed catalog of KPIs with automotive firms' practices, participants were asked to indicate whether the relevant indicators for overall performance management have been all considered in the proposed KPIs and if their companies are using other indicators that they found important to monitor OP. From the respondents, 80% confirmed that the appropriate metrics for OP management had been all considered. Likewise, the majority indicated that they use indicators from the proposed KPIs, and only 25% use other metrics such as flexibility and finance that they consider important. It is evident that some companies use specific indicators that they consider important for their overall performance depending on their special objectives in the short and long term. Yet, our final set of KPIs established on the basis of this survey can always be adapted to the case of the company that uses it by adding the indicators that it considers necessary for its overall performance management. Otherwise, if some users feel that the proposed set of indicators contains too many KPIs, they can obviously adapt it by eliminating indicators that they consider unnecessary for their OP measurement, based on the opinion of high-ranking managers and experts in sustainability implementation and OP measurement.

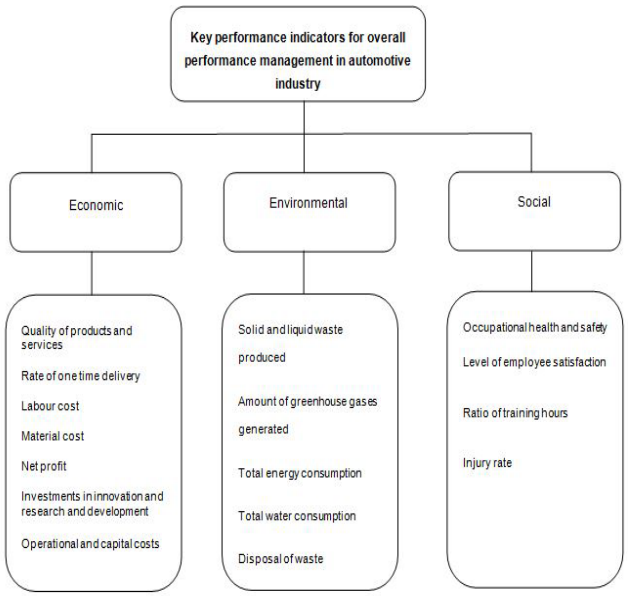


Figure 5: Proposed KPIs for OP management in automotive industry

3.5. Correlation of KPIs

To investigate the impact of the different KPIs on each other, the Pearson rank correlation was performed, using the statistical package SPSS, as shown in table 7. Results show that there generally exist substantial correlations between the various KPIs. For instance, quality is positively correlated with one time delivery, safety, solid and liquid waste produced, labour cost, employee satisfaction, and energy consumption. This emphasizes the importance of these indicators in achieving an automotive firm's quality. One time delivery is evenly correlated with safety, employee satisfaction, solid and liquid waste produced, amount of GHG generated, and investments in innovation and R&D. This implies that any increase/decrease in these indicators will significantly influence the delivery performance of the company. Safety also has a significant correlation with employee satisfaction. This may be attributed to the important impact of safety on people in the organization as advocated in [13]. Overall, it can be seen that most of the KPIs are significantly related to each other, which signifies that they forge substantial linkages and shall not be studied separately. In other words, these KPIs must be considered as diverse features of the same overall performance.

Table 7. Correlation test between the different KPIs

KPIs	Quality of products and services	Material cost	Labour cost	Operational and capital costs	Net profit	Rate of one time delivery	Investments in innovation and research and development	Total water consumption	Total energy consumption	Amount of greenhouse gases generated	Solid and liquid waste produced	Disposal of waste	Environmental certification	Injury rate	Occupational Health & Safety	Level of employee satisfaction	Investments in community development activities	Ratio of training hours
Quality of products and services	1																	
Material cost	0.196	1																
Labour cost	0.393**	0.392**	1															
Operational and capital costs	0.208	0.211	0.201	1														
Net profit	0.244	0.051	0.233	0.382**	1													
Rate of one time delivery	0.537**	0.081	0.514**	0.132	0.354*	1												
Investments in innovation and research and development	0.182	0.122	0.174	0.290*	0.211	0.324*	1											
Total water consumption	0.003	0.088	0.250	-0.007	-0.003	0.092	0.211	1										
Total energy consumption	0.322*	0.204	0.428**	0.428**	0.158	0.269	0.214	0.493**	1									
Amount of greenhouse gases generated	0.175	0.176	0.472**	-0.018	0.185	0.403**	0.328*	0,757**	0,511**	1								
Solid and liquid waste produced	0,395**	0,166	0,522**	0,368**	0,291*	0,429**	0,363**	0,390**	0,725**	0,481**	1							
Disposal of waste	0,189	0,005	0,286*	0,584**	0,102	0,253	0,370**	0,225	0,664**	0,339*	0,687**	1						
Environmental certification	0,065	0,071	0,193	0,076	0,284*	0,236	0,317*	0,488**	0,345*	0,595**	0,340*	0,168	1					
Injury rate	-0,179	-0,076	-0,073	-0,048	0,038	-0,036	0,181	0,510**	0,193	0,485**	,067	0,163	0,613**	1				
Occupational health and safety	0,430**	0,221	0,498**	0,146	0,257	0,564**	0,429**	0,155	0,371**	0,436**	0,597**	0,360**	0,299*	-0,004	1			
Level of employee satisfaction	0,325*	0,076	0,594**	0,098	0,113	0,544**	0,258	0,194	0,415**	0,450**	0,635**	0,432**	0,268	-0,035	0,669**	1		
Investments in community development activities	-0,105	-0,183	-0,018	0,133	-0,057	0,046	0,346*	0,537**	0,238	0,488**	0,211	0,368**	0,365**	0,436**	0,114	0,249	1	
Ratio of training hours	0,028	0,211	0,202	0,323*	0,133	0,142	-0,089	-0,142	0,335*	-0,134	0,178	0,366**	-0,167	-0,245	0,080	0,201	-0,082	1

** Correlation is significant at the 0.01 level (2-tailed)

* Correlation is significant at the 0.05 level (2-tailed)

4. Conclusions

Sustainability has become a strategic priority for the automotive industry worldwide. Companies preoccupied with adopting this strategy should necessarily manage and monitor their OP [48] and use key performance indicators that ensure they're achieving their preset targets and goals. However, many of these firms face difficulties in implementing and managing OP measurement systems and tend to surcharge themselves with a lot of KPIs [20]. This study investigates how OP measurement systems are perceived and managed in the Moroccan automotive sector and explores the KPIs relevant to this manufacturing field.

For this purpose, a literature review was first carried out to establish an initial catalog of KPIs commonly used in manufacturing to manage OP. A questionnaire-based survey was then conducted on 51 companies in the Moroccan automotive sector to get an overview about OP management and investigate the established KPIs in this industry.

Findings of the survey reveal a significant problem of effectiveness of OP measurement process in automotive companies' practice. Results showed that over half of the respondents utilize spreadsheets or manual processes to monitor OP, and only a few respondents use advanced applications. This indicates that there is a significant lack of use of the appropriate tools of overall performance management. Moreover, most respondents declared that "Too many indicators" followed by "Too many departments involved" are the challenges most affect them in the OP monitoring process. This is in line with what was advocated in the literature about the OP management complexity, which results, among others, from the multitude of indicators and departments involved, and causes important difficulties for decision-makers [8,11,20]. The unexpected result was that the most significant respondents have less than 25% of decisions and changes triggered by KPIs deployment. This is in contrast with KPIs' role supposed to support decision-making in the firm [13]. Further analysis reveals that the major part of managers is not satisfied with their companies' OP measurement process.

This last part of the study showed the importance of rigorous identification of KPIs practically needed to facilitate OP implementation and management in the automotive sector. For this purpose, we examined the use of the previously proposed catalog of KPIs in the studied sample of firms. The results showed that "Quality of products and services" is the most used indicator with a response rate of 98%, which is consistent with the results of [12] in which the author confirmed that automobile organizations put more focus on product quality according to client demand. In general, it was found that the automotive companies use the different OP dimensions metrics (social, economic, and environmental) in their practices. Despite this, it was seen that environmental certification and investments in community development activities are the least used indicators, probably due to the complexity of implementation and the high cost of such assets [8].

Conflict of Interest

The authors declare no conflict of interest.

References

[1] Ministry of Industry, The industry created 288,126 jobs between 2014 and 2017 and is positioned as a major job provider, 2017.

[2] J. Deveaux, Morocco: the automobile confirms itself as a major player in the economy, 2020.

[3] M. Diop, Automotive: How Morocco snatched Africa's number 1 spot from South Africa., 1, 2018.

[4] G. Scur, A.M. De Mello, L. Schreiner, F.J. das Neves, "Eco-design requirements in heavyweight vehicle development – a case study of the impact of the Euro 5 emissions standard on the Brazilian industry," Innovation & Management Review, **16**(4), 404–422, 2019, doi:10.1108/INMR-08-2018-0063.

[5] A.N. Nambiar, "Challenges in Sustainable Manufacturing," in Proceedings of the 2010 International Conference on Industrial Engineering and Operations Management, Dhaka, Bangladesh, January 9-10, 10–15, 2010.

[6] E.L. Psomas, C. Fotopoulos, D. Kafetzopoulos, "Motives, difficulties and benefits in implementing the ISO 14001 Environmental Management System," Management of Environmental Quality: An International Journal, **22**(4), 502–521, 2011, doi:http://dx.doi.org/10.1108/14777831111136090.

[7] I. Gavronski, G. Ferrer, E.L. Paiva, "ISO 14001 certification in Brazil: motivations and benefits," Journal of Cleaner Production, **16**(1), 87–94, 2008, doi:10.1016/j.jclepro.2006.11.002.

[8] M.L. Martín-Peña, E. Díaz-Garrido, J.M. Sánchez-López, "Analysis of benefits and difficulties associated with firms ' Environmental Management Systems : The case of the Spanish automotive industry," Journal of Cleaner Production, **70**(May), 220–230, 2014, doi:10.1016/j.jclepro.2014.01.085.

[9] M. Capron, F. Quairel, "An Assessment of Sustainable Development Strategies of the Firm: The attractive Utopia of Global Performance [Evaluer les stratégies de développement durable des entreprises : l'utopie mobilisatric]," Review of the Responsible Organization, **1**, 5–17, 2006, doi:10.3917/ror.001.0005.

[10] A. Renaud, N. Berland, "Companies overall performance measurement [Mesure de la performance globale des entreprises.]," Accounting and Environment, CD-Rom. halshs-00544875 HAL, 2007.

[11] A. Lamjahdi, H. Bouloiz, M. Gallab, "Toward a model to apprehend the complexity of manufacturing firm ' s overall performance," International Journal of Engineering Business Management, **12**, 1–20, 2020, doi:10.1177/1847979020901982.

[12] M. Ishaq Bhatti, H.M. Awan, "The key performance indicators (KPIs) and their impact on overall organizational performance," Quality and Quantity, **48**(6), 3127–3143, 2014, doi:10.1007/s11135-013-9945-y.

[13] P.N. Muchiri, L. Pintelon, H. Martin, A.M. De Meyer, "Empirical analysis of maintenance performance measurement in Belgian industries," International Journal of Production Research, **48**(20), 5905–5924, 2010, doi:10.1080/00207540903160766.

[14] G. Marsden, C. Kelly, C. Snell, "Selecting indicators for strategic performance management," Transportation Research Record, (1956), 21–29, 2006, doi:10.3141/1956-03.

[15] M.R. Galankashi, S.A. Helmi, P. Hashemzahi, "Supplier selection in automobile industry: A mixed balanced scorecard-fuzzy AHP approach," Alexandria Engineering Journal, **55**(1), 93–100, 2016, doi:10.1016/j.aej.2016.01.005.

[16] S. Singh, E.U. Olugu, S.N. Musa, "Development of Sustainable Manufacturing Performance Evaluation Expert System for Small and Medium Enterprises," Procedia CIRP, **40**, 609–614, 2016, doi:10.1016/j.procir.2016.01.142.

[17] E. Amrina, U. Andalas, S. Yusof, "Manufacturing performance evaluation tool for Malaysian automotive small and medium-sized enterprises," International Journal of Business Science and Applied Management, (January), 195–213, 2010.

[18] S. Toor, S.O. Ogunlana, "Beyond the ' iron triangle ': Stakeholder perception of key performance indicators (KPIs) for large-scale public sector development projects," International Journal of Project Management, **28**(3), 228–236, 2010, doi:10.1016/j.ijproman.2009.05.005.

[19] R.F. Cox, R.R.A. Issa, M. Asce, D. Ahrens, "Management ' s Perception of Key Performance Indicators for Construction," Journal of Construction Engineering and Management, (April), 142–151, 2003.

[20] K.K.B. Hon, "Performance and Evaluation of Manufacturing Systems," CIRP Annals - Manufacturing Technology, **54**(1), 139–154, 2005.

[21] L. Helleno, A. Jos, "Integrating sustainability indicators and Lean Manufacturing to assess manufacturing processes : Application case studies in Brazilian industry," **153**, 2017, doi:10.1016/j.jclepro.2016.12.072.

[22] E. Amrina, S.M. Yusof, "An AHP based-model for sustainable

manufacturing performance evaluation in automotive industry," in Proceedings of the International Symposium on the Analytic Hierarchy Process, 2013.
[23] M. Akbar, T. Irohara, "Scheduling for sustainable manufacturing: A review," Journal of Cleaner Production, 2018, doi:10.1016/j.jclepro.2018.09.100.
[24] A. Lamjahdi, H. Bouloiz, M. Gallab, "Overall performance indicators for sustainability assessment and management in mining industry," in 2021 7th International Conference on Optimization and Applications, ICOA, IEEE, Wolfenbüttel, Germany: 1–6, 2021, doi:10.1109/ICOA51614.2021.9442635.
[25] Z. Wang, N. Subramanian, A. Gunasekaran, M.D. Abdulrahman, C. Liu, "Composite sustainable manufacturing practice and performance framework: Chinese auto-parts suppliers' perspective," International Journal of Production Economics, **170**, 219–233, 2015, doi:10.1016/j.ijpe.2015.09.035.
[26] E. Chardine-Baumann, V. Botta-Genoulaz, "A framework for sustainable performance assessment of supply chain management practices," Computers and Industrial Engineering, **76**(1), 138–147, 2014, doi:10.1016/j.cie.2014.07.029.
[27] M.K. Lim, M. Tseng, K. Hua, T. Dat, "Knowledge management in sustainable supply chain management : Improving performance through an interpretive structural modelling approach," Journal of Cleaner Production, **162**, 806–816, 2017, doi:10.1016/j.jclepro.2017.06.056.
[28] R. Henao, W. Sarache, I. Gómez, "Lean Manufacturing and Sustainable Performance : Trends and Future Challenges," 2018, doi:10.1016/j.jclepro.2018.10.116.
[29] K. Govindan, R. Khodaverdi, A. Jafarian, "A fuzzy multi criteria approach for measuring sustainability performance of a supplier based on triple bottom line approach," Journal of Cleaner Production, **47**, 345–354, 2013, doi:10.1016/j.jclepro.2012.04.014.
[30] C. Hsu, A. Chang, W. Luo, "Identifying key performance factors for sustainability development of SMEs - integrating QFD and fuzzy MADM methods," Journal of Cleaner Production, **161**, 629–645, 2017, doi:10.1016/j.jclepro.2017.05.063.
[31] A. Huang, F. Badurdeen, "Metrics-based Approach to Evaluate Sustainable Manufacturing Performance at the Production Line and Plant Levels," Journal of Cleaner Production, 2018, doi:10.1016/j.jclepro.2018.04.234.
[32] C.B. Joung, J. Carrell, P. Sarkar, S.C. Feng, "Categorization of indicators for sustainable manufacturing," Ecological Indicators, **24**, 148–157, 2012, doi:10.1016/j.ecolind.2012.05.030.
[33] M.L. Tseng, "Modeling sustainable production indicators with linguistic preferences," Journal of Cleaner Production, **40**, 46–56, 2013, doi:10.1016/j.jclepro.2010.11.019.
[34] A. Huang, F. Badurdeen, "Sustainable Manufacturing Performance Evaluation: Integrating Product and Process Metrics for Systems Level Assessment," Procedia Manufacturing, **8**(October 2016), 563–570, 2017, doi:10.1016/j.promfg.2017.02.072.
[35] A. Cater-Steel, L. Al-Hakim, Information systems research methods, epistemology, and applications, 226–240, 2008, doi:10.4018/978-1-60566-040-0.
[36] H. El ouazzani, K. Rouggani, "Financing small and medium-sized enterprises: the case of morocco [Le financement des petites et moyennes entreprises : cas du Maroc]," in International conference: Finance, management control and overall performance of organizations, 2018.
[37] E. Amrina, S. Yusof, "Key Performance Indicators for Sustainable Manufacturing Evaluation in Automotive Companies," in Proceeding of the 2011 IEEE IEEM, 1093–1097, 2011.
[38] A.H. Azadnia, M.Z.M. Saman, K.Y. Wong, P. Ghadimi, N. Zakuan, "Sustainable Supplier Selection based on Self-organizing Map Neural Network and Multi Criteria Decision Making Approaches," Procedia - Social and Behavioral Sciences, **65**, 879–884, 2012, doi:10.1016/j.sbspro.2012.11.214.
[39] C.B. Joung, J. Carrell, P. Sarkar, S.C. Feng, "Categorization of indicators for sustainable manufacturing," Ecological Indicators, **24**, 148–157, 2013, doi:10.1016/j.ecolind.2012.05.030.
[40] E. Amrina, A.L. Vilsi, "Key Performance Indicators for Sustainable Manufacturing Evaluation in Cement Industry," Procedia CIRP, **26**, 19–23, 2015, doi:10.1016/j.procir.2014.07.173.
[41] H.X. Tan, Z. Yeo, R. Ng, T.B. Tjandra, B. Song, "A sustainability indicator framework for Singapore small and medium-sized manufacturing enterprises," in Procedia CIRP, Elsevier B.V.: 132–137, 2015, doi:10.1016/j.procir.2015.01.028.
[42] R. Usha, G. Angappa, S. Nachiappan, "Supply chain collaboration performance metrics : a conceptual framework," Benchmarking: An International Journal, **18**(6), 856–872, 2014, doi:10.1108/14635771111180734.
[43] M. Yurdakul, "Measuring a manufacturing system's performance using Saaty's system with feedback approach," Integrated Manufacturing Systems, **13**(1), 25–34, 2002, doi:10.1108/09576060210411486.
[44] A. Ehs, Occupational health and safety-Automotive Industry, 2019.
[45] S.R. Rao, G. Padmanabhan, "Application of Taguchi methods and ANOVA in optimization of process parameters for metal removal rate in electrochemical machining of Al/5% SiC composites," International Journal of Engineering Research and Applications (IJERA), **2**(3), 192–197, 2012.
[46] S.F. Sawyer, "Analysis of Variance : The Fundamental Concepts," Journal of Manual & Manipulative Therapy, **17**(2), 27–38, 2009, doi:10.1179/jmt.2009.17.2.27E.
[47] X.A. Koufteros, M.A. Vonderembse, W.J. Doll, "Developing measures of time-based manufacturing," Journal of Operations Management, **16**(1), 21–41, 1998, doi:10.1016/S0272-6963(97)00027-2.
[48] S. Arendt, M. Brettel, "Understanding the influence of corporate social responsibility on corporate identity , image , and firm performance," Management Decision, **48**(10), 1469–1492, 2010, doi:10.1108/00251741011090289.

ASTES

Advances in Science, Technology and Engineering Systems Journal Vol. 7, No. 3, 12-18 (2022)

www.astesj.com

ASTESJ

ISSN: 2415-6698

Analysis Methods and Classification Algorithms with a Novel Sentiment Classification for Arabic Text using the Lexicon-Based Approach

Bougar Marieme*, Ziyati El Houssaine

Laboratory c3s, higher School of technology, UHC, CASABLANCA, 20100, MAROC

ARTICLE INFO

Article history:
Received: 24 February, 2022
Accepted: 04 May, 2022
Online: 16 May, 2022

Keywords:
Sentiment analysis
Lexicon based Approach
Classification
Social networks

ABSTRACT

Social networks have become a valuable platform for tracking and analyzing Internet users' feelings. This analysis provides crucial information for decision-making in various areas, such as politics and marketing. In addition to this challenge and our interest in the field of big data and sentiment analysis in social networks, we have dedicated this work to combine different aspects of methods or techniques leading to the facilitation of feelings classification in social networks, including text analysis and sentiment analysis. We expose the approaches and the algorithms of supervised machine learning for the classification of feelings. We further our research to concisely present the methods of data representation and the parameters used to evaluate a sentiment analysis method in the context of social networks, with a section presenting our novel lexicon-based approach to give more accurate results in classifying Arabic text. The proposed approach has shown a promising accuracy percentage, especially the precision of the sentiment detected from text with F-Score up to 66%.

1. Introduction

Currently, the Internet allows billions of users to connect to each other, share information, communicate their ideas and opinions, and express their attitudes toward content through social networks. Thus, all actions generate high-volume, varied, high-velocity data called big social data [1].

Typically, this data is a set of opinions that can be processed to evaluate trends, audience preferences, and satisfaction related to a product, service, event, or even people.

Several areas are of interest in social networks, such as politics, health, and marketing. Because the particularity of data is unstructured texts, the data's exploitation makes text analysis an important factor for knowledge extraction and data mining.

In several of our published works, we have been particularly interested in the pre-processing of this data in the Arabic language [2], hence our interest in this research to move to the classification phase. Therefore, in this paper, we treat the analysis of textual data including the polarity of opinions (positive, negative, or neutral) and machine learning.

*Corresponding Author: Bougar Marieme, marieme.bougar7@gmail.com

2. Text mining

Textual analysis entails computer processing that extracts filtered and useful information from unstructured textual data. This analysis describes the content's structure and functions to extract patterns. This area of analysis includes techniques and algorithms, such as data mining and natural language processing. Text analysis occurs in two main actions: analyze and interpret.

2.1. The analysis phases

The analysis phase in Figure 1 consists of recognizing words, sentences, their relationships, and grammatical roles. This phase produces a standardization for the text through several methods or to automatically determine the language of a given content. We are interested in this field, and we have conducted several works in this direction especially for the Arabic language, which is recognized for its complex morphology.

The popular processes used in this phase are as follows:

https://dx.doi.org/10.25046/aj070302

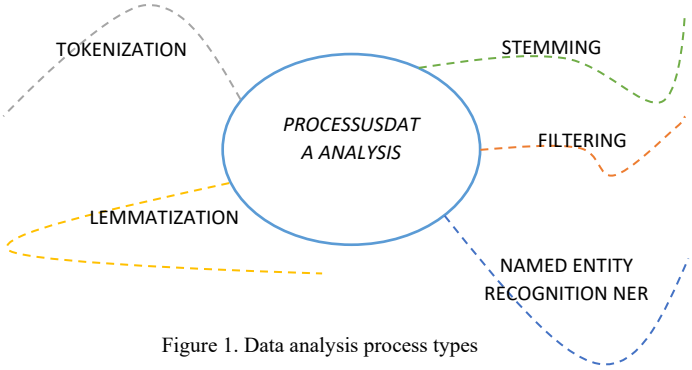


Figure 1. Data analysis process types

• Tokenization is the process of converting character strings into a list of symbols (tokens in English). Tokens are strings with meaning that eliminate spaces, punctuation, etc.
• The lemmatization operation assembles derivational variations (e.g., verbs, adjectives) or inflectional variations (e.g., plural, conjugations)
• Stemming is a process of transformation of bends or derivatives into their roots. The root of a word corresponds to the part of the word remaining once the prefixes and suffixes have been removed. There are specialized procedures for each language, for example, KHOJA Stemmer, which we treated for the stemming of the Arabic language. Stemming is especially advantageous because it is fast, it is based on precise and easy dictionaries and rules of derivation [3], and it allows for the treatment of the peculiarities of certain words.
• Filtering consists of applying filters that remove empty words
• Named entity recognition (NER) is a sub-task that extracts information from textual documents. This sub-task consists of labeling text with tags and searching for text objects that can be categorized into classes. Recognition is based on statistical systems, labeling a corpus that will serve as a learning tool, but these systems are expensive in human time.

2.2. The interpretation phases

Although data analysis is important to begin the interpretation phase to draw results and conclusions, this phase is based on data mining methods that establish reliable prediction models. This phase is a selection based on a lexical property, the presence or absence of a keyword, or other criteria. If it is a new element, we seek relationships that were not explicit between two distant elements in the text. For similarities, we try to discover texts that correspond most to a set of descriptors, such as the text's most frequent nouns and verbs.

As mentioned, the textual analysis process refers to different approaches, allowing data to be filtered, cleaned, and modeled afterward.

2.3. Sentiment analysis

Sentiment analysis appeared in the early 2000s and consists of classifying polarity into two opposing feelings, such as wanting/hating, positive/negative, or black/white.

Opinion analysis is used for different purposes and on a variety of corpora. This method has become essential in several areas, especially in marketing to evaluate a brand, analyze consumer opinions, or retain customers by detecting their emotional state and in politics to predict the results of presidential elections and the will of the public.

Classification begins with the retrieval of textual features in Figure 2:

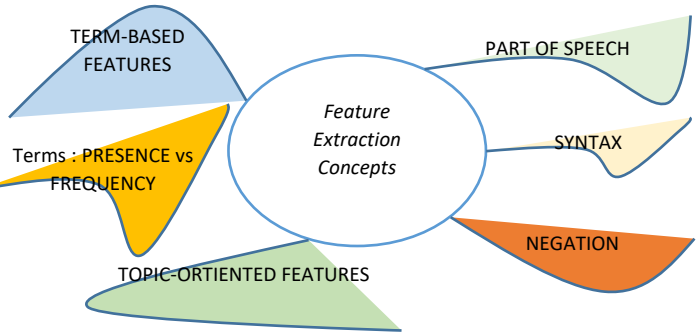


Figure 2. Feature extraction recapitulation

• Presence versus frequency of the term: According to a study by [4], the representation of a text by a vector, in which the elements indicate the existence of a term (1) or not (0) provides a better result than the frequency method (frequency of occurrence of the term) for the classification of polarity.
• Term-based features: In text, the position of a word (for example, in the middle or near the end of a document) can affect overall sentiment or subjectivity. Thus, position information is sometimes encoded in the characteristic vectors used [5].
• Part of speech analysis: This feature explains how a word is used in a sentence. There are several main parts of speech (also called word classes), including nouns, pronouns, adjectives, verbs, adverbs, prepositions, conjunctions, and interjections. Sentiment analysis by machine learning as well as by lexicon has an attraction toward adjectives [6].
• Syntax: This feature refers to the integration of syntactic relationships in the analysis and seems particularly relevant with short texts.
• Negation: In the bag-of-words method, for example, negation is not considered. The phrases "I hate" and "I like" are considered similar, hence the interest in the treatment of negation.
• Topic-oriented features: These features are the interaction between topic and opinion.

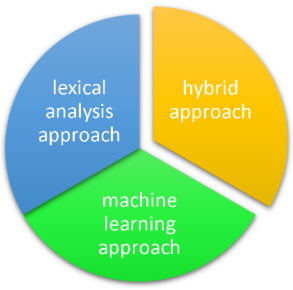


Figure 3. Sentiment analysis categories

There are three main categories of sentiment analysis as shown in Figure 3: a lexical analysis approach, a machine learning approach, and a hybrid approach that combines the first two.

3. Related work

3.1. Lexicon based approaches

Opinion extraction approaches based on lexical analysis consist of extracting the polarity of a sentence using a semantic analysis of words. Thus, a sentence is classified by its instances (words of opinion) for which emotions are already attributed. In the literature, words of opinion are also known as polar words or words carrying opinion. Positive opinion words are used to express certain wanted states, while negatives are used to express unwanted one. Examples of positive opinion words are good, genius, and appreciable. Examples of negative opinion words are scary, horrible, and mediocre.

In lexical analysis, the input text is converted into tokens. If the token has a positive match, its score is added to the total score of the input text. For example, if the word "perfect" has a positive match in the list of opinion words, the total score of the text is incremented by the associated weight. Otherwise, the score is decremented by the same amount when the word is labeled as negative.

To generate the list of opinion words, there are three main approaches resumed in Figure 4: a manual approach, a dictionary-based approach, and a corpus-based approach.

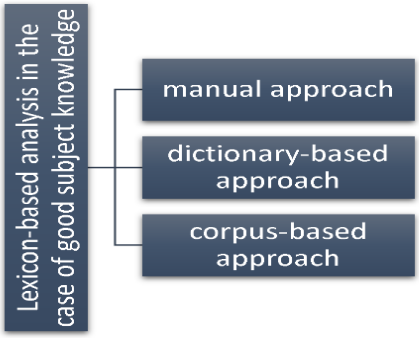


Figure 4. Lexical analysis approaches

• Manual approach: This method is precise, but it is time-consuming so is generally not used alone; instead, it is combined with automated approaches as a final verification.

• Dictionary-based approach: In this approach, a small set of annotated opinion words is collected manually and then developed by searching for their synonyms and antonyms in a dictionary. Newly found words are added to the starting list. This process is iterative and stops when no new words are found. Once the process is complete, a manual inspection is performed to correct any errors. Several researchers [7] have used this approach and generated lists of opinion words.

• Corpus-based approach: This method uses dictionaries to annotate words as well as the context for which polarity is valid. This approach begins with a list of words of opinion that is then expanded based on a large corpus. In [8] authors proposed a "sentiment coherence" technique that begins with a list of opinion adjectives and identifies additional adjective opinion words and their orientations using a set of linguistic constraints or sentence connections (e.g., OR, BUT NI). Another usual constraint is the conjunction "AND" indicates association with two similar orientations. For example, in the sentence "This man is brave and kind", if "brave" is known to be positive, "kind" is also positive because people generally express the same opinion on both sides of a conjuncture. The following sentence is rather unnatural: "This man is brave and authoritarian." If this sentence is changed to "This man is brave but authoritarian," it becomes acceptable. Learning is applied to a large corpus to determine whether two adjectives in the same sentence ("conjoined adjectives") have the same or different orientations. In practice, this is not always consistent. Indeed, in [9] authors demonstrated that the same word could indicate different orientations in different contexts, even in the same field. For example, in the laptop field, the word "long" expresses opposing opinions in these sentences: "Battery life is long" (positive) and "Startup time is long" (negative). Therefore, the generation of opinion words according to the domain becomes insufficient. For this issue, they suggest considering both the possible words of opinion and the aspects: use the couple (aspect, opinion word) as a context of opinion.

To link our contribution with the presentation of the aim methods of sentiment analysis, we present in the following the most relevant work of the SA employing the social networks and which are close to our contribution.

The method of [8] proposed a system based on corpus that retrieve automatically positive and negative semantic information using indirect information from a large corpus of adjectives. The system of prediction is based on linear regression, achieving 92% accuracy for the classification.

In [10] authors propose model based on the Arabic corpus created by hand where they annotated dataset and give steps to build their lexicon. They noted that more the lexicon is rich the performance is high. Authors propose approaches based on the Arabic corpus and on the lexicon, they created by hand an annotated dataset and they prove that SVM (Sector Vector Machine) method used for classifying text give the highest precision.

In [11], the authors used special approach based on correlation and indication of emotional signals and was tested on sets of twitter data. The results have shown the effectiveness of their approach and the importance of including the different emoticons in their analysis.

In [12], the authors developed an unsupervised approach to automate non-concatenative morphology, which they apply to generate a standard Arabic lexicon of roots. They use a recursive

notation based on hypothesized patterns and root frequencies. Their morphological analysis with the induced lexicon fulfils a root identification accuracy to 95%.

In [13], the authors developed classification model for the Egyptian dialect using Arabic tweets that are analyzed to identify their polarity (positive or negative), using Maching Learning approach that uses supervised classifiers and the semantic approach that requires the construction of a lexicon.

3.2. Machine learning-based approaches

In these approaches, the machine is trained to detect patterns in a corpus by having it learn on a first test corpus. In machine learning, the machine learns from data collected in the past, like human learned from their past experiences in real-world applications.

Machine learning involves five steps Figure 5: data collection, preprocessing, training, classification, and results.

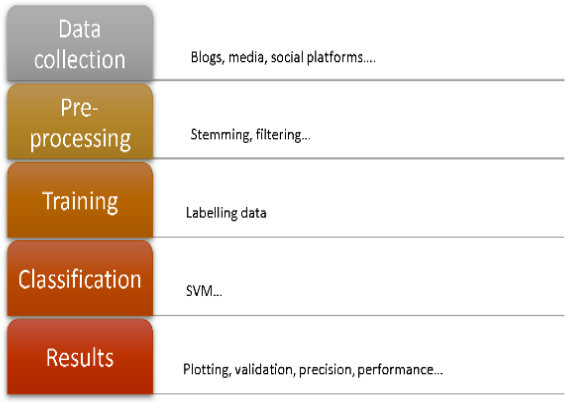


Figure 5. Machine learning steps

• Data collection: The data to be analyzed is collected from various sources, depending on the need of the application, the field, and the context studied.

• Pre-processing: The collected data is cleaned and prepared to be entered into the classifier. Pre-processing is a crucial step and has a direct impact on the quality of the classification operation. Textual data cleansing is completed in several steps and includes tokenization, stemming, or filtering.

• Training: This step consists of labeling a collection of data by hand to generate the training data. The most used method is crowdsourcing. This data is entered into the chosen algorithm for learning purposes.

• Classification: The classifier is trained to detect patterns or patterns in the corpus based on descriptors explained later. After completing the training and building the forecast model, the classifier is deployed on the new data to extract feelings.

• Results: The results are plotted according to the type of representation selected. Then, the classifier's performance is measured according to several methods, including accuracy, recall, F-score, and cross-validation, discussed later.

3.3. The evaluation of the classifier

As mentioned, once a classifier is chosen and built, we must evaluate it to measure its performance and accuracy. This is an important step for any classification project. There are several methods and measures to evaluate a classifier as shown in Figure 6, but the accuracy of the classification remains the main measure. This measure represents the number of documents in the test set that are properly classified, divided by the total number of documents in the test set. Next, we present other metrics and methods commonly used for the evaluation of classifiers.

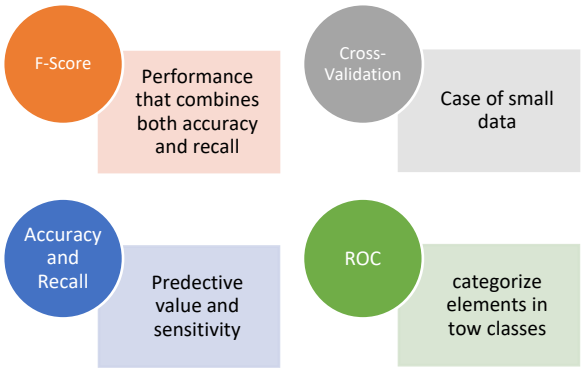


Figure 6. Methods to monitor the performance of a classifier

a) Cross-validation

This method is used especially when the data set is small. The goal of cross-validation is to define a set of data to test the model in the learning phase. In this evaluation method, the available data is partitioned into n disjoint subsets of equal size. Each subset is then used as a test set and the remaining n-1 subsets are combined as a learning set. This procedure is then performed n times, which gives n precision. The estimated final accuracy of learning from this data set is the average of the n precisions. In general, cross-validations 10 and 5 are most used. Cross-validation can also be used for parameter estimation [14].

b) Accuracy and recall

Accuracy and recall are two criteria for statistical measures evaluating classifiers, also known as predictive value and sensitivity. We note VP, which is the number of items correctly labeled positive (true positive), as well as FN, or the number of incorrect classifications of positive examples (false negatives); FP, or the number of items that were incorrectly labeled positive (false positives); and TN, the number of correct classifications of negative examples (true negative)

In an opinion classification task, the precision p of a class is the number of true positives divided by the total number of positive categorized elements:

$$p = VP/VP+FP \quad (1)$$

The recall r in this context is defined as the number of true positives divided by the total number of elements that belong to the positive class:

$$r = VP/VP+FN \quad (2)$$

An accuracy score of 1 for a class C means that each element associated with class C belongs to that class.

However, this score says nothing about the number of Class C items that have not been properly labeled).

A recall score means that each item belonging to Class C has been correctly labeled (but this score says nothing about the number of items that have been incorrectly associated with Class C). There is an inverse relationship between accuracy and recall: one may be increased at the expense of the other.

c) F-score

F-score is a popular measure of test performance that combines both accuracy and recall. This method is often used to compare different classifiers with a single measure.

$$F = 2 \times (p \times r / p + r) \quad (10)$$

F-score, also called F1-score or F-measure, is the weighted harmonic mean of accuracy and recall:

$$F = 2 / ((1/p) + 1/r) \quad (4)$$

d) The efficiency function of the receiver

The receiver efficiency function, more commonly referred to as the ROC curve or sensitivity/specificity curve, is a performance measure of binary classifiers (systems that categorize elements into two distinct classes). Graphically, this measure is represented by the rate of true positives compared to the rate of false positives. The true TPP positive rate refers to the fraction of positives detected and the TFP false positive rate refers to the fraction of negatives incorrectly detected.

$$TVP = VP/(VP+FN) \text{ and } TFP = FP/(VN+FP) \quad (5)$$

DVT is the reminder of the positive class and is also called sensitivity. Another measure, called specificity, represents the rate of true negative (TVN), or the recall of the negative class. TVN is defined as follows:

$$TVN = VN/(VN+FP) \quad (6)$$

An ROC space is defined by the axes x and y referring to TVP and TFP r, respectively, this representation demonstrates a state between true positive and false positive. Each prediction result is represented by a point in the ROC space.

In Figure 7 the points above the diagonal line represent accurate classification results and the points below the line represent poor results. A perfect classification would give the coordinate (0.1) in the ROC space, representing 100% sensitivity (no false negatives) and 100% specificity (no false positives) [15].

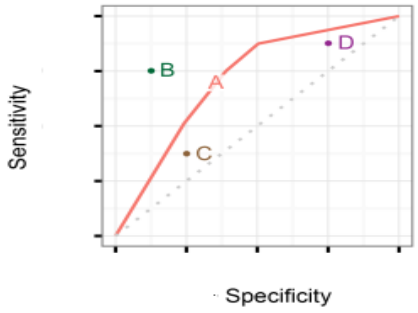


Figure 7. ROC space

4. Novel sentiment classification for Arabic text using the lexicon-based approach

The task of classifying feelings involves labeling the text with a sentiment class. There are two main families of approaches: supervised machine learning and lexicon based. Naturally, supervised methods use machine learning algorithms trained with labeled data (positive, negative, or neutral).

In this section, we propose solutions to overcome the difficulties and limitations related to the exploration of opinion in terms of contextual semantic orientation and adaptability. Adopting a lexicon-based methodology, we present a new adaptable approach that allows one to associate a polarity to words according to context through the construction of dictionaries based on instantiation rules. Our aim is to improve the finesse of the classification of feelings of an Arabic text.

4.1. Problematic

The Problematic of approaches based on lexical analysis generally involve the aggregation of polarity scores from generic repositories to classify text. These approaches are more flexible and therefore more suitable for the classification of feelings in the context of big social data, especially for morphologically complete languages such as Arabic.

Nevertheless, they themselves face challenges such as defining the semantic orientation of words that could be strongly influenced by the context, a word can be considered a negative word in a tweet and at the same time considered a positive word if the tweet is related to a different context. As a result, approaches based on lexical analysis do not give very good results if they are not contextualized. In addition, lexicon-based approaches and/or machine learning do not consider the informality of messages published in social networks. Indeed, these messages could contain special words such as those written in a repetitive or extended word. These special words can be used to weight the emotional load of posts. Therefore, they could be considered as intensifiers or diminutives of polarity.

In this paper, we seek to propose solutions to overcome the difficulties and limitations associated with the exploration of opinion facing semantic orientation presenting constraint. Based on lexicon-based approach, we propose a new approach that allows words to be assigned polarity based on context by creating lexicons assigning polarity, with extended rules to analyze a maximum of semantic orientation, improving the finesse of the classification of feelings of a text, and we focus on this work the sentiments expressed in standard Arabic.

To build a sentiment analysis model, we propose a methodology composed of three steps, as illustrated in Figure 8, which is detailed in the following.

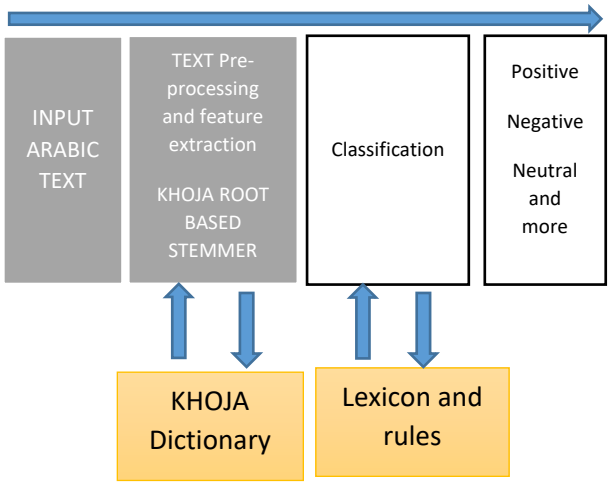


Figure 8. Novel lexicon-based approach

4.2. Preparing DATA

Extraction of data set: We use to scrap specific Tweets replies with Python to handle the verification of our result. The comments are written in Standard Arabic and about given opinion of a specific cosmetic product.

As a tweet could include more than one hashtag, the tweet could be extracted several times. Thus, we deleted duplicate tweets to avoid overweight.

Preprocessing Data: As illustrated in Table 1, this step consists first of translating the different words composing the comments in Arabic text pasted in their roots. To perform this step, we chose KHOJA, an algorithm for preprocessing data based on the roots of the words, we used KHOJA because it is an algorithm that we have perform in many previews work [16], it shows its effectiveness in term of accuracy, available and simple to implement. This algorithm processes the words under the following major steps:

- Tokenization: the process of converting character strings into a list of tokens. Tokens are strings with an assigned and identified meaning. Tokenization consists of eliminating "noises" from the source text, including comments, white space, and punctuation.
- Normalizing
- Stop word removal
- Root stemming: a process of transformation of bends (or sometimes derivatives) into a base or root form.

Table 1. Data preprocessing

Preprocessing step	Output
Basic text	من أحسن المنتوجات, أكثر من رااائع (One of the best products, the most amazing)
Tokenization	من-أحسن-المنتوجات-أكثر-من- رااائع
Normalizing	من- أحسن -المنتوجات-أكثر- من- رائع
Remove stop word	أحسن -المنتوجات -أكثر- رائع
Stemming	احسن -منتوج - اكثر- رائع

4.3. Lexicon dictionary Construction

Lexicon-based sentiment analysis involves extracting sentiment scores from a dictionary. The polarity of a position is therefore calculated by adding the score values {+1, -1} of the words.

To construct our lexicon dictionary, we associate each word to a weight (polarity). If the word does not exist in the lexicon its polarity is null. For feature construction, and to calculate the global polarity of the text, it suffices to aggregate the polarity score of each word constituting the text from the lexicon.

Our method is based on an ordinary polarity, which is founded on rules considering the intensity of the word appearing in the comment in its weight (positive, negative, or neutral), for example:

- Negation: a negative value is assigned to words that inverse the polarity of the word for example 1 turn to -1.
- Intensification: we define specific word that give unadded accent we add +1 to a positive word and −1 to a negative one. For example, the sentence (This cream is very smooth), has a polarity equals to 2 1 refers to the positive word smooth and +1 to the presence of the word very.
- Conflicting phrases: when two words with opposite polarity in the same comment the negative polarity reign.

The objective of this rule is to enhance performance to decide on our classification. Then, we present the results of our rankings according to several classes as follows.

We classified the tweets rated in seven classes according to their degree of polarity, limiting ourselves to this distinction according to the high weight result of polarity detected throw an empirical test on a sample of 1200 tweets to determine the limits of each classroom:

- Highly satisfied: polarity ≥ 7.
- Moderately satisfied: $4 \leq$ polarity ≤ 6.
- Slightly satisfied: $0 <$ polarity ≤ 3.
- Slightly unsatisfied: $-3 \leq$ polarity < 0.
- Moderately unsatisfied: $-6 \leq$ polarity ≤ 4.
- Strong negative: polarity ≤ -7
- If 0, the comment is considered neutral.

4.4. Results

In this stage, we compared results of the same data set using simple polarity and polarity with rules that give the ordinal scale with rules application. In the total text, we used 5000 words in Arabic text: 2000 positives and 3000 negatives. These are opinions of customer satisfaction on a cosmetic product.

Table 2. special words impacting polarity.

Rules	Word
Negation	لا ، لم ، ليس
Intensity	كثيرا ،جدا، قوي
Contradictory	غامض، عجيب

We present in Table 2 some specific words figuring in our training data set and make impact to word polarity.

Table 3. Examples of global polarity

Comment	Binary scale	Ordinal scale and rules
هدا منتوج رائع جدا (This is a great product)	+2	+6
صراحة لم يعجبني مطلقا (I don't like it at all)	-1	-3

Finally, we compared results of the same data set using simple polarity and polarity with rules that give ordinal scale with rules application.

Table 4. Comparing result of binary and ordinal scale with rules: accuracy and F-Score percentage

Method	Accuracy	F-Score
Binary scale	67.30%	61.11%
Ordinal scale and rules	69.08%	**66.30%**

Regarding to our related work [13] implemented sentiment classification for Arabic tweets and obtained the F-score about 65.4% and [14] their F-score obtained was equal to 59.6%.

Even if our F-score result is more interesting which exceeds 66% as shown in Table 4, the result of our training requires improvements, to make more exceptional and efficient output.

5. Conclusion

Currently, the main goal of applications using big social data is to make a machine able to identify emotions and feelings in various areas in real time. The analysis of big social data therefore creates many challenges, such as adaptability and processing of big data in real time.

The first objective of this work was to study and compare the different big data tools available and choose the appropriate tools to study and classify the nature of the data.

The second objective was to establish a methodology based on the lexicon. We present a new adaptable approach that allows assigning polarity to words based on context taking semantic constraint into account, through the construction of dictionaries based on rules. This method has remarkably improved the finesse of the results obtained.

6. Perspectives

The challenge that we face in our approach is to detect a ironic comment that can impact the performance of our classification process, so that we have to enrich our lexicon to consider Arabic words with both polarities (positive or negative).

References

[1] J. Ishwarappa, A.Anuradha, "Brief Introduction On Big Data 5vs Characteristics And Hadoop Technology" Procedia Computer Science, **48**, 319-324, 2015.

[2] M. Bougar, E. Ziyati, "Stemming Algorithm For Arabic Text Using Parallel Data Processing" Revue Méditerranéenne Des Télécommunications , **7**(2), 2017.

[3] S. Khoja, R.Garside, "Stemming Arabic Text. Computing Department," In 1999 Lancaster University, Lancaster, http://Zeus.Cs.Pacificu.Edu/Shereen/Research.Htm.

[4] B. Pang, L.Lee, Vaithyanathan, Shivakumar, "Thumbs Up? Sentiment Classification Using Machine Learning Techniques," Proceedings Of Emnlp, 2002.

[5] S. Kim, E.Hovy, "Extracting Opinions, Opinion Holders, And Topics Expressed In Online News Media Text," In Proceedings of the Workshop on Sentiment and Subjectivity in Text, 1-8. 2006. DOI: 10.3115/1654641.1654642

[6] V. Hatzivassiloglou, W. Janyce, "Effects of Adjective Orientation And Gradability On Sentence Subjectivity," In COLING 2000 Volume 1: The 18th International Conference on Computational Linguistics. 2000.

[7] A. Esuli, F.Sebastiani, "Sentiwordnet: A Publicly Available Lexical Resource For Opinion Mining, Proceedings Of The Fifth International Conference On Language Resources And Evaluation" (Lrec'06) May, 2006.

[8] V. Hatzivassiloglou, K.R. Mckeown, "Predicting The Semantic Orientation Of Adjectives," Proceedings Of The 8th Conference On European Chapter Of The Association For Computational Linguistics Madrid, Spain, 174-181, 1997.

[9] X. Ding, Xiaowen, B.Liu, P.Yu, "A Holistic Lexicon-Based Approach To Opinion Mining. Wsdm'08," Proceedings Of The 2008 International Conference On Web Search And Data Mining, 231-240, 2008, 10.1145/1341531.1341561.

[10] N. Bdulla, A.Ahmed, M.Shehab, M.Al-Ayyoub, "Arabic Sentiment Analysis: Lexicon-Based And Corpus-Based," In 2013 Ieee Jordan Conference On Applied Electrical Engineering And Computing Technologies (Aeect), Pp 1–6. Ieee, 2013.

[11] X. Hu, J.Tang, H.Gao, H.Liuunsupervised, "Sentiment Analysis With Emotional Signals,". Proceedings Of The 22nd International Conference On World Wide Web, 607–618, 2013.

[12] B. Khaliq, J.Carroll, "Induction Of Root And Pattern Lexicon For Unsupervised Morphological Analysis Of Arabic," 2013.

[13] A. Shoukry, A. Rafea, Sentence-Level Arabic Sentiment Analysis. In: International Conference On Collaboration Technologies And Systems (Cts), 546–550. Ieee. 2012

[14] L. Bing; " Data-Centric Systems And Applications Series Editors M.J. Carey S. Ceri Editorial Board P. Bernstein U. Dayal C. Falout." Https://Epdf.Tips/Web-Data-Mining-2nd-Edition-Exploring-Hyperlinks-Contents-And-Usage-Data.Html

[15] H. Mohsen, "Knowledge Discovery Considering Domain Literature And Ontologies : Application To Rare Diseases;" Computation And Language [Cs.Cl]. Université De Lorraine, English, Nnt: 2017lorr0092 Tel-01678860v2f, 2017.

[16] M. Bougar, E.Ziyati, (2019). Stemming Algorithm For Arabic Text Using A Parallel Data Processing: Icict 2018, London. 10.1007/978-981-13-1165-9_23.

ASTES

Advances in Science, Technology and Engineering Systems Journal
Vol. 7, No. 3, 19-31 (2022)
www.astesj.com
Special Issue on Multidisciplinary Sciences and Engineering

ASTES Journal
ISSN: 2415-6698

Towards a Model-based and Variant-oriented Development of a System of Systems

Sylvia Melzer[*1,2], Stefan Thiemann[3], Hagen Peukert[3], Ralf Möller[1]

[1] *Universität zu Lübeck, Institute of Information Systems, Lübeck, 23562, Germany*

[2] *Universität Hamburg, Centre for the Study of Manuscript Cultures, Warburgstraße 26, 20354 Hamburg, Germany*

[3] *Universität Hamburg, Center for Sustainable Research Data Management, Monetastraße 4, 20146 Hamburg, Germany*

ARTICLE INFO

Article history:
Received: 15 February, 2022
Accepted: 14 May, 2022
Online: 25 May, 2022

Keywords:
System of Systems
SysML
Information System
Variants

ABSTRACT

The development of an aggregated system consisting of autonomously developed components is usually implemented as a self-contained unit. If such an aggregation is understood as a system of systems (SoS) that communicates via interfaces with its autonomous subsystems and components, the interfaces and communication exchange should play a central role in the architectural design. In fact, complete and exact interface specifications simplify loose coupling of independent systems into an aggregation. Since an SoS consists of variant and non-variant subsystems, the main challenge in SoS development is the identification of all true variants and its deviating attributes within an SoS. If the system variants are identified at an early stage of the development process, redundant work in the interface design can be substantially reduced. This paper presents an efficient method to identify SoS variants with regard to life cycle management and it shows how to configure a variant-oriented SoS with a standardized communication interface. For the development, the forward-looking model-based systems engineering approach is recommended to create executable specification parts and to detect errors early on through simulations.

1 Introduction

The *Centre for the Study of Manuscript Cultures* (CSMC) at Universität Hamburg hosts a steadily growing number of autonomously developed database systems, e.g., for the projects *Epigraphic database of ancient Asia Minor* (EDAK) (https://www.epigraphik.uni-hamburg.de), *Going From Hand to Hand: Networks of Intellectual Exchange in the Tamil Learned Traditions* (NETamil) (https://www.manuscript-cultures.uni-hamburg.de/netamil/), and *Thesaurus Defixionum* (TheDefix) (www.thedefix.uni-hamburg.de). As a first priority requirement, the database schema reflect the high data variety of these research projects while maintaining the same overall structure. If these databases are now combined into an aggregated information system, new functionalities that are designed to the overall structure and not to the peculiarities of each schema can be defined as it is the case for federated searches. In addition, one can very well imagine that new database systems would like to connect to the aggregated information system later on as long as the structure remains clear.

An illustration for the usefulness of an aggregated information system are trove discoveries, whose associated parts, for some reason, are scattered at different places in the world as it often happens for old manuscripts. Such script fragments are administered in different information systems. As an example, one fragment AO 29196 [1] is located at the Louvre and the counterpart of this fragment, KUG 15 [2], is located in Germany. Indeed, both fragments were discovered without using federated search queries, but for analyzing data from different databases it would be desirable to find related data in an aggregated information system. This example highlights the need to combine, analyze, and query data from different database systems.

The requirement for the development of an aggregated system is, on the one hand, to develop autonomous systems in such a way that the variant parts are not implemented redundantly and, on the other hand, that external databases can be added to the aggregated system without much effort.

A systematic approach to model variant parts was developed at the *Institute of Product Development and Mechanical Engineering Design* (PKT) at the Hamburg University of Technology. The variant-oriented approach is called *integrated PKT* approach (see

[*]Corresponding Author: Sylvia Melzer, Universität Hamburg, Centre for the Study of Manuscript Cultures, Warburgstraße 26, 20354 Hamburg, Germany, sylvia.melzer@uni-hamburg.de

[3]). The integrated PKT approach aims at satisfying a wide variety of customer requirements while developing a component. Ideally, the approach applies to a product family that is developed within one organization, for which the marketed products are supposed to have as few variants as possible. Also, adding an external sub-product to the product family at a later date must be considered very early on in the planning phase. Early consideration of coupling systems or products can lead to modularization of systems to support the approach. The integrated PKT approach also supports modularization. Nevertheless, it is not always possible to extend the product family by a new variant. Another core idea of the integrated PKT approach, besides the variant-oriented development of products, is the combination of the database into an SoS and keeping the main focus on the development of a communication interface. While there are some good approaches to SoS development already, an approach that considers system variants during development of an SoS is to the best of our knowledge not yet available.

For the development of an SoS, model-based methods using the *Systems Modeling Language* (SysML) are increasingly used. Users also benefit from the general *Model-Based Systems Engineering* (MBSE) advantages, such as making complexity manageable. The first model-based SoS developing methods are evolving, cf. [4] as well as [5].

In this paper, we present a provident, model-based, and variant-oriented approach to develop new functions for an aggregated information system so that all functions can be used simultaneously to the benefit of all database systems.

The paper is structured as follows. In Section 2 and Section 3 we give an overview on related work and preliminaries for developing a model-based and variant-oriented SoS. In Section 4 we describe how to develop variant-oriented and sustainable information systems such that as many customer requirements as possible are considered while increasing the number of variants and reducing the necessity of redundant information system development.

In Section 5 we present, first, how variant and SoS relevant requirements are elicited during the requirements engineering process, and second, how to design the structure of an SoS. In Section 6 we describe modeling and simulating the systems' behavior to execute a federated search as an example. The application of our new approach and its results are presented in Section 7. We like to close in Section 8 with a summary and a preview of future work.

2 Related Work

Despite its frequent usage, there is little agreement nowadays on a concise and general definition of the term *system of systems* (SoS). Some approaches of SoS distinguish between SoS and traditional systems. These approaches elaborate specifically on the heuristics of SoS development. They also emphasize the differences to traditional systems. More particularly, it has been noted that the architecture of an SoS aims at optimal communication for the vast majority of all SoS (see e.g. [4, 6, 7]).

The application in [4] illustrates the development of a *Trusted Forwarder System* (TFS) for a secured air cargo transport chain as an SoS using a set of standards that enable useful communication between existing and newly developed components. For the TFS, a communication standard is used that satisfies the new requirements. Thus, after SoS integration, the systems and components are enabled to follow their original tasks without diversion.

In [7], the authors show how, from autonomously developed database systems, an aggregated information system enables relatively simple data exchange through a standardized communication interface. If the individual systems are combined into a federated system via a communication interface, new functions can be added easily, such as the federated search functionality. With the new search function, the individual systems can perform searches in all databases (as opposed to only one database) provided that access rights are correctly assigned. The new search feature can be considered as a new service relevant for a broader range of users.

Both aggregated systems, the TFS and the aggregated information system, were initially developed as single systems, cf. [8, 9]. The single system and the SoS development of the TFS were compared in [4]. It was observed that the SoS development approach mainly is advantageous for traceability beyond the system perspective to the service. The advantages of re-usability of a system, which was developed as SoS, are plain to see: Communication interfaces give more flexibility to add new functionalities or remove subsystems from the overall system.

In [6], the author recommends a stable architectural design for SoS. Such stability can be achieved by admitting independently, i.e. autonomously, developed systems in the architectural design together with a communication interface.

Model-based approaches, such as the *Variant Modeling with SysML* (VAMOS) presented in [10], pure::variant (`https://www.pure-systems.com/purevariants`), the *Variety Allocation Model* (VAM) (variant-oriented developing process of the integrated PKT approach) with SysML [11], exist to identify possible variants in the early phase of system development. The methods VAMOS and VAM with SysML can be represented in the SysML modeling tool *Cameo Systems Modeler* by extending the language elements. Pure::variant can be used as a stand-alone entity for variant modeling. In this paper we decided for VAMOS, since it was already applied in [7] for the development of a cross-domain information system.

Cameo Systems Modeler and the broker-based SysML Toolbox have been successfully used for simple modeling of communication networks in several projects such as SiLuFra [12], ConCabInO [13], KomKab [14], and KMUDigital [15].

3 Preliminaries

This section describes the languages, methods, and tools proposed for a model-based and variant-oriented development of an SoS.

3.1 Modeling Languages

According to the recommendations of MBSE, systems are described or documented using semi-formal modeling languages such as the *Unified Modeling Language* (UML) or the SysML.

Systems Modeling Language SysML was specified by the *Object Management Group* (OMG) to support the model-based devel-

opment of complex systems during the system development process. SysML is a subset of the standardized language UML 2 including some additional extensions. A SysML model can be used to describe the structure as well as the behavior of a system, and can be used to simulate the behavior of systems. In this paper, the focus of variant modeling is on structure. For variant behavior modeling, further challenges are to be expected (cf. [16]), which should be addressed separately due to the complexity of SoS development.

Variant Modeling with SysML A variant is characterized by a base model and differentiating parts, where the base model represents the core of the system and the differentiating parts represent the distinctions of the system components (see [17]). In [10], the author specified VAMOS to model variants with SysML. To use this, the existing model elements in SysML are extended (see Figure 1). The two model elements *Package* and *NamedElement* are extended with the stereotypes *Variant*, *Variation*, *VariationPoint*, and *VariationElement*. The *Variation* is a stereotype of the *Metaclass Package*, which contains all the elements of an option of *Variation*. A *Variation* is also a stereotype of the *Metaclass Package*, which contains multiple variation packages. A *VariationPoint* is a stereotype of the *Metaclass NamedElement*.

VAMOS is suitable for systems with some variability. Furthermore, VAMOS is applicable for the use of structural elements. For the description of variant system behavior, VAMOS can be applied conceptually. Practical applications usually exploit extensions of further SysML language elements related to the behavior necessitating the variant behavior description (see [16]).

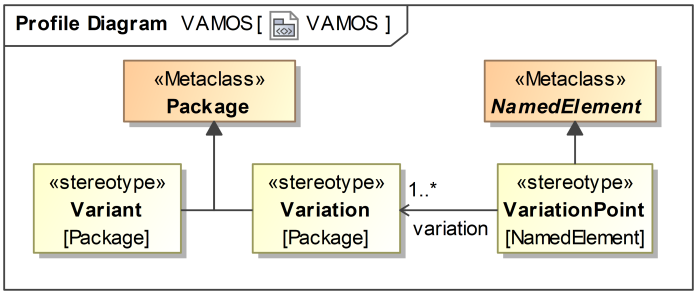


Figure 1: Profile diagram: Variant Modeling with SysML (VAMOS) for the development of a cross-domain information system

3.2 Methods

Context-Based Requirements Engineering The identification of variants should be done as early as possible, so it is necessary that views of all stakeholders involved are considered during the requirements engineering process. The associated systems have to be identified during the SoS context description process. In [18], the authors have defined the *Approach for Context-Based Requirements Engineering* (ACRE) ontology with the goal to capture the requirements of all stakeholders and manage them during the entire system development process. A fundamental approach of the ACRE ontology is the context, since it emphasizes and defines the view of use cases and requirements. Depending on the use cases, some new model-based systems engineering approaches add specific contexts defined for system or product development, e.g. the life phase modularization context in [8], the variety context in [11], and the system of systems context in [19]. In [20, 21], a supplemented ACRE ontology with the SoS aspect are presented and also introduce new terms for system development. Adding the variety context to the SoS approach results in a new variant-oriented approach, which is described in this paper.

V-model For all IT projects in the federal public administration, the V-model is a mandatory procedural standard. The processes of the V-model, inspired by the V-model XT (see www.v-modell-xt.de), can be described as follows: analysis of requirements, functional analysis, high-level design, low-level design, implementation, component test, system test, integration test, and acceptance test. Verification and validation also belongs to the processes.

We argue that the V-model is a good basis for system development, so we have used this approach to develop information systems. For a variant-oriented development and implementation of a communication interface, it is important to consider in the individual process steps like the high-level design, that, among other properties, combined approaches are used to develop the SoS efficiently and correctly. Approaches for the development of an SoS are described in Subsection 5.2.

Broker Federation The brokerage network enables the creation of message routing networks, in which messages in one broker are automatically routed to another broker. These routes may be defined, e.g., between exchanges in the source and destination brokers, or from a message queue in the source broker to an exchange in the destination broker [22]. The principle of coupling systems via a broker federation is a practically proven approach that is used in many applications. In this paper, broker federation is used to create a communication interface between the systems to develop the SoS.

3.3 Tools

Communication Tool The open-source message broker RabbitMQ (https://www.rabbitmq.com/) can be used to create communication networks. RabbitMQ uses the *Advanced Messaging Queuing Protocol* (AMQP) as a standardized communication technology. AMQP defines three components which are essential to implement a message-based architecture. 1) The *message queue* stores messages which can be consumed by client applications. 2) The *exchange* receives messages from publisher applications and routes these to message queues. 3) The *binding* defines a relationship between a message queue and an exchange. Using these components, classic communication paradigms can be implemented and used such as 1) send and receive, 2) work queues, 3) publish and subscribe, 4) routing, 5) topics, and 6) request and reply.

In [4, 7, 9, 15], the authors show that the developed communication interfaces with RabbitMQ can be used for implementing real software or hardware in the model with little effort. For this reason, we choose RabbitMQ to support a communication interface for the individual systems that become part of the SoS.

Modeling and Simulation Tool Cameo Systems Modeler (version 2021x) is a modeling and simulation tool that was originally

developed specifically for the development of systems using the SysML. To simulate behavioral diagrams, Cameo Systems Modeler uses a subset of the UML elements on the OMG *Foundation Subset for Executable Models* (fUML) and *W3C State Chart XML* (SCXML) standards. The broker-based SysML Toolbox is an extension of the Cameo Systems Modeler and provides the integration of real software and hardware [15]. The Toolbox also offers predefined SysML elements that can be used to create database interactions. The SysML Toolbox contains an implementation of these six messaging paradigms. These communication paradigms are implemented via the SysML element *opaque action* and the usage of the Java-like scripting language *BeanShell*.

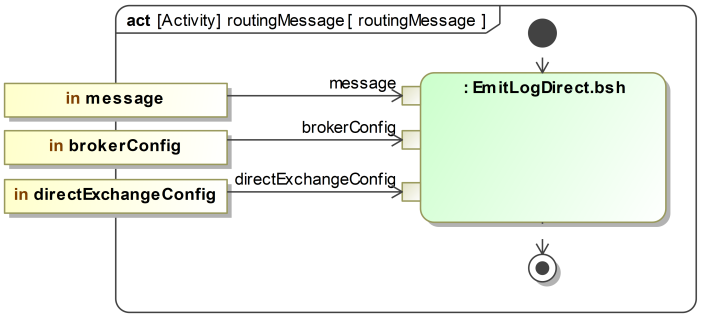


Figure 2: The opaque action *EmitLogDirect.bsh* for sending a routing message to the RabbitMQ server

An implementation of the routing communication paradigm as an opaque action element and the respective BeanShell code are presented in Figure 2 and in Figure 3. The opaque action is called *EmitLogDirect.bsh*. The other paradigms are also available as opaque actions. These opaque actions are implemented as drag-and-drop communication elements, with the aim to increase efficiency and to avoid coding effort.

Language:

BeanShell

Body:

EmitLogDirect.bsh(message, brokerConfig, directExchangeConfig)

```
18  try
19  {
20      ConnectionFactory factory = new ConnectionFactory();
21      factory.setHost(host);
22      factory.setVirtualHost(virtualHost);
23      factory.setPort(port);
24      factory.setUsername(userName);
25      factory.setPassword(pw);
26      Connection connection = factory.newConnection();
27      Channel channel = connection.createChannel();
28
29      try {
30         channel.exchangeDeclare(EXCHANGE_NAME, "direct", true);
31      } catch (IOException e1) {
32         // TODO Auto-generated catch block
33         e1.printStackTrace();
34      }
35
36      String severity = routingKey;
37
38      channel.basicPublish(EXCHANGE_NAME, severity, null, message.getBytes("UTF-8"));
39      print(" [x] Sent :'" + message + "'");
40
41      channel.close();
42      connection.close();
```

Figure 3: The source code for sending a routing message to the RabbitMQ server

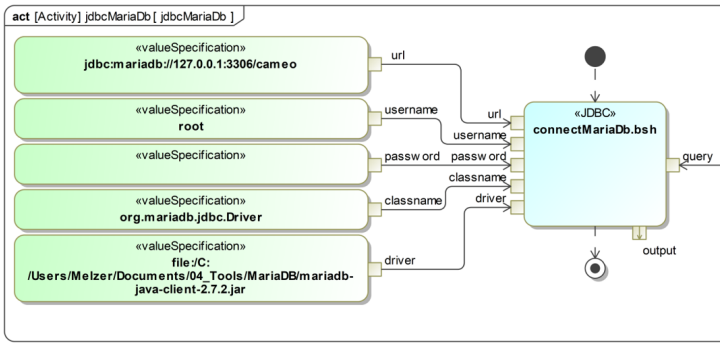


Figure 4: The opaque action *connectMariaDb.bsh* for sending a message to the database MariaDB, source: [23]

The following SysML blocks are used to define input and output parameters: *MessageBroker*, *MessageQueue*, and *MessageExchange*. The *MessageBroker* contains the properties: *host*, *virtualHost*, *port*, *username*, and *password*, which are input parameters for the opaque behavior *EmitLogDirect.bsh*. The properties of the MessageExchange are *exchangeName* and *routing key*. In order to set individual configurations, it is possible to create instances of the SysML blocks. An instance of the *MessageBroker* is *brokerConfig*. An instance of the *MessageExchange* is *directExchangeConfig* (see Figure 2). More details of the broker-based SysML Toolbox are given in [15].

For modeling database expressions, the extension of the broker-based SysML Toolbox can be used or replicated. The predefined database expressions for creating, manipulating, and querying databases are implemented as opaque actions. These predefined actions can also be used as drag-and-drop elements (cf. [23]).

Language:

BeanShell

Body:

connectMariaDb.bsh(url, username, password, classname, driver, query, output)

```
20  import org.mariadb.jdbc.Driver;
21
22  ConnectionMariaDB c = new ConnectionMariaDB();
23  conn = c.connectDb(classname, url, username, password, driver);
24
25  if(conn!=null) {
26      // create the beanshell statement
27      Statement st = null;
28      try {
29          st = conn.createStatement();
30      } catch (SQLException e) {
31          // TODO Auto-generated catch block
32          e.printStackTrace();
33      }
34      // execute the query, and get a beanshell resultset
35      ResultSet rs = null;
36      try {
37          rs = st.executeQuery(query);
38      } catch (SQLException e) {
39          // TODO Auto-generated catch block
40         e.printStackTrace();
41      }
```

Figure 5: The source code for sending a message to the database MariaDB

Figure 5 shows that the opaque action element *connectMariaDB.bsh* has the input values *classname*, *url*, *username*, *password*,

and *driver* to create a database connection. In addition, the BeanShell code is required to send a request to the database (Figure 5, line 29) and get a response (Figure 5, line 37).

The source code in the opaque behaviors is tested by running the simulation. The code can therefore be adopted when implementing, e.g., systems or SoS.

Database Management System The open-source web-based database management system Heurist was specially developed for the Humanities. Heurist allows researchers without prior IT knowledge to develop databases, store and search their data, and publish it on an automatically-generated website.

4 Information System Development

Data projects in the Humanities depict a perfect test scenario for information system development for two reasons. First, the projects tend to be comparatively small and, second, both data and usage show high degrees of heterogeneity. This phenomenon, known as the *long-tail* problem of the Humanities, is due to an institutional decision of most universities to subsume all kinds of subjects under one departmental unit called Humanities. Left the reasons aside, software architects and researchers alike find themselves in the situation to cope with the high variability of requirements, software quality attributes, and missing standards. From a point of view of information system development, one can think of several solutions for data heterogeneity. In fact, they can also all be viewed as an SoS. An analysis of the current situation reveals three strategic strains of data management. Firstly, isolated applications fully independent and maintained by decentralized units such as a single chair. Second, single data applications implemented with a set framework such as *My Content Repository* (MyCoRe) managed centrally. And third, a globally maintained platform with limited but extensive data curation functionality for archiving, publishing, and analyzing data such as Heurist. Since sooner or later, the isolated applications are transferred to one of the centralized data solutions, we will take a closer look at the two later approaches.

MyCoRe The framework MyCoRe (`https://www.mycore.de/en/`) contains all the functionality of a data repository. Some public institutions such as libraries and universities implement instances of MyCoRe to administer publication inventories and research data. As a typical client-server application, it can be used to host any kind of data. Among the main configurable components of the MyCoRe system are a Solr (`https://solr.apache.org/`) search engine, a data base access handler via Hibernate, a system management for user rights and access as well as a content store. Interfaces to external systems are restricted to library formats Z39.50, but also comprise REST, OAI and SWORD. Generally all documents and metadata are saved as XML, however, some information is stored in relational database tables for reasons of performance and modifiability. Other interfaces include information exchange to the application layer, that is, a layout engine rendering XSL stylesheets and some functionality to configure the data model as well as other system variables.

Although the structure of a MyCoRe database is known and could be used for automatic retrieval, the data models of a MyCoRe application are very flexible and represented without a standard as a XML schema definition. Its retrieval and analysis depend on how different data models are related to each other and which structural information on how to process the data is *hidden* in the application. Generally it is possible to parse the data model schema definitions and based on this information automate the data retrieval. Yet, for MyCoRe applications that make use of several data models whose interaction and processing became part of the business logic of the program, a semi-automated retrieval process seems to be the only doable solution.

It is a valid data management strategy to have these projects set up as independent MyCoRe instances if larger amounts of data need to be handled or if many users with many different tasks and views on the data require clear and comprehensible workflows. It ensures more flexibility while keeping data maintenance and server administration on an acceptable workload. Although the structure of the data, its formats and processing, is the same for all instances and it therefore has a lot of technical scalability potential, the operation of many MyCoRe instances still leads in the long run into maintenance problems if new versions have to be adjusted to the specific needs and the changing requirements of the project stakeholders. Thus, if specific needs such as a federated search are desired, this cannot be easily added. The implementation of a new function would have to be done for all instances. And if there are variant instances, a new function would have to be developed separately for each instance.

An elegant way around the growing maintainability dilemma is to find a new optimum between usability and scalability. More specifically, it means trading off the flexibility of front end layouts and some cut back on performance to integrate projects into one platform. Indeed, the tendency to focus on services rather than entire system development plays a role in the design decisions of SoS. A practical solution is to devise a system that allows for just so much adjustability as necessary for requirements satisfaction (variant-oriented system development), but leave the components responsible for all other quality requirements untouched. Heurist can be seen as such a way in the middle. Within the approach of SoS, one could push it a step further and classify data projects according to their requirements or one could also embrace all smaller projects into a new platform solution, such as Heurist, and leave the few projects with a large data inventory on MyCoRe instances to keep performance on an acceptable level.

Heurist The data management system Heurist is suitable for variant-oriented system development such as presented in [7, 9]. Even if the development of the systems, here database instances of Heurist, hold the same functions, these can be used to create a project-specific database autonomously. If the individual database instances were to be combined into an aggregated system, it would be possible to develop the complete system as a single system, as a product family or as an SoS. However, the system development of a single system has little flexibility to make extensions. Single systems cannot be used for different purposes as variants as effortless and cost-saving than SoSs.

With Heurist, for each project a project-specific web page can be constructed as a variant with the same functional range. In order

to create a website, the search area, the display of a result set, the display of the contents as well as the integration of a map can be straightforwardly arranged. The view can either be programmed with PHP or implemented via an editor interface.

To sum up, Heurist makes it possible to create a database instance as a variant and supports further development with individual properties.

4.1 Information Systems

The three information systems EDAK, TheDefix, and NETamil are autonomously developed information systems at the CSMC using the tool Heurist while the *Collection of Greek Ritual Norms* (CGRN) is an external application that was not modeled in Heurist.

The first three information systems mentioned above represent how systems can be developed in a variant-oriented manner. The CGRN system represents a system which becomes part of the SoS without being an instance of Heurist.

In practice, other systems are often developed as different instances, but they should also have the possibility to use the same functionality if required. Then, it is desirable that these systems can also be integrated into the existing overall structure without having a complete redesign of the SoS.

In what follows it is shown that both variant and non-variant systems can be part of the SoS and thus all these systems can use the federated search function.

NETamil During the project NETamil at the Universität Hamburg a repository was created containing digital images of classical Tamil manuscripts on palm leaves and on paper from Indian and European libraries along with a descriptive catalog, e-texts along with critical editions and annotated translations. The data was originally stored in a Word document.

In general, the database schemes can either be created individually or they can also be converted into well established XML standards such as *Text Encoding Initiative* (TEI) (`https://tei-c.org`). The TEI format is more common used in the humanities for data storage and exchange.

The automatic transformation of XML-encoded formats into a Heurist database instance has the feature of transferring a large data set into a new database instance in short time. In this paper, the created database systems have been automatically created using a word to TEI transformation process [24].

EDAK During the project EDAK the Department of History at the Universität Hamburg created an epigraphic database of ancient Asia Minor. This database contains a collection of Greek and Latin inscriptions in the area of modern-day Turkey. The data are stored in the format EpiDoc to enable easier data exchange between machines. EpiDoc is a widely used scheme for encoding scholarly and educational editions of ancient documents. It uses a subset of the TEI's standard for the representation of texts in digital form [25].

The EDAK Information system has been automatically created using an EpiDoc to Heurist transformation process.

TheDefix The database TheDefix contains curse inscriptions of the ancient world. The data are represented in a project-specific scheme. In Figure 6 the information system for the *TheDefix* project is presented: the search area is located at the left, in the middle is the result set, and on the very right the project specific data representation (text and map representation) are displayed.

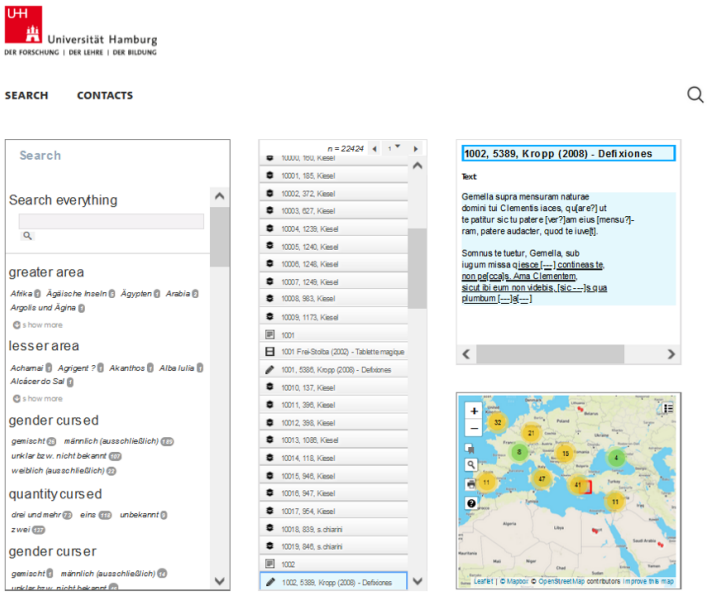


Figure 6: TheDefix Information System

CGRN The CGRN presents epigraphical data on a website. Its primary goal is to gather epigraphical material for the study of Greek rituals and to make these sources widely available [26]. The data are additionally stored in the EpiDoc format.

Merging information systems into an aggregated system, in general, requires addressing the complex issue of information integration. "Information integration is the merging of information from heterogeneous sources with differing conceptual, contextual, and typographical representations" (see [27]). For computers it is difficult to merge information without the knowledge of the syntax, semantics, model, and access of the data representations (see [28]). The approaches therefore require a standardized framework for representing data that, while supporting autonomy to some degree, can make heterogeneity manageable.

In the next chapters, we will reveal how to integrate different autonomously developed information systems, which can also be physically located in different places, as one SoS, taking into account that variant parts are not developed redundantly.

Product family The integrated PKT approach includes the VAM which, in a hierarchical approach of four levels, is used to develop a variant component for each custom-relevant differentiating property, whenever possible in a 1:1 relationship. This approach is very suitable if as many different customer requirements as possible have to be satisfied while still remaining competitive. The VAM approach was also transformed into a model-based approach, using VAMOS to represent the variants. It was observed that the structured package overview in the SysML model avoids redundancies

and improved transparency and traceability for large and complex projects (see [11]).

The integrated PKT approach aims at developing a product family comprising variants. The idea of developing an SoS in a standardized manner is obvious. However, it must be ensured that the development of product families also involves the development of variant hardware components and not just a communication interface. In addition, one has an influence on all systems with the development of product families. If, however, an information system were designed as an SoS consisting of both internal and external systems, it would be recommendable for a clear focus on the communication interface. This recommendation must be taken into account when the SoS is actually modeled on variants.

5 Variant-oriented SoS Development

In this section we present how context information relevant to the SoS and the variants are identified as part of the requirements engineering process ACRE. Additionally, it is described how to design the structure of an aggregated information system. Finally, this chapter depicts how to create a communication network as well as how to simulate the common function *federated search* using Cameo Systems Modeler and the broker-based SysML Toolbox.

5.1 Variant-Based Requirements Engineering

For successful system development it is essential that the needs of all stakeholders are sufficiently satisfied. Therefore, it is necessary to have identified all persons and institutions that have requirements or interest in the system. The respective requirements of all identified stakeholders are collected, documented, and structured according to the ACRE ontology, presented in [18], with the goal to identify the requirements of all stakeholders and to be able to manage them throughout the system development process.

For the variant-oriented development of an information system as an SoS, a lean version of the ACRE ontology was specified and used for modeling an aggregated information system. The ACRE ontology with SoS and variety contexts is presented in Figure 7.

An (abstract) requirement has the specializations business, functional, and non-functional requirements. A requirement description explains requirements, where some rules are applied. The rules could be that requirements have to be formulated in accordance with the ISO 29148:2011 [29] and RFC2119 [30] to use the linguistic syntax profitably. A requirement description is elicited from one or more *Source Element*(s). Source elements can be standards, conversations, requirement lists, and specifications among others.

The contexts are defined as:

- A system of systems context defines views on aggregated systems.
- A system context contains views of system, subsystem, assembly, and component.
- A stakeholder context defines views on different stakeholders.
- A variety context defines views on system variants.

Use cases are validated by one or more *Scenario*(s). A *Scenario* describes the “what ifs” in a semi-formal or formal way. SysML activity and sequence diagrams can present *Semi-formal Scenario*s to describe communication processes and interactions between elements in the system. SysML parametric diagrams present the “what ifs” formally. Both scenario types support the analysis of “what ifs” to validate the use cases. Through simulation, the modeled scenarios can test the interactions between all participants within the communication network.

The developed systems have a satisfying relationship with the requirements they meet. A *System of System* element is a generalization and has two or more systems as parts.

It should be noted that there are a number of other approaches to the requirements engineering process. However, it is crucial that the variety context will be considered during the requirements engineering process. Contextual information has to be added in all other approaches as well. The ACRE ontology has already been successfully applied in many projects over several years using the SysML [11, 13, 14]. Due to the well-known and proven approach of applying ACRE with a variety context in a model-based way during system development, the ACRE approach was chosen for the development of an SoS.

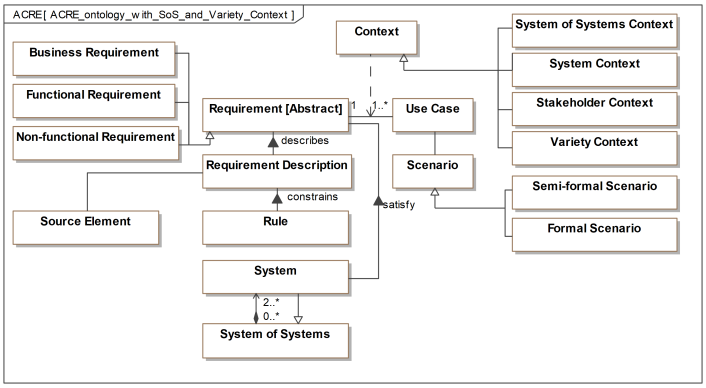


Figure 7: ACRE ontology with systems of system and variety contexts

5.2 SoS Development

Heurist can be used to create variant database instances and is realized as a client-server architecture. Although established design patterns are missing in the still evolving software, further development tends in the direction to have Heurist fully operational as an *Model-View-Controller (MVC)* application. The MVC separates an application into three main logical components: the model, the view, and the controller. Each of these components are built to handle specific development aspects of an application. In the context of creating an information system, the model represents a data scheme, the view a graphical user interface, and the controller accepts user inputs and converts it to commands for the model or view.

The new planned architectural approach is important when it requires adding another layer, the SoS layer. The development is currently still in the conceptual phase. As of now, Heurist is initially used for variant system development and the SoS layer is first tested out through simulations and prototype implementations.

In addition, the Universität Hamburg operates Heurist as a pub-

lic institution, which recommends the use of an adjusted version of the V-model. It follows that further adjustments will be made to Heurist in the area of verification and testing.

6 Modeling and Simulation of an SoS

Modeling and Simulating of an SoS using the SysML and the tool Cameo Systems Modeler has the advantage to test the system's behavior before implementation because the specification is executable. "This quality of executable specifications promises to remedy the most serious problem of software – its lack of correctness and reliability." [31]

In the following we present how to develop an executable specification for an SoS during the requirements engineering process.

6.1 Requirements Profile

For the special requirements (business, functional, and non-functional), new stereotypes were defined as an extension of the *Extended Requirement* stereotype. An *Extended Requirement* is a standard requirement extension that adds some properties to a requirement element. The requirements are devised in accordance with the ISO 29148:2011 and RFC2119.

6.2 SoS Profile in SysML

For representing Systems and SoS the new stereotypes *System* and *System-of-Systems* are defined as extensions of the *Metaclass Class*, see Figure 8.

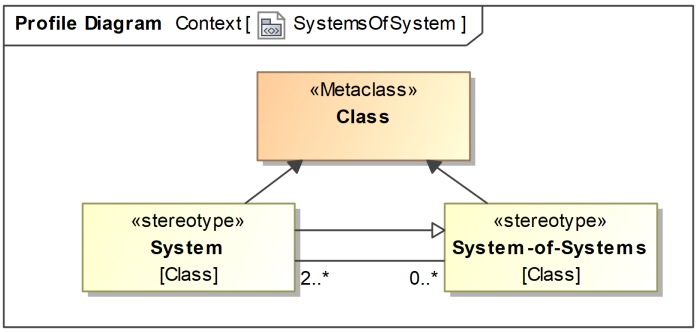


Figure 8: Profile Diagram: new stereotypes *Systems of System* is a specialization of a system and has an association to the stereotype *System*

6.3 Variety Profile

The VAMOS profile which is presented in Figure 1 is used for variant modeling. Figure 9 presents a concrete application of VAMOS.

The package *Variation 1* has the stereotype *Variation* and contains the System *Heurist*. The packages V1, V2, and V3 are variants of *Variation 1*. The variant *V1* contains the System EDAK, the variant *V2* contains the system TheDefix, and the variant *V3* contains the system NETamil, respectively.

One way to introduce a redundancy-reduced communication interface for all variants is to add a SysML port element to the Heurist system. All variants inherit the port via the specialization. However, if an external system were added to the aggregated system at a later point in time, a separate communication interface would have to be implemented for this external system. This is precisely the crux of the matter. If a communication interface is to be offered for internal variant systems as well as for external systems, the communication interface should be inherited by the systems via a specialization using an SoS element.

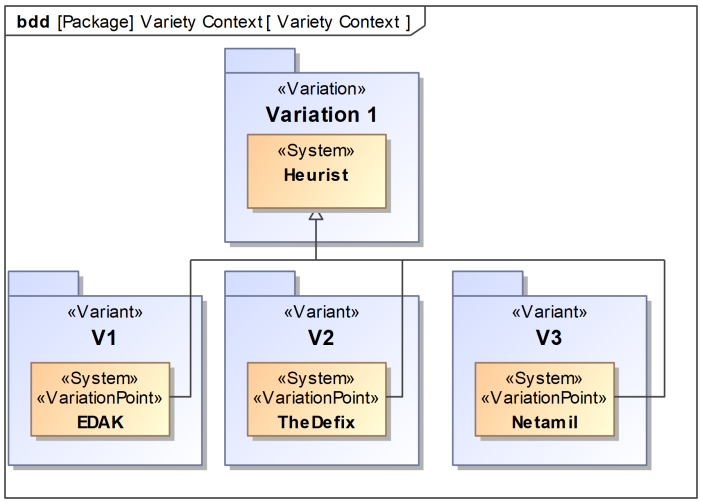


Figure 9: Representation of three variants using VAMOS

6.4 Use Cases

Figure 10 shows the representative use cases for different search functionalities while considering the variety and SoS contexts. The main actor is a CSMC user. The CGRN, EDAK, NETamil, and TheDefix users are specializations of the CSMC user.

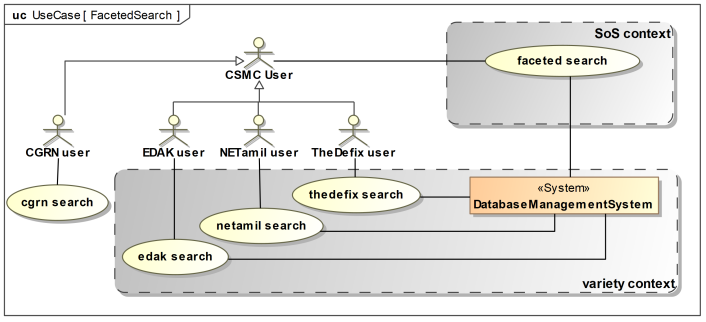


Figure 10: Use case diagram with variety and SoS context

One can see that CGRN users do not belong to the variant context as the other three users but all users also have an association to the use case faceted search. As described in Section 4, the three systems should be developed as variant systems using Heurist, while the development of the CGRN system was done externally. Nevertheless, all systems should be networked so that each system can use the federated search function.

A CSMC user can execute a faceted search. The specialized users can also execute this search while all users have (project-)specific search functionalities, e.g., EDAK users search for specific names mentioned in editions or for object types of inscriptions (use

case: *edak search*), NETamil users look up which word occurs in which poem and in which line (use case: *netamil search*), and TheDefix users want to know the curse id of curses (use case: *netamil search*). The different contexts are presented in SysML use case diagrams.

6.5 Scenarios

Each use case can be validated by one or more scenarios. The scenarios can be represented in behavioral diagrams such as activity or sequence diagrams. We use activity diagrams for modeling and simulating, e.g., the federated search functionality.

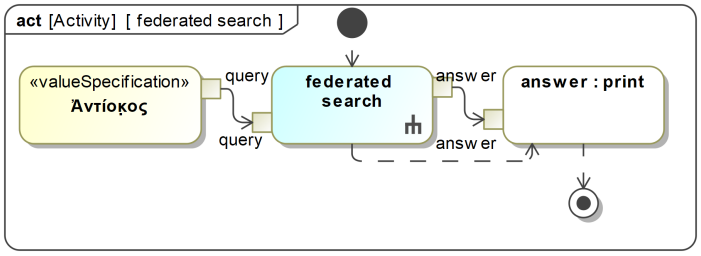


Figure 11: Search for the word “Antiochus” (engl.)

Figure 11 depicts the word “Antiochus” (engl.). It is the input value (=query) for the federated search activity. Behind the federated search activity is a more detailed federation process. As a result, the responses of all databases are printed.

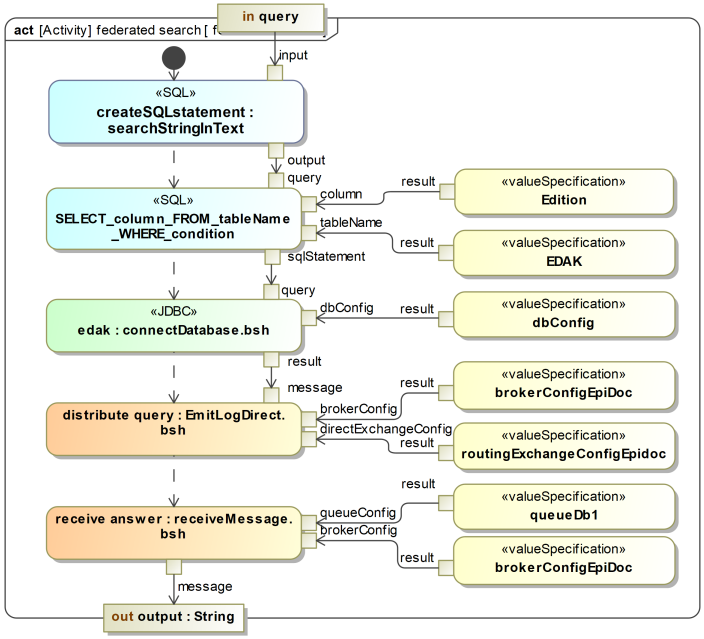


Figure 12: Faceted search actions

Figure 12 illustrates the faceted search process in more detail. The query is the input value. The first action *createSQLstatement:searchStringInText* creates the SQL statement *“Text LIKE ’%Antiochus%’;”*. The action *SELECT_column_FROM_tableName_WHERE_condition* creates the SQL expression *SELECT Edition FROM EDAK WHERE ’%Antiochus%’;*, which is the input value for the next action. The action *distribute query:EmitLogDirect.bsh* sends the SQL expression to a server. The action *receive answer:receiveMessage.bsh* sends a response from a server. The opaque action *EmitLogDirect.bsh* can also contain a forwarding process to another database. To realize a federated search, a script was implemented and must be active on the server side. In fact, the script calls the requests from the server (query queue), processes the schema mapping, and passes the response to the server (response queue). We implemented the server side scripts in Java. The source code is very similar to that of Beanshell (see https://www.rabbitmq.com/tutorials/tutorial-three-java.html). However, other programming and script languages can also be used such as Python, PHP, C#, or JavaScript (see https://www.rabbitmq.com/getstarted.html).

It should be noted here that the scenario at hand already incorporates decoupling of the systems using a communication interface. In a very early phase of system development, communication could take place directly with the database. And yet, communication interfaces are to be used in the development of SoS. Briefly put, this has already been taken into account in the scenarios. As intended by ACRE, the use cases were validated by the scenarios during the requirements engineering process.

6.6 Communication Interface

Communication interfaces ensure the coverage of the need for information and are used for data exchange. For creating communication networks, RabbitMQ is used as an *Application Programming Interface* (API) for SoSs. RabbitMQ offers broker federation and therefore allows the exchange between source and destination brokers, or from a message queue in the source broker to an exchange in the destination broker (see [15]). To model these communication interfaces the stereotype *interfaceBlock* is used. One port of the SoS has at least this communication interface to establish a communication network between the systems which are part of the SoS.

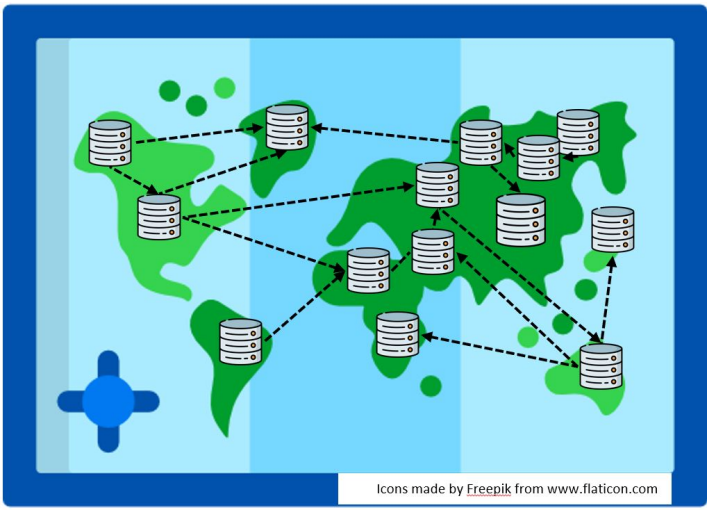


Figure 13: Federated Search Network

Figure 13 illustrates a communication network between loosely-coupled systems which are to be transferred to an aggregated system including a communication interface. By coupling the systems, e.g., federated searches can be realized. The idea is to provide each participant with its own RabbitMQ message broker to easily realize this communication network.

6.7 *Structure of the CSMC Information System*

Figure 14 illustrates the structure of the SoS named *CSMC Information System*. The SoS has the parts of systems *EDAK*, *TheDefix*, *NETamil*, and *CGRN*. These systems are also specializations of the SoS and inherit all activities of the SoS. In this case, federated search is part of each system. The SoS has a communication interface which is modeled as a port. The systems *EDAK*, *TheDefix*, and *NETamil* also inherit all elements of the system *Heurist*. The system *CGRN* is not a specialization of the system *Heurist* and thus does not inherit all activities of the Heurist system, but only those of the SoS.

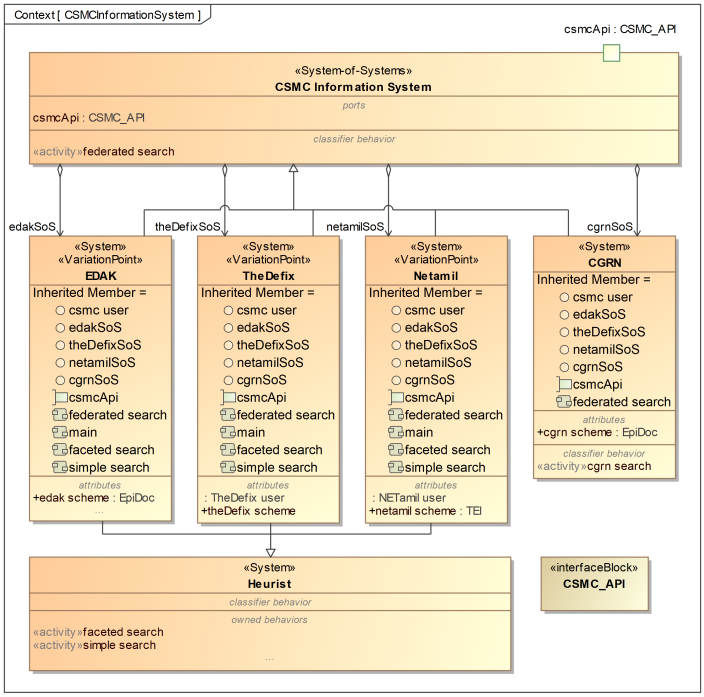


Figure 14: CSMC Information System

Figure 15 gives another overview about the dependencies *generalization* and *inherit* members between the SoS and the systems. In the allocation diagram it can be seen at a glance which systems inherit which activities or which do not. When adding more activities to an SoS or when adding more external systems, this overview can be used to quickly determine which elements will be added to a system when it becomes part of an SoS.

In the development of interfaces, the allocation diagram is an excellent way to illustrate the dependencies of all the systems involved. In the diagram, the separation between the interfaces of the SoS or other interface dependencies can be clearly highlighted.

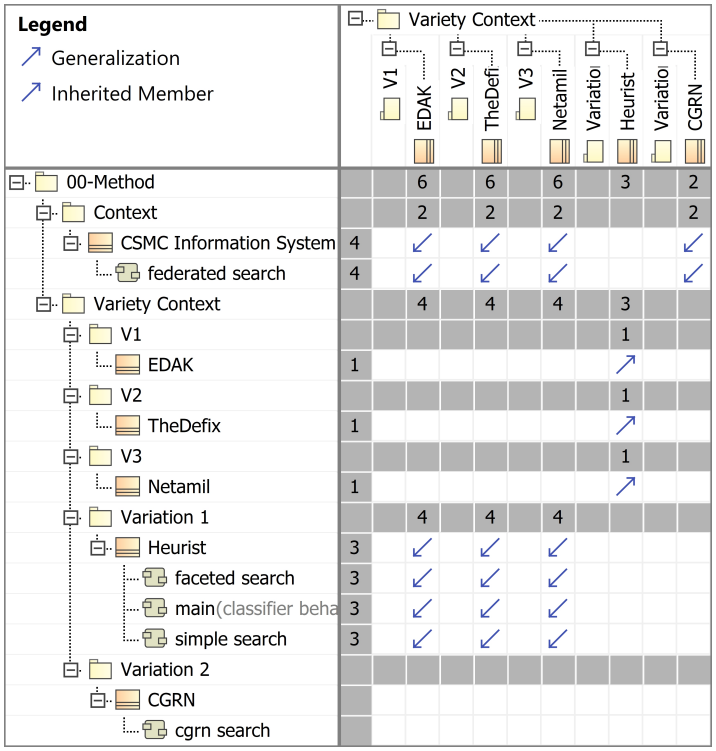


Figure 15: Allocation diagram which represents the dependencies *generalization* and *inherit* members

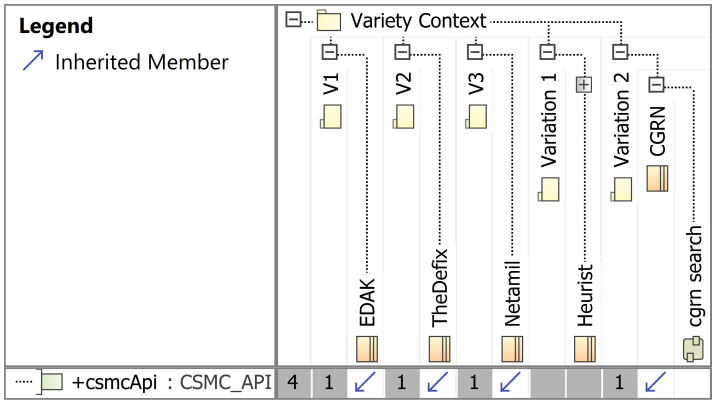


Figure 16: Allocation diagram which represents the systems which have the SoS communication interface

7 Application and Results

We evaluate our approach by a feasibility study. For this purpose, we use a notebook where the tool Cameo Systems Modeler (version 2021x) and the broker-based SysML Toolbox, a RabbitMQ server (version 3.8.9), and MariaDB (10.5.6) are installed. We emulate the databases EDAK, and NETamil on the database MariaDB which represents the Heurist database instances. On a Raspberry Pi 4 we also installed a RabbitMQ server and MariaDB where the database CGRN is simulated. Both RabbitMQ servers are configured with particular message queues, exchanges, and bindings as follows.

The message queues *queueDb1* for EDAK, *queueDb2* for NETamil, and *queueDb3* for CGRN are defined. They are all in the

same virtual host *dbFederatation* (see Figure 17).

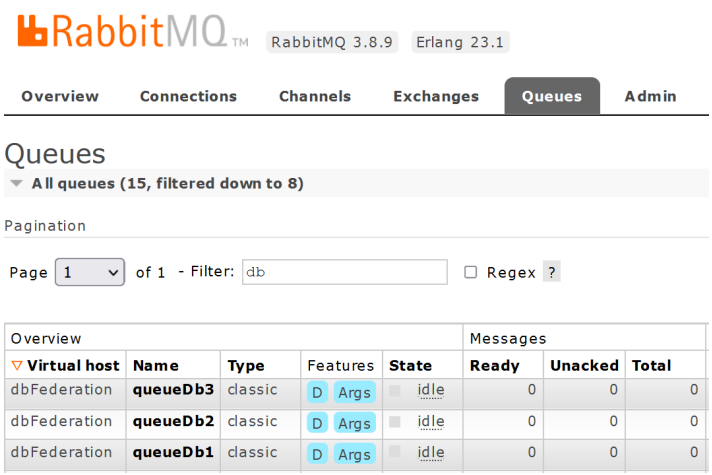


Figure 17: Defined queues on a RabbitMQ server

The exchange is called *db.direct*. The bindings with the particular routing key are: queueDb1 → epiDoc, object, query; queueDb2 → object, queueDb3 → epiDoc, object.

The EDAK data model is represented by the entity type "description." A description has the attributes "identifier", "description_id", "edition", "category", "region", "location", "find spot", "text", and "date." Each "description" has the unique identifier "description_id."

The CGRN data model is represented by the entity type "description." A description has the attributes "idno", "date", "provenance", "support", "layout", "bibliography", "text", "translation", "traduction", "commentary", "publication", "authors", and "project director."

The NETamil data model is represented by the entity types "poem", "commentary", and "dictionary." A poem has the attributes "edition", "transliteration", "word_by_word_translation_into_english", "translation_into_english", and "source."

We simulate federated searches, such as presented in Figure 12. During the simulation the query *SELECT Edition FROM EDAK* is sent to the EDAK database via the opaque action *edak:connectDatabase.bsh*. The SQL expression is published via the opaque action *distribute query EmitLogDirect.bsh*. The databases EDAK receives this expression via the opaque action *receive answer:receiveMessage.bsh*. The database CGRN is queried with the same SQL expression because of the defined routes in the RabbitMQ servers. For the search query "Antiochus", written in Greek language, (*SELECT Edition FROM EDAK WHERE '%Antiochus%';*) we received 1 answer from EDAK and 0 answers from CGRN. For the search query "Zeus", written in Greek language, (*SELECT Edition FROM EDAK WHERE '%Zeus%';*) we received 5 answers from EDAK and 1 answer from CGRN (see Figure 18).

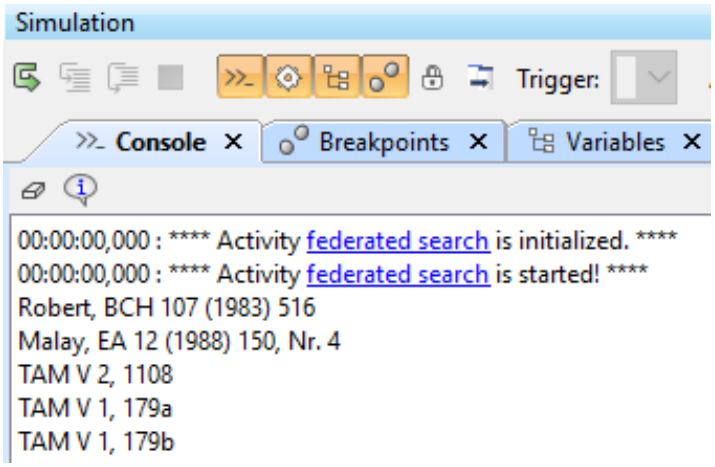


Figure 18: Database results for the search query "Zeus" written in Greek language

For the search query *SELECT COUNT (e.Date) AS Number, e.Date AS Date FROM EDAK e GROUP BY e.Date* we receive the following results (excerpt):

```
+--------+----------------+----------+
| Number | Date           | Database |
+--------+----------------+----------+
|     31 | 4. Jh. v. Chr. | EDAK     |
|    200 | 1. Jh. n. Chr. | EDAK     |
|    818 | 2. Jh. n. Chr. | EDAK     |
|    642 | 3. Jh. n. Chr. | EDAK     |
|    156 | 4. Jh. n. Chr. | EDAK     |
+--------+----------------+----------+
|      1 | ca. 250-200 BC | CGRN     |
|      2 | ca. 350-300 BC | CGRN     |
+--------+----------------+----------+
```

The responses returned by EDAK and CGRN show that the date is differently represented in both databases. The date differs in language and representation (indication as century or year). When the query is filtered by year, one of the two databases returns an empty result set as response. A translation of the date representations can lead to a complete answer. A mapping between the representation of the date is required to ensure *correct* query results. Schema mapping is generally required when defining federated search queries.

This simulation example also presents that queries from EDAK are answered using both the EDAK and the CGRN databases. NETamil is not involved in this specific query process because of the missing routing key in the RabbitMQ configurations. At this point it makes no sense in terms of content. If one wants to compare another repository with Tamil poems, a route can be defined via the RabbitMQ configuration that supports the sensible federated search. In this example, the Cameo System Modeler's console represents the CSMC information system, which receives all responses from the various databases from the federated search. If either a Heurist database instance or an externally developed system is to be aggregated to the SoS, this can be realized by installing a RabbitMQ broker, programming a script for publishing and receiving messages from the broker, and setting the broker configurations.

In this way, new databases can be added so that our principle "*bring your own database* is supported. Then all new systems of the

SoS will also benefit from the federated search. Consider that the challenges of information integration must be resolved for mapping to use federated searches successful.

In the humanities, as well as other fields, it is important that existing functions such as federated searches can be used without large expenditure of resources. After all, resources are limited. Existing data can thus be enriched with further information in a short time and users can focus more on editing content. Using the allocation diagram to keep an overview of all systems to be aggregated also helps to keep track of the growing number of systems (cf., Figure 15 and Figure 16). All systems, whether variants or not, can be specifically developed and integrated into the overall SoS. Without considering this overview, existing systems could be produced mistakenly from scratch simply because they are unknown and as such a variant is regarded as an external (unregistered) system. As a positive effect of the present approach, one has the advantage of being able to cooperate with external parties whether their system can also be developed as a variant of one's own system, so that the same communication interface (*csmcInterface*) can be used.

At the Universität Hamburg Heurist has already been set up and it is being used for the autonomous development of information systems. More than ten information systems have already been created. When changing functions, such as the web page design, the allocation diagram can be used to see which systems are affected.

The special feature of the model-based documentation of SoSs and the variants with VAMOS make it convenient for the developers to get an overview between the different diagram types and the filter properties for displaying data. Developers also see which systems are relevant and whether changes have an impact on the system properties or not. It is self-explanatory to switch between a display of specific SoS or variant elements, or view everything together in one diagram, as shown in Figure 15. The model-based and variant-oriented approach also ensures that the interfaces of the systems are not developed redundantly. The realization of a communication interface using RabbitMQ supports the aggregation of decoupled systems by implementing message scripts that are publicly available.

It is planned to transfer the prototypically implemented CSMC Information System with the presented communication interface into a product. Regarding the product development, the variant-oriented approach points in the right direction as to even more relevant parameters, such as performance and security attributes, that should be tested in advance.

The approach at hand also fits in other areas of system development, e.g., in the aviation industry. There, it has already been shown that the networking of systems after the digitization of business processes can be helpful to automate ordering processes between supply chain tiers [32]. It would be conceivable to design the ordering process as part of the SoS and thereby identify the variants in the ordering process as well as the external systems that want to become part of the SoS. As an assumption, there will be more individual solutions that will be aggregated into an SoS. Here, too, the approach offers the advantage of keeping an overview of all systems and working towards a common interface in a targeted manner so that the connection to the SoS can be made with little resource effort.

8 Conclusion and Outlook

In this paper, we presented how to develop an aggregated system, which, understood as an SoS, was put into practice in a model-based and variant-oriented way. The aim was to identify the number of variants easily at an early stage of the requirements engineering process so that the development of elements has neither to be done holistically nor redundantly. For this purpose, we used the ACRE ontology with the extended contexts on SoS and variety, and applied this approach with the SysML tool Cameo Systems Modeler as well as the VAMOS profile. In addition, we defined an SoS profile which helped us to distinguish between developing variants and merging variants as well as non-variants into an SoS. For the implementation of a communication interface, RabbitMQ was used as a message broker, which allows loosely coupled systems to be brought together in a simple way. The variant database systems were developed with Heurist which on the one hand supports the development of automated database systems and on the other hand keeps the heterogeneity under control, often resulting from the many requirements. The prototype implementation showed us that this path is promising and should be further pursued.

The advantage of merging multiple database systems is that functions such as a federated search can be implemented, however, the problem of data integration between all the database instances must be solved beforehand so that a search query does not lead to a faulty response. Therefore it must be ensured that the data or their representations have the same syntax, semantics, model, and access. At the CSMC, a feasible study is currently in progress.

Conflict of Interest The authors declare no conflict of interest.

References

[1] E. Werner, "Clay Tablet (AO 29196) from the Louvre in Paris (3D model)," 2020, doi:http://doi.org/10.25592/uhhfdm.918.

[2] E. Werner, "Clay Tablet (KUG 15) from the University Library Giessen (3D model)," 2020, doi:https://doi.org/10.25592/uhhfdm.766.

[3] D. Krause, G. Beckmann, S. Eilmus, N. Gebhardt, H. Jonas, , R. Rettberg, "Integrated Development of Modular Product Families - a Methods Toolkit," in In T.W. Simpson, J. Jiao, Z. Siddique, K. Hölttä-Otto (Eds.): Advances in product family and product platform design: Methods & applications, 245–269, Berlin Springer, 2014.

[4] O. C. Eichmann, S. Melzer, R. God, "Model-based Development of a System of Systems Using Unified Architecture Framework (UAF): A Case Study," in Proceedings of 2019 IEEE International Systems Conference, IEEE, 2019.

[5] C. B. Nielsen, P. G. Larsen, J. Fitzgerald, J. Woodcock, J. Peleska, "Systems of Systems Engineering: Basic Concepts, Model-Based Techniques, and Research Directions," 2015.

[6] M. W. Maier, "Architecting principles for systems-of-systems," Systems Engineering, **1**, 267–284, 1998.

[7] S. Melzer, H. Peukert, H. Wang, S. Thiemann, "Model-based Development of a Federated Database Infrastructure to support the Usability of Cross-Domain Information Systems," in Proceedings of 2022 IEEE International Systems Conference, IEEE, 2022.

[8] O. C. Eichmann, S. Melzer, M. Hanna, R. God, D. Krause, "A Model-Based Approach for the Development of Modular Product Families Considering Different Life Phases," in Proceeding EMEA Systems Engineering Conference, EMEASEC 2018 / TdSE 2018, 2018.

[9] S. Melzer, S. Thiemann, R. Möller, "Modeling and Simulating Federated Databases for early Validation of Federated Searches using the Broker-based SysML Toolbox," in Proceedings of 2021 IEEE International Systems Conference, IEEE, 2021.

[10] T. Weilkiens, Variant Modeling with SysML, MBSE4U Booklet Series, 2016.

[11] T. Bahns, S. Melzer, R. God, D. Krause, "Ein modellbasiertes Vorgehen zur variantengerechten Entwicklung modularer Produktfamilien," in Tagungsband zum Tag des Systems Engineering (Eds.: Chr. Muggeo, S.O. Schulze), Carl Hanser Verlag GmbH & Co. KG, 2015.

[12] R. God, S. Melzer, U. Wittke, "SiLuFra Schlussbericht - Sichere Luftfracht-Transportkette: Konzepte, Strategien und Technologien für sichere und effiziente Luftfracht-Transportketten; Teilvorhaben: Modellbasierte Architektur- und Lösungsspezifikation; Laufzeit des Vorhabens: 01.07.2013 - 31.08.2016," 2016.

[13] R. God, S. Melzer, "Teilvorhaben: Spezifikation und Integration cyber-physischer Betriebs- und Geschäftsprozesse : Schlussbericht : Laufzeit des Vorhabens: 01.05.2016-30.09.2019 : Berichtzeitraum: 01.05.2016-30.09.2019, Spezifikation und Integration cyber-physischer Betriebs- und Geschäftsprozesse," Technical report, Technische Universität Hamburg, Institut für Flugzeug-Kabinensysteme, Hamburg, 2020, doi:10.2314/KXP:1726105857.

[14] R. God, U. Wittke, S. Melzer, C. Witte, "KomKab Schlussbericht - Kommunizierende Kabine; Teilvorhaben: Digitaler Ramp-Agent; Laufzeit des Vorhabens: 01.01.2016 - 31.03.2019," 2019.

[15] S. Melzer, J. P. Speichert, O. C. Eichmann, R. God, "Simulating cyber-physical systems using a broker-based SysML Toolbox," in AST 2019 - 7th International Workshop on Aircraft System Technologies, Hamburg University of Technology, 2019.

[16] D. Arndt, S. Melzer, R. God, M. Sieber, "Konzept zur Verhaltensmodellierung mit der Systems Modeling Language (SysML) zur Simulation varianten Systemverhaltens," in Tagungsband zum Tag des Systems Engineering (Eds.: S.O. Schulze, C. Tschirner, R. Kaffenberger, S. Ackva), Carl Hanser Verlag, 2017.

[17] S. Melzer, R. God, T. Kiehl, R. Möller, M. Wessel, "Identifikation von Varianten durch Berechnung der semantischen Differenz von Modellen," in Tagungsband zum Tag des Systems Engineering (Eds.: M. Maurer, S. O. Schulze), Carl Hanser Verlag GmbH & Co. KG, 2014.

[18] J. Holt, S. A. Perry, M. Brownsword, Model-Based Requirements Engineering, volume 9 of *Professional Applications of Computing Series*, Institution of Engineering and Technology, Stevenage, 2012.

[19] O. C. Eichmann, S. Melzer, F. Giertzsch, R. God, "Stakeholder Needs and Requirements Definition During Service Development in a System of Systems," in Proceedings of 2020 IEEE International Systems Conference, IEEE, 2020.

[20] J. Holt, S. Perry, R. Payne, J. Bryans, S. Hallerstede, F. Hansen, "A Model-Based Approach for Requirements Engineering for Systems of Systems," IEEE Systems Journal, **9**(1), 252–262, 2015, doi:10.1109/JSYST.2014.2312051.

[21] J. Holt, S. Perry, SysML for Systems Engineering: A Model-Based Approach, Computing, Institution of Engineering and Technology, 2018.

[22] A. Qpid, "Messaging built on AMQP," https://qpid.apache.org/releases/qpid-cpp-master/cpp-broker/book/chap-Messaging_User_Guide-Broker_Federation.html, 2015, accessed January 22, 2022.

[23] S. Melzer, O. C. Eichmann, H. Wang, R. God, "Modeling and Simulation of Database Interactions," Tag des Systems Engineering 2021 (TdSE2021), 2021, doi:10.25592/uhhfdm.9696.

[24] S. Schiff, S. Melzer, E. Wilden, R. Möller, "TEI-based Interactive Critical Editions," in 15th IAPR International Workshop on Document Analysis Systems, Lecture Notes in Computer Science (LNCS), Springer, 2022.

[25] T. Elliott, G. Bodard, E. Mylonas, S. Stoyanova, C. Tupman, S. Vanderbilt, et al., "EpiDoc Guidelines: Ancient documents in TEI XML (Version 9)." Available: https://epidoc.stoa.org/gl/latest/., (2007-2022), accessed January 22, 2022.

[26] J.-M. Carbon, S. Peels, V. Pirenne-Delforge, "A Collection of Greek Ritual Norms (CGRN)," Liège, http://cgrn.ulg.ac.be, consulted in 2020, 2016–2020, online; accessed 10 December 2021.

[27] W. Hao, S. De-wen, F. Xujian, X. Haitao, "Application of information fusion technologies for multi-source data," Journal of chemical and pharmaceutical research, **5**, 2013.

[28] U. Leser, F. Naumann, Informationsintegration: Architekturen und Methoden zur Integration verteilter und heterogener Datenquellen, dpunkt.verlag GmbH, 2007.

[29] ISO/IEC/IEEE 29148:2011(E), "Systems and software engineering - Life cycle processes - Requirements engineering," https://wiki.unix7.org/_media/ict/lib/iso-iec-ieee-29148-2011.pdf, accessed January 22, 2022.

[30] S. Bradner, "Key words for use in RFCs to Indicate Requirement Levels," Harvard University, https://datatracker.ietf.org/doc/html/rfc2119, 2017, accessed January 22, 2022.

[31] N. E. Fuchs, "Specifications are (preferably) executable," Softw. Eng. J., **7**, 323–334, 1992.

[32] H. Wang, S. Melzer, "Simulation of Ordering Processes across different Supply Chain Tiers in the Aviation Industry," in 2022 IEEE International Systems Conference (SysCon) (IEEE SysCon 2022), Montreal, Canada, 2022.

ASTES

Advances in Science, Technology and Engineering Systems Journal Vol. 7, No. 3, 32-43 (2022)

www.astesj.com

Special Issue on Multidisciplinary Sciences and Engineering

ASTESJ

ISSN: 2415-6698

Encompassing Chaos in Brain-inspired Neural Network Models for Substance Identification and Breast Cancer Detection

Hanae Naoum[1,*], Sidi Mohamed Benslimane[1], Mounir Boukadoum[2]

[1]*LabRI-SBA Lab., Ecole Superieure en Informatique, Sidi Bel Abbes, 22000, Algeria*

[2]*Design and fabrication of Microsystems Research Laboratory, Department of Computer Science, University of Quebec At Montreal, Montreal, H3C 3P8, Canada*

ARTICLE INFO

Article history:
Received: 28 January, 2022
Accepted: 06 May, 2022
Online: 25 May, 2022

Keywords:
Pattern recognition
Artificial Neural Networks
Chaos
Medical diagnosis aid systems
Breast cancer detection
Substance identification
Noise resilience

ABSTRACT

The main purpose in this work is to explore the fact that chaos, as a biological characteristic in the brain, should be used in an Artificial Neural Network (ANN) system. In fact, as long as chaos is present in brain functionalities, its properties need empirical investigations to show their potential to enhance accuracies in artificial neural network models. In this paper, we present brain-inspired neural network models applied as pattern recognition techniques first as an intelligent data processing module for an optoelectronic multi-wavelength biosensor, and second for breast cancer identification. To this purpose, the simultaneous use of three different neural network behaviors in the present work allows a performance differentiation between the pioneer classifier such as the multilayer perceptron employing the Resilient back Propagation (RProp) algorithm as a learning rule, a heteroassociative Bidirectional Associative Memory (BAM), and a Chaotic-BAM (CBAM). It is to be noted that this would be in two different multidimensional space problems. The later model is experimented on a set of different chaotic output maps before converging to the ANN model that remarkably leads to a perfect recognition for both real-life domains. Empirical exploration of chaotic properties on the memory-based models and their performances shows the ability of a specific modelisation of the whole system that totally satisfies the exigencies of a perfect pattern recognition performance. Accordingly, the experimental results revealed that, beyond chaos' biological plausibility, the perfect accuracy obtained stems from the potential of chaos in the model: (1) the model offers the ability to learn categories by developing prototype representations from exposition to a limited set of exemplars because of its interesting capacity of generalization, and (2) it can generate perfect outputs from incomplete and noisy data since chaos makes the ANN system capable of being resilient to noise.

1. Introduction

This paper is an extension of a work originally presented in the First International Conference on Cyber Management and Engineering (CyMaEn`21) [1].

During more than 300 years, there were only two kinds of movements known in simple dynamical systems: the uniform and the accelerated movements. Maxwell and Poincare were among the minority of scientists who disagreed with those facts. It was only in the last quarter of the 20th century that the third kind of movement appeared: chaos [2].

The existence of dynamics and nonlinearity in the brain has been the topic of numerous research investigations since the 1980s. In [3], it was revealed in neurosciences that the activity of the olfactory bulb of rabbits is chaotic and, at any time, it may switch to any perceptual state (or attractor). In fact, the experimentations assessed that when rabbits inhale an odorant, their Electroencephalograms (EEGs) display gamma oscillations, signals in a high-frequency range [4, 5]. The odor information represents then an aperiodic pattern of neural activity that could be recognized whenever there was a new odor in the environment of after a session of training.

*Corresponding Author: Hanae Naoum, Email: h.naoum@esi-sba.dz

Furthermore, during the same period of Freeman's research, other works figured out the existence of chaos in the temporal structure of the firing patterns of squid axons, of invertebrate pacemaker cells, and in temporal patterns of some brain disorders such as schizophrenia and human epileptic EEGs [4-7]. Moreover, in red blood cells, chaotic dynamics of sinusoidal flow were determined by 0-1 test. In fact, numerous simulations identified the existence of chaotic dynamics and complexity in the sinusoidal blood flow [8]. In addition, the exploration of dreaming through the application of concepts from chaos theory to human brain activity during Rapid-Eye-Movement state (REM-state) sleep/dreaming proved that chaos is on the flow of thoughts and imagery in the human mind [8-10]. Finally, chaos is ubiquitous in the brain operations and cognitions according to cognitive sciences, linguistics, psychology, philosophy, medical sciences, and human development [11-15]. The later issues are still addressed in depth in the context of research on the complex systems, to which the brain obviously belongs. This is at this level that the ultimate goal of AI has to be considered. Indeed, creating a machine exhibiting human-like behavior or intelligence, cannot be, with keeping the chaos properties aside.

Moreover, being an offshoot of Artificial Intelligence (AI) paradigms, pattern recognition techniques focus on the identification of regularities in data in an automated process [16]. It is worth noting the fact that, pattern recognition is a cognitive functionality in the brain. In fact, in real life, human beings are capable of recognizing and recalling patterns of different natures and forms (not necessarily perfect patterns) and in different conditions, naturally without significant effort. An intelligent pattern recognition system must thus include brain properties, such as, the presence of chaos.

Furthermore, ANNs represent a discipline of AI that has successfully been applied on different nature of pattern recognition problems. In fact, ANNs models were employed for data compression, data classification, data clustering, feature extraction, etc. Data classification is particularly one of the most active search and application fields in connectionism [16-18]. Consequently, ANN approaches encompasses potential techniques to face pattern recognition problems.

Several works can be noticed in the literature, that focus on the construction of ANN models that implement NDS properties [19-23]. Those proposed models are challenging the classical kinds of ANNs in terms of biological plausibility [24-28], and in some cases, even in terms of computational efficacy of the model [29-32]. Except that, most of the proposed models were developed including the stabilities of attractors with no attention to the ongoing instabilities. The present work shows the chaos potentials in a recurrent ANN model in comparison with two other conventional ANN models. The potentials are such as a perfect pattern recognition accuracy and an excellent resilience to noise. In addition, the model proposed in this paper faces two aspects in the biological plausibility formerly mentioned; the resilience to noise, and as a matter of fact that chaotic properties are actually present in the brain.

In the present work, three different ANN models are investigated as pattern recognition systems in two different real-life classification domains: substance identification and breast cancer identification.

1.1. Substance identification

Sensing technology encloses various instrumentation techniques for variable characterization in diverse aspects of human life [33]. From a hybridization of chemical and physical measurement devices, results the construction of biosensors. Those devices are capable of converting a chemical or physical characteristic of a particular analyte into a measurable signal [34]. Those devices offer a great potential for several integrated applications for rapid and low-cost measurement and were widely used in contrastive scientific practice, certainly owing to their remarkable outcome [33-39]. Biosensors were developed throughout various applications and principally put an accent on the construction of sensing components and transducers. Those applications fall under a multiple substance analysis for diagnosis [40-44], and estimations [45-48].

The technology of sensor devices led to remarkable achievements that are undeniable. Except that, it remains important in any sensing process to create a steady and precise pattern recognition model admitted to the sensory system for substance detection [33]. The raw data that are collected from the sensor need to be analyzed. For that purpose, and to offer a complete integrated instrument, the classical method has suggested incorporating optical filters to the basic sensor, also, the use of statistical and threshold-value based techniques for data processing. Recent researches offer a potential and more efficient alternative: the use of AI paradigms. Plenty of applications can be found in the literature that use these pattern recognition methods for substance detection such like, Decision Trees [38], random forest [34, 43], K Nearest Neighbor [34, 40, 42], Support Vector Machine (SVM) [40-43], and ANNs [36-47].

The authors use in [47] a biochemical sensor to acquire fluorescence measurements from a variety of substances at different concentrations. The sensor prototypes utilized Light Emitting Diodes (LEDs) as excitation sources, as detailed in [47], and LEDs and/or photodiodes as photodetectors.

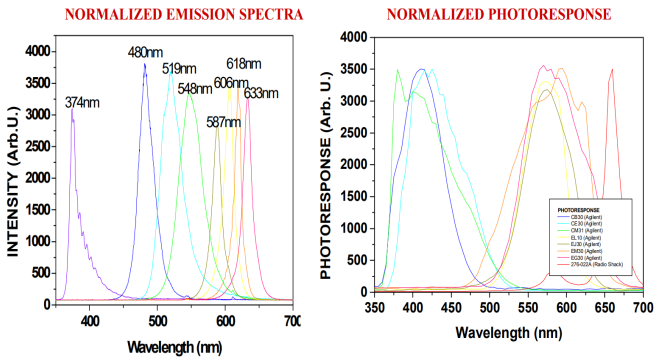


Figure 1: Sample mission and photo response spectra of the color LEDs [47]

Basically, the excitation light procured an interaction with the tested analyte of various aspects such as fluorescence, reflection, absorption, and scattering in a synchronous manner. The resulting light was quantified by the photodetectors having distinct spectral sensitivities. The LEDs were excited synchronously, one at a time, detecting the resultant fluorescence with the remaining LEDs. That was the process followed to collect an amount of data for each analyte at a specific concentration .This process generated a data collection that characterizes a singular spectral signature for a compound at a specific concentration, as illustrated in Figure 1. That process was repeated for all components at different

concentrations. Then, after detecting and amplifying the data with a multi-wavelength sensor front end, they are used to train the ANN and, upon satisfactory training, the network is assigned to the identification of other data collected by the biosensor. The ability to determine very low substance concentration levels using the ANN dramatically increases the specificity of the biochemical sensor.

The focus of classification techniques is, for given input patterns, to detect target classes, determined to define a particular substance concentration pairs. In this context, the authors in [47] developed a MultiLayered neural network, that was trained with the collected data from the biosensor, to process the classification phase on data reserved to test the network. The topology of the network model is basically a Multi-Layered Neural Network (MLNN) which consists of two hidden layers with 56 processing neurons for each layer, and a single output neuron. The RProp algorithm was employed as a learning rule on the network to process the training phase. The learning algorithm in multilayered network models consists basically of two phases. At the beginning, an input pattern is randomly selected from the training dataset and is assigned to the input layer of the ANN. Then, the network propagates that pattern from layer to layer until a corresponding output pattern is computed by the output layer. In case there is a difference between the resulting pattern and the desired output, the error is estimated and then propagated in the opposite direction through the network, from the output to the input layers. In the meanwhile, the weights' values are readjusted, as the error value is propagated backward [16].

In [47], the authors developed the MLNN to detect four fluorescent organic compounds at different concentrations, as one can notice on Table 1. The resulting performance attests a good classification capacity of the network, reaching more than 94% of perfect analyte detection. The error curves for both the training and the recall phases are plotted bellow in Figure 2.

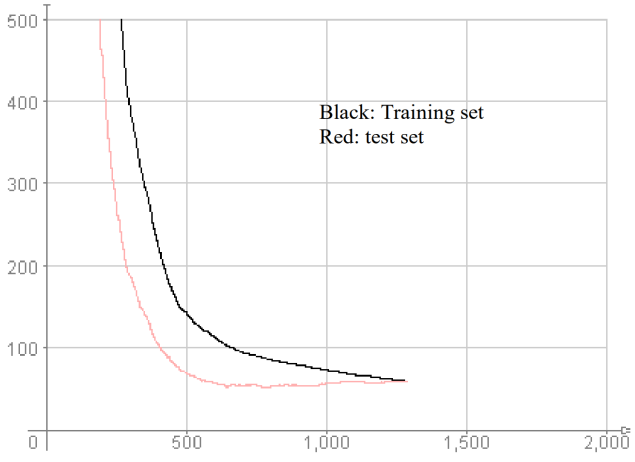


Figure 2: Sum-of-Squares classification error curves versus the number of cycles for the 56-56-56-1 MLNN topology.

Dealing with the same substance identification problem, the authors in [48] developed an evolutionary AI approach, based on Particle Swarm Optimization (PSO), viewing the detection problem as an optimized search. Indeed, the same datasets used in [47] were employed in the experimental set-up. The obtained results with the PSO model enhanced the recognition accuracy reaching 98% of exactitude.

The results of the latter two works and the same data collection were reserved in the present work, with the aim of trying to enhance even more the recognition rates obtained, by incorporating the properties of chaos in ANN models. We investigate a BAM and a CBAM recurrent models to process fluorescence data for the substance identification task.

1.2. Breast cancer detection

Cancer is a disease that might attack numerous human organs. A scourge that continues spreading all over the world with alarming new statistics each year. Statistics report that, breast cancer is the 2nd dangerous disease all over the world, in fact, the rate of mortality from this disease is overwhelming.

The WHO (*World Health Organization*), states that breast cancer affects more than 2 million women every one year across-the-board [49, 50]. In 2020, 2.3 million women were diagnosed with breast cancer and globally 15% of all death among women from cancer was from breast cancer disease. An early diagnosis of the disease, increases the chances of survival for the patient.

The main purpose of medical diagnosis aid systems for breast cancer disease is the detection of non-cancerous and cancerous tumors [49-53]. The only valid prevention approach for breast cancer disease remains the early diagnosis [54-56]. In the 1980s, in developed countries, with the establishment of early detection protocols and a set of treatment processes generate enhancements in survival rates. For a prevention purposes, the National Breast Cancer Foundation (NBCF) prescribe a mammogram once a year for women that are over 40 years old.

Technologies based on AI paradigms are getting more accurate and reliable results than conventional ones. AI tools such as pattern recognition techniques [49, 51, 53, 54, 55, 56], are estimated for being of great help in the medical diagnosis aid field. In fact, as part of breast cancer detection, doctors need to be able to differentiate between categories of tumors through a reliable procedure of examination. Specialists assess the fact that tumors' diagnosis is a task that is considered to be very hard to accomplish. It is thus crucial to diagnose breast cancers in an automated manner to overcome that difficulty.

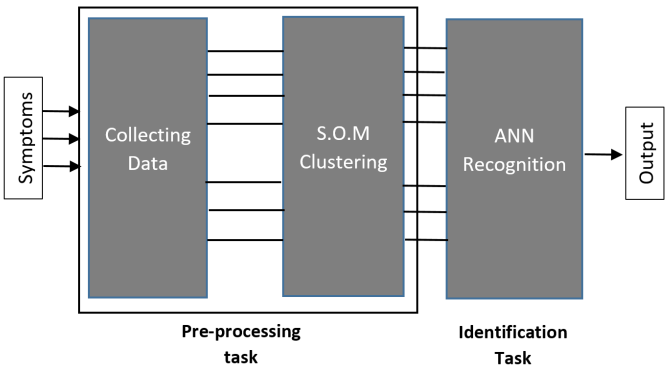


Figure 3: ANN for medical diagnosis aid system

In the field of breast cancer detection, numerous paradigms were employed to construct medical diagnosis aid systems. Those techniques use pattern recognition tools as Random forest [49-55], K-Nearest-Neighbor [52, 54], Logistic regression [51, 53], Naïve Bayes [49, 52, 53, 56], Decision Tree [49, 51-55]; not only but also ANN models [49-55], in particular, Support Vector Machines (SVM), Multi-Layer Neural Networks, Convolutional Neural

Networks, or BAM neural network model. All those techniques have the same aim: the automation of breast cancer identification to assist medical diagnosis protocols.

As for the first pattern recognition task, the substance identification problem, the same three different ANN models are used in this work to face a breast cancer identification problem, such as, a MLNN, a BAM and a CBAM model. Indeed, a different real-life pattern recognition problem is investigated with a different dataset to highlight the potential of chaos in an ANN model's performance. The proposed ANN models are operating as breast cancer diagnosis aid systems according to the process shown in Figure 3. First, symptoms are collected from the diagnosed patients, thereby creating a raw database. Subsequently, raw data undergoes a preprocessing phase before being assigned to the ANN model for the identification phase. The preprocessing phase is detailed in the *Data Acquisition* section.

The rest of this paper is organized as follows. Section 2 proposes briefly an overview about models theory. Section 3 defines the properties and the pre-processing methods used on the two different datasets employed in the experimental set up, such as, the fluorescence based measurements and the breast cancer dataset. Section 4 presents the different parameters' details and implementations' descriptions of the developed ANN models with their respective results. The obtained results are discussed in the last section, the conclusion.

2. Theory of Models

Among the various AI techniques applied in pattern recognition problems, connectionist approaches proved their good ability to process classification, which represents the most active field in the research on ANNs [16, 18, 57]. We present in the rest of this section an overview of research works employing multilayered neural network models and memory-based ones. The kind of ANN models that we are about to present in this paper.

2.1. Multi-Layered Neural Network (MLNN)

The MLNN is built in a multilayer Perceptron fashion. The architecture of such a model is composed of an input layer, one or more hidden layers of computing neurons, and an output layer generating an output pattern corresponding to the input assigned formerly to the network [18]. The MLNN performs a supervised learning. In addition, most of MLNNs processes their learning phase according to the basis of the pioneer backpropagation (Backprop) algorithm or one of its variants. In fact, an error is calculated according to the difference between an actual and a desired output patterns, which is propagated backward again through the network layers and the values of the weights are then updated to reduce it [16].

Basically, the backpropagation algorithm represents a chain rule to estimate the effect of each weight value in the network according to an arbitrary error-function [58]. Once all the weights of the connections are computed, the purpose is to make the error value as small as possible through an error-function. The commonly used error-function is the simple gradient descent including a learning rate parameter. The choice of that parameter has a considerable impact on the number of learning epochs needed for the ANN to converge. On the one hand, the smaller is the learning parameter the greater is the number of learning cycles. On the other, under a large value of the learning rate, the network risks to generate oscillation, which makes difficult to diminish the error value.

Given the limits of the classical Backpropagation rule, it has gone through several improved versions [16, 18, 58, 59], among which, introducing a momentum-term. A parameter that was supposed to make the learning algorithm more stable and the learning convergence quicker. However, it turns out experimentally that the optimal value of the momentum parameter is equally problem dependent as the learning rate, and finally, no general improvement can be carried out.

Later, numerous algorithms were proposed to face the problem of appropriate update of the weight values by using an adaptation parameter in the training epochs. That parameter is used actually to estimate the weight-step. The adaptation algorithms are approximately grouped into two classes, local and global rules [58]. On the one hand, local adaptation rules employ the partial derivative to adjust weight-specific parameters. On the other hand, global ones employ the information concerning the state of the whole ANN model, meaning, it uses the orientation of the previous weight-step, to update global parameters. Basically, the local rules fit better the conception of learning in ANNs. Again, the adaptive enhanced version of the Backpropagation algorithm has certain limitations. Concretely, the impact of the chosen value of the adapted learning parameter is very sensitive to the partial derivative [16, 18, 58].

Finally, the weaknesses of all the aforementioned backpropagation variants took over the conception of the Rprop. The fact that this algorithm updates the size of the weight-update directly and without taking into account the partial derivative's size, keeps the system away from the '*blurred adaptivity'* phenomenon. All the modified verisons of the backpropagation algorithm have the aim to accelerate the neural network convergence, and through various experimentations, the Rprop have proven to be more useful than the others. The Rprop learning scheme offers a great efficacy compared to the classical backpropagation algorithm and its above-mentioned modifications [58]. Basically, that learning rule processes the weight-step adaptation according to a local gradient information. In addition, the contribution made by the Rprop rule is that, the introduction of an individual update-value for each weight avoids the effort of adaptation to be blurred by the gradient behavior. Basically, the individual update-value determines the size of the weight-update. The value of the update parameter is estimated while the training is processed on the network, and that estimation is based on its error-function local information.

Through a set of experimentations on several MLNN models [47], the global error performance of the MLNN employing the RProp algorithm was the smallest. In addition, the speed of convergence of the model was the quickest one. Consequently, the resilient backpropagation is implemented in the learning phase for the MLNN model in the present work. Apart from that, the sigmoid function [59] is employed as the neuron output function of the multilayered model.

2.2. Bidirectional Associative Memory (BAM)

The first ANN model offering a learning process operating in a heteroassociative scheme, was proposed in the 1970s [19]. That ANN was designed with a focus on constructing a formal system that demonstrates the way that brain associates different patterns.

In fact, when training the brain perceive something (input pattern), another one is recalled (output pattern, or a category). The pioneer memory-based ANN model is linear [19], whether through its training rule or its the unit activation function. That in fact limits the recall capacity of the network, especially in case correlation occurs in input patterns. That weakness of the model has pushed the research on the field to evolve towards the construction of recurrent auto-associative and interpolative models of memory that includes nonlinearity [20]. The later models generate dynamic functionalities through a nonlinear output feedback function. That function nonlinearity offers to the system the possibility to converge to stable fixed-points. Consequently, if the networks learned training patterns in correspondence with given fixed-point attractors, it would be capable of recalling them despite the presence of noise in data. The author in [60], used those characteristics to incorporate the nonlinear feedback of the Hopfield model to a hetero-associative memory model. Finally, the BAM came up, a new kind of brain-inspired connectionist models.

A Multilayered ANN model has different functionalities from a memory-based ANN model. Instead of propagating the signal in a layer-by-layer fashion from the input layer to output layer, the BAM consists of generating feedback loops from its output to its inputs; in fact, this model has a recurrent topology [18]. In addition, the learning process in memory-based ANN models are a brain-inspired artificial approach. Indeed, that learning process allows to develop attractors for each pattern since the recurrent architecture offers the ability of feedback connections [21]. Furthermore, learning with BAMs demonstrates a remarkable stability and adaptability against noise and a great capacity of generalization. The memory-based model has also exhibited a great potential for pattern recognition especially given its capacity to be trained under a supervised or an unsupervised scheme. The Hebbian rule is the common learning algorithm used for the unsupervised trainings in the BAM models [16]. If two units in the network are activated in a simultaneous way on either side of a connection, the corresponding weight is then increased. Otherwise, if two units in the BAM are activated in an asynchronous way on either side of a connection, the corresponding weight is thus decreased. Concretely that is the Hebb's law basis. Being the fact that the Hebbian learning consists of an unsupervised learning, the process is local to the network and is performed independently of any external interaction.

The BAM training is originally performed with a classical Hebb's law [16, 20]. Because of its multiple limitations, among which, pattern-correlation, there were numerous enhanced versions of the hebbian learning principle. The first memory-based model employed a nonlinear activation function (the Signum function) in the recall phase. Again, the latter learning rule had some limitations such as it is accomplished offline and the network is limited to bipolar/binary input patterns. In addition to, the BAM generates numerous inaccurate attractors and its memorization aptitude is limited.

To confront those limits, the learning algorithm was modified with the use of a projection matrix following the principle based on least mean squared error minimization. Other alternatives were put forward in the literature tempting to enhance the learning algorithm behaviors. In fact, the proposed models tried to overcome the classical learning rule by increasing the model's storage capacity and his performance, but also by reducing the number of inaccurate states. Unfortunately, most of the proposed processes increases the neural network complexity [20].

In the present work, the authors use a BAM model that allows either an offline or an online learning of the patterns, and most of all, a model that is not limited to memorize binary or bipolar patterns. Indeed, the memory-based ANN model has to be capable of learning real-valued to deal with both of our real-life problems, substance identification and breast cancer detection. In the present work, the learning rule used in the BAM model is derived from the Hebbian/anti-Hebbian rule detailed in [43, 44].

$$W_{[k+1]} = W_{[k]} + \eta(y_{[0]}x_{[0]}^T + y_{[0]}x_{[t]}^T - y_{[t]}x_{[0]}^T - y_{[t]}x_{[t]}^T \quad (1)$$

$$V_{[k+1]} = V_{[k]} + \eta(x_{[0]}y_{[0]}^T + x_{[0]}y_{[t]}^T - x_{[t]}y_{[0]}^T - x_{[t]}y_{[t]}^T \quad (2)$$

W and V in equations (1) and (2) are the weight matrices for both network directions, $x_{[0]}$ and $y_{[0]}$ are the initial inputs to be associated. The variable η represents the training parameter, whole k represents the number of learning cycles. Through $x_{[t]}$ and $y_{[t]}$, a feedback from a nonlinear activation function is included in the learning algorithm; which offers to the network the ability to learn online and then contributes to the convergence of the BAM's behavior. Given those particularities, we opt to develop this learning function on the BAM model in the present work.

It is worth noting that, the cubic map detailed in [20, 22], is used as the unit output function of the memory-based model. The cubic map is employed for the BAM model under a non-chaotic mode, as detailed in section IV.

The training process of the BAM was performed under the basis of the following algorithm:

1) *Selecting randomly a pattern pair from the learning dataset;*
2) *Computing X_t and Y_t according to the output function employed (Cubic-map);*
3) *Computing the adjusted values of the weight matrix according to (1) and (2);*
4) *Reiteration of steps 1) to 3) until the weight marix converges;*

This same learning rule is used further in the third ANN developed model, the C-BAM.

2.3. Chaotic Bidirectional Associative Memory (C-BAM)

Since chaotic patterns were found in the brain [11, 12, 13, 14, 24, 25], numerous research works were proposed in the literature, tempting to include dynamic properties in ANN models. It must be noted that, time and change are the two properties that particularly defines the impressive properties of the NDS approach. As a result, those proposed NDS-based models are confronting the classical doctrines on brain functionalities and most of the theories that were assessed since the inception of neuroscience and cognitive sciences. This comes up with, challenging also the disciplines that focus on the construction of brain-inspired models such like artificial intelligence, and specifically ANNs paradigms.

Furthermore, most of proposed models are of a computational nature and leave dynamic principles aside. Moreover, concerning the models that encompasses NDS characteristics, only fixed points are taken into account to store and retrieve information. As a result, characteristics of the NDS approach are kept aside [46, 50]. Basically, perceived as a nuisance, chaos is, most of the time, excluded from the models.

The main purpose in the construction of brain-inspired AI models, particularly ANNs; cannot however ignore principles of nonlinear dynamical systems.

Moreover, in spite the various existing models in connectionism, not all those models are applicable to include NDS properties. Indeed, memory-based models such as BAMs offer the ability to develop nonlinear and dynamic behaviors. In fact, their recurrent architecture offers characteristics that allow an ANN model to generate oscillations [20, 21, 22, 23, 24, 25].

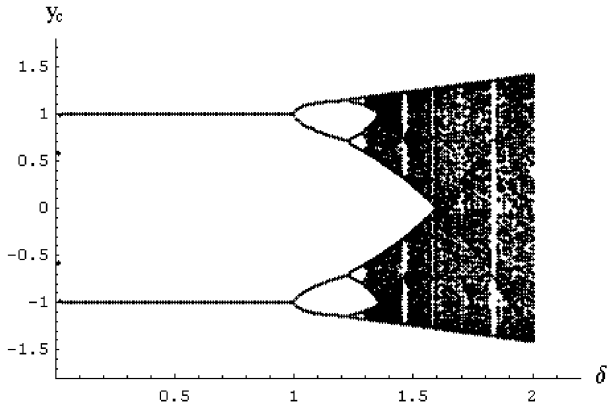


Figure 4: Bifurcation diagram of the cubic map [20].

The parameters employed in the CBAM model are as follows: the learning rule employed in the former BAM model, derived from the Hebbian/antiHebbian algorithm; also, the BAM's topology is kept the same; and concerning the output function, 23 chaotic maps (including the former cubic-map) operating in a chaotic mode are used after training, at the recall phase.

Chaotic maps characteristics

We test a set of 22 chaotic function defined in detail in [2, 61, 62, 63, 64, 65], among which, the Spikin maps family, the Tent maps, the Mira group, the Bernoulli map, and the Henon map. The last map is the one that completes the CBAM model's performance to perfect, at performing both the substance identification task and the breast cancer identification problem, as detailed further in this paper in the Experimentation section.

The 23rd function is the same cubic map used in the former BAM model, except that we set its parameters this time so that it can work in a chaotic mode. The function parameters are set according to its bifurcation diagram in the Figure 4, as detailed further in section IV. As one can notice, the process is capable of leading the system to stable attractors for value $\delta < 1$, although an aperiodic behavior can occur when the parameter value exceeds 1, and then, switching the system into a chaotic phase (black areas in the bifurcation diagram). The 23 output functions tested on the CBAM model are listed in Table 2.

The Henon map

This function is detailed in [2] as follows:

$$f(x, y) = (\alpha - x^2 + \beta.y, x)$$

The Henon map has to inputs, such as x and y, and two outputs, the new values of x and y [2]. The use of the control parameters' values α = 1.28 and β = -0.3, we notice that the orbit converges to a 2-period attractor (as shown in Figure 5), when at the value α=1.4 the attractor becomes fractal [2].

Table 1: The performances of the CBAM with the 23 maps, the BAM and the MLNN model.

ANN Model	Breast Cancer Recognition	Substance Identification	Map ID.
CBAM-Henon	100%	100%	1
CBAM-Bernoulli	100%	75.35%	2
CBAM-Logistic3	98.59%	77.32%	3
CBAM-Mira1	98.59%	72.16%	4
CBAM-Spikin Map3	96.48%	76.83%	5
CBAM-Spikin Map	95.43%	76.66%	6
CBAM-Logistic2	94.38%	74.89%	7
CBAM-Tent	91.92%	75.78%	8
CBAM-Logistic	89.47%	68.77%	9
CBAM-Spikin Map2	84.90%	76.78%	10
CBAM-PWAM2	77.89%	55.93%	11
CBAM-TailedTent1	75.78%	53.52%	12
CBAM-Logistic1	72.62%	51.95%	13
CBAM-PWAM4	70.17%	49.23%	14
CBAM-PWAM3	65.96%	46.34%	15
CBAM-Tent1	62.45%	44.87%	16
CBAM-PWAM1	53.32%	40.21%	17
CBAM-Logistic-Cubic	44.21%	72.98%	18
CBAM-Mira2	22.45%	31.44%	19
CBAM-Spikin Map1	18.94%	45.76%	20
CBAM-Ideka	18.94%	45.89%	21
CBAM-Mira-Gumolski	12.27%	23.91%	22
CBAM-Tent2	4.56%	18.67%	23
BAM	96.56%	90.17%	---
MLNN	89.32%	94.79%	---

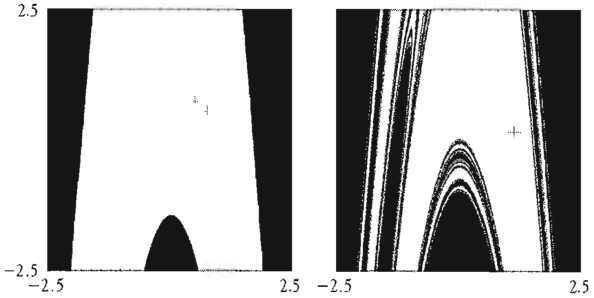


Figure 5: The attraction bassin of the Henon map, β = -0.3 [65].

On the one hand, the black zones in the Figure 5 represent the initial values whose trajectories diverge towards infinity. On the other hand, the white zones represent the initial values that are pulled by the 2-periods attractor. The basin limit is a curve that moves from the inside to the outside of initial values' space.

The Bernoulli map

This chaotic function is defined in [31] as follows:

$x[n] = y[n-1]$,

$y[n] = mod(\delta.x[n-1], 1)$

Where δ is a control parameter set to 1.99 so the map can operate in a chaotic mode [64]. *x* and *y* represent the input and the output of the system respectively.

The Mira maps group

The original Mira map is defined as follows [61]:

$x_{[n]} = y_{[n-l]}$,

$y_{[n]} = y_{[n-l]} - ax_{[n-1]}$ if $x_{[n-1]} < 6$,

$y_{[n]} = y_{[n-l]} + bx_{[n-1]} - 6\ (a + b)$, otherwise

where *x* and *y* represent the initial conditions for the trajectory of the map. After numerous trials on the values of the control parameters of the Mira map, we set them during the final experimentation to: a=1.05; and b=2. The modified Mira map is used to imitate the spiking phenomenon of the biological neurons [32], this map is defined as follows:

$f(x) = y$,

$f(y) = a\,x + b\,x^2 + y^2$

We use this map with the following control parameters' values:

- Mira_map1, a = 0.8, b = 1, where the map has one breast cancer fixed-point with two positive eigenvalues.
- Mira_map2, a = -0.8, b = 0.2, where the map has a stable set.

The dynamic characteristics of the Mira map and its versions are detailed in [47, 48].

The Spiking maps group

The original Spiking_map is defined in [63] as follows:

$$X_{[n]} = f\left(X_{[n-1]}, Y_{[n-1]}\right),$$

$$Y_{[n]} = Y_{[n-1]} - \mu\left(X_{[n-1]} + 1\right) + \mu\sigma,$$

$$f(x,y) = \begin{cases} \dfrac{\alpha}{(1-x)} + y, & if\ x \le 0 \\ \alpha + y, & if\ 0 < x < \alpha + y, \\ -1, & if\ x \ge \alpha + y \end{cases}$$

Where, $x_{[n]}$ is the fast dynamical variable, μ is a constant value set at 0.001, $y_{[n]}$ is the slow dynamical variable, its moderate evolution is on account of to the small value of the parameter μ. The map's control parameters are the variables α and σ. We use also the Spiking Map with three different modifications. We experiment the parameters' values used in [63] so that the map operates in a chaotic mode:

- Spiking_Map1: $\mu = 0.001$; $\alpha = 5.6$ and $\sigma = 0.322$
- Spiking_Map2: $\mu = 0.001$; $\alpha = 4.6$ and $\sigma = 0.16$
- Spiking_Map3: $\mu = 0.001$; $\alpha = 4.6$ and $\sigma = 0.225$

In the present work, we aim at experimenting 23 chaotic maps with the CBAM model, among which we have presented few ones. These selected functions are among the ones that have laid the best accuracy rates.

Furthermore, we have relied on useful mathematical tools to set the different chaotic maps' parameters, among which, the Lyapunov exponent. It represents a logarithmic estimation for the mean expansion rate per cycle of the existing distance between two infinitesimally close trajectories [2]. The interest in the present work concerns particularly the case where that value is positive. It must be noted that, a NDS with a positive Lyapunov exponent characterizes the fact that the system is chaotic. That system is in particular sensible to initial conditions. Figure 6 shows a bifurcation diagram of the Henon map.

For each vertical slice shows the projection onto the *x*-axis of an attractor for the map for a fixed value of the parameter α.

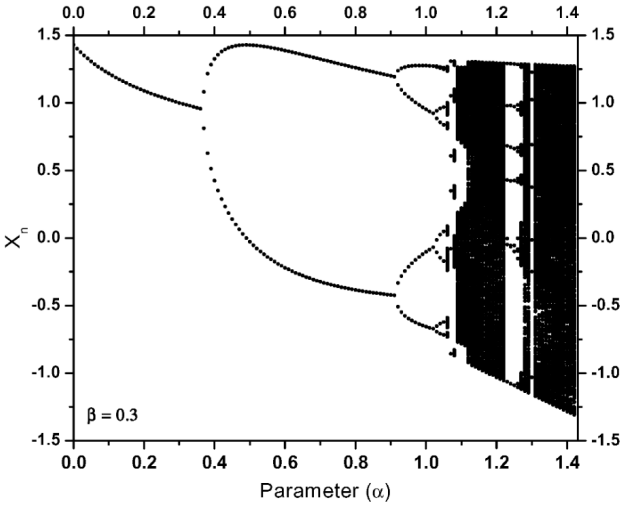


Figure 6: The bifurcation diagram for the Henon map [65].

Whereas, a stable movement has a negative Lyapunov exponent. An example is illustrated in the Figure 7, which concerns the Henon map Lyapunov Exponent.

Both of those mathematical tools, the bifurcation diagram and the Lyapunov exponent, are useful in our experimental set up to have control over the dynamical behaviors of the performed output unit functions.

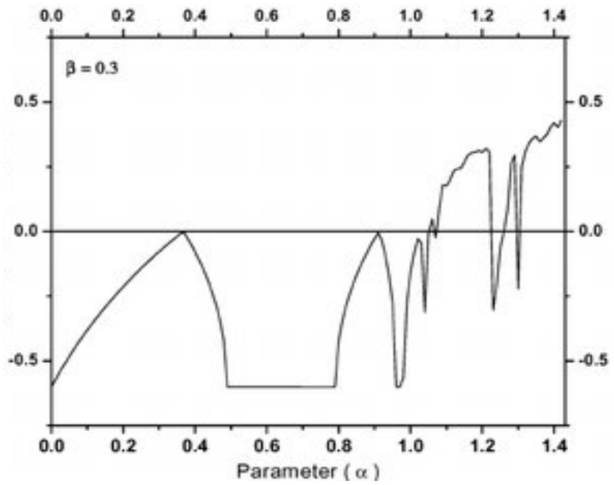


Figure 7: The Lyapunov exponent [65].

3. Data Acquisition

It is a common routine to prepare data before performing the learning and the recall phases on an ANN model. Numerous preprocessing techniques exist in the literature varying according to the nature of the data. We detail in the following subsections the preprocessing procedure applied on both the set of fluorescence-based measurements, and the breast cancer data collection. It is worth noting here that, noise and missing data are two characteristics that are kept in the datasets.

3.1. Fluorescence based measurements

A set of fluorescence-based measurements was collected from the optoelectronic biosensor for each analyte at a specific concentration through the procedure detailed in [47]. The first pattern recognition problem to aboard in the present work concerns the identification of different analytes at different concentrations. Table 1 illustrates the test compounds under their different concentrations.

Table 2: Test substances with their concentrations

Compound	Concentration	Category
Chlorophyll	10-4 M	1
	10-5 M	2
	10-6 M	3
	10-7 M	4
	10-8 M	5
Coumarin	10-3 M	6
	10-4 M	7
	10-5 M	8
	10-6 M	9
	10-7 M	10
Rhomadine B	10-4 M	11
	10-5 M	12
	10-6 M	13
	10-7 M	14
Erythrosin B	10-4 M	15
	10-5 M	16
	10-6 M	17
	10-7 M	18
	10-8 M	19

We have 19 classes among which each class represents one compound at a specific concentration.

Concretely, each measurement in the dataset is composed of 64 values forming an 8x8 matrix. The row in the matrix provides the outputs of 7 photodetectors LEDs [47], and one more output corresponding to the excitation LED which is fixed to 0, and then removed leading to the resulting matrix 8x7. Consequently, the size of each vector in the data collection is 56. Furthermore, a random division of the dataset was employed to get two distinct ones, the first one is dedicated to the learning process (two thirds of the entire collection), while the second one is reserved for the testing phase (the remaining third). The resulting datasets contains 2103, and 1051 vectors respectively.

3.2. Breast Cancer database

Concerning the second pattern recognition problem investigated in the present work, the authors employ the *Yougoslavia Breast Cancer* dataset in the experimentations. Clinical data have been collected by Matjaz Zwitter & Milan Soklic at the oncological institute of the university medical center of Ljubljana in Yougoslavia. The entire data collection contain 286 tumor cases. Each tumor case is represented as a vector of 10 attributes: the tumor frequency, the patient age, the type of the menopause, the tumor-size, inv-node, node-caps, deg-malig, breast position, breast-quad, the irradiation value. It must be noted here that few attributes are missing in the breast cancer dataset. The authors apply scaling on the data collection with dividing the values by 10. The digit 0 was excluded to prevent the system from instable fixe points. As a result, the normalized dataset is represented by digits in the interval [1, 13].

Finally, the data collection is composed of 286 tumor cases coded in a set of rows of dimension 10.

Moreover, the different tumor cases in the breast cancer dataset are not identified. A categorization phase must be accomplished to determine them. As a consequent, the authors developed a Self-Organized Map (SOM), an ANN model detailed in [66]. The designed SOM model consists of a map that is composed of 400 neurons (20*20 cells). Computing units are represented through the matrix cells. As one can notice in Figure 8, the numbers displayed in the 73 cells consist of the resulting categories' identifiers processed by the SOM network.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
1	16	28	94		159	172	147	47	148	56	9	100	78	97		91			89	115
2	19			104		98	93		107			27	80	1		149	87	156		90
3	12		163					20	22		69	133				143	92	162		166
4	7	38		18	74		55	129	23		135	110			175	112				105
5		70	68		102		17			76	32		36			101	154	127		145
6	41		165			146		15	117											103
7	174		26	108		122	88		111		21	150	120				167	86		153
8								24	161	168	72		79		71	109				126
9	54				119								134	4		53		171		164
10		57	4		66			77									62			158
11	106			65		29		10	114	3	81		2	67						169
12	142	152	125	113			116	40			43	42		128		45	96			140
13	46				49	31	130				75		51	59			50			141
14						64			25					11					137	
15							84	37	60		35	96		5	33	39	85	34		
16							13						48					82		131
17			52					8	121	61	30	63						58		
18								44										83		
19							6		139		73	173	123		160	99				155
20								132	136	138	170		144	118	157	124		151		

Figure 8: The SOM breast cancers' categorization

The obtained map indicates that each neuron in the network has the medical specificities of the category that it represents. As a result, the 175 classes (plotted in Figure 8) generated by the SOM are used for breast cancer identification.

For the experimentation needs, and according to the procedure followed in the substance identification task, the breast cancer dataset was separated into two subsets. The first set contains the equivalent of two-thirds (191 vectors), while the second set contains the remaining third (95 vectors). The first dataset is reserved for the training phase, and the second for the recall phase.

The cross-validation method is used to determine the accuracy rates for the different ANN models developed, and this, whether in substance identification task or in the breast cancer detection problem.

4. Experimentations and Results

It is worth noting that, the topology, the activation function and the learning rule are the main parameters that characterizes an ANN model. We detail in the following subsections those parameters in each of the MLNN, the BAM and the C-BAM models.

4.1. MLNN model

The authors in [47] investigate a set of experimentations on several MLNN architectures and different parameters and converged on the multilayered model employed for the substance identification task. The Stuttgart Neural Networks Simulator (SNNS) of the Stuttgart University in Germany was used for the different MLNN experimentations. The best network performance obtained was with the MLNN topology consisting of an input layer of the size 56 according to the size of the input vectors, two hidden layers with the same size as the input layer, and one output unit generating the output pattern corresponding to input assigned to the network. The aforementioned learning algorithm was used as the learning rule, the RProp. The sigmoid was the output function of the MLNN model.

We kept the same conception principle for the breast cancer detection task. The topology of the resulting MLNN model consists of an input layer containing 10 units according to the size of the tumor-case vector, two hidden layers with 10 computing neurons for each, and one output. The MLNN topology for the breast cancer detection task is plotted in the Figure 9 bellow.

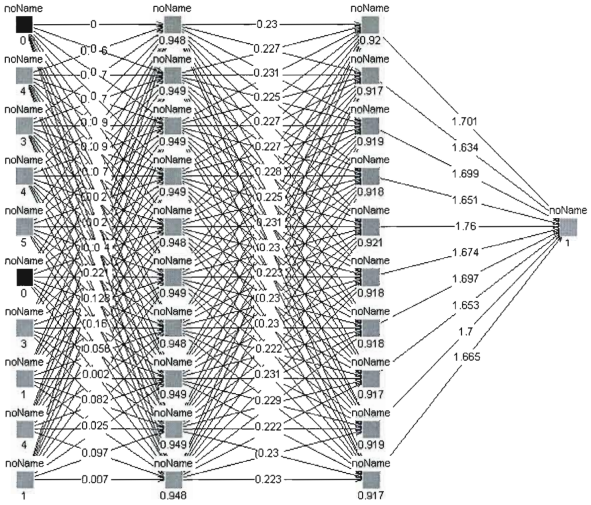


Figure 9: The MLNN topology 10-10-10-1.

The implemented ANN models are tested in the recall phase with the patterns that were not used during the learning process. Meaning that those patterns were not affected to the network during its learning phase, but during the recall only. In addition, the cross validation technique is used to determine the different ANNs' recognition accuracies. Finally, once the tests are achieved, the average of the three experimentations on each of the ANN models is considered as its overall classification exactitude. Empirically, the MLNN model reached 94,79% of good overall recognition for the substance identification task, 89.32% of exactitude for the breast cancer identification problem.

4.2. BAM model

The memory-based network architecture is the second ANN model developed in the present work and it is composed of two Hopfield-like neural networks interconnected in head-to-tail fashion, as one can notice in Figure 10. We employ in the BAM model the parameters experimented in [20, 21].

The network topology describes an interconnection that allows a recurrent flow of information that is processed bidirectionally. In that way, the vectors composing the pairs to be learned do not have to be specifically of same dimensions and that, contrary to the conventional BAM designs, the weight matrix from one side is not necessarily the transpose of that from the other side.

The unit activation function employed in the BAM model is the cubic map described in [21]. Figure 4 illustrates the bifurcation diagram of that function according to δ, the parameter that dictates the dynamic behavior of the outputs.

Fundamentally, this cubic function has three fixed points, -1, 0, and 1, of which both the values-1 and 1 are stable fixed points. They offer to the memory the possibility to develop two attractors at these values. The cubic output function takes several time steps to converge. First, the given stimulus is projected from the network space to the stimuli space. Second, in the following time steps, the stimulus is progressively pushed toward one of the stimuli space corners.

Furthermore, according to Figure 4, one can notice that the learning rule leads to stable attractors for value $\delta < 1$, when it gets an aperiodic behaviour when exceeding 1 before leading the system into a chaotic phase (black areas in the bifurcation diagram).

The experimentations on the BAM were realized with the cubic map operating in a fixed-points mode. The Hebbian/antiHebbian learning rule includes a feedback from the nonlinear output function via the couple of patterns to be associated; which allows the BAM to learn online, thus contributing to the convergence of the weight connections.

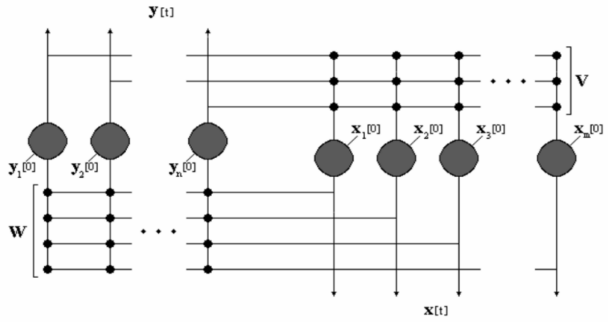


Figure 10: The BAM architecture.

The breast cancer recognition accuracy of the BAM model is plotted in Figure 11.

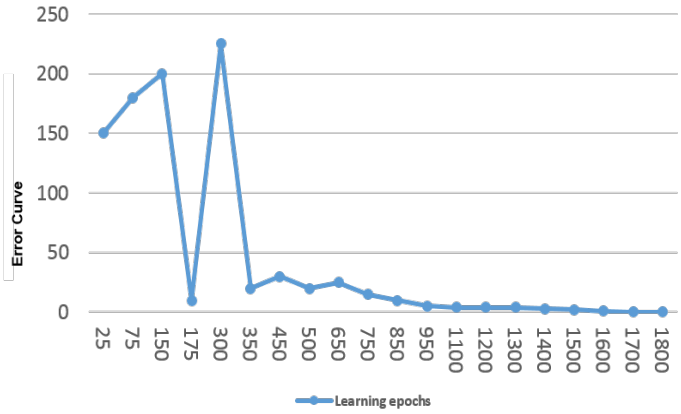


Figure 11: Error curve relative to the number of learning epochs.

The network could reach a steady state after a reasonable number of learning cycles. The error was less than 0,0005 after 1700 epochs. The BAM could correctly categorize 85 tumors cases from the testing dataset. That means that the accuracy rate of the network reached 90,17% at the recall phase. It is worth here that, that recognition rate was accomplished inspite the missing values in the data, which proves a good capacity of generalization on the one hand and a good resilience to noise on the other.

The accuracy increased with the substance identification task to 96,56 % of exactitude. That is certainly due to the better conditions of the data collection despite the larger problem space.

4.3. CBAM model

The third model developed in the present work is the chaotic BAM model. The same main BAM's parameters are kept in the C-BAM network, consisting of the same topology and learning rule as the ones of the former BAM, detailed in [20], except that, the neuron activation function was replaced by chaotic functions. Distinction between the two memory-based ANN models throughout that parameter offers the possibility to concretely estimate the network pattern recognition performance with, and without chaos. The first chaotic output function tested on the C-BAM is the same cubic map employed in the BAM model, operating in a chaotic mode during recall. For that purpose, the value of δ was set to 0.1 during training and to 1.5 during recall.As one can notice in Figure 4, those values correspond to a fixed-point and chaotic behavior respectively for that output map. Subsequently, the experimentation of 22 other chaotic maps on the CBAM model were investigated. Indeed, after reaching a steady state in the learning process, the parameters of each function were fixed so it can operate in a chaotic mode, according to the bifurcation diagram of each map [2, 61, 62, 63, 64].

The recognition accuracy rates are illustrated in Table 2 for both substance identification and the breast cancer identification tasks. For the purpose of results comparison, the MLNN's and the classical BAM's performances are also mentioned.

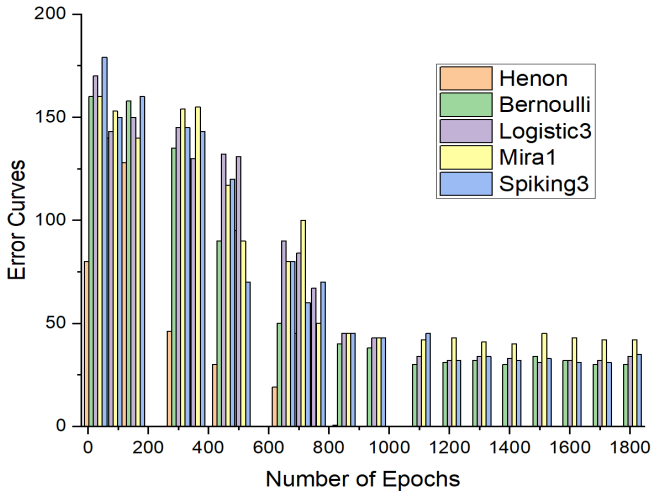


Figure 12: The CBAM error curves for the substance identification's task with the best 5 performing chaotic maps.

The Figure 12 and the Figure 13 show the best recognition accuracies of the CBAM model with five particular maps, for the substance identification task, and the breast cancer identification task, respectively.

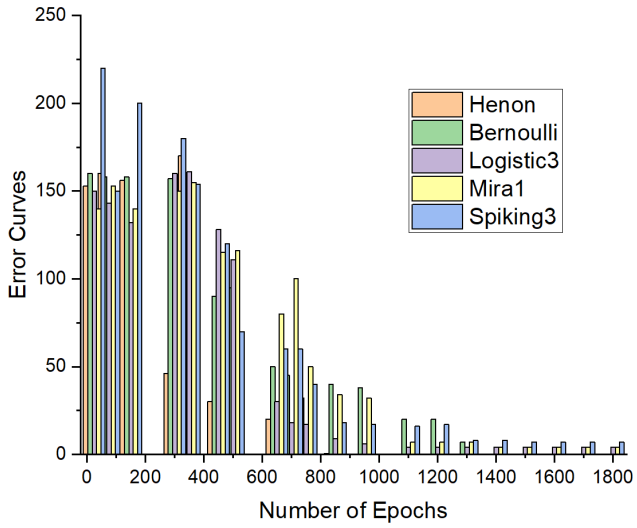


Figure 13: The CBAM error curves for the breast cancers' recognition task with the best 5 performing chaotic maps.

As one can notice in the above graphs, the accuracy is total in both problem domains, particularly with the Henon function.

5. Conclusion

Three different ANN models were developed in the present work to deal with two different real life problems. Both of those problems focus on pattern recognition, the first task concerns substance identification while the second is about breast cancer detection. On the one hand, the MLNN model reached a good performance at recalling the substance identification data despite the problem of the huge space dimension (56 is the vectors' size) and the multiclass criterion (19 different classes). The rate was less good for breast cancer identification with the same model in spite of this; the problem of space dimension was diminished to 10. This fact is indicative of the poor generalization capacity of the Multi-Layered Neural Network when data contains noise. In fact, as mentioned in the *Data Characteristics* section, noise and missing values are two properties that are kept in the datasets. On the other hand, the BAM recurrent model provides a good overall recall for breast cancer identification seeking more that 96% of exactitude. However, the error increased with the multi-class problem relative to fluorescent-based measurements. The BAM results highlight its good resilience to noise; nevertheless, it was less good at facing the large problem space of the fluorescence-based measurements.

In addition, the presence of chaos in the brain-inspired memory-based model provided remarkable results particularly with certain chaotic maps. The recognition accuracy of the CBAM facing the breast cancer detection task was acceptable with six functions employed in a chaotic mode [70% to 89%]; while eight maps varied from good to perfect, reaching a 100% of correct recognition with the Henon and Bernoulli map. The performance was less good with regard to the substance identification problem compared to the first one, and that is with almost all the chaotic maps except for the Henon map that kept the overall accuracy total. The pattern recognition system employing that particular map was remarkable dealing with both substance identification and breast cancer detection problems. Accordingly, we can state that this particular model encompasses assets which allow it an excellent generalization capacity and a great resilience to noise leading to perfect pattern recognition performance.

It is crucial to line up future artificial neural networks investigations with the dynamical characteristics of the Henon chaotic function, among which, the fractal dimensions of the Henon attractors. In essence, fractals and modern chaos theory radically question the dynamical concepts in all contexts and more particularly in nature and its mimetic artificial systems, among which, artificial neural networks. Besides, and to conclude, since it is the era of big data; chaos must imperatively be in the perspectives of deep learning techniques in artificial neural network models' analytics.

Conflict of Interest

The authors declare no conflict of interest.

Acknowledgment

This study was funded by the Algerian ministry of higher education and scientific research and the General Directorate for Scientific Research and Technological Development (DG-RSDT). This work was also possible thanks to the financial support of excellence scholarships foundation of the University of Quebec at Montreal (UQAM) for graduate studies, Montreal, Quebec, Canada. Many thanks to the Texas-center of Superconductivity and Advanced Materials (TcSAM) in the Physics Department of the University of Houston, Houston, Texas, USA, for allowing us to use their data in our experimentations.

References

[1] H. Naoum, S. Benslimane, M. Boukadoum, "Classical and Brain-inspired Neural Networks for Substance Identification and Breast Cancer Detection: The Chaos Challenge," The first international conference on Cyber Management and Engineering (CyMaEn`21), IEEE, 1–6, 2021, doi: 10.1109/CyMaEn50288.2021.9497280.

[2] K. T. Alligood, T. D. Sauer, J. A. Yorke, "Chaos: An Introduction to Dynamical Systems," Textbook in Mathematical Sciences. Springer, New York, NY, 105–147, 1996, doi:10.1007/b97589.

[3] W. J Freeman, "Simulation of chaotic EEG patterns with dynamic model of the olfactory system," Biological Cybernetics, **56**(2–3), 139–150, 1987.

[4] H. Korn, P. Faure, "I there chaos in the brain? Experimental evidence and related models," Neurosciences, **326**(9), 787–840, 2003.

[5] A. Combs, S. Krippner. W. Freeman, "III and the chaotic nature of deams," Nonlinear Dynamics, Psychology and life sciences, **21**(4), 475-484, 2017.

[6] V. V. Kozlova, V. A. Galkin, M. A. Filatov, "Diagnostics of brain neural network states from the perspective of chaos," Journal of physics: conference series. 1889 052016, 2021, doi:10.1088/1742-6596/1889/5/052016.

[7] M. AR. Thabet, "Quantum chaos and the brain," IBCHN - Imperial Collage - Michael Crawford's Lab, 2020, doi:10.13140/RG.2.2.20160.48645.

[8] B. Yan, S. Mukherjee, A. Saha, "Exploring noise-induced chaos and complexity in a red blood cell system," Springer, The European Physical Journal Special Topics, **230**, 1517–1523, April 2021, doi:10.1140/epjs/s11734-021-00030-2.

[9] N. B. Harikrishnan, N. Nagaraj, "When noise meets chaos: stochastic resonanace in neurochaos learning," Elseivier, **143**, 425–435. Special Issue, 2021, doi:10.1016/j.neunet.2021.06.025.

[10] G. Eason, B. Noble, I. N. Sneddon, "On certain integrals of Lipschitz-Hankel type involving products of Bessel functions," Phil. Trans. Roy. Soc. London, **A247**, 529–551, April 1955, doi:10.1098/rsta.1955.0005.

[11] A. Babloyantz, C. Lourenc-o, "Computation with chaos: A paradigm for cortical activity," Proceedings of the National Academy of Sciences, **91**, 9027–9031, 1994, doi:10.1073/pnas.91.19.9027.

[12] M. P. Dafilis, D. T. J. Liley, P. J. Cadusch, "Robust chaos in a model of the electroencephalogram: Implications for brain dynamics," Chaos, **11**, 474–478, 2001, doi:10.1063/1.1394193.

[13] H. Korn, P. Faure, "Is there chaos in the brain? II. Experimental evidence and related models," Comptes Rendus Biologies, **326**(9), 787–840, 2003, doi:10.1016/j.crvi.2003.09.011.

[14] M. A Rozhnova, E. V Pankratova, S. V Stasenko, V. B Kazantsev, "Bifurcation analysis of multistability and oscillation emergence in a model of brain extracellular matrix," Elseivier, Chaos, Solitons & Fractals, **151**, October 2021, doi:10.1016/j.chaos.2021.111253.

[15] A. Wu, Y. Chen, Z. Zeng, "Quantization synchronization of chaotic neural networks with time delay under event-triggered strategy," Springer Verlag, Cognitive Neurodynamics, **15**, 897–914, 2021, doi: 10.1007/s11571-021-09667-0.

[16] M. Negnevitsky, "Artificial Intelligence: A guide to Intelligent Systems," Addison Wesley, 3rd edition, 2011.

[17] K. Saravanan, S. Sasithra, "A review on Classification Based on Artificial Neural Networks," International Journal of Ambient Systems and Applications (IJASA), **2**(4), 11–18, 2014, doi:10.5121/ijasa.2014.2402.

[18] S. Haykin, "Neural networks: A comprehensive foundation," Englewood Cliffs, NJ: Prentice-Hall, 1999.

[19] T. Kohonen, "Correlation matrix memories," IEEE Trans. Comput. , **C-21**, 353–359, Dec. 1972, doi: 10.1109/TC.1972.5008975.

[20] S. Chartier, M. Boukadoum, "A bidirectional Heteroassociative Memory for binary and Grey-Level Patterns, " IEEE Transactions on Neural Networks, **17**(2), March 2006, doi: 10.1109/TNN.2005.863420.

[21] S. Chartier, M. Boukadoum, "Encoding static and temporal patterns with a bidirectional heteroassociative memory,"Journal of applied mathematics, **20011**, 1–34, 2011, doi: 10.1155/2011/301204.

[22] S. Chartier, S. Helie, M. Boukadoum, R. Proulx, "SCRAM: statistically converging recurrent associative memory," IEEE International Joint Conference on Neural Networks, IJCNN, 2005, doi: 10.1109/IJCNN.2005.1555941.

[23] S. Chartier, M. Renaud, M. Boukadoum, "A nonlinear dynamic artificial neural network model of memory," New Ideas in Psychology, **26**(2), 252–277, 2008, doi:10.1016/j.newideapsych.2007.07.005.

[24] M. Adachi, K. Aihara, "Associative dynamics in a chaotic neural network. Neural Networks," **10**(1), 83–98, 1997, doi:10.1016/S0893-6080(96)00061-5.

[25] K. Aihara, T. Takabe, M. Toyoda, "Chaotic neural networks," Physics Letters A, **144**(6–7), 333–340, 1990, doi:10.1016/0375-9601(90)90136-C.

[26] H. Imai, Y. Osana, M. Hagiwara, "Chaotic analog associative memory," Systems and Computers in Japan, **36**(4), 82–90, 2005, doi: 10.1109/IJCNN.2001.939522.

[27] R. S. T. Lee, "e-associator: A chaotic auto-associative network for progressive memory recalling," Neural Networks, **19**(5), 644–666, 2006, doi: 10.1016/j.neunet.2005.08.017.

[28] Y. Osana, M. Hagiwara, "Knowledge processing system using improved chaotic associative memory," Proceeding of the International Joint Conference on Neural Networks (IJCNN'00), **5**, 579–584, 2000.

[29] U. Ozdilek, "Value order in disorder," Springer, International Journal of Dynamics and Control, 2022, doi:10.1007/s40435-021-00903-3.

[30] H. Lin, C. Wang, Q. Deng, C. Xu, Z. Deng, C. Zhou, "Review on chaotic dynamics of memristive neuron and neural network," Nonlinear Dynamics, **106**, 959–973, 2021, doi:10.1007/s11071-021-06853-x

[31] Y. Zhang, Y. He, F. Long, "Augmented two-side-looped Lyapunov functional for sampled-data-based synchronization of chaotic neural networks with actuator saturation," Elsevier, Journal of Neurocomputing, **422**, 287–294, 2021, doi:10.1016/j.neucom.2020.09.018.

[32] C. Chen, A. Abbott, D. Stilwell, "Multi-Level generative chaotic recurrent network for image inpainting," Proceedings of the IEEE CVF winter conference on applications of computer vision (WACV), 3626-3635, 2021, doi: 10.1109/WACV48630.2021.00367.

[33] H. Kaur, A. Bhosale, S. Shrivastav, "Biosensors: Classification, fundamental characterization and new trends: A Review," International Journal Of Health Sciences and Research, **8**(6), 315–333, 2018.

[34] M. H. Mozaffari, L. Tay, "A review of 1D Convolutional Neural Networks toward Unknown Substance Identification in Portable Raman Spectrometer," arXiv:2006.10575 [eess.SP] (2020).

[35] R. Fleh, M. Othman, S. Gomri, "WO3 sensors array coupled with pattern recognition method for gases identification," 13th International Multi-Conference on systems, Signals and Devices, IEEE, 147–152, 2016.

[36] D. Karakaya, O. Ulucan, M. Turkan, "Electronic nose and its applications: A survey," International Journal of Automation and Computing, **17**, 179–209, 2020, doi:10.1007/s11633-019-1212-9.

[37] B. Podola, M. Melkonian, "Genetic programming as a tool for identification of analyte-specificity from complex response patterns using a non-specific whole-cell biosensor," Biosensors and Bioelectronics **33**, 254-259, 2012, doi: 10.1016/j.bios.2012.01.015.

[38] M. Kukade, T. Karve, D. Gharpure, "Identification and classification of spices by Machine Learning," IEEE International Conference on Intelligent Systems and Green Tchnology (ICISGT), 2019, doi:

10.1109/ICISGT44072.2019.00015.
[39] A. L. Vazquez, M. M. Domenech Rodriguez, T. S. Barrett, S. Schwartz, N. G. Amador Buenabad, M. N. Bustos Gamino, M. L. Gutierrez Lopez, J. A. Villatoro Velazquez, "Innovative Identification of Substance Use Predictors: Machine Learning in a National Sample of Mexican Children," Journal of Society for Prevention Research, Springer-Verlag, 2020, doi: 10.1007/s11121-020-01089-4.
[40] F. L. Melquiades, A. Mattos Alves da Silva, "Identification of sulphur in nail polish by pattern recognition methods combined with portable energy dispersive X-ray fluorescence spectral data," Journal of Analytical Methods, **8**, 3920–3926, 2016, doi:10.1039/C6AY00195E.
[41] Z. Almheiri, M. Meguid, T. Zayed, "Intelligent Approaches for predicting failure of water mains," Journal of Pipeline Systems Engineering and Practice, **11**(4), 1949–1190, 2020, doi:10.1061/(ASCE)PS.1949-1204.0000485.
[42] F. Hu, M. Zhou, P. Yan, K. Bian, R. Dai, "PCANet: A common solution for laser-induced fluorescence spectral classification," IEEE Access **7**, 2169–3536, 2019, doi: 10.1109/ACCESS.2019.2933453.
[43] L. G. Zhang, X. Zhang, L. J. Ni, Z. B. Xue, X. Gu, S X. Huang, "Rapid identification of adulterated cow milk by non-linear pattern recognition methods based on near infrared spectroscopy," Food Chemistry **145**, 342–348, 2014, doi:10.1016/j.foodchem.2013.08.064.
[44] P. T. Hernandez, S. Hailes, I. P. Parkin, "Cocaine by-product detection with metal oxide semiconductor sensor arrays," Royal Society of Chemistry, **10**, 28464–28477, 2020, doi:10.1039/D0RA03687K.
[45] Y. Hui, X. Xue, Z. Xuesong, W. Yan, Z. Junjun, "Bacteria strain identification with fluorescence spectra," Applied Mechanics and Materials, **865**, 630–635, 2017, doi:10.4028/www.scientific.net/AMM.865.630.
[46] L. Poryvkina, V. Aleksejev, S. M. Babichenko, T. Ivkina, "Spectral pattern recognition of controlled substances in street samples using artificial neural network system," Optical Pattern Recognition, Proceedings of SPIE, **8055**, 2011, doi:10.1117/12.883408.
[47] H. Naoum, M. Boukadoum, C. Joseph, D. Starikov, A. Bensaoula, "Intelligent Classifier Module for Fluorescence Based Measurements," Proc. International Workshop on Signal Processing and its Applications (WoSPA 2008), Sharjah (UAE), 18–20 2008.
[48] N. Nouaouria, M. Boukadoum, "A Particle Swarm Optimization Approach for Substance Identification," The Genetic and Evolutionary Computation Conference (GECCO), 1753–1754, 2009, doi:10.1145/1569901.1570142.
[49] A. R. Vaka, B. Soni, R. K. Sudheer, "Breast Cancer Detection by leveraging Machine Learning," The Korean Institute of Communication and Information Sciences (KICS), 10–1016, 2020, doi:10.1016/j.icte.2020.04.009.
[50] H. Jouni, M. Issa, A. Harb, G. Jacquemod, Y. Leduc, "Neural Network Architecture for Breast Cancer Detection and Classification," IEEE International Multidisciplinary Conference on Engineering Tchnology (IMCET), 987-1-5090-5281-3, 2016, doi: 10.1109/IMCET.2016.7777423.
[51] S. Sharma, A. Aggarwal, T. Choudhury, "Breast Cancer Detection Using Machine Learning Algorithms," International conferfence on computational techniques, electronics and mechanical systems (CTEMS), IEEE, 2019, doi: 10.1109/CTEMS.2018.8769187.
[52] J. Sivapriya, V. Aravind Kumar, S. Siddarth Sai, S. Sriram, "Breast cancer prediction using machine learning," International Journal of Recent Technology and Engineering (IJRTE), **8**(4), 2019, doi:10.35940/ijrte.D8292.118419.
[53] K. Santhosh, T. Daniya, J. Ajayan, "Breast Cancer Prediction Using Machine Learning Algorithms," International Journal of Advanced Science and Technology, **29**(3), 7819–7828, 2020.
[54] B. Karthikeyan, G. Sujith, H. V. Singamsetty, P. V. Gade, S. Mekala, "Breast cancer cetection using machine learning," International Journal of Advanced Trends in Computer Science and Engineering, **9**(2), 981–984, 2020, doi:10.30534/ijatcse/2020/12922020.
[55] A. E. Bayrak, P. Kirci, T. Ensari, "Comparison of machine learning methods for breast cancer diagnosis," Scientific Meeting on Electrical-Electronics and Biomedical Engineering and Computer Science (EBBT), Istambul, Turkey, 1-3, 2019, doi: 10.1109/EBBT.2019.8741990.
[56] M. Karabatak, " New classifier for breast cancer detection based on Naïve Bayesian," Measurement. **72**, 32–36, 2015.
[57] B. Li, C. Delpha, D. Diallo, A. Migan-Dubois, "Application of artificial neural networks to photovoltaic fault detection and diagnosis: A review," Elsevier, Renewable and Sustainable Energy Reviews, **138**, 2021, doi:10.1016/j.rser.2020.110512.
[58] M. Riedmiller, H. Braun, "A direct adaptive method for faster backpropagation learning: the RPROP algorithm,"IEEE International Conference on Neural Networks, 1993, doi: 10.1109/ICNN.1993.298623.
[59] J. Han, C. Moraga, "The influence of the sigmoid function parameters on the speed of backpropagation learning," Springer Verlag, International Workshop on Artificial Neural Networks, 195–201, 1995, doi:10.1007/3-540-59497-3_175.
[60] B. Kosko, "Bidirectional associative memories," IEEE Trans. Syst., Man, Cybern., **18**(1), 49–60, Jan.-Feb, 1988, doi: 10.1109/21.87054.
[61] I. Gumowski, C. Mira, "Recurrences and discrete dynamic systems," Lecture notes in mathematics, book series, **809**, 1980, doi:10.1007/BFb0089135.
[62] J. P. England, B. Krauskopf, H. M. Osinga, "Computing One-Dimensional Stable Manifolds and Stable Sets of Planar Maps without the Inverse," SIAM J. Applied Dynamical Systems, Society for Industrial and Applied Mathematics, **3**, 161–190, 2004, doi:10.1137/030600131.
[63] A. L. Shilnikov, N. F. Rulkov, "Origin of Chaos in a Two-Dimensional Map Modeling Spiking-Bursting Neural Activity," International Journal of Bifurcation and Chaos (IJBC), 13(11), 3325–3340, 2003, doi:10.1142/S0218127403008521.
[64] S. J. Baek, E. Ott, "Onset of synchronization in systems of global1y coupled chaotic maps," Physical Review Letters, 69(6), 066210 2004, doi: 10.1103/PhysRevE.69.066210.
[65] V. Patidar, G. Purohit, K.K Sud, "Dynamical behavior of q-deformed Hénon map". International journal of bifurcation and chaos, **21**(05), 1349–1356, 2011, doi:10.1142/S0218127411029215.
[66] T. Kohonen, "Self-Organizing Maps," In: Sammut C., Webb G.I. (eds) Encyclopedia of Machine Learning. Springer, 2011, doi:10.1007/978-0-387-30164-8_746.

Advances in Science, Technology and Engineering Systems Journal Vol. 7, No. 3, 44-50 (2022)

www.astesj.com

Special Issue on Multidisciplinary Sciences and Engineering

ASTESJ

ISSN: 2415-6698

Effectiveness of Gamified Instructional Media to Improve Critical and Creative Thinking Skills in Science Class

Neni Hermita [*,1], Rian Vebrianto[2], Zetra Hainul Putra[1], Jesi Alexander Alim[1], Tommy Tanu Wijaya[3], Urip Sulistiyo[4]

[1]*Faculty of Teacher Training and Education, Universitas Riau, Pekanbaru, 28154, Indonesia,*

[2]*Fakultas Tarbiyah, UIN SUSKA Riau, Pekanbaru, 28154, Indonesia*

[3]*School of Mathematical Sciences, Beijing Normal University, Beijing, 065001, China*

[4]*Faculty of Teacher Training and Education, Universitas Jambi, Jambi, 36131, Indonesia*

ARTICLE INFO

Article history:
Received: 05 December, 2021
Accepted: 04 May, 2022
Online: 25 May, 2022

Keywords:
Critical thinking skills
Creative thinking skills
Gamification
Genially
Heat transfer concept

ABSTRACT

Gamified Instructional Media has recently been widely used in the education sector to improve students' abilities. Using Gamified Instructional Media at the elementary school level becomes more interesting because it is in accordance with the way children learn K1-K6. The research aims to identify the gamified instructional using Genially to improve students' critical and creative thinking skills. A quasi-experimental method was applied using a nonequivalent control group research design. The research subject is 40 students of Public Primary School in Pekanbaru. The results show a significant effect of the gamified instructional learning using Genially toward students' critical and creative thinking skills. Besides, there is a significant difference in students' critical and creative thinking skills between the control and experimental group. This study implies that gamified instructional media with Genially can support teachers and teaching practices.

1. Introduction

Artificial intelligence contributes to education, particularly to the implementation of educational process, particularly in the teaching and learning scheme in the industrial revolution of the 21st century. It is necessary to conduct an online or blended learning program to respond the health situation we live in [1]. Despite the students appear to miss the face-to-face meetings interaction in a physical classroom setting, they can also adapt well to the sudden changes from offline to online settings [2]. This phenomenon further indicates that most of the current university students are ready to participate in an innovative educational procedure primarily based on blended learning activities. The strength of such educational setting is that the students are able to learn technical skills in their personal environment without any pressure as offered by the online platform, while simultaneously obtaining social resources in a classroom environment [3]. There were various 21st-century skills that students can improve through online platform, including critical and creative thinking skills [4, 5].

According to [6,7], critical thinking requires three components: (1) a disposition to thoughtfully analyze the issues and subjects experienced by someone, (2) knowledge concerning logical exploration and argumentation procedures, and (3) certain competence in using those approaches. Paul and Binker (1990) state that critical thinking skill is the ability and disposition to critically evaluate a belief, what assumptions are based on it and on what basis these assumptions can survive [8]. In [9, 10], the author stated that critical thinking skill is defined as the ability to make decisions by considering the facts available, the situation's context, and the concept raised. In [11], the authors then added that critical thinking is a form of a rational reflective thinking process that focuses on determining what to believe or what to do. Students are expected to have the ability to think critically if they are able to ask something and find information appropriately. Based on the analysis of the information and knowledge they have, students try to answer problems logically and creatively with conclusions that are acceptable to common sense [12].

Creative thinking skill is the ability of students to identify, solve or find solutions, and solve various problems by looking for alternative problem solving based on their own abilities and thoughts [13–15]. The ability to think creatively can be measured

*Corresponding Author: Neni Hermita, Email: neni.hermita@lecturer.unri.ac.id

by several indicators including (1) fluency which means the capacity of sharing ideas, (2) flexibility which means the skill to suggest several problem solutions, (3) originality which is the capability to develop new ideas as a result of their own thought process, and (4) elaboration which is the capability to describe something in detail [16].

In practice, there are some factors causing students' low critical and creative thinking skills one of which is teacher-centered learning approach that the teacher function in the classroom is more like a lecturer presenting instructional materials and the students are expected to passively receive the knowledge being presented. Science teachers should applied student-centered learning approach in which students are allowed to hone and use their critical and creative thinking skills [17]-[20]. In the industrial revolution of the 21st century, direct learning model increasingly makes students not accustomed to think critically and creatively. Therefore, by using digital technology teachers can present science instructional materials which can allow the students to have a better critical thinking skill [21].

In the Industrial revolution of the 21st century, teachers usually use technology in presenting the instructional materials. The role of technology in education is very significant [16]. Almost all learning processes in elementary schools today involve technology such as using applications as instructional media. This is commonly known as blended learning, and the use of technology aims to support in achieving learning objectives and in creating a different learning experience for students.

Gamification is the use of game in a non-game environment to increase students' learning motivation [22]. Gamification increases students' participation in the learning process [23]. Children generally enjoy playing games and such phenomenon may be applied in learning context, prompting the development of a new teaching technique: gamification [24–27]. There are several applications that can be implemented to innovate interactive learning media one of which is Genially.

Genially is a learning media creation platform that has been widely used in education. vidergor [28] uses genially to design a digital escape room to increase elementary school students' collaboration and motivation. Genially has the advantage of being easy to use and accessible [29, 30]. The features in Genially are suitable for beginner developers, so teachers at schools can design learning media using Genially according to their needs [31]. Genially can also increase engagement in learning and allow students to may share their knowledge and improve their communication abilities [32].

The study conducted by [24] found that a game increased the knowledge of Serbian fifth graders in recognizing plants. In education, Gamification may help students improve their computational thinking skills and motivate them to develop their learning capacities on their own by assisting them in increasing their insight, processing the information, communication, and community awareness skills [33]. During the learning process, students get innovation and new insight. In other words, students get a new atmosphere in the learning process. Therefore, students easily understand the learning materials [34]. Science learning involves students directly in acquiring knowledge as a result of student curiosity. The effectiveness of gamification is different from each student. Gamification must be studied and implemented with care, by paying attention to several factors such as including individual learning styles and personality characteristics [35]. Summary of the findings reported Gamified learning experiences developed using the software package Genially were shown to enhance the students' critical thinking competence.

Through information technology, blended learning is able to enhance students' critical thinking skills [36]. In this case, the teachers conduct the learning process by employing learning media in the forms of technology [37]. In particular, gamification can improve the performance of academic skills [38]. Academic skills can be developed through oral discussion, critical thinking, vocabulary development, oral interpretation, creative acting, observation and recording, information research, graphs and graph interpretation, and summaries. In a classroom setting, a physical and intellectual setting provided has the ability to supports the development critical thinking through a spirit of discovery [39]. Critical and creative thinking skill are two skills that must be included in school curriculum in the 21st century.

Gamification such as digital escape room was implemented as a teaching approach, particularly in science classes for the fourth-graders students of primary school. This approach has been proven to affect students effectively, so that they become more motivated and able to solve problems they face [40]. In addition, both critical and creative thinking skills significantly affected the cognitive learning outcomes [41]. Compared from urban environments, higher levels of realism were believed to boost the restorative effects of viewing natural environments and promote creative thinking [42].

The learning objective of learning science in primary schools is to develop students' process skills in investigating nature, solving problems and making decisions, and applying the learning experiences gained in the previous learning process that has been done [43]. Science learning at the fifth-grade primary school level includes several materials that are human and animal organs, green plants, adaptation of living things and their environment, the building blocks of objects and their properties, changes in the properties of objects, forces, simple machines, light, earth, and the universe. The instructional materials of heat transfer are learnt in the sub unit of changes in the properties of objects. The instructional materials are basic materials at the primary school level so that the instructional materials must be presented as attractively as possible [44]. The presentation of the heat transfer materials should be done in an interesting way. For example, it can be done by using Genially.

Gamification is expected to make students able to remember science concepts and be able to apply learning in everyday life [38]. According to [45], Improvements in critical thinking abilities can be identified both from novel biological and non-biological daily issues, indicating that thinking skills can be implemented in various aspects. Furthermore, knowledge test, the experimental students outperformed the control group, implying that "knowledge of facts" and "learning to think" both as educational purposes should be able to interact with each other.

In science learning, teachers can design learning that can improve the critical and creative thinking abilities of the students by using gamified instructional media using Genially since it has many features that support interactive learning. Therefore, by using gamified instructional media with Genially to teach science, it is expected to improve students' critical and creative thinking skills.

In accordance with the background above, this study aimed to used Genially to create interactive science learning media in elementary schools. The researchers make an educational video game that helps students in learning science and in improving their critical thinking. The research question underpinning this study is: Can gamified instructional using Genially improv students' critical and creative thinking skills?

2. Methodology

This study investigated the effectiveness of gamified instructional media with Genially to improve fifth-grade primary school students' critical and creative thinking skills in science class. The researchers applied a quasi-experimental nonequivalent control group research design to conduct this study. The researchers taught the experimental class by using gamified instructional media with Genially while the control class had conventional learning. The sample comprised 40 students of Public Primary School in Pekanbaru, 20 students in the experimental class and 20 students in the control class. This study was conducted from March to April 2021. To collect the data, the researchers used an instrument consisted of 7 short-answered questions to assess the students' critical thinking and 4 short-answered questions to assess their creative thinking skills.

The researchers conducted a pretest and posttest before and after the treatment by using the instrument. The validity and reliability of the instrument had been tested before it was used. After assessing the students' critical thinking and creative thinking skills, the data collected was analyzed. The link to the gamified instructional media is; https://view.genially/605214085ec620fd0b41406/interactive-content-hasil-final.

The data obtained were analyzed by using SPSS, a computer program used for statistical analysis, to know the effectiveness of gamified instructional media with Genially to improve critical and creative thinking skills of the students in fifth grade of primary school in science class on the topic of heat. This research was conducted from March to April 2021 in the fifth grade of Public Primary School in Pekanbaru. Among these populations, samples were selected based on those who have obtained legality based on the certificate of doing research no: 422/SDN192PKU/2021/277.

3. Results and Discussion

The results and discussion sections are divided into several sections, analyze of critical thinking skills, analyze of creative thinking skills and the correlation between creative thinking skill and critical thinking skill to get more in-depth data. In the end the study results are discussed to analyze the effects of Gamified Instructional Media more deeply.

3.1. Analysis of Critical Thinking Skills

Pretest

After the researchers had conducted normality and homogeneity test, the researchers conducted an independent sample t-test. Table 1 below presents the results of the independent sample t-test.

Table 1: The output table of the independent sample t-test

Independent Samples Test										
		Levene's Test for Equality of Variances		t-test for Equality of Means						
									95% Confidence Interval of the Difference	
		F	Sig.	t	Df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
Critical Thinking Skills	Equal variances assumed	8.758	.005	-.871	38	.389	-.300	.345	-.997	.397
	Equal variances not assumed			-.871	27.428	.391	-.300	.345	-1.006	.406

Table 2: The output table of the independent sample t-test

Independent Samples Test										
		Levene's Test for Equality of Variances		t-test for Equality of Means						
									95% Confidence Interval of the Difference	
		F	Sig.	t	Df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
Critical Thinking Skills	Equal variances assumed	1.055	.311	4.887	38	.000	3.450	.706	2.021	4.879
	Equal variances not assumed			4.887	36.929	.000	3.450	.706	2.020	4.880

Levene's Test for Impartiality of Discrepancies had an implication rate of 0.005 0.05. Because the value indicated that the two variances were not the same, the interpretation of the above output table was based on the values in the equal variances not assumed column. In the equal variances assumed column, the Sig. (2-tailed) value was 0.391> 0.05. There was no substantial difference in critical thinking skills among students in the experimental and control groups, according to the results (pretest). In other words, before treatment, the critical thinking skills of students in the experimental and control classes were the same.

Posttest

After the researchers had conducted normality and homogeneity test, the researchers conducted an independent sample t-test. Table 2 below offers the outcomes of the independent sample t-test.

The consequence rate of Levene's Test for Equivalence of Discrepancies was 1.055 > 0.05. The value indicated that the two variances were identical, so the interpretation of the output table above is based on the values in the identical variances assumed column. The Sig. (2-tailed) value in the equal variances assumed column was 0.000 < 0.05. It indicated a significant difference in critical thinking skills between students in experimental and control classes (posttest). In other words, the critical thinking skills between students in experimental and control class before treatment was dissimilar.

This study proved that the gamified instructional media with Genially media can progress students' critical thinking skills. This benefit arises from agreeing on the separate mastery of procedural skills in the secluded and stress-free situation provided by the operational stage and admission to social properties in the classroom situation. Instruction and education consuming online platform have carried confident influences specifically in emerging the 21st-century assistances. Critical thinking skills are one of the 21st-century skills that must be included in the world of education. However, the main challenge is how to teach thinking or critical thinking and how to stimulate students to reflect on their own thinking ways. [46].

3.2. Analysis of Creative Thinking Skills

Posttest

After the researchers had conducted normality and homogeneity test, the researchers conducted independent sample t-test. Table 3 below presents the results of the independent sample t-test.

The significance value of Levene's Test for Equality of Variances was 0.005 0.05. The value indicated that the two variances were not the same, so the interpretation of the output table above was based on the values in the equal variances not assumed column. The Sig. (2-tailed) value in the equal variances assumed column was 0.391> 0.05. It revealed that there was no significant difference in critical thinking skills between students in the experimental and control groups (pretest). In other words, before treatment, the critical thinking abilities of students in the experimental and control classes were the same.

According to [47], games can be used as a supplement to traditional teaching methods to improve learners' learning experiences while also teaching other skills such as following rules, adaptation, problem solving, interaction, critical thinking skills, creativity, teamwork, and good sportsmanship. Furthermore in [48], contend that digital games can assist students in developing higher-order thinking skills and 21st-century skills, as well as making learning more enjoyable and engaging.

3.3. Analysis of the correlation between students' critical and creative thinking skills Experiment class

To know if there was a correlation between students' critical and creative thinking skills, the researchers conducted a (Pearson) bivariate correlation. Table 4 below is the output table of the (Pearson) bivariate correlation.

Table 3: The output table of the independent sample t-test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
									95% Confidence Interval of the Difference	
		F	Sig.	t	Df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
Creative Thinking Skills	Equal variances assumed	4.672	.037	2.148	38	.038	1.550	.722	.089	3.011
	Equal variances not assumed			2.148	27.782	.041	1.550	.722	.071	3.029

Table 4: the output table of the (Pearson) bivariate correlation

		Critical thinking skills	Creative thinking skills
Critical thinking skills	Pearson Correlation	1	.588**
	Sig. (2-tailed)		.006
	N	20	20
Creative thinking skills	Pearson Correlation	.588**	1
	Sig. (2-tailed)	.006	
	N	20	20

**. Correlation is significant at the 0.05 level (2-tailed).

Table 5: The output table of the (Pearson) bivariate correlation

Correlations			
		Critical thinking skills	Creative thinking skills
Critical thinking skills	Pearson Correlation	1	.425
	Sig. (2-tailed)		.062
	N	20	20
Creative thinking skills	Pearson Correlation	.425	1
	Sig. (2-tailed)	.062	
	N	20	20

**. Correlation is significant at the 0.05 level (2-tailed).

Based on the output table, the correlation coefficient was 0.588, and the Sig. (2-tailed) value was 0.006 < 0.05. Therefore, there was a positive correlation between students' critical thinking skills and creative thinking skills, and the correlation was at a moderate level.

Control class

To know if there was a correlation between students' critical and creative thinking skills, the researchers conducted a (Pearson) bivariate correlation. Table 5 below is the output table of the (Pearson) bivariate correlation.

The correlation coefficient was 0.425, and the Sig. (2-tailed) value was 0.062 > 0.05, according to the output table. As a result, there was no positive correlation between students' critical and creative thinking skills, and the correlation was moderate.

This study demonstrated that students in the experimental class demonstrated a positive correlation between critical thinking skills and creative thinking skills, whereas students in the control group demonstrated no positive correlation between critical thinking skills and creative thinking skills.

This finding is supported by [41] claim that there is a significant correlation between critical thinking skills and creative thinking skills on cognitive learning outcomes. The restorative properties of nature are most visible for creativity when viewing stimuli indoors; however, being outdoors in general may be enough to stimulate creativity, regardless of whether it is surrounded by nature or a busy urban environment [42, 49]. The application of Digital Escape Room with Science Teaching in Primary School has Problem Solving Ability. Especially in science subjects [50].

Figure 1: Interesting animation

There is a significant effect because gamification of instructional media makes the learning fun and interesting (see figure 1) and helps students develop higher-order thinking skills [48]. Moreover, [44] and [51] state that presenting instructional materials by using interactive instructional media provides convenience to students. In addition, the animations displayed and interactive simulations that must be done by students through discussion sheets can train students' logical thinking in physics problems solving related to the concepts of temperature and heat. That evidence proves that there is a significant role of using gamified instructional media with Genially to teach science on the topic of heat transfer.

Based on the test results above, the results of this study have the same results with the study proposed by [44] showing on the topic of temperature and heat, interactive instructional media affects the students' conceptual understanding and critical thinking skills. Gamified instructional media can help students improve their computational thinking competence and motivate them to develop their learning capacities on their own by assisting them in expanding their insight, processing information they have, communication, and community awareness skills [33].

Figure 2: interesting games in the learning media

Learning media on heat transfer material with genius has an interesting game. In addition, another previous research project also claimed that games can also be implemented teach other skills including critical thinking, problem solving, sportsmanship, interaction and peers-collaboration [47]. Moreover, great potential is provided by games for training because they affect the learning process of users significantly [52]. Hikmah and Ngazizah (2020) state that creative thinking skills are part of higher order thinking

skills. Therefore, gamified instructional media with Genially can improve students' creative thinking skills which belongs to higher order thinking skills.

The existence of technology and various innovations helps overcome learning problems, especially at the primary school level [40]. That indicates that gamified instructional media helps students to identify and solve problems. Instructional media is one of the most important things that support learning. Gamified instructional media is physical means in delivering instructional materials from teachers to students in a more sophisticated and efficient way [53], [54]. Therefore, gamified instructional media is necessary to be considered carefully as it has long-term effects on education [55].

In conclusion, gamified instructional media using Genially in science class, particularly on heat transfer topic improve students' critical and creative thinking skills in terms of the ability to generate ideas, the ability to propose various solutions to problems, and the ability to create different ideas.

4. Conclusions and Suggestions

This study discusses in depth the effects of Gamified Instructional Media on critical thinking and creative thinking skills, then analyse of correlation between critical thinking and creative thinking skills. The results of the study prove that Gamified Instructional Media has an influence on the critical thinking and elementary school students' creative thinking skills. While the education is focusing on improving high order thinking skills, this study adds new literature on Gamified Instructional Media at the elementary school level which can convince teachers that using Gamified Instructional Media is to increase HOTS at the elementary school level effectively.

Based on the results, the researchers suggest that teachers have to improve their technology literacy so that they can involve technology in their teaching. With technology, teachers can develop many forms of instructional media. The researchers also suggest facilitating teachers to develop instructional media with technology to others stakeholders in education.

Acknowledgements

The authors express gratitude to the Ministry of Research and Technology/National Research and Innovation Agency Indonesia for supporting this study under the Grant of DRTPM 2022

References

[1] García-Peñalvo, F.J, A.J. Mendes, "Exploring the computational thinking effects in pre-university education," Computers in Human Behavior Journal, 2018.

[2] C. Giovannella, "Effect Induced by the Covid-19 Pandemic on Students' Perception About Technologies and Distance Learning," In Ludic, Co-Design and Tools Supporting Smart Learning Ecosystems and Smart Education, 105–116, 2021, doi:10.1007/978-981-15-7383-5.

[3] L. Warren, D. Reilly, A. Herdan, Y. Lin, "Self-efficacy, performance and the role of blended learning," Journal of Applied Research in Higher Education, 2019, doi:10.1108/JARHE-08-2019-0210.

[4] R. Vebrianto, R.U. Rery, K. Osman, "BIOMIND Portal for Developing 21st Century Skills and Overcoming Students' Misconception in Biology Subject," International Journal of Distance Education Technologies, **14**(4), 55–67, 2016, doi:10.4018/IJDET.2016100105.

[5] T.T. Wijaya, W. Hidayat, Y. Zhou, "Development of Interactive Learning Video on Linear Program," Universal Journal of Educational Research, **8**(12A), 7530–7538, 2020, doi:10.13189/ujer.2020.082537.

[6] E. Glaser, "An experiment in the development of critical thinking," Teachers College Record, **43**(5), 409–410, 1942.

[7] T.T. Wijaya, J. Tang, L. Li, A. Purnama, "Implementing Dynamic Mathematics Software in Calculus II for Engineering Students: Quadratic Surfaces," Software Engineering and Algorithms, **230**, 480–491, 2021, doi:10.1007/978-3-030-77442-4_41.

[8] R.W. Paul, A.J.A. Binker, "Critical thinking: What every person needs to survive in a rapidly changing world," Center for Critical Thinking and Moral Critique, Sonoma State University, Rohnert Park, CA 94928., 1990.

[9] P.A. Facione, "Critical Thinking: What It Is and Why It Counts," Insight Assessment, **2007**(1), 1–23, 2011.

[10] M. Hutajulu, T.T. Wijaya, W. Hidayat, "the Effect of Mathematical Disposition and Learning Motivation on Problem Solving: an Analysis," Infinity Journal, **8**(2), 229, 2019, doi:10.22460/infinity.v8i2.p229-238.

[11] R. Ennis, "Critical thinking: Reflection and perspective Part II," Inquiry: Critical Thinking across the Disciplines, **26**(2), 5–19, 2011.

[12] R. Anjani, "Analisis Kemampuan Berpikir Kritis Siswa Gaya Belajar Accomodator dalam Menyelesaikan Soal Pemecahan Masalah Matematika di Kelas VIII Smp Negeri 6 Muaro Jambi," Jurnal Pendidikan Matematika, **3 Oktober**, 2017.

[13] R. Nuraini, Suparman, "Deskripsi Kemampuan Berpikir Kritis dan Kreatif Siswa Melalui Penerapan Pendekatan Saintifik," in Prosiding Seminar Nasional Etnomatnesia ISBN: 978-602-6258-07-6, 2019.

[14] N. Hermita, A. Suhandi, E. Syaodih, A. Samsudin, Isjoni, H. Johan, F. Rosa, R. Setyaningsih, Sapriadil, D. Safitri, "Constructing and Implementing a Four Tier Test about Static Electricity to Diagnose Pre-service Elementary School Teacher' Misconceptions," Journal of Physics: Conference Series, **895**(1), 2017, doi:10.1088/1742-6596/895/1/012167.

[15] J. Pereira, T. Jianlan, T.T. Wijaya, A. Purnama, Neni, Hermita, M. Tamur, "Using Hawgent Mathematics Software to Help Primary School Students to Read Clocks," Journal of Physics: Conference Series, **2049**(1), 2021, doi:10.1088/1742-6596/2049/1/012049.

[16] Yustina, W. Syafii, R. Vebrianto, "The Effects of Blended Learning and Project-Based Learning on Pre-Service Biology Teachers' Creative Thinking Skills through Online Learning in the Covid-19 Pandemic," Jurnal Pendidikan IPA Indonesia, **9**(3), 2020, doi:10.15294/jpii.v9i3.24706.

[17] N. Azriani, N. Islami, N. Hermita, M. Nor, E. Syaodih, H. Handayani, Z. Zulirfan, A. Suhandi, A. Malik, K. Mahbubah, A. Samsudin, "Implementing inquiry learning model to improve primary school students' critical thinking on earth and universe concept," Journal of Physics: Conference Series, **1227**(1), 2019, doi:10.1088/1742-6596/1227/1/012033.

[18] A.R. Ningsih, A. Suhandi, E. Syaodih, B. Maftuh, N. Hermita, A. Samsudin, "Fourth-grade elementary students critical thinking skills: A preliminary study on magnetic force," in Journal of Physics: Conference Series, 2019, doi:10.1088/1742-6596/1157/3/032045.

[19] E. Syaodih, L. Kurniawati, H. Handayani, D. Setiawan, I. Suhendra, N. Hermita, "Critical Thinking Skills of Fifth Grade Elementary School Students in Bandung City on the Topic of Water Cycle in Natural Science Subjects," in Journal of Physics: Conference Series, 2019, doi:10.1088/1742-6596/1351/1/012073.

[20] D.A.I. Wijayanti, K. Pudjawan, I.G. Margunayasa, "Analisis Kemampuan Berpikir Kritis Siswa Kelas V dalam Pembelajaran IPA di 3 SD Gugus X Kecamatan Buleleng," E-Journal PGSD Universitas Pendidikan Ganesha, **3**(1), 2015.

[21] F.Z. Firdaus, Suryanti, U. Azizah, "Pengembangan Multimedia Interaktif Berbasis Pendekatan SETS untuk Meningkatkan Kemampuan Berpikir Kritis Siswa Sekolah Dasar," Jurnal Basicedu, **4**(3), 681–689, 2020, doi:10.31004/basicedu.v4i3.417.

[22] M. Kuo, T. Chen, T.-Y. Chuang, B.-Y. Cheng, Y.-N. Su, "What Are the Better Gamification Tools for Elementary School Teachers?," International Congress on Advanced Applied Informatics (IIAI-AAI), (1), 346–349, 2018, doi:10.1109/IIAI-AAI.2018.00074.

[23] Q. Zhang, Z. Yu, "A literature review on the influence of Kahoot! On learning outcomes, interaction, and collaboration," Education and Information Technologies, 2021.

[24] E. Borsos, "The gamification of elementary school biology : a case study on increasing understanding of plants increasing understanding of plants," Journal of Biological Education, 1–14, 2018, doi:10.1080/00219266.2018.1501407.

[25] M.A. Çiftçi, M. Aykaç, "The effect of creative drama activities in early childhood on the executive functions of children," Education 3-13, **0**(0), 1–14, 2020, doi:10.1080/03004279.2020.1849343.

[26] S.B. Kert, M.K. Büyükimdat, A. Uzun, "Comparing active game-playing scores and academic performances of elementary school students," **4279**(March), 2016, doi:10.1080/03004279.2016.1140800.

[27] K. Jayantilal, N.O. Leary, "The factors influencing two primary teachers ' interpretation of games," Education 3-13, **0**(0), 1–17, 2020,

doi:10.1080/03004279.2020.1810094.
[28] H.E. Vidergor, "Effects of digital escape room on gameful experience, collaboration, and motivation of elementary school students," Computers & Education, **166**(February), 104156, 2021, doi:10.1016/j.compedu.2021.104156.
[29] M. Khoiron, Harmanto, A. Kasdi, A.R. Wardani, "Development of digital social studies teaching materials in the era of pandemic emergency learning," The Indonesia Journal of Social Studies, **4**(1), 36–44, 2021.
[30] Â. Musskopf, D.N.F.D.N.F. Barbosa, P.B.S.P.B.S. Bassani, A. Jefferies, "Using Digital Resources to Boost English Writing Development," Communications in Computer and Information Science, **1011**, 337–348, 2019, doi:10.1007/978-3-030-20798-4_29.
[31] P.M. Manuel, A.M. Pilar, R.M. María Dolores, D. MP, P. Sara, M.J. M. Pilar, R.M. Dolores, D. MP, P. Sara, M.-J.M. Pilar, "Characterization of biodiesel using virtual laboratories integrating social networks and web app following a ubiquitous- and blended-learning," Journal of Cleaner Production, **215**, 399–409, 2019, doi:10.1016/j.jclepro.2019.01.098.
[32] G. Cheung, K. Wan, K. Chan, "Efficient use of clickers: A mixed-method inquiry with university teachers," Education Sciences, **8**(1), 1–15, 2018, doi:10.3390/educsci8010031.
[33] E. Choi, Y. Jung, N. Park, "Strategies to Teach Elementary School Students the Principles of Blockchain Technology by Implementing Gamification," Ilkogretim Online, **20**(3), 1205–1211, 2021, doi:10.17051/ilkonline.2021.03.134.
[34] M.T. Alshammari, "Evaluation of Gamification in E-Learning Systems for Elementary School Students," TEM Journal, **9**(2), 806–813, 2020, doi:10.18421/TEM92.
[35] P. Buckley, E. Doyle, "Individualising gamification: An investigation of the impact of learning styles and personality traits on the efficacy of gamification using a prediction market," Computers & Education, **106**, 43–55, 2017, doi:10.1016/j.compedu.2016.11.009.
[36] B.N. Nirbita, S. Joyoatmojo, S. Sudiyanto, "ICT Media Assisted Problem Based Learning for Critical Thinking Ability," International Journal of Multicultural and Multireligious Understanding, **5**(4), 341–348, 2018.
[37] W. Rajagukguk, E. Simanjuntak, "Problem-Based Mathematics Teaching Kits Integrated With ICT to Improve Student's Critical Thinking Ability In Junior High Schools In Medan," Cakrawala Pendidikan, (3), 2015.
[38] Ü. Cakiroğlu, B. Başibüyük, M. Güler, M. Atabay, B.Y. Memiş, "Gamifying an ICT Course: Influences on Engagement and Academic Performance," Computers in Human Behavior, 2017, doi:10.1016/j.chb.2016.12.018.
[39] B. Potts, "Strategies for Teaching Critical Thinking," Practical Assessment, Research, and Evaluation, **4**(4), 1995.
[40] S. Huang, Y. Kuo, H. Chen, "Applying digital escape rooms infused with science teaching in elementary school : Learning performance , learning motivation , and problem-solving ability," Thinking Skills and Creativity, **37**(129), 100681, 2020, doi:10.1016/j.tsc.2020.100681.
[41] J. Siburian, A.D. Corebima, Ibrohim, M. Saptasari, "The Correlation Between Critical and Creative Thinking Skills on Cognitive Learning Results," Eurasian Journal of Educational Research, **81**, 99–114, 2019, doi:10.14689/ejer.2019.81.6.
[42] A. Palanica, A. Lyons, M. Cooper, A. Lee, Y. Fossat, "A comparison of nature and urban environments on creative thinking across different levels of reality," Journal of Environmental Psychology, **63**, 44–51, 2019.
[43] N. Hendracipta, "Menumbuhkan Sikap Ilmiah Siswa Sekolah Dasar Melalui Pembelajaran IPA Berbasis Inkuiri," JPSD, **2**(1), 109–116, 2016, doi:10.31226/osf.io/etg5n.
[44] S. Husein, L. Herayanti, Gunawan, "Pengaruh Penggunaan Multimedia Interaktif terhadap Penguasaan Konsep dan Keterampilan Berpikir Kritis Siswa pada Materi Suhu dan Kalor," Jurnal Pendidikan Fisika Dan Teknologi, **I**(3), 2015.
[45] A. Zohar, Y. Weinberger, P. Tamir, "The Effect of the Biology Critical Thinking Project on the Development of Critical Thinking," Journal Of Research in Science Teaching, **31**(2), 183–196, 1994.
[46] E. AlJaafil, M. Sahin, "Critical Thinking Skills for Primary Education: The Case in Lebanon," Online Submission, **1**(1), 1–7, 2019.
[47] V. Zirawaga, A. Olusanya, T. Maduki, "Gaming in education: Using games a support tool to teach History," Journal of Education and Practice, **8**(15), 55–64, 2017.
[48] Y. An, L. Cao, "The Effects of Game Design Experience on Teachers' Attitudes and Perceptions regarding the Use of Digital Games in the Classroom," TechTrends, 2016, doi:10.1007/s11528-016-0122-8.
[49] H. Huang, G.-J. Hwang, "Facilitating inpatients' family members to learn: A learning engagement-promoting model to develop interactive e-book systems for patient education," Educational Technology and Society, **22**(3), 74–87, 2019.
[50] M. Kalogiannakis, S.J. Papadakis, "Evaluating pre-service kindergarten teachers ' intention to adopt and use tablets into teaching practice for natural sciences Evaluating pre-service kindergarten teachers ' intention to adopt and use tablets into teaching practice for natural sciences Mic," International Journal of Mobile Learning and Prganisation, **13**(1), 24–28, 2019, doi:10.1504/IJMLO.2019.096479.
[51] A. Syawaludin, G. Gunarhadi, P. Rintayati, "Development of Augmented Reality-Based Interactive Multimedia to Improve Critical Thinking Skills in Science Learning," International Journal of Instruction, **12**(4), 2019, doi:10.29333/iji.2019.12421a.
[52] P.-M. Noemí, S.H. Máximo, "Educational games for learning," Universal Journal of Educational Research, **2**(3), 230–238, 2014, doi:10.13189/ujer.2014.020305.
[53] S. Hikmah, N. Ngazizah, "Profil Kemampuan Higher Order Thinking Skills dan Karakter Siswa pada Materi Panas dan Perpindahannya pada Kelas 5 Sekolah Dasar," in Seminar Nasional Pendidikan Dasar, 2020.
[54] N. Hermita, H.S. Ningsih, J.A. Alim, M. Alpusari, Z.H. Putra, T.T. Wijaya, "Developing Science Comics for Elementary School Students on Animal Diversity. Solid State Technology," **63**(1s), 2020.
[55] M. Kalogiannakis, S. Papadakis, A.-I. Zourmpakis, "Gamification in Science Education. A Systematic Review of the Literature," Education Sciences, **11**(22), 2021.

ASTES

Advances in Science, Technology and Engineering Systems Journal Vol. 7, No. 3, 51-61 (2022)

www.astesj.com

Special Issue on Innovation in Computing, Engineering Science & Technology

ASTESJ

ISSN: 2415-6698

Generalized Linear Model for Predicting the Credit Card Default Payment Risk

Lu Xiong[*,1,2], Spendylove Duncan-Williams[3]

[1]*Department of Mathematical Sciences, Middle Tennessee State University, Murfreesboro, TN 37132, USA*

[2] *Computational and Data Science Doctoral Program, Middle Tennessee State University, Murfreesboro, TN 37132, USA*

[3] *Master of Science in Professional Science Program, Middle Tennessee State University, Murfreesboro, TN 37132, USA*

ARTICLE INFO

Article history:
Received: 13 February, 2022
Accepted: 04 May, 2022
Online: 25 May, 2022

Keywords:
Credit Card Default Risk
Generalized Linear Model
Decision Tree
Random Forest

ABSTRACT

Predicting the credit card default is important to the bank and other lenders. The credit default risk directly affects the interest charged to the borrower and the business decision of the lenders. However, very little research about this problem used the Generalized Linear Model (GLM). In this paper, we apply the GLM to predict the risk of the credit card default payment and compare it with a decision tree, a random forest algorithm. The AUC, advantages, and disadvantages of each of the three algorithms are discussed. We explain why the GLM is a better algorithm than the other two algorithms owing to its high accuracy, easy interpretability, and implementation.

1. Introduction

The credit card debt crisis has been a major concern in the capital market and among card-issuing institutions for many years. Credit cards and cash-card debts are abused by most card users, regardless of their payment capabilities. This crisis poses a great threat to both cardholders and banks. The payment default means the failure to pay the credit card bills. Researchers have attempted to forecast the credit card customers' default payments using machine learning techniques [1]. The public's lack of understanding of basic financial principles, as seen by the recent financial crisis, has proven that they are unable to make sound financial judgments.

Individuals' minimum monthly credit card payments should be researched in depth to discover the link between consumer income, payment history, and future default payments [2]. Increasing consumer finance confidence, to avoid delinquency is a big challenge for cardholders and banks as well. In a well-established financial system, risk assessment is more important than crisis management [2]. With historical financial data, for example, business financial statements, client transaction and reimbursement records, etc., we can predict business execution or individual clients' default risk and lessen the harm and instability. To make accurate customer risk assessments for their credit services department, banks are required by having sophisticated credit scoring systems to automate the credit risk scoring tasks [3]. Management of the credit risk for the banking sector and financial organizations have extensively started to gain importance. Developing an automated system to accurately forecast the probability of cardholder's future default, will help not only to manage the efficiency of consumer finance but also effectively handle the credit risk issues encountered in the banking sector [3], [4].

Individual firms have been gathering massive amounts of data every day in the era of big data. Finding the relevant information from data and turning that information into meaningful outcomes is a big issue for businesses. As a result, in this article, we investigate the risk of failure to pay the minimum credit card amount, which is a transaction that should be done monthly. The Generalized Linear Model (GLM), single classification tree, and random forest are compared to predict the credit default risk. Prediction accuracy and interpretability are the two main factors we consider when selecting the final model. The principal component analysis (PCA) is used to reduce the data dimensionality.

2. Literature Review

Machine learning is the act of automatically or semi-automatically exploring and analyzing massive amounts of data to uncover significant patterns and rules [5]. It has been utilized in different types of financial analysis such as predicting money laundering, stock analysis, detection of bankruptcy, the decision of loan approval, etc. [6, 7]. Machine learning algorithms are

[*]Corresponding Author: Lu Xiong, Email: lu.xiong@mtsu.edu

inadequately used for detecting default payments of credit card users. Different kinds of research are ongoing to improve the accuracy of machine learning algorithms in predicting default payment of credit card users, and a small portion of improvement plays a vital role in the economic developments of the related organizations [8, 9]. Before the 1980s, some statistical methods such as Linear Discriminant Analysis (LDA) [10], and Logistic Regression (LR) [11] were used to estimate the credit default probability. Starting from the 1990s, machine learning methods such as K-nearest neighbor (KNN) [12], neural network (NN) [13], genetic algorithm [14], and support vector machine (SVM) [15] were used to assess the default risk of credit card users. In 2018, researchers compared 5 data mining methods on credit card default prediction: logistic regression, SVM, neural network, Xgboost, and LightGBM [16]. In 2020, AdaBoost was employed to build credit default prediction models [17]. In 2020, researchers investigated the credit card default prediction in the imbalanced data sets [18].

However, the GLM is rarely used in credit card default prediction. GLM has the advantage of easy interpretability and implementation. We believe it's worthwhile to apply the GLM in the research of credit card payment default prediction. In this paper, we will compare the GLM with the decision tree and random forest and explain its advantages.

3. Data Description and Exploratory Analysis

3.1. Data description

The data set came from the credit cardholders from a major bank in Taiwan in October 2005, with a total of 25000 records. The target variable to be predicted in this data is a binary variable – default payment (Yes = 1, No = 0). Table 1 lists the 25 variables in the data including all the predictors and the target variable (default status). The first 5 of the predictors are demographic characteristics. The next 6 variables are about the status of past payments. The further 6 variables are about the amount of the past bill statement. The next 6 variables are about the amount of paid bills and the last 2 predictors are the limit balance of individuals and the default status of the client. This credit payment data has a typical characterization of imbalanced datasets in terms of the target variable, for 5529 records (22.12%) are the records of default and 19471 records (77.88%) are non-default payments.

Table 1: Data dictionary

VARIABLE NAME	DESCRIPTION	FACTOR LEVELS
ID	The ID of each client	Values
SEX	Gender	1=male 2=female
EDUCATION	The level of education of each client	0= Unknown 1=graduate school 2=university 3=high school 4=others
MARRIAGE	Marital Status	1=married 2=single, 3=others
AGE	Age in years	Ages
LIMIT_BAL	The amount of given credit in NT dollars includes individual and family credit	Numerical NT Dollars
PAY_0	Repayment status in September 2005	-2=payment made two months earlier, -1=pay duly, 0=paid right before due 1=payment delay for one month, 2=payment delay for two months, … 8=payment delay for eight months,)
PAY_2	Repayment status in August 2005	scale same as above
PAY_3	Repayment status in July 2005	scale same as above
PAY_4	Repayment status in June 2005	scale same as above
PAY_5	Repayment status in May 2005	scale same as above
PAY_6	Repayment status in April 2005	scale same as above
BILL_AMT1:	amount of bill statement in September 2005	Numerical NT dollars
BILL_AMT2:	Amount of bill statement in August 2005	Numerical NT dollars
BILL_AMT3	Amount of bill statement in July 2005	Numerical NT dollars
BILL_AMT4	Amount of bill statement in June 2005	Numerical NT dollars
BILL_AMT5	Amount of bill statement in May 2005	Numerical NT dollars
BILL_AMT6	Amount of bill statement in April 2005	Numerical NT dollars
PAY_AMT1	Amount of previous payment in September 2005	Numerical NT dollars
PAY_AMT2	Amount of previous payment in August 2005	Numerical NT dollars
PAY_AMT3	Amount of previous payment in July 2005	Numerical NT dollars
PAY_AMT4	Amount of previous payment in June 2005	Numerical NT dollars
PAY_AMT5	Amount of previous payment in May 2005	Numerical NT dollars
PAY_AMT6	Amount of previous payment in April 2005	Numerical NT dollars
default.payment.next.month	Default Status	1=Yes 0=No

3.2. Explore the relationship between the predictors and the target variable

In this section, relations between the predictors vs the target variable will be explored using graphs.

Figure 1 is the stacked histogram that shows the distribution of the credit limit (limit balance) among defaulting records (blue bars) and non-defaulting records (pink bars).

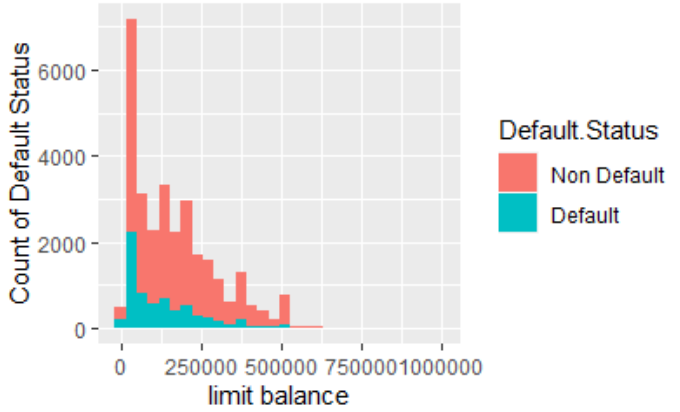


Figure 1: The stacked histogram of the credit limit in default records and non-default records.

From Figure 1, we can observe the default records have a higher proportion of Lower LIMIT_BAL values than non-defaulting records. To confirm this observation, we did a two-sample t-test with a null hypothesis mean of LIMIT_BAL in the non-default group is less than in the default group. The p-value is less than 10e-15 which means we can confidently accept the alternative hypothesis: the mean of LIMIT_BAL in the non-default group is higher than in the default group. This confirms our observation in Figure 1. This conclusion matches the reality that lower credit balances are usually issued to credit users with higher default risk. This distinctive characteristic of the predictor LIMIT_BAL indicates it's a good variable to predict default.

Table 2: Some details of hypothesis tests.

Alternative Hypothesis	Type of Hypothesis Test	P-value
The mean of LIMIT_BAL in the non-default group is greater than in the default group	T-test	$<2.2\times 10^{-16}$
The chance of default at age 25-40 is lower than at other ages.	Two-proportions z-test	$<2.2\times 10^{-16}$

Figure 2 shows the distribution of the age among default credit users and non-default users, using a density histogram. We observe the non-default group has a higher proportion of age 25-40, but a lower proportion of other ages. This makes us infer the age 25-40 has a lower chance of default. To confirm this as the alternative hypothesis, we did a two-proportions z-test. The P-value as shown in Table 2 is very significant and confirms our hypothesis. The reason age 25-40 has a lower default rate could be people in this age range are in the working-age, healthy, and don't have much financial burden from their family yet. In contrast, people younger than 25 may have not started their careers yet, therefore with little to no income. People older than 40 could be less healthy, retired, or need financial support from their families such as children's college education. The above discussion indicates the variable Age can be a good predictor.

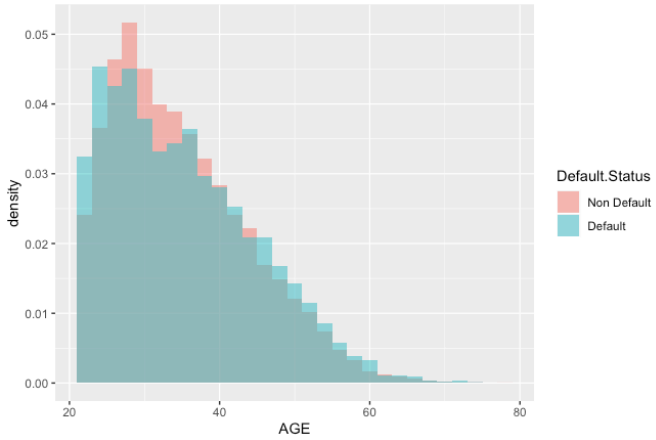


Figure 2: The stacked histogram of age in default records and non-default records.

Figure 3 represents the default status distribution within different gender. The percentage of defaults in males (24%) is slightly higher than in females (20%). To confirm this difference is statistically significant, we did a two-proportion z-test with the alternative hypothesis that male has a higher default chance than female. The p-value of this z-test is 2.47×10^{-12}, which means we should accept the alternative hypothesis. This confirms our observation. This difference is probably because females are more conservative when managing their personal finance than males, which results in lower default risk. This difference indicates the variable Sex is a good predictor.

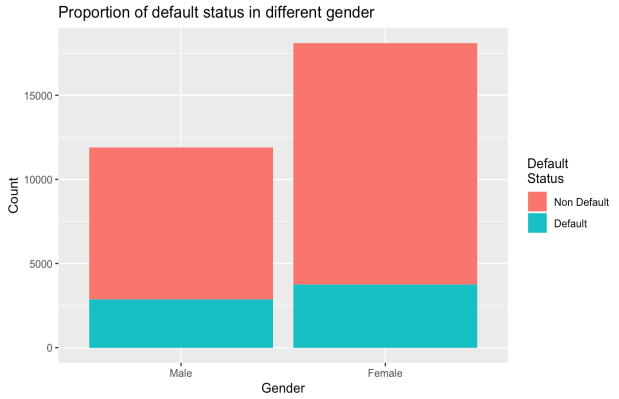


Figure 3: Distribution of default and non-default records in males and females.

Figure 4 displays the proportion of education levels in non-default and default records. It shows that the non-default records have a larger proportion of higher educated individuals from graduate school and university levels. This also matches our intuition that higher education levels associate with higher income

and lower default risk. To statistically confirm this observation, we did a two-proportion z-test with an alternative hypothesis that non-default users have a higher proportion of education level of university or graduate school. The p-value of this test is 0.00226, which confirms our observation.

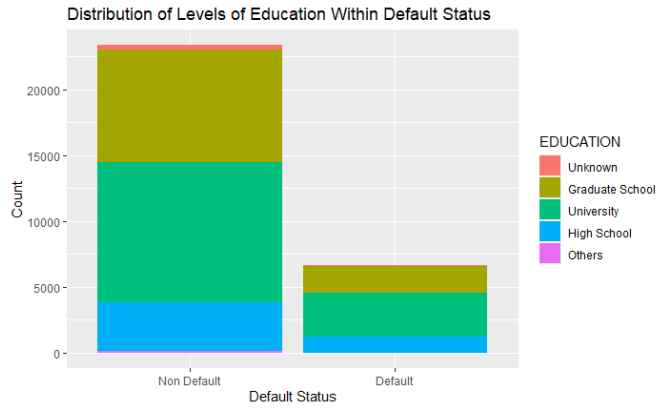


Figure 4: The distribution of education level in default and non-default.

Figure 5 displays the proportion of different lateness in the last statement payment (PAY_0) in default records and non-default records. The variable PAY_0 has more proportion of "On Time" or "Early" values in the non-default records than in default values. We also did a two-proportions z-test to confirm this observation as the alternative hypothesis in the z-test. The p-value of the test is less than 2.2×10^{-16}, which confirms the alternative hypothesis. This means that being current or ahead of payments is associated with non-defaulting in the following month. Therefore the lateness in the last statement payment (PAY_0) is a good predictor to predict the default. For the lateness of the last 2nd payment to the last 7th payment (PAY_1 to PAY_6), we observed the same pattern. That is, non-default records have a higher proportion of "On Time" or "Early" payments in each of the past 7 months than default records. This suggests the payment statuses in the past months have potentially strong predictive power on the default risk.

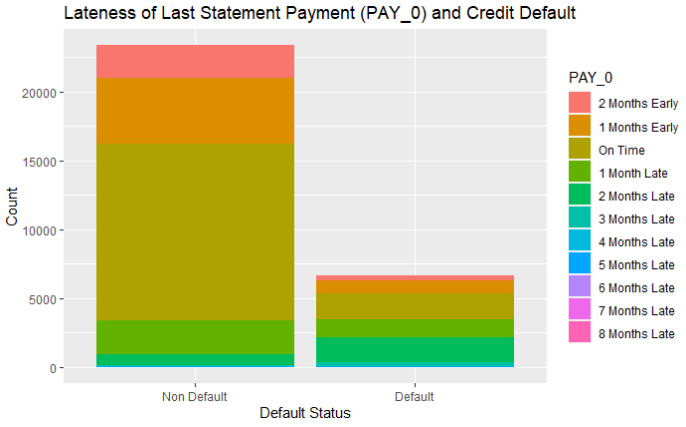


Figure 5: The distribution of PAY_0 values in non-default and default values. Variable PAY_0 is the lateness of the last statement payment

3.3. Reduce the levels of factor variables to increase the predictive power

If a factor variable in the data has too many different values (levels), its predictive power is usually lower. This is because the factor variable is often expanded as the combination of multiple dummy variables when predicting the target variable, where each dummy column is one possible value of the factor variable. If there are many possible values in the variable, then this expanded data will have too many columns versus the number of rows. The data rows will not be long enough compared with its dimensionality to train a highly accurate and robust algorithm. To reduce the data dimensionality, we need to reduce the levels of the factor variables by combining similar levels, so the prediction accuracy can be improved.

Table 3 shows the number of non-default and default records in the factor variable Education. According to this table, the university and high school education have higher proportions of defaults than the other values. We combine these 2 levels as one level and name it "HighSchoolUniversity". Since "others" and "unknown" have a lower proportion of default, they will be combined and named 'others'. The factors are then reduced to HighSchoolUniversity, Graduate School, and Others.

Table 3: Data dictionary

EDUCATION	Non-default	Default	Total	Proportion of default
Unknown	319	26	345	8%
Graduate School	8549	2036	10585	19%
University	10700	3330	14030	24%
High School	3680	1237	4917	25%
Others	116	7	123	6%

Table 4: The proportion of non-default in each level of predictor PAY-0 before its levels combined.

PAY _0	Non Default	Default	Total	Proportion of default
2 Months Early	2394	365	2759	13%
1 Month Early	4732	954	5686	17%
On Time	12849	1888	14737	13%
1 Month Late	2436	1252	3688	34%
2 Months Late	823	1844	2667	69%
3 Months Late	78	244	322	76%
4 Months Late	24	52	76	68%
5 Months Late	13	13	26	50%
6 Months Late	5	6	11	55%
7 Months Late	2	7	9	78%
8 Months Late	8	11	19	58%

Table 4 shows the number of non-default and default records in the factor variable PAY_0. The values of non-default in PAY_0

= "2 Months Early", "1 Month Early", and "On Time" have a similar lower proportion of defaults than the other PAY_0 values. We combine these 3 levels as one level and name it "Pay duly". Also, the PAY_0 = "2 Months Late" or more have a similarly high proportion of default and it will be reasonable to combine them as one level. PAY_0 = "1 Month Late" stands as the other level. Therefore the number of levels in predictor PAY_0 is reduced to 3.

3.4. Using Kernel Principal Component Analysis to generate a new feature

The predictors in this data may not be independent of each other. The predictors SEX, EDUCATION, and MARRIAGE might be related to each other. For instance, the male may have higher education, but a lower percentage of getting married compared with female. We can use the kernel principal component analysis (Kernel-PCA) [19] to generate new variable(s) to represent the related variables so that the dimensionality of the data can be further reduced. The kernel-PCA works better than PCA when nonlinear relation exists between initial variables in the data. Since there is an interaction between the variables according to our later discussion in session 4.2, such nonlinear relation exists. We run the kernel-PCA on the data that only contains the three variables SEX, EDUCATION, and MARRIAGE to find out the principle components among them. Table 5 is the kernel-PCA results showing the first 4 principal components (PC). The first 4 principal components combined can describe 87% of the variation.

Table 5: The importance of the first 2 principal components.

	PC1	PC2	PC3	PC4
Proportion of variance	0.270	0.240	0.207	0.153
Cumulative Proportion	0.270	0.511	0.718	0.871

The new features PC1, …, PC4 will replace the total 8 levels in the variables SEX, EDUCATION, and MARRIAGE in the later predictive modeling to predict the target variable.

4. Methodology

In this session, we will introduce the main idea of GLM and apply it to the data pre-processed in previous sessions. The decision tree and random forest algorithms will also be applied to compare the GLM. The data was split into training (70%) and test (30%) set for each of the algorithms.

4.1. Introduction of GLM

The ordinary linear regression assumes the target variable follows the normal distribution. However, the target variables in many real-world data don't follow the normal distribution. For instance, personal income follows a lognormal distribution with the majority of people in middle income and few are super-rich. In the credit data used in this paper, the target variable is a binary variable that obviously doesn't follow the normal distribution. For the non-normal distributed target variable, the ordinary linear regression no longer works. The GLM is the generalized version of the linear regression that works for a broader range of target variables including non-normal and normal distributed. The GLM uses a link function to transform the non-normal target variable to normal distributed, then the ordinary regression method can be used.

4.2. Select an Interaction

Even though the generalized linear model (GLM) is a linear model, it can still deal with the non-linear relation. That's done by including the interaction term. The interaction means the relationship between one predictor and the target changes when the value of another predictor changes. This suggests the two predictors are not independent of each other but exist in interaction between them.

In Figure 6, we look at how education and marriage interact. Individuals who are married and are in high school have a larger proportion of non-default than those who are single. However, individuals who are single and are in higher education have a higher proportion of non-default than married people. This might be because single people spend less than married people and have a higher proportion of non-default. It is also worth noting that there are more singles in higher education than married people who are in higher education, as well as a large number of married high school graduates. We observe that at each level of education the number of people who are married and will default varies with no pattern hence we will select this interaction between marriage and education.

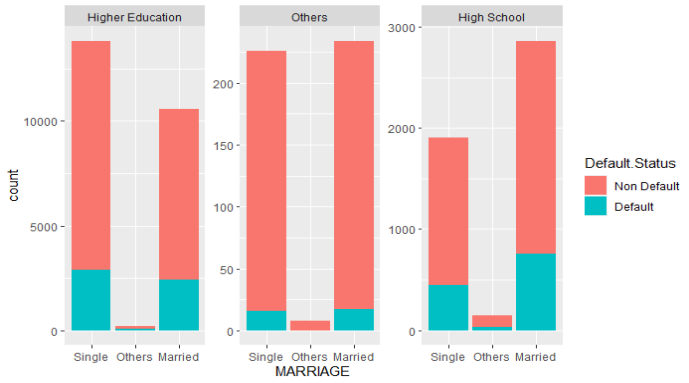


Figure 6: The distribution of default status with Education for each marital level.

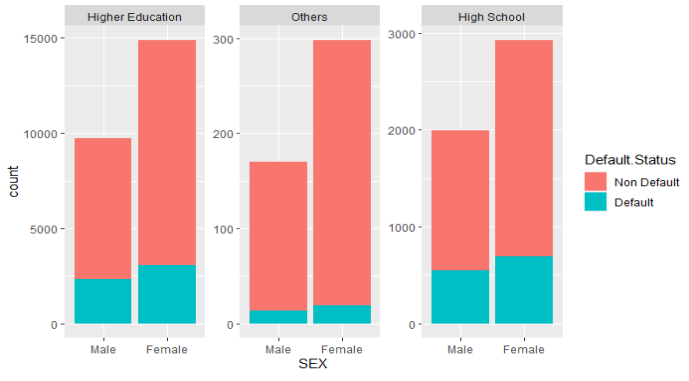


Figure 7: The distribution of default status with Education for each gender.

The interaction between education and sex is explored in Figure 7. There are much more females in all three levels of education who will default than males. Females, on average, have a greater degree of education than males and are less likely to

default. There is an interaction between sex and education because of the pattern it has in each educational level.

We do the same thing to explore the interaction between MARRIAGE and SEX using Figure 8. When SEX = "male", the married male has a significantly lower default risk than the single males. However, when SEX = "female", the default risk is not significantly different between the single females and married females. To explain this, we can regard it's often the men who pay more bills in the marriage. Married men should take more financial responsibility after getting married than single men. Therefore they have higher default risk. While for females, such financial responsibility in marriage is not so significant, thus there is not much difference in terms of default risk between married females and single females. Therefore, the SEX value impacts.

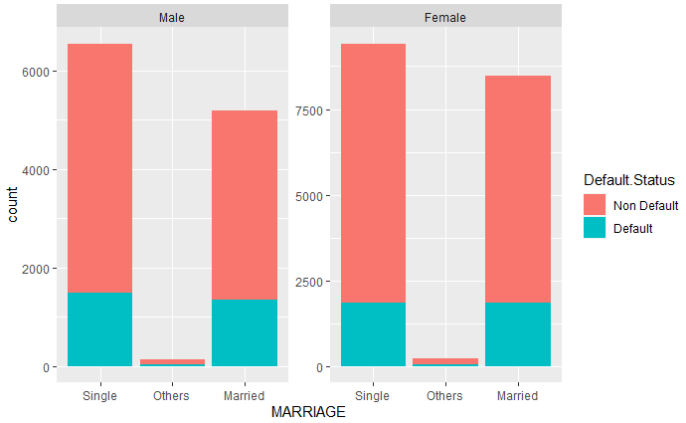


Figure 8: The distribution of default status with Gender for Marital status

4.3. Link Function Selection

We need to specify the distribution of the target variable and what link function to be used. Our target variable is binary, which is either "default" or "not default", the natural choice of the target distribution is binomial.

For the link function, it must map the prediction to the zero to one range because we will first predict the probability of default, then classify it. The binomial distributed GLM has four candidates for the link functions: logit, probit, Cauchit, and cloglog. They will all map the regression result to a value between 0 and 1, which can be regarded as the payment default probability. Table 6 provided details about these 4 link functions.

In Table 6, the $\boldsymbol{X^T\beta}$ is the regression equation. p is the target variable. With the link function, the target variable is transformed into a normal distributed variable that can be regressed using the regular least squares method. The AIC is used as an important metric when we decide which link function to choose. Each link function is tried on the training data and the AIC of its GLM is listed in the last column of Table 6. The Cauchit link and Cloglog link have higher AIC, so we exclude them. The Logit and Probit have similar small AIC, so either of them can be chosen as the link function. Since the Logit link is the canonical link for binary family and it's more widely used, we decide to choose Logit function as the link function.

Table 6: Four link functions for binary distribution and its AIC on our training data.

Name	Link Function	Response Probability	AIC on training data
Logit	$\log\left(\frac{p}{1-p}\right) = \boldsymbol{X^T\beta}$	$p = \frac{\exp(\boldsymbol{X^T\beta})}{1+\exp(\boldsymbol{X^T\beta})}$	18388
Probit	$\Phi^{-1}(p) = \boldsymbol{X^T\beta}$	$p = \Phi(\boldsymbol{X^T\beta})$	18386
Cauchit	$\tan(\text{pv} - \text{p}/2) = \boldsymbol{X^T\beta}$	$p = \frac{1}{\pi} arctan(\boldsymbol{X^T\beta}) + \frac{1}{2}$	18451
Cloglog	$\log(-\log(1 - \text{p})) = \boldsymbol{X^T\beta}$	$p = 1 - exp(-exp(\boldsymbol{X^T\beta}))$	18468

4.4. Feature selection

Feature selection (also known as variable selection, attribute selection, or variable subset selection) is the technique of selecting a subset of relevant features (predictors and variables) for use in the development of a model. It is the automatic selection of the most significant and relevant qualities contained in the data for predictive modeling [20]. Forward selection and backward selection are the two main types of feature selection methods. Forward selection is an iterative approach in which the model starts with no variables. This method keeps adding up the variable that improves the model the most (measured by AIC) in each iteration until adding a new variable no longer enhances the model's performance. The AIC of forward selection is 18380. The backward selection begins with all the variables (full model) and removes the least significant variable one after the other until its AIC no longer decreases. The AIC of backward selection is 18390. The best model solely depends on the defined evaluation criterion of which the AIC was used. Since forward selection has a lower AIC, it's used as the feature selection method. This choice removed variables Age, PC2, PC3, PC4, BILL_AMT3,4,5, PAY_AMT3,5 by not selecting them. The area under curve (AUC) for the test data is reduced to 0.7637 with an AIC decreasing to 18442.

4.5. Data Sampling and G-K-Fold Cross-validation.

To ensure the distribution is unchanged after data is spitted into the training set and testing set, we use stratified sampling. Both sets contain the same portion of credit default data after the data partition. To analyze the accuracy and stability of different algorithms, the G-K-fold stratified cross-validation is used. In K-fold stratified cross-validation, the data is stratified partitioned into K equal parts, where each part of the data has the same distribution for the target variable. 1 part of this data is defined as the testing data, the remaining K-1 parts of the data are the training data. G-K-folder stratified cross-validation will do the K-folder stratified cross-validation for G times, to generate enough results for algorithms performance evaluation statistically.

4.6. Interpretation of the GLM Results

Table 7 listed the results of the regression coefficients of the GLM. It's the trained GLM model generated from the training data

using the features we selected from the step-AIC method based on the new features built from kernel-PCA and variables interaction considered. The "Estimate" column contains the regression coefficients. According to this table, the following variables, levels or interactions are significantly important in predicting the default payment, due to the small p-value and relatively large estimated coefficients:

- PAY_0 = 1 month late
- PAY_0 = More than a month late
- PAY_2 = More than a month late
- PAY_3 = More than a month late
- PAY_4 = More than a month late
- PAY_5 = More than a month late
- PAY_6 = More than a month late
- EDUCATIONOthers:MARRIAGESingle

This makes sense. The PAY_0 is more important than the other payment status variables because it's the most recent payment status. The PAY_0=More_than_a_month_late has a larger estimated coefficient than PAY_0=1_month_late because the former level is associated with a higher probability of default payment, and both levels result in a higher chance of default compared with PAY_0 paid in time. The negative coefficients of interaction levels between EDUCATION and MARRIAGE mean they indicate a lower probability of default payment.

Table 7: Summary of the GLM results.

	Estimate	Std. Error	z value	Pr(>|z|)	Signif. Codes
(Intercept)	-10.51	22.62	-0.47	0.64	
BILL_AMT1	<1E-4	<1E-4	-2.69	0.01	**
BILL_AMT2	<1E-4	<1E-4	3.78	<1E-3	***
BILL_AMT6	<1E-4	<1E-4	-3.10	<1E-2	**
PAY_AMT1	<1E-4	<1E-4	-5.32	<1E-4	***
PAY_AMT2	<1E-4	<1E-4	-2.55	0.01	*
PAY_AMT4	<1E-4	<1E-4	-0.82	0.41	
Limit.Balance	<1E-4	<1E-4	-6.86	<1E-4	***
PAY_AMT6	<1E-4	<1E-4	-2.93	0.00	**
PAY_01 month late	0.78	0.06	13.10	< 2E-16	***
PAY_0More than a month late	2.00	0.06	30.92	< 2E-16	***
PAY_21 month late	-1.26	1.05	-1.20	0.23	
PAY_2More than a month late	0.23	0.07	3.17	<1E-3	**
PAY_31 month late	-10.33	228.91	-0.05	0.96	
PAY_3More than a month late	0.27	0.07	3.79	<1E-4	***
PAY_41 month late	0.87	397.31	0.00	1.00	
PAY_4More than a month late	0.21	0.08	2.76	0.01	**
PAY_5More than a month late	0.27	0.08	3.20	<1E-3	**
PAY_6More than a month late	0.33	0.07	4.49	<1E-4	***
PC1	0.72	2.04	0.35	0.72	
PC2	-2.57	2.57	-1.00	0.32	
PC3	-2.66	4.09	-0.65	0.52	
PC4	-4.28	4.00	-1.07	0.28	
EDUCATIONHigher Education:MARRIAGESingle	11.16	22.68	0.49	0.62	
EDUCATIONOthers:MARRIAGESingle	-51.98	35.15	-1.48	0.14	
EDUCATIONHigh School:MARRIAGESingle	-4.97	13.57	-0.37	0.71	
EDUCATIONHigher Education:MARRIAGEOthers	-45.27	55.01	-0.82	0.41	
EDUCATIONOthers:MARRIAGEOthers	-88.72	127.20	-0.70	0.49	
EDUCATIONHigh School:MARRIAGEOthers	-53.17	38.30	-1.39	0.17	
EDUCATIONHigher Education:MARRIAGEMarried	16.43	34.10	0.48	0.63	
EDUCATIONOthers:MARRIAGEMarried	-48.24	41.72	-1.16	0.25	
EDUCATIONHigh School:MARRIAGEMarried	NA	NA	NA	NA	

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

The trade-off between sensitivity (or TPR) and specificity (1 – FPR) is depicted by the ROC curve. Classifiers with curves that are closer to the top-left corner perform better. A random classifier is supposed to offer diagonal points (FPR = TPR) as a baseline. The test becomes less accurate as the curve approaches the 45-degree diagonal of the ROC space to the left. The area beneath the ROC curve is referred to as the AUC [21]. To generate the ROC curve, we use the roc() function in R's pROC library by comparing the model prediction probability vs the real target value. The ROC curve of this GLM method on the test data is provided in figure 9 and it has an AUC of 0.757.

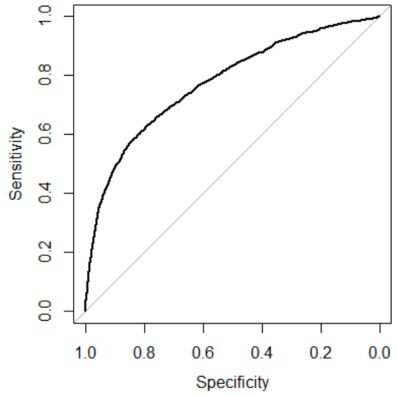


Figure 9: ROC curve of the GLM model

4.7. Classification Tree

Classification tree methods (i.e., decision tree methods) are recommended when the machine learning task contains

classifications or predictions of the outcome. A Classification tree labels, records, and assigns variables to discrete classes. A Classification tree is built through a process known as binary recursive partitioning. This is an iterative process of splitting the data into partitions and then splitting it up further on each of the branches. Before constructing a tree, the data was split into training (70%) and testing (30%) sets.

We use the rpart package in R to implement the decision tree. With its tree structure, the decision tree can automatically incorporate the non-linear relation between variables. Therefore, we don't need to add the interaction term (SexEduMarriage) as the input variable. All the variables in the original data are included as the input variables to build the decision tree. Gini impurity is used to split the tree. Figure 10 is the classification tree produced.

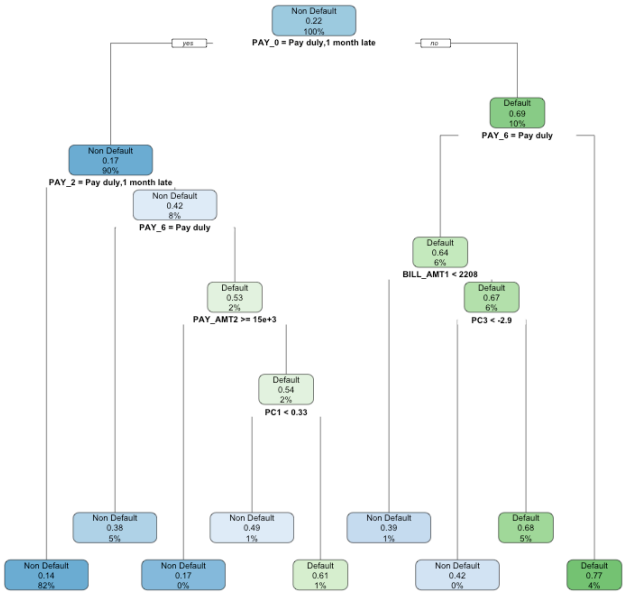


Figure 10: Classification tree result.

A confusion matrix is used to check the performance of a classification model on a set of test data for which the true values are known. Most performance measures such as precision, and recall are calculated from the confusion matrix. We can observe in Table 8 that the confusion matrix of our test data has the true positive to be 6633 with a small false positive of 1253 and there is a false negative of 376 and a true negative of 737, with a classification accuracy of 0.8188.

Table 8: Confusion Matrix of decision tree on the test data

		Reference	
		Non Default	Default
Prediction	Non Default	6633	1253
	Default	376	737

The classification tree in figure 10 indicates that the variable PAY_0 is the most important variable to predict the default risk because it's the variable used to split the first node in the tree. This classification tree algorithm has an accuracy rate of 0.819 and AUC of 0.6966 which is lower than the GLM. Figure 11 provides the ROC curve of this algorithm.

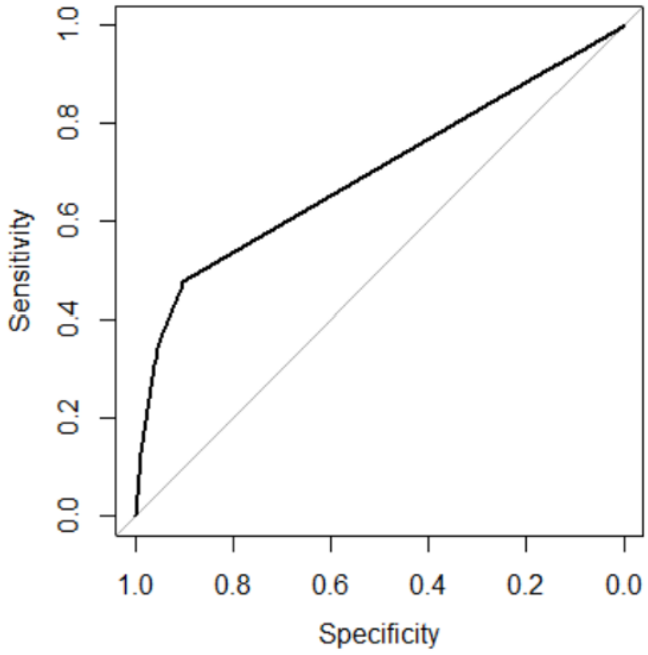


Figure 11: ROC curve for classification tree

4.8. Random Forest

Random forest, as its name implies, consists of many individual decision trees that operate as an ensemble. Each tree in the random forest splits out a class prediction and the class with the most votes becomes our model's final prediction. The random forest can also return the importance of each predictor as shown in Table 9. The results are scaled, so the numbers indicate relative importance. Generally, variables that are used to make the split more frequently and earlier in the trees in the random forest are determined to be more important. According to Table 9, Pay_0 is the most important variable to predict the default payment. This matches the GLM results in Table 7, where the coefficients of the two levels of Pay_0 are highly significant and have large values. The new variable SexEduMarriage that was created by us has an importance of 2.263, which means it was not used frequently in any of the trees. This agrees with the single tree built previously, where the variable SexEduMarriage is not used in any of the split nodes.

Table 9: The relative importance of each predictor variable

	Relative Importance
AGE	392.04
PAY_0	651.78
PAY_2	238.91
PAY_3	189.40
PAY_4	116.50
PAY_5	115.35
PAY_6	77.57
BILL_AMT1	424.41

BILL_AMT2	391.97
BILL_AMT3	376.45
BILL_AMT4	354.90
BILL_AMT5	353.91
BILL_AMT6	356.20
PAY_AMT1	378.61
PAY_AMT2	343.74
PAY_AMT3	330.01
PAY_AMT4	318.30
PAY_AMT5	313.88
PAY_AMT6	328.04
Limit.Balance	371.19
PC1	139.72
PC2	132.30
PC3	149.37
PC4	142.92

The ROC curve of this random forest algorithm is given in figure 12. Its AUC is 0.7639.

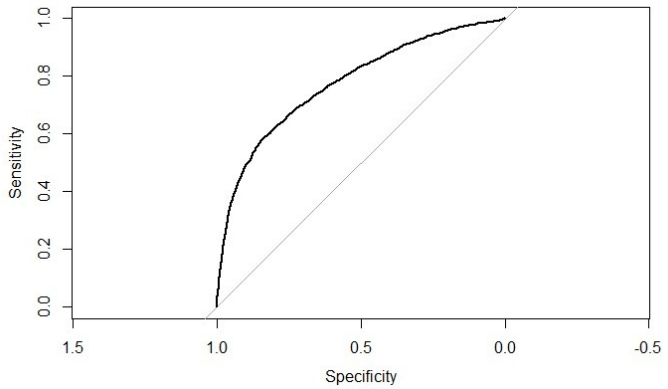


Figure 12: ROC curve for Random Forest

5. Results Comparison and Conclusion

As we mentioned in session 4.5, the G-K-fold stratified cross-validation is used to collect the AUCs and other metrics of each algorithm. We set G=10, K=10. That is in total 100 runs for each algorithm based on different training and testing data partitions. To compare the 3 algorithms, we consider the mean of AUCs, accuracy, sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV), run time with the corresponding standard deviation, and 95% confidence interval for each classifier. They are listed in Table 10.1-10.3. The metrics of GLM are significantly better than the decision tree and comparable to the random forest.

All 3 algorithms are implemented in R and run on a MacBook Pro 2018 with 2.2 GHz 6-Core Intel Core i7, 16 GB 2400 MHz DDR4 memory, and macOS Catalina. We found the GLM runs 7.7 times faster than the decision tree, and 45.7 times faster than the random forest. This is also a big advantage of GLM versus the other 2 algorithms. The details of run time are also listed in tables 10.1-10.3.

Table 10.1: Mean of metrics for 3 algorithms

	Mean		
	GLM	Decision Tree	Random Forest
AUC	0.764	0.693	0.762
Accuracy	0.820	0.820	0.817
Sensitivity	0.357	0.350	0.361
Specificity	0.952	0.954	0.946
PPV	0.678	0.683	0.656
NPV	0.839	0.838	0.839
Run time	0.108	0.830	4.931

Table 10.2: Standard deviation of metrics for 3 algorithms

	Standard Deviation		
	GLM	Decision Tree	Random Forest
AUC	0.0051	0.0047	0.0057
Accuracy	0.0027	0.0027	0.0028
Sensitivity	0.0092	0.0191	0.0104
Specificity	0.0022	0.0052	0.0029
PPV	0.0116	0.0165	0.0123
NPV	0.0020	0.0034	0.0021
Run time	0.0245	0.0649	0.2310

Table 10.3: Standard deviation of metrics for 3 algorithms

	95% Confidence Interval		
	GLM	Decision Tree	Random Forest
AUC	[0.756, 0.776]	[0.685, 0.703]	[0.75, 0.77]
Accuracy	[0.814, 0.826]	[0.815, 0.826]	[0.812, 0.822]
Sensitivity	[0.342, 0.376]	[0.312, 0.381]	[0.343, 0.38]
Specificity	[0.948, 0.956]	[0.944, 0.963]	[0.941, 0.952]
PPV	[0.655, 0.7]	[0.65, 0.716]	[0.637, 0.678]
NPV	[0.836, 0.843]	[0.832, 0.844]	[0.836, 0.843]
Run time	[0.088, 0.136]	[0.734, 0.977]	[4.658, 5.682]

DeLong test is used to compare the GLM, decision tree, and random forest classifiers AUCs. We did the DeLong test for GLM vs decision tree with alternative hypnosis "GLM has higher AUCs than decision tree", another DeLong test for GLM vs random forest with alternative hypnosis "GLM has lower AUCs than random forest". The ROC curves for each pair of algorithms obtained in each run of the G-K-fold crossed validation are used as the input for the DeLong test. Therefore, for each pair of algorithms, G×K DeLong test is computed. The p-values of these tests are recorded, and their mean, standard deviation are reported in Table 11. The p-value of the DeLong test of GLM vs decision tree is extremely small, so we should accept its alternative hypothesis that GLM has higher AUCs than the decision tree. For the GLM vs random forest

DeLong test, its p-value is 0.8628, so we should accept its null hypothesis that GLM has higher AUCs than random forest. A boxplot is given in Figure 13 to summarize the comparison of AUCs obtained from the cross-validation for all 3 algorithms. Both DeLong tests and the boxplot show the superiority of GLM vs the other 2 algorithms.

Table 11: DeLong test to compare AUCs.

	GLM vs Decision Tree	GLM vs Random Forest
Alternative hypothesis	GLM has higher AUCs than the decision tree	GLM has lower AUCs than random forest
Mean of p-value	4.71E-38	0.6617
SD of p-value	4.70E-37	0.2530

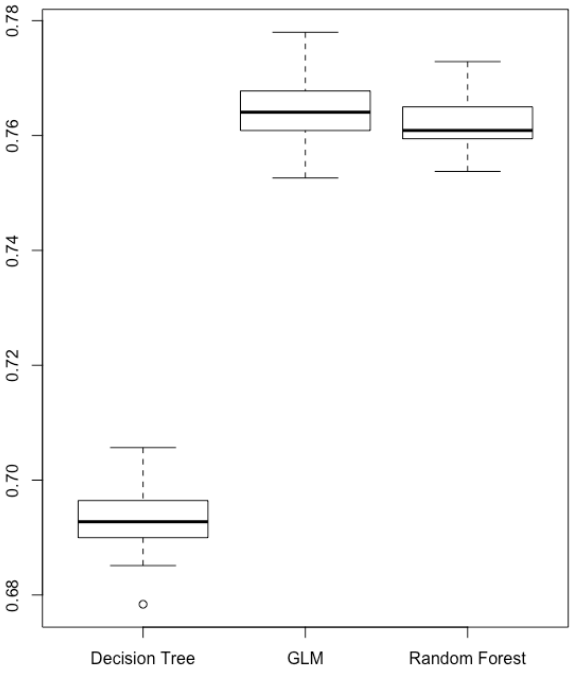


Figure 13: Boxplot to summarize the AUCs comparison.

To discuss the stability of different classifiers, we use the coefficient of variation (CV) of AUC. The CV is defined as the standard deviation divided by the mean. A better classifier is the one with a smaller standard deviation of AUCs and higher average AUCs. The smaller CV of AUC means the algorithm is more stably accurate. We collected the AUCs for 3 algorithms from the G-K-fold cross-validation and calculate the CV of each algorithm's AUCs. The result is listed in Table 12. It shows the 3 algorithms have the similar CV of AUCs.

Table 12: Coefficient of variation (CV) of AUCs.

	GLM	Decision Tree	Random Forest
CV of AUCs	0.00671	0.00676	0.00749

To summarize, the average AUCs for the single decision tree was 0.693 on the test data set, whereas the GLM had 0.764 and the random forest had 0.762. We don't recommend the single decision tree as the final model because of its low AUC. For the remaining two algorithms, the GLM has several advantages over the random forest, one of which is its ease of implementation. The GLM regression formula can be clearly written using an algebra expression. The probability of a default payment can be easily calculated even with simple tools like a calculator or spreadsheet. However, for the random forest, it requires computer programming to obtain the results. Also, the GLM is easier to comprehend and interpret than the random forest, making it easier for credit card companies to grasp and make further business or management decisions upon it. One disadvantage of a GLM over the random forest is that it does not automatically incorporate variable interactions. In GLM, we need first to explore which variables could interact with each other and then manually add those interactions into the regression. The random forest can achieve this automatically. However, the random forest is prone to overfitting, especially when a tree is particularly deep. A small change in the data value could lead to totally different results if using the tree algorithms. Furthermore, the GLM uses fewer features than random forest because of the stepAIC forward selection applied for GLM. Last but not least, the GLM runs much faster than the random forest. Table 13 summarizes the comparison. Based on the above reasons, we conclude GLM model is a better model to predict the credit card default than the decision tree and random forest.

Table 13: Comparison of GLM with other two algorithms on credit card default prediction.

	GLM	**Decision Tree**	**Random Forest**
Average AUC	0.764	0.693	0.762
Interpretability	High	Medium	Low
Implementation	Easiest	Easy	Difficult
Overfitting risk	Low	Medium	High
Algorithm speed	Fast	Slower	Slowest

Conflict of Interest

The author declares no conflict of interest.

Acknowledgment

The author would like to thank the reviews for their valuable time spent on reviewing this paper, and for helpful suggestions to improve this paper.

References

[1] M. Chou, "Cash and credit card crisis in Taiwan." Business weekly: 24-27, 2006.

[2] A. Subasi, S. Cankurt, "Prediction of default payment of credit card clients using Data Mining Techniques," In 2019 International Engineering Conference (IEC), 115-120, IEEE, 2019, doi: 10.1109/IEC47844.2019.8950597.

[3] A. Verikas, Z. Kalsyte, M. Bacauskiene, A. Gelzinis, "Hybrid and ensemble-based soft computing techniques in bankruptcy prediction: a survey," Soft Computing, **14**(9), 995-1010, 2010, doi: 10.1007/s00500-009-0490-5.

[4] F.N. Koutanaei, H. Sajedi, M. Khanbabaei, "A hybrid data mining model of feature selection algorithms and ensemble learning classifiers for credit scoring," Journal of Retailing and Consumer Services, **27**, 11-23, 2015, doi: 10.1016/j.jretconser.2015.07.003.

[5] M. Berry, G. Linoff, Mastering data mining: The art and science of customer relationship management. New York: John Wiley & Sons, Inc, 2000.

[6] P. Giudici, "Bayesian data mining, with application to benchmarking and credit scoring," Applied Stochastic Models in Business and Industry, **17**(1), 69-81, 2001, doi:,10.1002/asmb.425.

[7] L. C. Thomas, "A survey of credit and behavioral scoring: Forecasting financial risk of lending to consumers," International Journal of Forecasting, **16**, 149–172, 2000, doi: 10.1016/S0169-2070(00)00034-0.

[8] J. N. Crook, D. B. Edelman, L. C. Thomas, "Recent developments in consumer credit risk assessment," European Journal of Operational Research, **183**(3), 1447– 1465, 2007, doi: 10.1016/j.ejor.2006.09.100.

[9] G. Derelioğlu, F. G. Gürgen, "Knowledge discovery using neural approach for SME's credit risk analysis problem in turkey," Expert Systems with Applications, **38**(8), 9313– 9318, 2011, doi: 10.1016/j.eswa.2011.01.012.

[10] E. I. Altman, "Financial ratios, discriminant analysis and the prediction of corporate bankruptcy," The journal of finance, **23**(4), 589–609, 1968, doi: 10.2307/2978933.

[11] C. Verma, V. Stoffová, Z. Illés, S. Dahiya, "Binary logistic regression classifying the gender of student towards computer learning in European schools," In The 11th Conference Of Phd Students In Computer Science (p. 45), 2018.

[12] W. Henley, D. J. Hand, "A K-Nearest-Neighbour Classifier for Assessing Consumer Credit Risk," Journal of the Royal Statistical Society: Series D (The Statistician), **45**(1), 77-95, 1996, doi: 10.2307/2348414.

[13] D. West, "Neural network credit scoring models," Computers & Operations Research, **27**(11), 1131–1152, 2000, doi: 10.1016/S0305-0548(99)00149-5.

[14] S. Oreski, G. Oreski, "Genetic algorithm-based heuristic for feature selection in credit risk assessment," Expert systems with applications, **41**(4), 2052-2064, 2014, doi: 10.1016/j.eswa.2013.09.004.

[15] T. Harris, "Credit scoring using the clustered support vector machine," Expert Systems with Applications, **42**(2), 741–750, 2015, doi: 10.1016/j.eswa.2014.08.029.

[16] S. Yang, H. Zhang, "Comparison of several data mining methods in credit card default prediction," Intelligent Information Management, **10**(5), 115-122, 2018, doi: 10.4236/iim.2018.105010.

[17] Y. Yu, "The application of machine learning algorithms in credit card default prediction," In 2020 International Conference on Computing and Data Science (CDS), 212-218, IEEE, 2020, doi: 10.1109/CDS49703.2020.00050.

[18] T. M. Alam, K. Shaukat, I. A. Hameed, S. Luo, M. U. Sarwar, S. Shabbir, J. Li, M. Khushi, "An investigation of credit card default prediction in the imbalanced datasets," IEEE Access, 8, 201173-201198, 2020, doi: 10.1109/ACCESS.2020.3033784.

[19] H. Hoffmann, "Kernel PCA for novelty detection," Pattern recognition, **40**(3), 863-874, 2007, doi: 10.1016/j.patcog.2006.07.009.

[20] V. Kumar, S. Minz. "Feature selection: a literature review," SmartCR, **4**(3), 211-229, 2014, doi: 10.6029/smartcr.2014.03.007.

[21] J. Huang, C. X. Ling, "Using AUC and accuracy in evaluating learning algorithms," IEEE Transactions on knowledge and Data Engineering, 17(3), 299-310, 2005, doi: 10.1109/TKDE.2005.50.

ASTES

Advances in Science, Technology and Engineering Systems Journal Vol. 7, No. 3, 62-69 (2022)

www.astesj.com

ASTESJ

ISSN: 2415-6698

A New Technique to Accelerate the Learning Process in Agents based on Reinforcement Learning

Noureddine El Abid Amrani [1,2*], Ezzrhari Fatima Ezzahra[1], Mohamed Youssfi[1], Sidi Mohamed Snineh[1], Omar Bouattane[1]

[1]*2IACS LAB, ENSET, Hassan II University of Casablanca, Morocco*

[2]*Institut supérieur du Génie Appliqué (IGA), Casablanca, Morocco*

ARTICLE INFO

Article history:
Received: 08 January, 2022
Accepted: 01 May, 2022
Online: 25 May, 2022

Keywords:
Markov Decision Process
Multi-agent System
Q-Learning
Reinforcement Learning

ABSTRACT

The use of decentralized reinforcement learning (RL) in the context of multi-agent systems (MAS) poses some difficult problems. The speed of the learning process for example. Indeed, if the convergence of these algorithms has been widely studied and mathematically proven, they suffer from being very slow. In this context, we propose to use RL in MAS in an intelligent way to speed up the learning process in these systems. The idea is to consider the MAS as a new environment to be explored and the communication, between the agents, is limited to the exchange of knowledge about the environment. The last agent to explore the environment has to communicate the new knowledge to the other agents, and the latter have to build their knowledge bases taking into account this knowledge. To validate our method, we chose to evaluate it in a grid environment. Agents must exchange their tables (Qtables) to facilitate better exploration. The simulation results show that the proposed method accelerates the learning process. Moreover, it allows each agent to reach its goal independently.

1. Introduction

In the literature, an agent is defined according to the type of application for which it is designed. In [1]-[3], the authors define an agent is an entity that can be considered as perceiving its environment by means of sensors and acting on this environment by means of effectors.

In this paper, we have chosen to limit the communication between agents to the exchange of knowledge because we are interested in the autonomy of the agent. In [4], [5], the authors specify that the autonomy of the agent is related to its structure: for example, for a cognitive agent that plans to reach its goals, we can talk about autonomy by planning.

For an agent deployed in a MAS, the agent's autonomy depends on the objective to be achieved, which is:

- Global: this requires cooperation with other agents and limits the agent's decision-making and behavior and makes the RA algorithm difficult.
- Individual (agent-specific): in this case, the agent can achieve his objectives alone. He does not need to cooperate with other agents. And if it decides to do so, it is to enrich its knowledge base. In this context, the authors of [6] propose characteristics of the autonomous agent:
  - An autonomous agent has its objectives;
  - He can make decisions about his objectives;
  - It can decide autonomously when to adopt the objectives of other agents;
  - He sees cooperation with other agents as a way to enrich his knowledge base to achieve his objectives;

In this context, and concerning our problem and the aspect we want to address, we adopt for our agents the properties proposed by the authors of [6].

The structure of this paper is as follows: after establishing the necessary background and notation in Section 1, we briefly introduce in Section 2 some approaches that use Q-Learning on non-Markovian environments. Then, in Section 3, we give a brief overview of Markov decision process. In Section 4, we give an overview of the mathematical formulas used in reinforcement learning. We then describe our new technique for accelerating the agent learning process based on reinforcement learning in Section 5. Finally, we present some conclusions and suggestions for future work in Section 6.

* Corresponding Author: Noureddine El Abid Amrani, Morocco, +212673551661, noureddine.elabid@iga.ac.ma

2. Related Work

In MAS, interaction with the environment, both for perception and action operations and for message sending operations, is synchronous for the agent. On the other hand, the communication between agents remains asynchronous, preventing any assumption on the state of an agent following the sending of a message. The main disadvantage of this domain is that the behavior of an agent will depend on the other agents and thus, we go out of the classical framework of reinforcement learning. This is a problem that has been raised by several researchers:

In [7], the authors show that Q-Learning cannot converge for non-Markovian problems. Indeed, in theory, the convergence of Q-Learning is only possible if the policy used is fixed during the exploration phase of the environment. This policy can be stochastic, but it must be fixed: for a given state, the probability of choosing an action must be constant. This is generally not the case with classical methods. If the exploration policy is fixed, the Q-Learning algorithm converges to the utility estimate of the fixed policy used. This requires the introduction of new techniques in the reinforcement learning algorithm. Several solutions have been proposed:

In [8], the authors present a hybrid Q-learning algorithm named CE-NNR which is springed form the CE-Q learning and NNR Q-learning is presented. The algorithm is then well extended to RoboCup soccer simulation system.

In [9], the authors propose a decentralized progressive reinforcement learning that, based on classical RL (Reinforcement Learning) techniques, allows to endow agents with stochastic reactive behaviors. Indeed, with these stochastic behaviors, the agents will be more efficient in this framework of partial perceptions of the global system.

In [10], the authors first present the principles of multi-agent systems and their contribution to the solution of certain problems. They then propose an algorithm that allows to solve in a distributed way a problem posed globally to the community of agents based on reinforcement learning in the multi-agent framework.

In [11], the authors present a new algorithm for independent agents that allows learning the optimal joint action in games where coordination is difficult. The approach is motivated by the decentralized character of this algorithm which does not require any communication between agents and Q-tables of size independent of the number of agents. Conclusive tests are also performed on repeated cooperative games, as well as on a chase game.

In [12], the authors of this article propose an algorithm called adaptive Q-learning to judge whether the strategy of a particular agent affects the benefit of the corresponding joint action depending on the TD error. In a multi-agent environment, this algorithm makes it possible to adjust the dynamic learning rate between the agents, the coordination of the strategies of the agents can be carried out.

3. Markov Decision Process (MDP)

The problem of learning [13]-[17], like planning, is to control the behavior of an agent. We, therefore, need to model this agent in its environment. The model used considers on the one hand the possible states of the system (=agent + environment) and on the other hand the possible actions of the agent in its environment.

Since the environment is uncertain, it is easy to think of stochastic modeling of the system. The basic tool that will be used for the learning and the general functioning of the agents is the Markov Decision Process (MDP) and its derivatives. This is a class of Markov models, themselves belonging to stochastic processes.

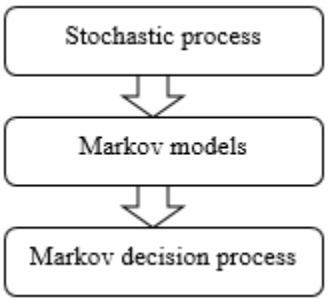


Figure 1: Evolution of the models [18]

The starting idea of a stochastic process is to use as a representation a graph in which the nodes are the possible states of the system and the arcs (oriented and annotated) give the probabilities of passage from one state to another.

3.1. Stochastic process

Any family of random variables X_t is called a stochastic process or random process. This means that to any $t \in T$ is associated a random variable taking its values in a numerical set E. We denote the process X_t. If T is countable, we say that the process is discrete; if T is an interval, we say that the process is permanent.

3.2. Markov Model

In the Markov model, at state t, X(t) depends only on the n previous states (memory of the process in a model of order n). At order 1, as it is often the case:

$$\boldsymbol{P(X(t) = si|X(t-1) = sj, X(t-2) = sk, \dots) = P(X(t) = si|X(t-1) = sj)} \quad (1)$$

The Markov model is said to be stationary if the transition probability between two states does not vary with time:

$$\boldsymbol{\forall t, k\ P(X(t) = si|X(t-1) = sj) = P(X(t+k) = si|X(t+k-1) = sj)} \quad (2)$$

This leads to define a transition probability matrix A = [aij] where the aij are defined by:

$$\boldsymbol{aij = P(X(t) = sj|X(t-1) = si)\ 1 \le i \le N, 1 \le j \le N} \quad (3)$$

The matrix A being stochastic, i.e. verifies:

$$\boldsymbol{\forall i, j\ aij \ge 0, \forall i\ \sum_{j=1}^{N} aij = 1} \quad (4)$$

where N is the number of states of the system.

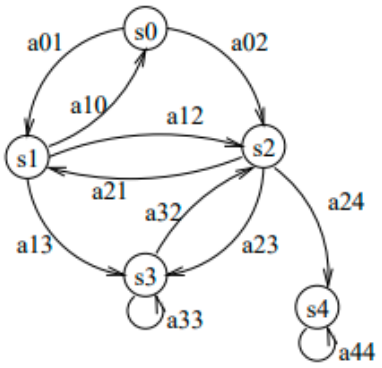


Figure 2: Graph of a Markov model [18].

aij = transition probability of X(i) → X(j).

The agents will be immersed in a world where, starting from a given state, the same action does not always lead to the same state. This may be because the state is only incompletely known (only a partial observation is known), or that simultaneous actions of other agents may take place. Markov models allow modeling such situations.

The different actions performed and the different states encountered will bring any agent again or a cost. The agents must therefore try to act in the best way. Their policies can be diverse depending on whether they consider the more or less long term, different estimates of the expected future gain being possible.

One last point to note is that an agent may not know at the beginning the laws governing its environment. It will then have to learn them more or less directly to improve its chances of winning.

A Markov decision process is a quadruple $(\boldsymbol{S},\ \boldsymbol{A},\ \boldsymbol{P_a},\ \boldsymbol{R_a})$, where:

- $\boldsymbol{S}$ is a set of states called state space,
- $\boldsymbol{A}$ is a set of actions called action space (alternatively, $\boldsymbol{A_s}$ is the set of actions available from state $\boldsymbol{S}$),
- $\boldsymbol{P_a(s,s') = P_r\ \ (\ s_{t+1} = s' \mid s_t = s,\ a_t = a)}$ is the probability that action $\boldsymbol{a}$ in state $\boldsymbol{s}$ at time $\boldsymbol{t}$ will lead to state $\boldsymbol{s'}$ at time $\boldsymbol{t+1}$,
- $\boldsymbol{R_a(s,s')}$ is the immediate reward (or expected immediate reward) received after the transition from state $\boldsymbol{s}$ to state $\boldsymbol{s'}$, due to action $\boldsymbol{a}$

Example:

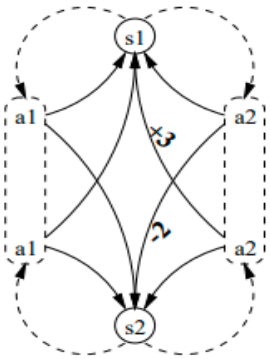


Figure 3: An example of a Markov decision process [18]

3.3. Use

Starting points: The aim is to maximize the reward in the more or less long term. It is, therefore, necessary to measure the expectation of gain, by taking less and less account of the future (less confidence, thanks to the coefficient: γ ϵ [0, 1]):

$$\boldsymbol{E\left(\sum_{t=0}^{\infty} \gamma^t r_t\right)} \qquad (5)$$

Other calculation methods are possible, but only the previous formula will be used in the following.

In each state, it will be necessary to determine the optimal action to take. We thus define a policy p:

$$\boldsymbol{p: S \rightarrow A} \qquad (6)$$

Values used: the expectation of gain will depend on the chosen policy. We can thus define, for a fixed policy p, the utility of a state:

$$V_p(S_t) = E\left(\gamma^k E\left[r\left(S_{t+k}, p(S_{t+k})\right)\right]\right) | (S_t = s) \qquad (7)$$

For an optimal policy: V* (s) = Supp Vp(s), the utility of an optimal policy can be written recursively as the Bellman equation:

$$\boldsymbol{V^*(s) = \max_{a \epsilon A}\left(E_{r \times \pi(s,a)}[r(s,a) + \gamma V^*(s')]\right)} \qquad (8)$$

We obtain an optimal deterministic policy by looking for the action for which we obtain this maximum.

In planning, the algorithms called Value Iteration and Policy Iteration allow, starting from any function V, to converge to V* and thus to find an optimal policy.

Another greatness [15], Q (s, a), can be used, obtained from V (s). It will provide the "quality" of the action performed in the state:

$$\boldsymbol{Q(s,a) = E_{r \times \pi(s,a)}[r(s,a) + \gamma V^*(s')]} \qquad (9)$$

3.4. Others models

From the Markov model, variants are defined which are used in pattern recognition as well as in planning (like MDPs). The general goal is to guess the present or future state of a system. Here are some derived Markovian models:

- HMM: in a Hidden Markov Model, the state of the system is not known. On the other hand, we have an observation that is linked to the states by probabilistic laws. We cannot be sure of a state with a perception of the external world, but a series of perceptions can refine a judgment. This is the main tool for pattern recognition based on Markov models.
- MMDPs (Multiple MDPs) is a variant of MDPs adapted to the case of multi-agent systems, as are DEC-MDPs (decentralized MDPs) and Markov games.
- SMDP: the model, called Semi-Markovian, aims at improving time management, considering that the passage in a state can be of variable duration (according to stochastic laws).

POMDP: this is a mixture of HMMs and MDPs, these Partially Observable MDPs add the idea that an agent has only a partial perception of its environment so that it knows only an observation and not a complete state.

4. Reinforcement learning

4.1. Principle

Reinforcement learning is one of the most widely used machine learning methods in the world of data science. This technique allows the computer to perform complex tasks on its own. The machine learns from its experiences through a system of penalties or rewards. Reinforcement learning involves an algorithm with great potential: Q-learning. This algorithm focuses on learning a policy that maximizes the total reward. Q-learning focuses on the usefulness of the action to be performed to obtain a reward. The Markov Decision Process is the formal model used for reinforcement learning [19], [20].

A MDP is a tuple$(\boldsymbol{S}, \boldsymbol{A}, \boldsymbol{P}, \boldsymbol{R})$ with:

- $\boldsymbol{S}$ is a finite set of states,
- $\boldsymbol{A}$ is a set of actions,
- P: $\boldsymbol{S} \times \boldsymbol{A} \times \boldsymbol{S} \rightarrow [\boldsymbol{0}, \boldsymbol{1}]$, The probability that the agent moves from one state to another is called the transition probability.
- $\boldsymbol{R}$: $\boldsymbol{S} \times \boldsymbol{A} \rightarrow \boldsymbol{R}$, the reward function.

A very important assumption of reinforcement learning, called the Markov property, is defined on the transition function of the environment states. An environment is Markovian if the transition function to a state depends only on the previous state.

Remember that the agent's objective is to maximize the total rewards received. In this context, we introduce the notion of a discount coefficient, denoted γ, which allows us to give more importance to nearby rewards rather than to more distant ones. The sum of the rewards is then: $\boldsymbol{R_t} = \sum_{\boldsymbol{k=0}}^{\mathrm{T}} \boldsymbol{\gamma}^{\mathrm{k}} \boldsymbol{E}(\boldsymbol{r_{t+k+1}})$, with $\boldsymbol{\gamma \epsilon [0, 1[}$. In the special case where $\boldsymbol{\gamma = 0}$, the agent maximizes the immediate reward, whereas when $\boldsymbol{\gamma \rightarrow 1}$, the agent increasingly considers long-term rewards. T represents the duration of an episode that corresponds to a sub-sequence of interaction between the agent and the environment.

The value function $\boldsymbol{V^{\pi}(s)}$ represents the quality of state s in terms of the expected reward from following policy π. Formally, this value function is defined by:

$$\boldsymbol{V^{\pi}(s)} = \sum_{\boldsymbol{t=0}}^{\infty} \boldsymbol{\gamma' E(r_t | \pi, s_t)} \qquad (10)$$

Similarly, we define the value function $Q_{\pi}(s, a)$ as the expectation of the reward if the agent prefers the action in state s and follows the policy π thereafter. Finding a solution to a reinforcement learning problem comes down to finding a policy that maximizes the sum of long-run rewards. Bellman's optimal equation for $\boldsymbol{V^*(s)}$ describes the fact that the optimal value of a state is equal to the expected reward for the best action in that state. More formally:

$$\boldsymbol{V^*(s)} = \boldsymbol{max_{a \in A}} \{\boldsymbol{r} + \boldsymbol{\gamma} \sum_{\boldsymbol{s' \epsilon S}} \boldsymbol{p(s'|s, a) V(s', \pi')}\} \qquad (11)$$

The relationship between $\boldsymbol{V(s)}$ and $\boldsymbol{Q(s, a)}$ is represented by the equation $\boldsymbol{V^*(s) = max_a Q^*(s, a)}$. From the optimal value function $\boldsymbol{Q^*(s, a)}$, we define the optimal policy as $\boldsymbol{\pi^*(s) = arg\ max_a Q^*(s, a)}$. If the dynamics of the system is known, it is possible to solve the problem using dynamic programming to find the optimal policy. In cases where the dynamics of the system are not known, one must either use estimation or use a set of methods, called time-difference learning (TD-Learning), which are able to evaluate the optimal policy through the experiences generated by an interaction with the system. One of the main algorithms for computing the optimal policy, Q-Learning, uses the Q-value function. This algorithm works by interacting with the environment in the following way: in each episode, the agent chooses an action a based on a policy $\boldsymbol{\pi}$ derived from the current values of Q. It executes action a, receives the reward, and observes the next state. Then it updates the values of $\boldsymbol{Q(s, a)}$ by the formula:

$$\boldsymbol{Q(s_t, a_t)} = \boldsymbol{Q(s_t, a_t)} + (\boldsymbol{1} - \alpha)(\boldsymbol{r_t} + \boldsymbol{\gamma} \max_{\boldsymbol{a}} \boldsymbol{Q(s_{t+1}, a_t)} - \boldsymbol{Q(s_t, a_t)}) \qquad (12)$$

where r is the immediate reward, s' is the new state, α is the learning rate. The convergence of this algorithm to the optimal value function has been demonstrated in [21]. The choice of action at each step is made using exploration functions. Among these methods we can mention the voracious, e-voracious and softmax approach which choose an action at random according to a certain probability. For a more exhaustive description of exploration functions, one can refer to[8], [22].

Hence the algorithm:

Algorithm 1: Q-Learning

Initialize Q0 (s,a) arbitrarily;
Choose a starting point s0;
while the policy is not good enough **{**
  Choose at as a function of Q_t (S_t,.);
  In return: s_{t+1} and r_t
  $Q_{t+1}(s_t, a_t) = (1 - \alpha_t) Q_t(s_t, a_t) + \alpha_t(r_t + \gamma * max_{a \in A} Q_t(s_{t+1}, a')$;
}

The coefficient α should then be adjusted to gradually fix the learned policy. γ allows for its part to modulate the importance of expected future rewards.

4.2. Implementation

Two factors must be defined to apply the Q-learning algorithm: first, the learning factor α, which determines to what extent the newly calculated information outperforms the old one. The latter must be between 0 and 1, and decrease towards 0 because if $\alpha = 0$, the agent learns nothing. On the other hand, if $\alpha = 1$, the agent always ignores everything it has learned and will only consider the most recent information. Second, the discount factor γ determines the magnitude of future rewards. A factor of 0 would make the agent myopic considering only current rewards, while a factor close to 1 would also imply more distant rewards. In our approach, we chose: $\alpha = 0.1$ and $\gamma = 0.9$.

It remains to choose the actions to be carried out according to the knowledge acquired.

5. Proposed approach

In our proposed approach, the agent can achieve its goals alone. The cooperation with other agents is limited to the exchange of Qtable learning tables to enrich its knowledge base. To do so, after each new exploration, the agent has to communicate its learning table Qtable to the other agents, and the latter have to build their knowledge bases taking into account this Qtable. And each agent that wants to explore the environment again must use its knowledge base, i.e. the values of the last Qtable communicated

by the environment, as initial values of its Qtable. This allows the agent to gain a large exploration step and subsequently accelerate the speed of the learning process.

5.1. The class diagram

In the class diagram shown in Figure 4, agents are autonomous entities hosted in a graphical environment. Each node of this graph represents a state, and the links between these nodes represent the transitions between these states. There are three types of states in this environment: with penalties, without penalties and obstacles.

In this case, reinforcement learning is applied as follows:

- The transition that allows access to the final state (the goal) is represented by a reward equal to 100.
- The transition that leads to an obstacle is represented by a reward equal to -1.
- The transition that allows access to other states: -10 for states with penalties and 0 for normal states.

The agent must maximize its rewards which allows it to reach the optimal path to the final state.

In the learning phase, the experience of agents who have already explored the environment can be used to improve the speed of the adaptation and learning process of other agents. Indeed, before the start of learning, the function is initialized by the Q values of the last agent having explored the environment. After learning, the agent shares the new Q-values with the other agents in order to use them as initialized values for a new exploration.

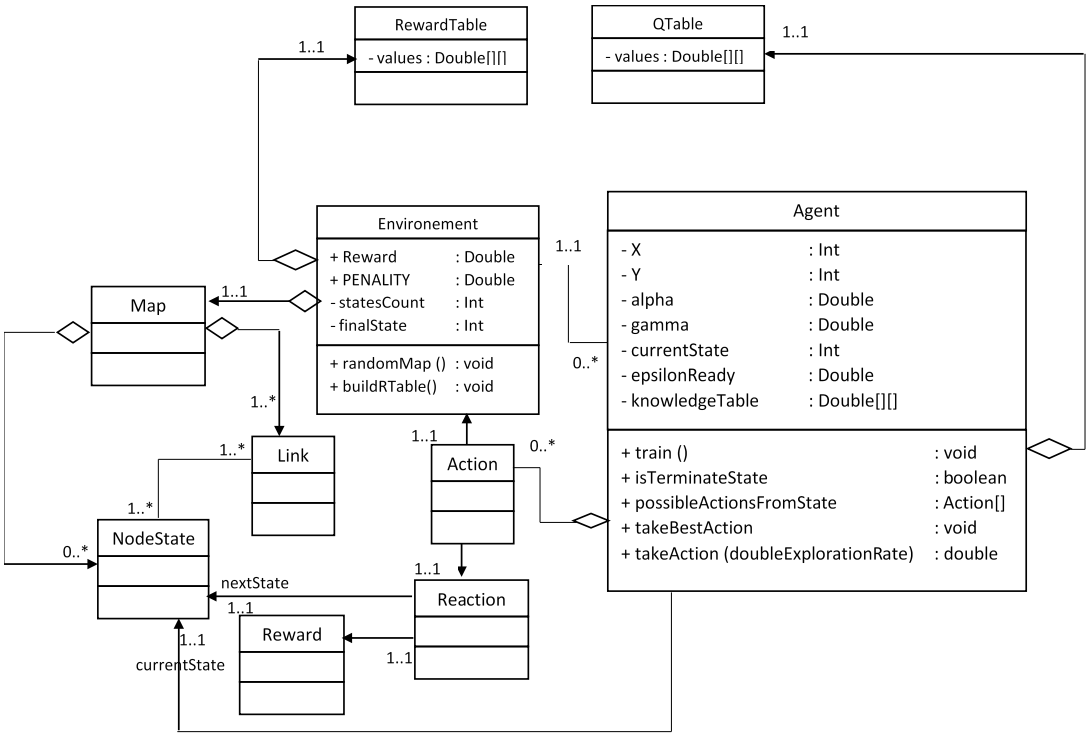


Figure 4: The class diagram of our approach

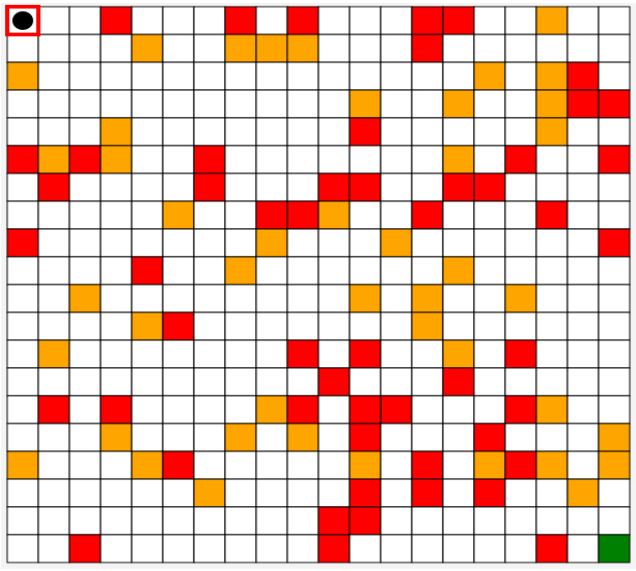


Figure 5: An example of the environment generated in the simulator we have developed.

5.2. Experiment

We have chosen a maze-like environment to test our model. The number of rows and the number of columns of this environment are chosen randomly using the following function:

Algorithm 2:

```
this.map=new char[rows][cols];
for (int i = 0; i < rows; i++){
for (int j = 0; j < cols; j++) {
    double rnd=Math.random();
     if (rnd<0.75)
        this.map[i][j]='0';
    else{
            this.map[i][j] = (Math.random() > 0.5) ? 'P' : 'W';}
}
}
```

To test our application, we choose our environment as a maze where 75% of the cells represent normal states and the rest represent obstacle and penalty states (see Figure 5).

To build the rewards table, the application calls the buildRewardTable function as follows:

Algorithm 3:

```
public void buildRTable(){
    for (int k = 0; k < statesCount; k++) {
        int row=k/gridWidth;
        int col=k%gridWidth;
        for (int s = 0; s < statesCount; s++) {
            System.out.println(s+"->"+k);
            R[k][s]=-1;
            }
        if (map[row][col]!='F'){
            for(int[] d :directions){
                int c=col+d[1];
                int r=row+d[0];
                if ((c>=0 && c<gridWidth)&&(r>=0 &&
                r<gridHeight)){
                    int st=r*gridWidth+c;
                    if (map[r][c]=='0')
                        R[k][st]=0;
                    else if (map[r][c]=='W')
                            R[k][st]=-1;
                        else if (map[r][c]=='F')
                                R[k][st]=EnvGrid.REWARD;
                            else if (map[r][c]=='P')
                                    R[k][st]=EnvGrid.PENALT
                                    Y;
                }
            }
        }
    }
    }
```

The learning of the agents is described by the code below; it allows the construction of the learning table Qtable. The values of this table are computed as follows:

- The agent tries at each learning period to find a solution to achieve the objective.
- Each time, the agent repeats random movements between the states by performing the possible actions.
- For each transition, the value of the Qtable is calculated.

Algorithm 4:

```
public void train(){
    Random random=new Random();
    for (int i = 0; i <1000 ; i++) {
        this.currentStat=random.nextInt(envGrid.statesCou
nt);
        while (!isTerminalState()){
            int[]
            possibleActionsFromState=possibleActionsFro
mState(currentStat);
            int
index=random.nextInt(possibleActionsFromState.l
ength);
            int
nextState=possibleActionsFromState[index];
            double q=Q[currentStat][nextState];
            double maxQNextState=maxQ(nextState);
            double r=envGrid.R[currentStat][nextState];
            double
value=q+alpha*(r+gamma*maxQNextState-q);
            Q[currentStat][nextState]=value;
            currentStat=nextState;
        }
    }}
public int[] possibleActionsFromState(int state){
        List<Integer> states=new ArrayList<>();
        for (int i = 0; i <envGrid.statesCount ; i++) {
            if(envGrid.R[state][i]!=-1) states.add(i);
        }
        return states.stream().mapToInt(i->i).toArray();
    }
```

5.3. Result

The graph in Figure 6 shows the evolution of an agent's learning process that started with a randomly defined knowledge base and no experience:

This agent must communicate his experience (his Qtable) to the other agents. The latter will use it as an initial Qtable for new exploration.

The graph in figure 7 shows the evolution of the learning process of an agent who used the experience of another agent who had already explored the environment, i.e. he initially used the latter's Qtable as a knowledge base.

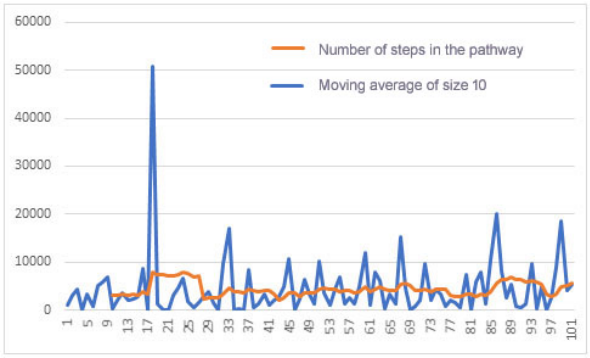


Figure 6: Number of learning path states for the first 100 epochs of an agent with an empty knowledge base (initial Qtable).

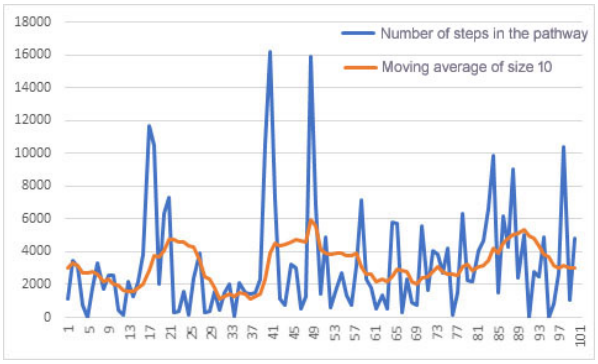


Figure 7:Number of learning path states for the first 100 epochs of an agent

Figure 8 show the paths taken by the last agent for three trials at different initial positions. We see that the agent correctly takes the optimal path with sometimes the preference to cross a penalty point in a minimum number of moves.

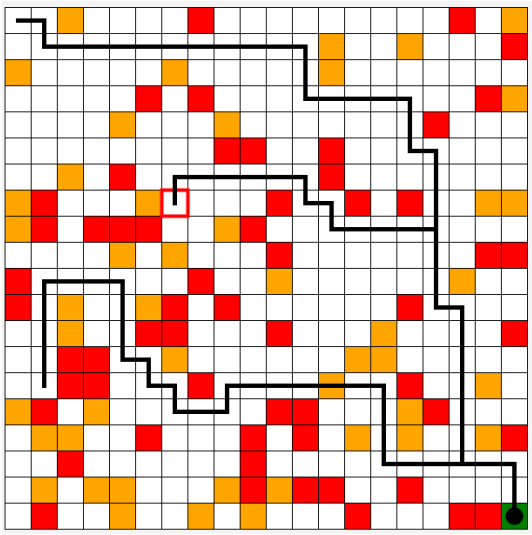


Figure 8: Execution.

6. Conclusion and Future Work

In this paper, we have proposed an approach based on reinforcement learning to solve the problem of cooperation between agents in multi-agent systems. To do so, we proposed in the agent structure that cooperation is limited to the exchange of new knowledge. Indeed, an agent who has just explored the environment must send the new knowledge collected to the other agents, and the latter use this knowledge for new exploration. To validate our method, we chose to evaluate it in a grid environment. Agents must exchange their tables (Qtables) to facilitate better exploration. According to the results, the experiences of the agents that have already explored the environment help to accelerate the reinforcement learning process in MAS. This allows the agent to take an important exploration step and then reach its goal faster.

Ensuring communication between agents in different environments is the goal of our future work. Ontologies can meet this need, that is to enrich the knowledge base of agents with new knowledge provided by other environments. And this requires migrating to other environments, our idea is to use a mobile agent that can migrate to other environments to collect new knowledge and share it with other agents. The mobile agent receives rewards in return.

References

[1] N. Carlési, F. Michel, B. Jouvencel, J. Ferber, “Generic architecture for multi-AUV cooperation based on a multi-agent reactive organizational approach,” IEEE International Conference on Intelligent Robots and Systems, 5041–5047, 2011, doi:10.1109/IROS.2011.6048686.

[2] P. Ciancarini, M. Wooldridge, “Agent-oriented software engineering,” Proceedings - International Conference on Software Engineering, 816–817, 2000.

[3] Z. Papp, H.J. Hoeve, “Multi-agent based modeling and execution framework for complex simulation, control and measuring tasks,” Conference Record - IEEE Instrumentation and Measurement Technology Conference, **3**(section 2), 1561–1566, 2000, doi:10.1109/imtc.2000.848734.

[4] X. Yang, Z. Feng, G. Xu, “An active model of agent mind: The model of agent’s lived experience,” Proceedings - 2009 International Conference on Information Engineering and Computer Science, ICIECS 2009, 3–5, 2009, doi:10.1109/ICIECS.2009.5364300.

[5] A. Kumar, A. Tayal, S.R.K. Kumar, B.S. Bindhumadhava, “Multi-Agent autonomic architecture based agent-Web services,” Proceedings of the 2008 16th International Conference on Advanced Computing and Communications, ADCOM 2008, 329–333, 2008, doi:10.1109/ADCOM.2008.4760469.

[6] M. Huang, P.E. Caines, R.P. Malhamé, “A locality generalization of the NCE (Mean Field) principle: Agent specific cost interactions,” Proceedings of the IEEE Conference on Decision and Control, 5539–5544, 2008, doi:10.1109/CDC.2008.4738976.

[7] S. Filippi, O. Cappé, A. Garivier, “Optimally sensing a single channel without prior information: The Tiling Algorithm and regret bounds,” IEEE Journal on Selected Topics in Signal Processing, **5**(1), 68–76, 2011, doi:10.1109/JSTSP.2010.2058091.

[8] W. Chen, J. Guo, X. Li, J. Wang, “Hybrid Q-learning algorithm about cooperation in MAS,” 2009 Chinese Control and Decision Conference, CCDC 2009, 3943–3947, 2009, doi:10.1109/CCDC.2009.5191990.

[9] Y. Xu, W. Zhang, W. Liu, F. Ferrese, “Multiagent-based reinforcement learning for optimal reactive power dispatch,” IEEE Transactions on Systems, Man and Cybernetics Part C: Applications and Reviews, **42**(6), 1742–1751, 2012, doi:10.1109/TSMCC.2012.2218596.

[10] P. Zhou, H. Shen, “Multi-agent cooperation by reinforcement learning with teammate modeling and reward allotment,” Proceedings - 2011 8th International Conference on Fuzzy Systems and Knowledge Discovery, FSKD 2011, **2**(4), 1316–1319, 2011, doi:10.1109/FSKD.2011.6019729.

[11] L. Matignon, G.J. Laurent, N. Le Fort-Piat, “Hysteretic Q-Learning : An algorithm for decentralized reinforcement learning in cooperative multi-agent teams,” IEEE International Conference on Intelligent Robots and Systems, 64–69, 2007, doi:10.1109/IROS.2007.4399095.

[12] M.L. Li, S. Chen, J. Chen, “Adaptive Learning: A New Decentralized Reinforcement Learning Approach for Cooperative Multiagent Systems,” IEEE Access, **8**, 99404–99421, 2020, doi:10.1109/ACCESS.2020.2997899.

[13] R. Bellman, The Theory of Dynamic Programming, Bulletin of the American Mathematical Society, **60**(6), 503–515, 1954, doi:10.1090/S0002-9904-1954-09848-8.

[14] D.P. Bertsekas, C.C. White, “Dynamic Programming and Stochastic Control,” IEEE Transactions on Systems, Man, and Cybernetics, **7**(10), 758–759, 2008, doi:10.1109/tsmc.1977.4309612.

[15] R. ABellman R., and Kalaba, “Dynamic Programming and Modem Control Theory,” Academic Press, New York; 1965., **4**, 1–23, 2016.

[16] S.E. Dreyfus, “Dynamic Programming and the Calculus of Variations,” Academic Press, New York, 1965.

[17] D.T. Greenwood, "Principles of Dynamics," Prentice-Hall, Englewood Cliffs, N. J., 1991.

[18] G. Mongillo, H. Shteingart, Y. Loewenstein, "The misbehavior of reinforcement learning," Proceedings of the IEEE, **102**(4), 528–541, 2014, doi:10.1109/JPROC.2014.2307022.

[19] A.O. Esogbue, "Fuzzy dynamic programming: Theory and applications to decision and control," Annual Conference of the North American Fuzzy Information Processing Society - NAFIPS, 18–22, 1999, doi:10.1109/nafips.1999.781644.

[20] E.F. Morales, J.H. Zaragoza, "An introduction to reinforcement learning," Decision Theory Models for Applications in Artificial Intelligence: Concepts and Solutions, 63–80, 2011, doi:10.4018/978-1-60960-165-2.ch004.

[21] F. Khenak, "V-learning," 2010 International Conference on Computer Information Systems and Industrial Management Applications, CISIM 2010, **292**, 228–232, 2010, doi:10.1109/CISIM.2010.5643660.

[22] J. Fiala, F.H. Guenther, "Handbook of intelligent control: Neural, fuzzy, and adaptive approaches," Neural Networks, **7**(5), 851–852, 1994, doi:10.1016/0893-6080(94)90107-4.

ASTES

Advances in Science, Technology and Engineering Systems Journal Vol. 7, No. 3, 70-81 (2022)

www.astesj.com

ASTESJ

ISSN: 2415-6698

High Performance SqueezeNext: Real time Deployment on Bluebox 2.0 by NXP

Jayan Kant Duggal*, Mohamed El-Sharkawy

Department of Electrical and Computer Engineering, IoT Collaboratory, Purdue School of Engineering and Technology, IUPUI, Indianapolis, INDIANA, USA

ARTICLE INFO

Article history:
Received: 13 February, 2022
Accepted: 04 May, 2022
Online: 25 May, 2022

Keywords:
Bluebox 2.0
Convolution Neural Networks (CNNs)
Deep Learning
Deep Neural Networks (DNNs)
Modified SqueezeNext
Real-time deployment
SqueezeNext

ABSTRACT

DNN implementation and deployment is quite a challenge within a resource constrained environment on real-time embedded platforms. To attain the goal of DNN tailor made architecture deployment on a real-time embedded platform with limited hardware resources (low computational and memory resources) in comparison to a CPU or GPU based system, High Performance SqueezeNext (HPS) architecture was proposed. We propose and tailor made this architecture to be successfully deployed on Bluexbox 2.0 by NXP and also to be a DNN based on pytorch framework. High Performance SqueezeNext was inspired by SqueezeNet and SqueezeNext along with motivation derived from MobileNet architectures. High Performance SqueezeNext (HPS) achieved a model accuracy of 92.5% with 2.62MB model size at 16 seconds per epoch model using a NVIDIA based GPU system for training. It was trained and tested on various datasets such as CIFAR-10 and CIFAR-100 with no transfer learning. Thereafter, successfully deploying the proposed architecture on Bluebox 2.0, a real-time system developed by NXP with the assistance of RTMaps Remote Studio. The model accuracy results achieved were better than the existing CNN/DNN architectures model accuracies such as alexnet_tf (82% model accuracy), Maxout networks (90.65%), DCNN (89%), modified SqueezeNext (92.25%), Squeezed CNN (79.30%), MobileNet (76.7%) and an enhanced hybrid MobileNet (89.9%) with better model size. It was developed, modified and improved with the help of different optimizer implementations, hyper parameter tuning, tweaking, using no transfer learning approach and using in-place activation functions while maintaining decent accuracy.

1. Introduction

The dream of achieving a true human experience lies within the domain of cybernetics, machine learning, deep learning and AI. AI is currently responsible for transcending hard coded application based programmed machines to artificially intelligent machines with some situational awareness.

All the existing CNN or DNN models trained and tested on large datasets occupy extensive computational and memory resources. In the last couple of years, with the introduction of new CNN or DNN based macro architectures such as ViT [1], CaiT, BiT [2], EfficientNetv2 [3], LaNet [4,5], GPipe [6], enhanced MobileNets [7], SqueezeNet [8], SqueezeNext [9], etc., deep learning became better and more efficient in terms of CNN/DNN model performance than the traditional ones [10,11]. The model efficiency, model performance and its ability to be deployed on limited resource constraint [12] real-time platform was attenuated majorly due to following factors such as design space exploration (DSE) of DNNs [13], hyper parameter tuning and tweaking, different optimizers [14], and activation functions implementation, regularization methods, and powerful hardware accelerators. These existing architectures were never tailor made for deployment on real-time embedded systems with limited resources. This research also makes an effort to develop a new architecture with an impressive model size under 5 MB while maintaining an impressive model accuracy.

In this research, a new architecture called High Performance SqueezeNext [15] was developed in order to attenuate the succeeding various deployment problems of DNN based architectures such as DNN deployment on resource constrained real-time platforms [16,17], DNN model compression, over fitting,

*Corresponding Author: Lu Xiong, Email: lu.xiong@mtsu.edu

https://dx.doi.org/10.25046/aj070308

maintaining a competitive model accuracy without major compromises and developing a hardware aware DNN architecture. This DNN based architecture was inspired and derived from the valuable insights of SqueezeNext and SqueezeNet architectures. The fundamental blocks of High Performance SqueezeNext architecture were derived from the fire modules of SqueezeNet [8,18] and bottleneck modules of SqueezeNext architectures [9].

This architecture tries to attenuate the SqueezeNext architecture problems namely model compression, deploy ability on a real-time embedded platform, efficiently working with resource constraint embedded platforms rather than GPUs or CPUs, incorporating newly developed leaning rate techniques such as cosine annealing, step-based decay, cyclic, and cosine annealing warm restarts. Baseline SqueezeNext architecture was a Caffe based architecture and was not able to utilize the power Pytorch based libraries and functions such as the above-mentioned learning rate decay functions, newly introduced optimizers such as Adaboost, Adabound, in-place activation functions and also some new activation functions in contrast to High Performance SqueezeNext architecture [15] that was developed entirely on a Pytorch based framework.

After strenuous training and testing of High Performance SqueezeNext architecture on multiple datasets such as CIFAR-10 [19] and CIFAR-100 [19], implementing several optimizers, activation functions, incorporating and replacing regular operations with in-place operations, reducing stride in subsequent layers, using preliminary data augmentation and some model compression, it achieved impressive model performance. Also, one of several important factors contributing to the success of this architecture was training and testing High Performance SqueezeNext [15] without any form of transfer learning along with some model compression. Finally, High Performance SqueezeNext was deployed on a real time embedded system, Bluebox 2.0 [16] with the assistance of RTMaps software platform.

This research was focused to deploy the proposed High Performance SqueezeNext [15] comprehensively on real-time embedded platform, Bluebox 2.0 [16] by NXP, explore the major hyperparameter tuning with no transfer learning [13,18], develop a Pytorch framework DNN in order to be deployed on Bluebox 2.0 and compare the proposed architecture with several other pytorch based CNN/DNN based architecture.

2. Literature Review

Deep learning transformed the artificial and machine learning domain with the introduction of deep convolutional neural networks. CNNs/DNNs are tweaked and tuned with the hyper parameters, newly introduced large datasets, powerful hardware, model compression, and data augmentation [20] to attain better results. Also, batch normalization [21] is observed to be a major contributor for improving DNN performance. Other prominent factors include use of skip connections [9], data preprocessing techniques, regularization, and number of pooling layers. CNNs/DNNs are used to develop image classifiers [10,22,23], object detectors, object recognizers and object segmentation. In order to solve the problem of real time embedded system DNN deployment with limited resources, a requirement for CNN/DNN architectures is introduced. Recently introduced macro architectures such as SqueezeNet, SqueezeNext, and Shallow SqueezeNext [17] were successfully able to overcome computation and memory resources limitations of CNNs and DNNs.

The research in this paper is heavily oriented on the following discussed architectures in detail: SqueezeNet [8] and SqueezeNext [9] architectures. Some motivation is also derived from MobileNet architecture [24]-[26]. Other techniques used are hyperparameter tuning and tweaking, implementation of different optimizers [14], activation functions and regularization, data augmentation and some data compression (width and depth wise compression) that was introduced within SqueezeNext architecture [9]. High Performance SqueezeNext [15] architecture is majorly based on SqueezeNext followed by SqueezeNet architecture. The fundamental blocks and block structures of High Performance SqueezeNext of the proposed architecture were inspired by fire modules (figure 1), basic blocks (figure 6) of various architectures, and bottleneck modules [9], respectively. The proposed architecture, HPS, was modified and improved with the help of no transfer learning approach, use of in-place functions that helped to reduce mathematical operations being performed with each layer and blocks of the architecture.

2.1. SqueezeNet

The first baseline CNN architecture, baseline SqueezeNet [8] was utilized in this research for inspiration of High Performance SqueezeNext architecture [15]. SqueezeNet baseline architecture is made up of 1x1 and 3x3 convolutions, fire modules, max pooling layers, Relu and Relu in place activation layers, softmax activation, and kaiming uniform initialization.

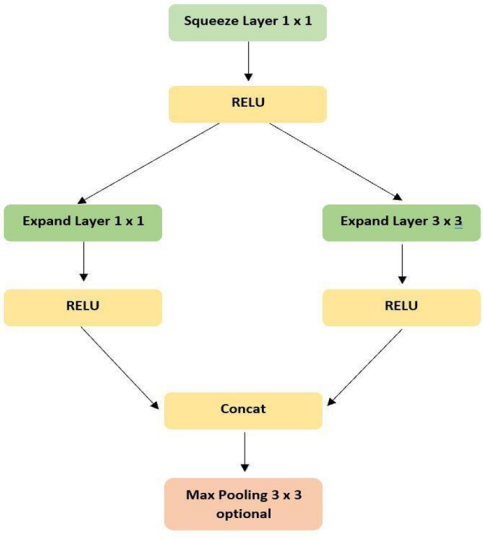


Figure 1. Fire module

Fire module [8] (figure 1.) is the main component of this architecture that consists of one squeeze layer, s2 (1x1), two expand layers, e1 (1x1) and e3 (3x3). This architecture consists of following highlighted factors such as 3x3 convolutions replaced with 1x1 convolutions, number of input channels decreased to 3x3 convolutions, and down sampling max pooling late down the CNN.

Fire module (figure 1.) is the main component of this architecture that consists of one squeeze layer, s2 (1x1), two expand layers, e1 (1x1) and e3 (3x3). This architecture consists of following highlighted factors such as 3x3 convolutions replaced with 1x1 convolutions, number of input channels decreased to 3x3 convolutions, and down sampling max pooling late down the CNN.

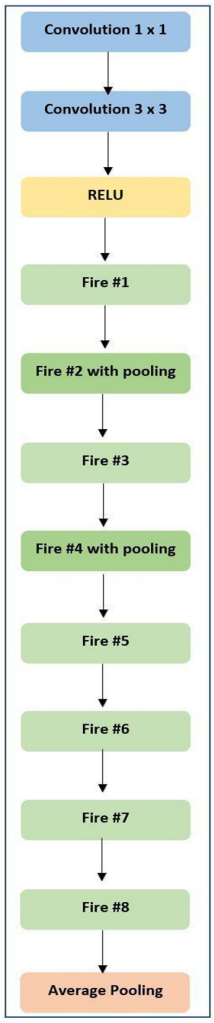


Figure 2. SqueezeNet baseline architecture

SqueezeNet provides some valuable insights relevant to fire modules in accordance with design significance of the fundamental building blocks of CNNs/DNNs and its effect on the entire architecture itself affecting a DNN architecture performance. Squeezed CNN [18] was compared in this research with the proposed High Performance SqueezeNext [15] architecture. Squeezed CNN is an architecturally compressed version of SqueezeNet based CNN architecture that was previously successfully deployed on the Bluebox 2.0 real time embedded platform.

2.2. SqueezeNext

SqueezeNext baseline architecture [9], another architecture used for major development and laying the foundation of High Performance SqueezeNext architecture. It is also used for comparison with the proposed architecture called High Performance SqueezeNext. Key factors of the SqueezeNext baseline architecture:

- Aggressive channel reduction by a two-stage squeeze module.
- Use of separable 3x3 convolutions.
- Use of an element-wise addition skip connection.

In SqueezeNext architecture, two stage squeeze model channel reduction, 3x3 separable convolution, and an element-wise skip connection [9] techniques are used to drastically reduce the total number of parameters and computation resource usage. Baseline SqueezeNext architecture comprises of bottleneck modules [9] with four stage implementation (figure 3.), batch normalization layers [21], Relu and Relu (in-place) nonlinear activation layers, max pooling and average pooling layers, Xavier uniform initialization, a spatial resolution layer and a fully connected layer. All these techniques are also utilized in High Performance SqueezeNext architecture.

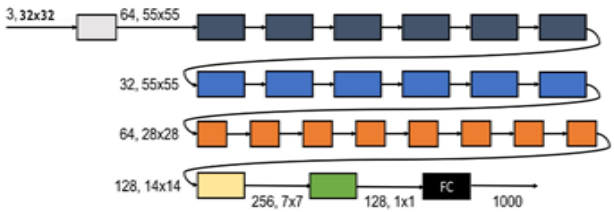


Figure 3. SqueezeNext baseline architecture basic block with [6,6,8,1] four stage implementation configurations trained and tested on CIFAR-10 dataset with no transfer learning [15].

Bottleneck modules [9] are responsible for huge parameters reduction. The consecutive different colored blocks: dark blue, blue, orange, and yellow blocks after the first convolution represent the four-stage configuration implementation referring to a low level (dark blue), medium level (blue and orange), and high-level features (yellow), respectively. Green block represents spatial resolution layer. The baseline SqueezeNext architecture achieves 112x fewer parameters than the AlexNet top-5 performance and 31x fewer parameters than VGG-19 performance.

2.2.1. Modified SqueezeNext

Modified baseline SqueezeNext is a modified form of the above baseline SqueezeNext architecture. It is derived and modified in order to make a fair comparison with the proposed architecture, High Performance SqueezeNext [13,15] based on a pytorch framework instead of caffe framework. Modified SqueezeNext architecture is trained and tested on CIFAR-10 [19] from scratch and is developed to be implemented with the Pytorch framework.

The fundamental block of baseline SqueezeNext architecture is made up of a convolution layer followed by batch normalization in place, scale in place and ReLU in place layers. In contrast to a fundamental building block of modified SqueezeNext architecture that consists of a convolution layer, batch normalization and a ReLU layer.

The basic blocks depict each of the first individual blocks of the four-stage configuration within the SqueezeNext architectures (the first dark blue, blue, orange and the last

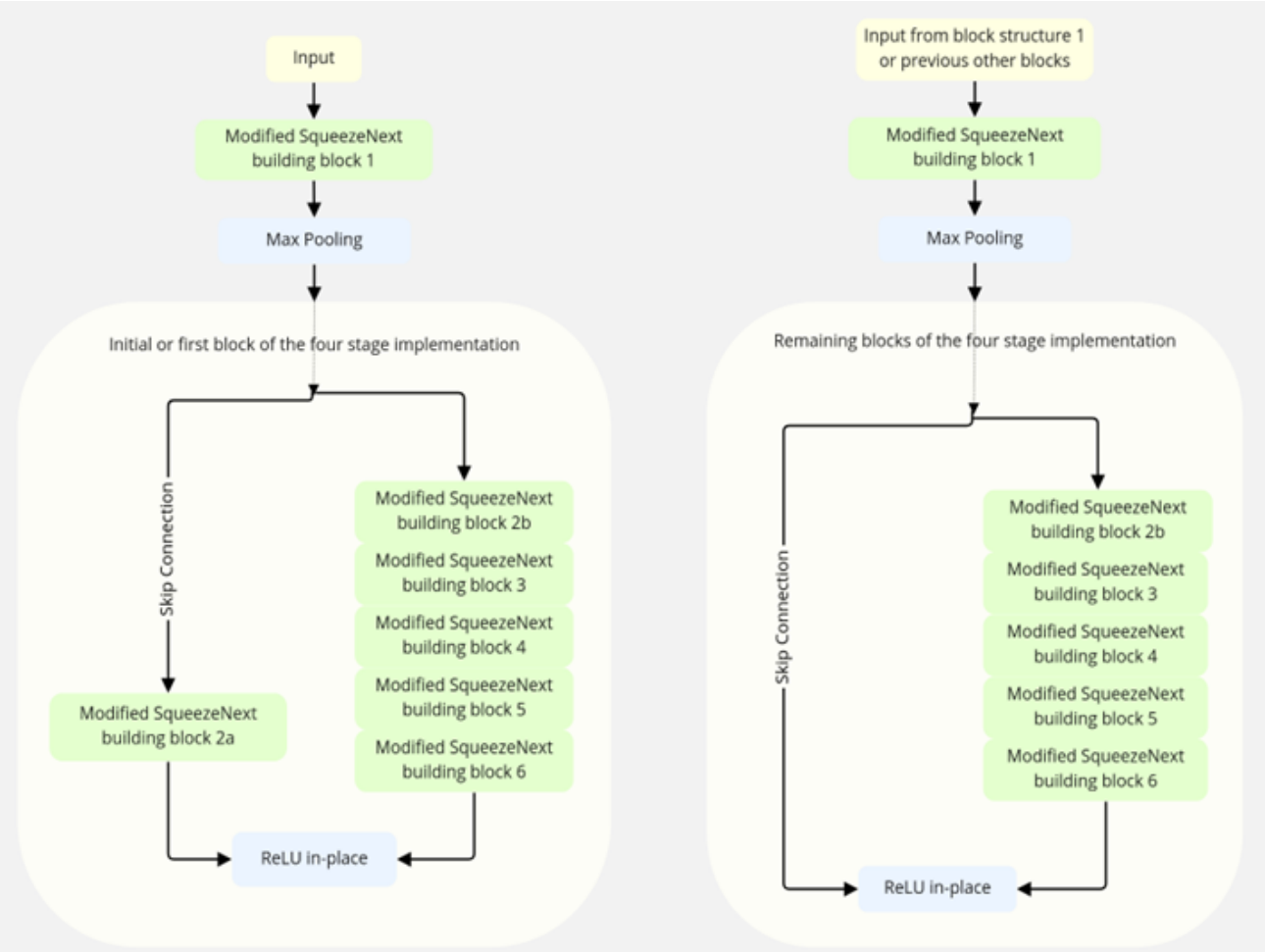


Figure 4: Modified SqueezeNext baseline architecture block structure 1 & block structure 2.

yellow block of the four-stage configuration). All the five SqueezeNext Block modules inside both of the block structures (Figure 4.) are exactly the same fundamental building blocks (convolution layer, batch norm and an activation function layer) described in the previous paragraph.

The efficient organization of these two block structures within a DNN architecture makes a better and efficient DNN. This modified SqueezeNext architecture is trained and tested with datasets such as CIFAR-10 [19] and CIFAR-100 [19]. It was also modified with the help of data augmentation, data compression and different optimizer functions [14] are implemented in order to improve the performance of the modified SqueezeNext architecture [13].

3. Hardware & Software Used

- Intel i9 8th generation processor with 32 GB RAM.
- Required memory for dataset and results: 4GB.
- NVIDIA RTX 2080Ti GPU.
- Spyder version 3.7.1.
- Pytorch version 1.1.
- RTMaps Remote Studio
- Linux BSP
- SD card 8GB minimum
- RTMaps Embedded

4. Bluebox 2.0 real-time embedded platform

Bluebox 2.0 [16] is a real time development platform by NXP for developing self-driving car applications. It delivers the performance required to analyze ADAS systems or environments. ASIL-B and ASIL-D compliant real time embedded hardware.

Figure 5. Real time embedded platform, Bluebox 2.0 by NXP

It includes three independent SoCs that are S32V234: a vision processor, LS2084A: a compute processor, and S32R274: a radar microcontroller. This research paper was also a way to analyze and test the capability of Bluebox 2.0 as an autonomous embedded platform system for real-time autonomous applications. It can be used or utilized for implementing Level 1- Level 3 autonomous applications. The detailed description of all SoCs is discussed in the following subsections.

4.1. Vision Processor (S32V234)

S32V234 [24] is a micro processing unit consisting of an ISP, powerful 3D GPU, automotive-grade reliability, dual APEX-2 vision accelerators, and functional safety. It provides good

computation support for ADAS, NCAP front camera, object detection and recognition, image processing, machine and deep learning, and sensor fusion applications. The 32-bit based Arm Cortex-A53 S32V processors are supported by S32 Design Studio IDE for development. The software platform studio includes a compiler, debugger, linux BSP vision SDK, and graph tools.

It is a vision-based processor that comprises ISP available on all MIPI-CSI camera inputs. It supports and provides the functionality to integrate multiple cameras. APEX-2 vision accelerators, GPU along with vision accelerators, four ARM Cortex-A53 cores, and an arm M4 core are used for embedded related applications and computer vision functions. It operates on linux BSP, Ubuntu 16.04 LTS and NXP vision SDK. The processor boots up from the SD card with the help of linux BSP.

4.2. LS-2084A

LS2 processor [13,16] embedded in bluebox 2.0 is a high-performance computing processor platform. It comprises ARM Cortex-A72 cores, 10 Gb Ethernet ports, DDR4 memory, and a PCIe expansion slot. It is also a convenient platform to develop the arm-based application or features.

It makes use of *an SD card* interface enabling *its* processor to run linux BSP, Ubuntu 16.04 LTS on the platform. The software enablement on the LS2084A and S32V234 SoC is done with the help of *Linux* BSP. *It* is a complete, developer-supported system *with* eight core QorIQ LS2084A and the four core LS2044A. This *multi-core* processors-based system offers advanced, high-performance data path and network peripheral interfaces required for networking, datacom, wireless infrastructure, military, and aerospace applications.

4.3. Real-time Multisensor applications (RTMaps)

RTMaps [13,17] is an efficient and easy-to-use framework for fast and robust developments. It is a high-performance platform that is the easiest way to develop, test, validate, benchmark, and execute applications. It is used for fusing the data streams in real-time. It consists of several independent modules that can be used in different scenarios. It is described as follows:

- **RTMaps Runtime Engine** is an easily deployable, multi-threaded, highly optimized module that is designed to be integrated with third-party applications. It is also accountable for all base services such as component registration, buffer management, time stamping threading, and priorities.
- **RTMaps Component Library** consists of the software module that can be easily interfaced with the automotive and other related sensors and packages responsible for the development of an ADAS application.
- **RTMaps Remote Studio** is a graphical modeling environment with the functionality of programming using Python packages. It is available for both, windows and ubuntu based operating system platforms. Applications are developed by using the modules and packages available from the RTMaps component library.
- **RTMaps Embedded** is a framework consisting of a component library and the runtime engine with the capability of running on an embedded x86 or ARM capable platform.

The connection between the computer running RTMaps Remote Studio and Bluebox 2.0 platform can be accessed via a static TCP/IP. The detailed approach of High Performance SqueezeNext deployment is explained in the following section 6.

5. High Performance SqueezeNext

High Performance SqueezeNext architecture [15] is a compact DNN, heavily inspired from architectures such as baseline SqueezeNet and baseline SqueezeNext with some insights taken from MobileNet architecture [24]. The basic block shown in figure 6 (extreme right) consists of a convolution layer, Relu in place layer, and batch normalization layer forms the building blocks of the High Performance SqueezeNext. This basic block in figure 6 are the blocks in figure 7 that form a bottleneck module.

This bottleneck module forms the blocks, arranged together in a four-stage implementation configuration along with a dropout layer. It is concluded with the descriptions of the two-model shrinking hyper parameters such as the width multiplier and resolution multiplier which are explained below towards the end of this section.

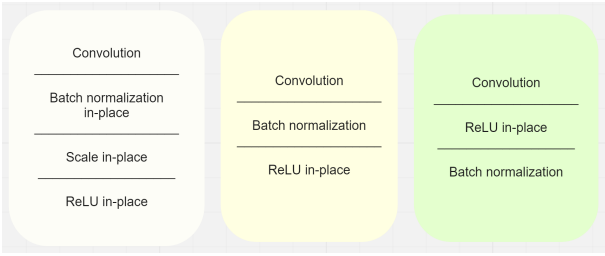


Figure 6: Comparison of baseline SqueezeNext block, Modified SqueezeNext block, High Performance SqueezeNext block, in left to right sequence.

High Performance SqueezeNext is based on the following strategies:

- Using resolution and width multipliers.
- Using only in place operations sandwiching ReLU in-place between convolutional and batch normalization layers.
- An element-wise addition skip connection.
- Adding a drop out after the average pooling layer.
- Minimizing use of any pooling layers.

In detail, the proposed High Performance SqueezeNext architecture comprises bottleneck modules (basic blocks arranged in four stage configuration), a spatial resolution layer, average pooling layer and a fully connected layer. In order to fine tune and tweak the hyperparameters [13] of the proposed architecture different optimizers are implemented.

Modified bottleneck module, shown in the figure 7, comprises a 1x1 convolution, second 1x1 convolution, 3x1 convolution, 1x3 convolution, and then a 1x1 convolution for the proposed High Performance SqueezeNext architecture forms the High Performance SqueezeNext block module. In figure 7, the blocks in figure refer to adaptive forms of High Performance SqueezeNext basic building blocks (right most block in figure 6) that depict the first basic block with 1x1 convolution is High Performance SqueezeNext (HPS) basic building block with 1x1 convolution as a first layer, then, another 1x1 convolution based second HPS basic building block, a third similar block but with first 1x3 convolution layer instead of 1x1 convolution layer. Followed by a similar 3x1 convolution HPS basic building

block and in the end again a 1x1 convolution based HPS building block.

Figure 8 illustrates High Performance SqueezeNext architecture with block structure-1 (top left), four stage implementation and block structure-2 (bottom left). This block structure-1 is responsible for feature extraction from the initial input blocks with the different colors of the four stage implementation of this proposed architecture. Figure 8 illustrates block structure-2 (bottom left) of the proposed High Performance SqueezeNext architecture. These block structures form the rest of the colored feature blocks of the four-stage implementation of the High Performance SqueezeNext.

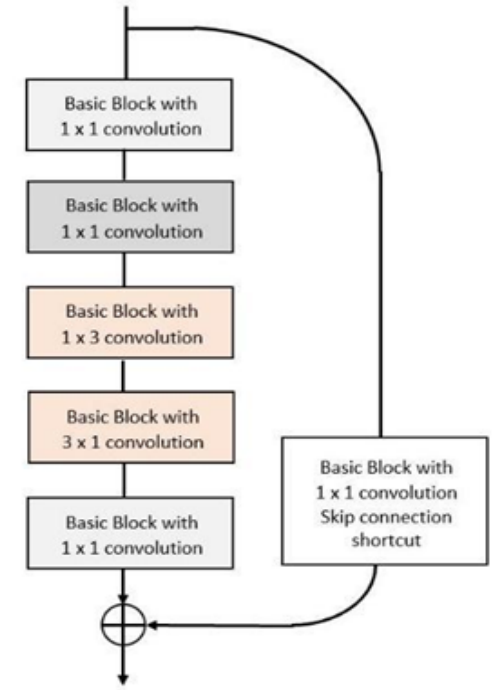


Figure 7. Modified bottleneck module for High Performance SqueezeNext architecture.

5.1. Resolution Multiplier

This hyper-parameter, resolution multiplier [9,13], is used to reduce the computational cost of deep neural networks (DNN). It reduces the computational cost and number of parameters.

5.2. Width Multiplier

Width multiplier [9,13] is used to construct compact and less computationally expensive models. It is used to thin DNN at each layer, further reducing the number of parameters to roughly twice the power of the width multiplier term.

5.3. Architecture optimization

Other few factors which contributed towards the improvement of the proposed High Performance SqueezeNext architecture [15] are:

- *Rectified Linear Units (ReLU) in place and ELU in place function operations:* In place operations activation functions help to reduce the number of parameters by performing in place element wise operations. It changes the content of a given linear operator without actually making a copy of it.
- *Different optimizers implementations for training:* In this research, different optimizers were implemented in order to find a tuned and tweaked proposed architecture. The different optimizers implemented here are SGD, SGD with nesterov and momentum, Adabound, Adagrad, Adamax, Adam, Rprop, and RMSprop.
- *Dataset specific training with no transfer learning:* According to this approach, we do not use any transfer model using Python pickle module bound to the specific classes and the exact directory structure used when the model is saved. It saves a path file containing the class.
- Use of adaptive average pooling would help set stride and kernel-size, especially in adaptive pooling the stride and kernel size are set automatically.
- Real-time embedded platform deployment, Bluebox 2.0 [24, 27].

Specifically, we train data on a powerful GPU based system generating a checkpoint file for testing. This testing file is further used within Bluebox 2.0 for deployment of the DNN architecture.

6. Bluebox 2.0 architecture deployment

The proposed architecture High Performance SqueezeNext [15] is initially, trained on system with RTX 2080ti equipped gpu. After successfully testing on an average of three times for a model, generate a checkpoint file using save and load checkpoint method for pytorch. For testing, the python-based module is connected to a real-time embedded system, Bluebox 2.0 with the help of RTMaps Remote studio connector.

RTMaps Remote studio and RTMaps Embedded provide support for the pytorch framework that currently is empowered by a huge collection of libraries for machine learning and deep learning support. Figure 9 represents the process of High Performance SqueezeNext architecture deployment on the real-time embedded system platform, Bluebox 2.0 by NXP. RTMaps studio initiates a connection to the execution engine using TCP/IP that runs the software on Linux BSP and then, installed on Bluebox 2.0. RTMaps provides a python block to create and deploy python Pytorch code.

Python code for RTMaps comprises three function definitions: birth (), core (), and death (). Due to organized structure, and flexibility within RTMaps, it makes it easy to develop a modular code. LS2084A processor is used for the maximum utilization of available 8 ARM Cortex-A72 cores to run RTMaps. The deployment process for High Performance SqueezeNext on the real time platform, Bluebox 2.0 by NXP.

Birth (): It is executed once, initially, for setting up and initializing the python environment.

Core (): It is an infinite loop function to keep the code running continuously.

Death (): It is used to perform cleanups and memory release after the python code terminates.

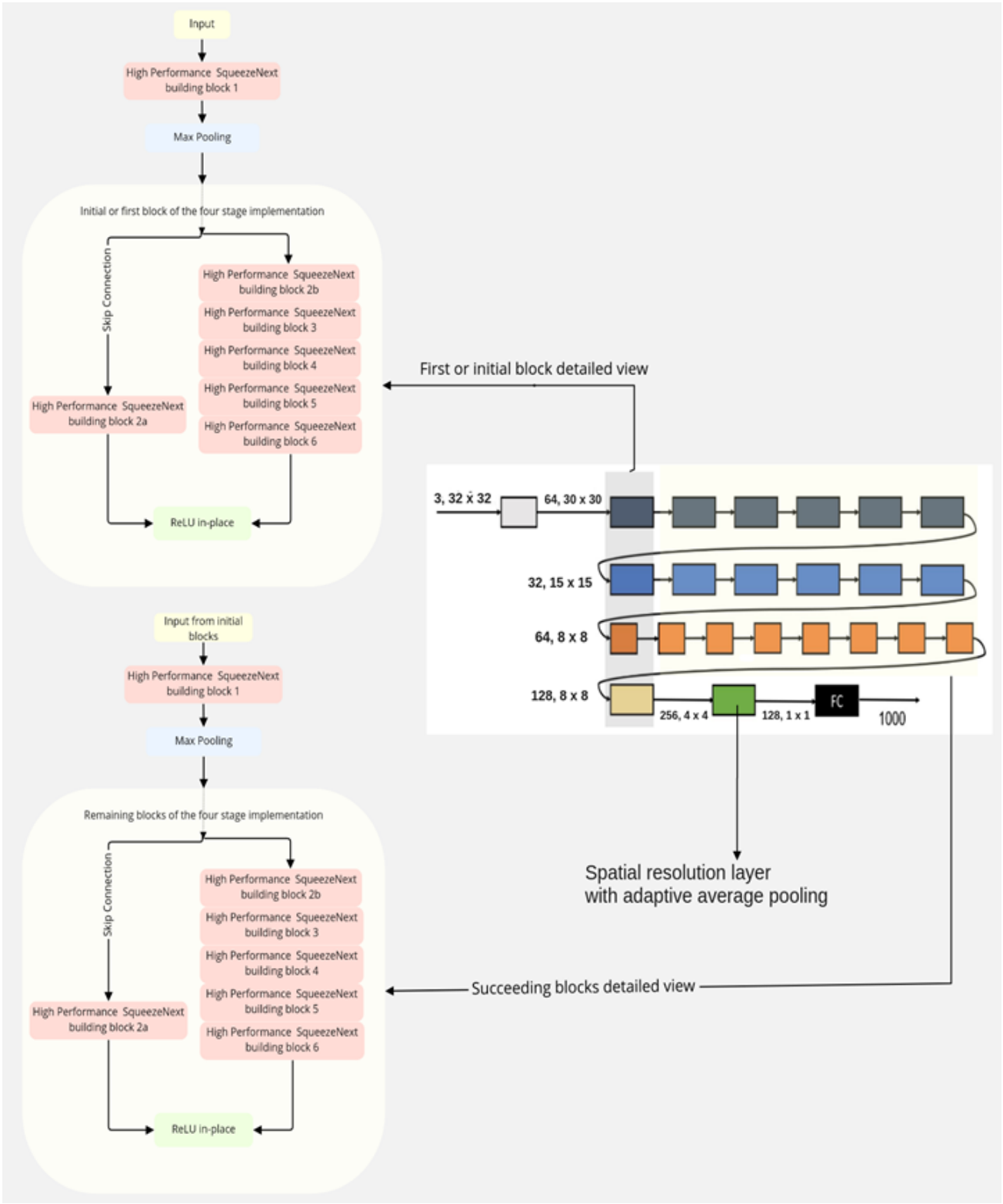


Figure 8. High Performance SqueezeNext block structure 1, High Performance SqueezeNext four stage implementation and HPS block structure 2.

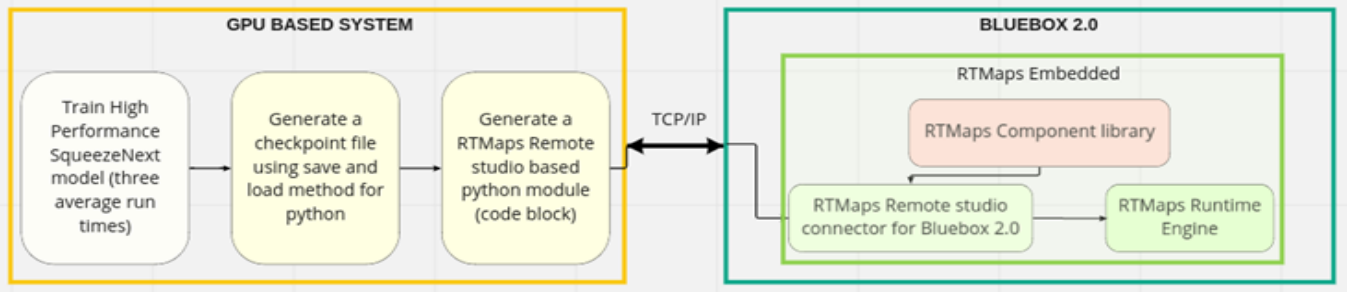


Figure 9: Bluebox 2.0 deployment process for High Performance SqueezeNext architecture overview.

7. Results

7.1. High Performance SqueezeNext Results

High Performance SqueezeNext architecture is implemented with the development and optimization approaches mentioned in the section 5. This research led to different variants of High Performance SqueezeNext models. Also, it leads to another compressed and shallow depth based DNN architecture called Shallow SqueezeNext [17,25,26].

The High Performance SqueezeNext model size ranges from 10.8MB to a small size of 370KB as shown in Table1 and Table2 with model accuracy between 70% to 93% and model speed of approximately under 25 seconds per epoch for the experimental models. In the following tables, only a few of the several better model's results were shown out of the total 500 models or experiments.

The nomenclature for the proposed High Performance SqueezeNext model in all the tables in section 7 illustrates the proposed High Performance SqueezeNext architecture version name followed by width and resolution multiplier, and version number.

7.1.1 Model Accuracy Improvement

We can infer from Table 1. that a better High Performance SqueezeNext model accuracy is achieved that is 92.50% with a model size of 2.6 MB and 18 seconds per epoch.

High Performance SqueezeNext -21-1x-v3 model has 12.91% and 14.95% better accuracy along with little decrease of 0.399MB and 0.347MB model size with respect to baseline SqueezeNet- v1.0 and baseline SqueezeNet-v1.1, respectively. High Performance SqueezeNext -21-1x-v3 model has 3.35% and 2.02% better accuracy with respect to baseline SqueezeNext-1x-v1 and baseline SqueezeNext-2x-v1, respectively. All the results obtained in this paper were implemented with the following common hyper parameter values: LR: 0.1, batch size: 128, weight decay: 5e-4, total number of epochs: 200, standard cross entropy loss function and livelossplot package.

Other existing algorithms and methodologies have attained better accuracy than the baseline SqueezeNet and SqueezeNext architecture. However, all the other machine learning or deep learning algorithms use transfer learning techniques, in that the respective model is first trained on a large dataset maybe such as ImageNet and then a pre-trained model is fine-tuned on a smaller datasets like CIFAR-10 [19], CIFAR-100 [19]. Also, these architectures are deeper [10,11,22] and expensive in terms of computation and memory resources [12,13]. Table 1, figure 10 (a), (b) and (c) compares the DNN results for baseline SqueezeNet, baseline SqueezeNext, modified SqueezeNext and High Performance SqueezeNext.

Table 1: High Performance SqueezeNext accuracy improvement

Model type	Model Acc.%	Model size (MB)	Model speed (sec)
Baseline SqueezeNet-v1.0	79.59	3.013	04
Baseline SqueezeNet-v1.1	77.55	2.961	04
Baseline SqueezeNext-23-1x-v1	87.15	2.586	19
Baseline SqueezeNext-23-2x-v1	90.48	9.525	22
High Performance SqueezeNext -21-1x-v3	92.50	2.614	18
High Performance SqueezeNext -23-1x	92.25	5.14	29
High Performance SqueezeNext -21-1x-v2.0	92.05	2.60	16
High Performance SqueezeNext -06-0.5x-v1	82.44	0.37	07
High Performance SqueezeNext -06-1x	86.82	1.24	08

+*Model Acc. – Model Accuracy*

These graphs depict that the overfitting problem is reduced and becomes less problematic from baseline SqueezeNet architecture [8] to baseline SqueezeNext architecture [9]. Then, again, the overfitting problem got reduced and became better further from modified SqueezeNext architecture in comparison to proposed High Performance SqueezeNext. Due to this reason, we can infer that High Performance SqueezeNext [15] is amongst the better DNN architecture by overcoming the overfitting problem of DNNs along with better resource usage and a competitive accuracy of 92.50%.

7.1.2. Model Size and Model Speed Improvement

The model speed in this paper refers to the time taken per epoch to train and further test the DNN architecture. High Performance SqueezeNext architecture is initially trained with the help of powerful GPUs (GTX 1080 and RTX 2080 Ti).

In general, more powerful hardware (better GPU or multiple GPUs), architecture pruning, and other methods can be implemented to improve the performance of a DNN. To train and test the High Performance SqueezeNext DNN in a better manner, we train and test the proposed DNN on CIFAR-10, CIFAR-100 datasets [13] that are quite small as compared to ImageNet dataset. For better performance of proposed High Performance SqueezeNext architecture model depth as well as model width are modified.

High Performance SqueezeNext architecture is implemented with resolution multiplier and width multiplier, in-place operations [13], and no max-pooling layers used after the four stage implementation but only one adaptive average pooling layer just before the fully connected convolutional layer. This architecture is tuned using hyper parameters such as SGD optimizer with momentum and nesterov values, a step learning rate decay schedule with an exponential learning rate update. We observe in table 2, when High Performance SqueezeNext is trained and tested on CIFAR-10 dataset with no transfer learning approach with a nesterov based SGD optimizer. HPS attains a model accuracy of 92.50% with a decent 2.614MB model size. The results published in table 2 were derived with the help of a gpu based system. Each result entry in table 2 is an average run of three training and validation cycles here.

Table 2: High Performance SqueezeNext Model Speed and Model Size Improvement for CIFAR-10

Model type	Mod Acc.%	Mod. Size (MB)	Mod. speed (sec)	Resol.	Width
Baseline SqueezeNet-v1.0	79.59	3.013	04	-	-
Baseline SqueezeNet-v1.1	77.55	2.961	04	-	-
Baseline SqueezeNext-23-1x-v1	87.15	2.586	19	6 6 8 1	1.0
Baseline SqueezeNext-23-2x-v1	90.48	9.525	22	6 6 8 1	2.0

High Performance SqueezeNext -06-0.5x	82.44	0.37	07	1 1 1 1	0.5
High Performance SqueezeNext -06-1x	86.82	1.24	08	1 1 1 1	1.0
High Performance SqueezeNext -06-0.75x	82.86	1.20	08	1 1 1 1	0.75
High Performance SqueezeNext -08-0.5x	87.37	1.41	09	1 2 2 1	0.5
High Performance SqueezeNext -14-1.5x	89.86	3.24	18	1 2 8 1	1.5
High Performance SqueezeNext -21-1x-v2	92.05	2.60	16	2 2 14 1	1.0
High Performance SqueezeNext -21-1x-v3	92.50	2.614	18	2 4 12 1	1.0
High Performance SqueezeNext -23-1x	92.25	5.14	29	2 2 16 1	1.0

**Mod. Acc.: Model Accuracy; Mod. Size: Model Size; Mod. Speed: Model Speed, Resol: Resolution multiplier, and Width: Width multiplier.*

Table 3: High Performance SqueezeNext Model Speed and Model Size Improvement for CIFAR-100

Model type	**Mod Acc. %**	**Mod. Size (MB)**	**Mod. speed (sec)**	**Resol.**	**Width**
Baseline SqueezeNet-v1.0	51.27	6.40	04	-	-
Baseline SqueezeNext-23-1x-v1	60.37	5.26	19	6 6 8 1	1.0
High Performance SqueezeNext -06-0.4x	55.89	0.95	06	1 1 1 1	0.4
High Performance SqueezeNext -06-0.575x	60.68	0.95	07	1 1 1 1	0.575
High Performance SqueezeNext -09-0.5x	62.72	1.10	08	1 1 4 1	0.5
High Performance SqueezeNext -14-1x	68.46	5.00	14	1 2 8 1	1.0
High Performance SqueezeNext -14-1.5x	69.70	10.80	18	1 2 8 1	1.5
High Performance SqueezeNext -23-1x	68.20	7.70	25	2 2 16 1	1.0
High Performance SqueezeNext -25-1x	70.10	7.80	25	2 4 16 1	1.0

**Mod.Acc.: Accuracy; Mod. Size: Model Size; Mod. Speed: Model Speed, Resol: Resolution multiplier, and Width: Width multiplier.*

In this paper validation is used interchangeably between testing and validation because CIFAR-10 and CIFAR-100 datasets are in comparison small datasets to imagenet. Testing along with training and validation cycles provides inconclusive and non-reliable results.

Therefore, the whole research in this paper only performs training and testing (referred as validation in case of graphs in Figure 10). It also reflects the significance of resolution and width multiplier used in the proposed architecture.

Table 3. refers to the set of results obtained in the case of training and testing of High Performance SqueezeNext on CIFAR-100 dataset with no transfer learning. Again, each result entry in table 3 is an average run of three training and validation cycles here. It also reflects the importance of width and resolution multipliers and its effect on the proposed High Performance SqueezeNext architecture.

We can also observe the model accuracy significantly drops by 20% on average regardless of any architecture deployment, therefore, verifying that the small datasets do perform poorly with a CNN/DNN based architecture with no transfer learning.

7.2. Bluebox 2.0 Results

High Performance SqueezeNext architecture is deployed on the BlueBox 2.0 platform by NXP. For real-time DNN deployment on Bluebox 2.0 [13,16], DNN parameters are saved and loaded using Pytorch method from a checkpoint file generated during training of DNN with GPU. Then, this saved checkpoint file is loaded with the help of RTMaps on the Bluebox 2.0 using the BSP Linux OS dependencies.

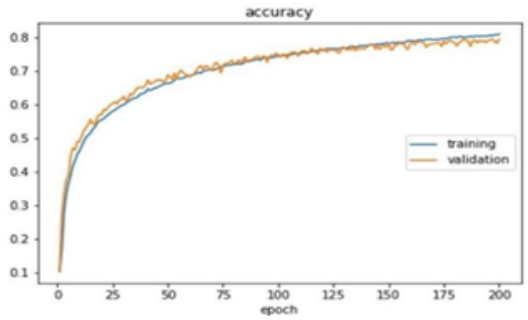


Figure 10: (a) SqueezeNet baseline architecture training and validation (testing) graph representation implemented on CIFAR-10 dataset.

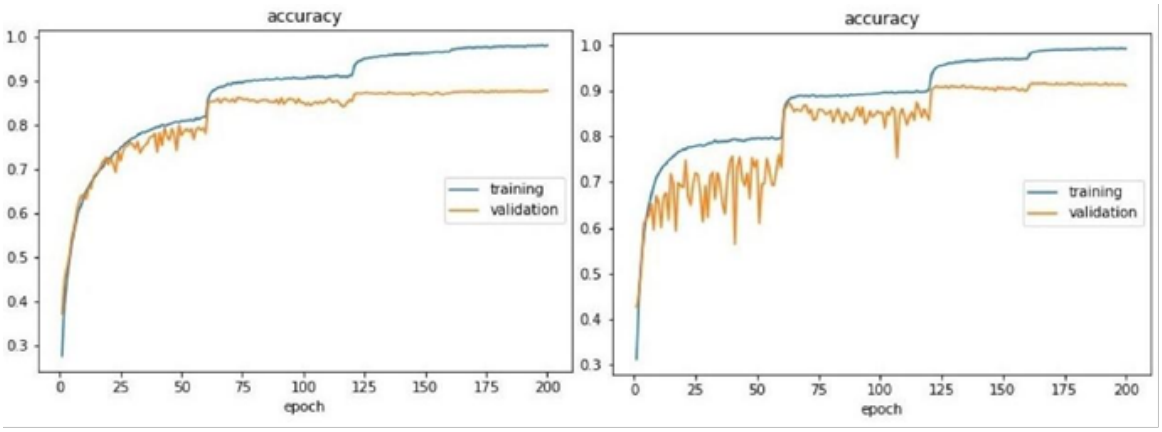


Figure 10: (b) Graph comparison between SqueezeNext baseline architecture and High Performance SqueezeNext architecture training and validation implemented on CIFAR-10 dataset.

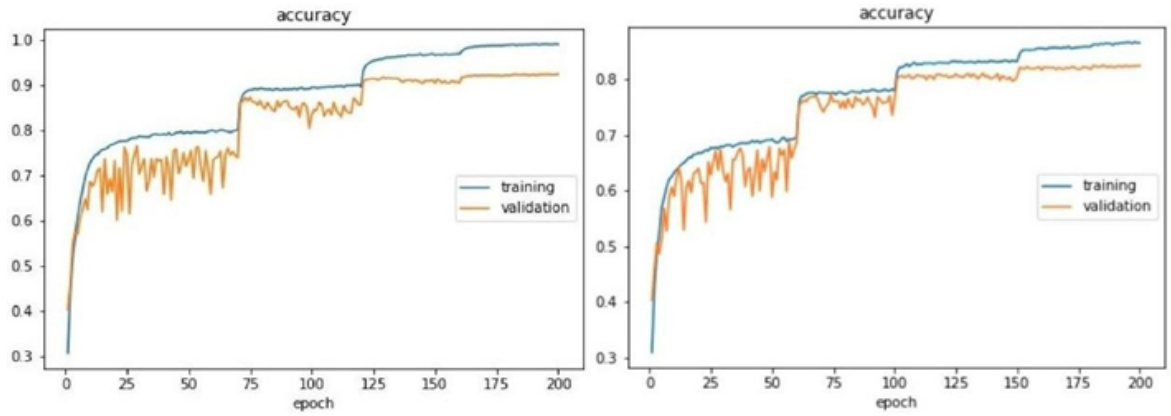


Figure 10: (c) Graph comparison between Modified SqueezeNext baseline architecture and High Performance SqueezeNext architecture training and validation implemented on CIFAR-10 dataset.

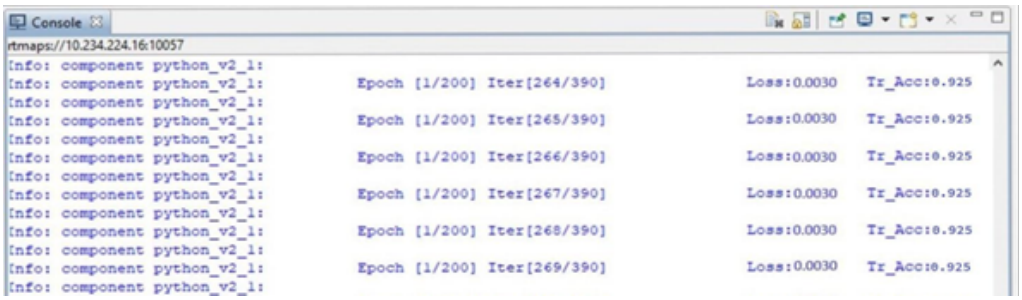


Figure 11: High Performance SqueezeNext architecture deployment on Bluebox 2.0 by NXP with the help of RTMaps Remote Studio.

Table 4: High Performance SqueezeNext deployment results on Bluebox 2.0

Model type	Mod. Acc. %	Mod. Size (MB)	Mod. Speed (sec)
Squeezed CNN (Baseline SqueezeNet based CNN)	79.30	12.90	11
Modified SqueezeNext architecture	92.25	5.14	28
High Performance SqueezeNext-21-1x-v3	92.50	2.62	16
High Performance SqueezeNext-06-1.0x	86.82	1.24	08
High Performance SqueezeNext-06-0.50x	82.44	0.37	06

**Mod. Acc.: Model Accuracy; Mod. Size: Model Size; Mod. Speed: Model Speed.*

Table 4 illustrates the results obtained for the proposed High Performance SqueezeNext architecture, baseline SqueezeNext and squeezed CNN architecture [18] representing that the proposed High Performance SqueezeNext performs better and more efficiently. Table 4 illustrates that High Performance SqueezeNext-21-1x-v3 has 92.50% accuracy that is 13.20% better accuracy than Squeezed CNN (SqueezeNet Implementation for Pytorch framework) and 0.25% better accuracy than modified SqueezeNext-23-1x. Further, High Performance SqueezeNext-21-1x is 5x better than Squeezed CNN (SqueezeNet Implementation) and 2x better than modified SqueezeNet.

Also, High Performance SqueezeNext-06-0.5x model attained 0.37MB with 6 seconds per epoch, minimum model size and model speed in comparison to squeezed CNN [18] and modified SqueezeNext models deployed on the Bluebox 2.0 by NXP. All models are trained initially on GPU and then tested on real time platform Bluebox2.0 by NXP with datasets such as CIFAR-10 and CIFAR-100, individually.

Table 5: High Performance SqueezeNext Architecture composition

Layer name	**Input Size** (Wi x Hi x Ci)	**Padding** (Pw x Ph)	**Stride**	**Filter size** (Kw x Kh)	**Output size** (W0 x H0 x C0)	**Parameters**	**Repeat**
Convolution 1	32 x 32 x 3	0 x 0	1	3 x 3	30 x 30 x 64	1792	1
Convolution 2	30 x 30 x 64	0 x 0	1	1 x 1	30 x 30 x 16	1040	1
Convolution 3	30 x 30 x 16	0 x 0	1	1 x 1	30 x 30 x 8	136	1
Convolution 4	30 x 30 x 8	0 x 1	1	1 x 3	30 x 30 x 16	400	1
Convolution 5	30 x 30 x 16	1 x 0	1	3 x 1	30 x 30 x 16	784	1
Convolution 6	30 x 30 x 16	0 x 0	1	1 x 1	30 x 30 x 32	544	1
Convolution 32	30 x 30 x 32	0 x 0	2	1 x 1	30 x 30 x 32	1056	1
Convolution 33	15 x 15 x 32	0 x 0	1	1 x 1	15 x 15 x 16	528	1
Convolution 34	15 x 15 x 16	0 x 1	1	1 x 3	15 x 15 x 32	1528	1
Convolution 35	15 x 15 x 32	1 x 0	1	3 x 1	15 x 15 x 32	3104	1
Convolution 36	15 x 15 x 32	0 x 0	1	1 x 1	15 x 15 x 64	2112	1
Convolution 37	15 x 15 x 64	0 x 0	1	1 x 1	15 x 15 x 32	2080	1
Convolution 38	15 x 15 x 32	0 x 0	1	1 x 1	15 x 15 x 16	528	1

Convolution 39	15 x 15 x 16	1 x 0	1	3 x 1	15 x 15 x 32	1568	1
Convolution 40	15 x 15 x 32	0 x 1	1	1 x 3	15 x 15 x 32	3104	1
Convolution 41	15 x 15 x 32	0 x 0	1	1 x 1	15 x 15 x 64	2112	1
Convolution 62	15 x 15 x 64	0 x 0	2	1 x 1	15 x 15 x 64	4160	1
Convolution 63	8 x 8 x 64	0 x 0	1	1 x 1	8 x 8 x 32	2080	1
Convolution 64	8 x 8 x 32	1 x 0	1	3 x 1	8 x 8 x 64	6208	1
Convolution 65	8 x 8 x 64	0 x 1	1	1 x 3	8 x 8 x 64	12352	1
Convolution 66	8 x 8 x 64	0 x 0	1	1 x 1	8 x 8 x 128	8320	1
Convolution 67	8 x 8 x 128	0 x 0	1	1 x 1	8 x 8 x 64	57792	7
Convolution 68	8 x 8 x 64	0 x 0	1	1 x 1	8 x 8 x 32	14560	7
Convolution 69	8 x 8 x 32	1 x 0	1	3 x 1	8 x 8 x 64	43456	7
Convolution 70	8 x 8 x 64	0 x 1	1	1 x 3	8 x 8 x 64	86464	7
Convolution 71	8 x 8 x 64	0 x 0	1	1 x 1	8 x 8 x 128	58240	7
Convolution 102	8 x 8 x 128	0 x 0	2	1 x 1	8 x 8 x 128	16512	1
Convolution 103	4 x 4 x 128	0 x 0	1	1 x 1	4 x 4 x 64	8256	1
Convolution 104	4 x 4 x 64	0 x 1	1	1 x 3	4 x 4 x 128	24704	1
Convolution 105	4 x 4 x 128	1 x 0	1	3 x 1	4 x 4 x 128	49280	1
Convolution 106	4 x 4 x 256	0 x 0	1	1 x 1	4 x 4 x 256	65792	1
Convolution 107 Spatial Resolution	4 x 4 x 256	0 x 0	1	1 x 1	4 x 4 x 128	32896	1
Adaptive Average Pool	4 x 4 x 256	-	-	-	4 x 4 x 256	-	1
FCC	1 x 1 x 128	0 x 0	1	1 x 1	1 x 1 x 10	1290	1

+First column Wi x Hi x Ci refer to input width x input height x input number of channels; Second column, Pw x Ph refer to width and height of padding; third column refers to the number of stride used; fourth column, Kw x Kh refer to width and height of the kernel; fifth column, W0 x H0 x C0 refer to width, height and number of channels for the output; sixth column represents number of parameters for the particular layer, and last column depicts the number of times a layer is repeated in the four stage implementation configuration.

8. Conclusion

In this paper, the existing macro architectures such as baseline SqueezeNet, baseline SqueezeNext and a family of MobileNet architectures had laid foundation and motivated the development of the proposed High Performance SqueezeNext. Fine hyper parameter tuning and tweaking, compression using width and resolution multipliers, implementing different optimizers, step-based learning decay rate, data augmentation, save and load method for python, dataset specific training with no transfer learning approach and real-time embedded platform deployment contributed towards the improvement of the proposed DNN architecture. This research also initiated and encouraged the DSE of DNNs with the help of experiments implementing different activation functions and various optimizers.

Hence, these insights helped to build a better understanding of various optimizers, model compression, learning rate scheduling methods, save and load checkpoint, dataset specific training and testing. High Performance SqueezeNext architecture is one of the several new CNNs/DNNs that have been discovered while broadly exploring the DSE of DNN architectures. Detailed composition of High Performance SqueezeNext is shown in Table 5.

This architecture has 15x and 13x better model accuracy than baseline SqueezeNet and baseline SqueezeNext, respectively. It has a minimal 0.370MB model size, in other words, it is 8x and 7x smaller than baseline SqueezeNet and baseline SqueezeNext baseline. All the results discussed in this paper demonstrate the trade-off between model accuracy, model speed and model size with different resolution and width multipliers. SGD with momentum and nesterov is proposed as a suggested optimizer to be implemented on any DNN architecture. It is expected that with the incredibly small model size of 370KB for High Performance SqueezeNext, referring to Table 2, with an accuracy of 92.05%, referring to Table 1, High Performance SqueezeNext model can be easily deployed on a real time embedded platform.

The proposed, High Performance SqueezeNext architecture is trained and tested from scratch on datasets such as CIFAR-10 and CIFAR-100, individually without any transfer learning. The proposed DNN was successfully deployed on Bluebox 2.0 by NXP with DNN model accuracy of 92.50%, 16 seconds per epoch model speed and 2.62MB of model size.

High Performance SqueezeNext attains model accuracy 15.8% better than MobileNet (76.7%), 13.2% better than Squeezed CNN (79.30%), 10% better than alexnet_tf (82% model accuracy), 3.50% better than DCNN (89%), 2.6% better than enhanced hybrid MobileNet (89.9%), 1.85% better than Maxout networks (90.65%), and 0.25% better than modified SqueezeNext (92.25%) with better model size. Hopefully, this research will inspire design space exploration (DSE) of DNNs in a more intrinsic and aggressive manner.

Conflict of Interest

The authors declare no conflict of interest.

Acknowledgment

We would like to acknowledge all the IoT Collaboratory lab members and other colleagues for their continuous support, reviews and regular feedback.

References

[1] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly, J. Uszkoreit, N. Houlsby, "An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale," in International Conference on Learning Representations, 2021.

[2] A. Kolesnikov, L. Beyer, X. Zhai, J. Puigcerver, J. Yung, S. Gelly, N. Houlsby, Big Transfer (BiT): General Visual Representation Learning,

ECCV 2020:, Springer, Cham, 2020, doi:10.1007/978-3-030-58558-7_29.
[3] M. Tan, Q. V. Le, "EfficientNetV2: Smaller Models and Faster Training," in International Conference on Machine Learning, PMLR: 10096–10106, 2021.
[4] L. Wang, S. Xie, T. Li, R. Fonseca, Y. Tian, "Sample-Efficient Neural Architecture Search by Learning Action Space," arXiv, 2019, doi:10.48550/ARXIV.1906.06832.
[5] L. Wang, S. Xie, T. Li, R. Fonseca, Y. Tian, "Sample-Efficient Neural Architecture Search by Learning Actions for Monte Carlo Tree Search," IEEE Transactions on Pattern Analysis and Machine Intelligence, **01**, 1–1, 2021, doi:10.1109/TPAMI.2021.3071343.
[6] Y. Huang, Y. Cheng, A. Bapna, O. Firat, M.X. Chen, D. Chen, H. Lee, J. Ngiam, Q. V. Le, Y. Wu, Z. Chen, "GPipe: Efficient Training of Giant Neural Networks using Pipeline Parallelism," NIPS'19: Proceedings of the 33rd International Conference on Neural Information Processing Systems: 103–112, 2019, doi:10.48550/ARXIV.1811.06965.
[7] H.Y. Chen, C.Y. Su, "An Enhanced Hybrid MobileNet," in 9th International Conference on Awareness Science and Technology, iCAST 2018, Institute of Electrical and Electronics Engineers Inc., Fukuoka: 308–312, 2018, doi:10.1109/ICAwST.2018.8517177.
[8] F.N. Iandola, S. Han, M.W. Moskewicz, K. Ashraf, W.J. Dally, K. Keutzer, "SqueezeNet: AlexNet-level accuracy with 50x fewer parameters and <0.5MB model size," arXiv, 2016, doi:10.48550/ARXIV.1602.07360.
[9] A. Gholami, K. Kwon, B. Wu, Z. Tai, X. Yue, P. Jin, S. Zhao, K.K. Eecs, "SqueezeNext: Hardware-Aware Neural Network Design," in 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), IEEE: 1719–171909, 2018, doi:10.1109/CVPRW.2018.00215.
[10] K. Simonyan, A. Zisserman, "Very Deep Convolutional Networks for Large-Scale Image Recognition," in International Conference on Learning Representations, 2014.
[11] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, "Rethinking the Inception Architecture for Computer Vision," in Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR), IEEE Computer Society, Las Vegas: 2818–2826, 2016, doi:10.1109/CVPR.2016.308.
[12] K. He, J. Sun, "Convolutional Neural Networks at Constrained Time Cost," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), IEEE Computer Society: 5353–5360, 2015, doi:10.1109/CVPR.2015.7299173.
[13] J.K. Duggal, DESIGN SPACE EXPLORATION OF DNNS FOR AUTONOMOUS SYSTEMS, Purdue University, 2019, doi:https://doi.org/10.25394/PGS.8980463.v1.
[14] S. Ruder, "An overview of gradient descent optimization algorithms," in arXiv preprint arXiv:1609.04747, arXiv, 2016, doi:10.48550/ARXIV.1609.04747.
[15] J.K. Duggal, M. El-Sharkawy, "High Performance SqueezeNext for CIFAR-10," in Proceedings of the IEEE National Aerospace Electronics Conference, NAECON, Institute of Electrical and Electronics Engineers Inc.: 285–290, 2019, doi:10.1109/NAECON46414.2019.9058217.
[16] S. Venkitachalam, S.K. Manghat, A.S. Gaikwad, N. Ravi, S.B.S. Bhamidi, M. El-Sharkawy, "Realtime applications with rtmaps and bluebox 2.0," in Proceedings on the International Conference on Artificial Intelligence (ICAI), The Steering Committee of The World Congress in Computer Science, Computer: 137–140, 2018.
[17] J. Kant Duggal, M. El-Sharkawy, "Shallow SqueezeNext: Real Time Deployment on Bluebox2.0 with 272KB Model Size," Journal of Electrical and Electronic Engineering, **8**(6), 127, 2020, doi:10.11648/j.jeee.20200806.11.
[18] D. Pathak, M. El-Sharkawy, "Architecturally Compressed CNN: An Embedded Realtime Classifier (NXP Bluebox2.0 with RTMaps)," in 2019 IEEE 9th Annual Computing and Communication Workshop and Conference (CCWC), 331–336, 2019, doi:10.1109/CCWC.2019.8666495.
[19] A. Krizhevsky, V. Nair, G. Hinton, "Cifar-10 (canadian institute for advanced research)," 2019.
[20] T.B. Ludermir, A. Yamazaki, C. Zanchettin, "An optimization methodology for neural network weights and architectures," IEEE Transactions on Neural Networks, **17**(6), 1452–1459, 2006, doi:10.1109/TNN.2006.881047.
[21] S. Ioffe, C. Szegedy, "Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift," in Proceedings of the 32nd International Conference on Machine Learning - Volume 37, PMLR: 448–456, 2015.
[22] B. Recht, R. Roelofs, L. Schmidt, V. Shankar, "Do CIFAR-10 Classifiers Generalize to CIFAR-10?," arXiv, 2018, doi:10.48550/ARXIV.1806.00451.
[23] H. Touvron, M. Cord, A. Sablayrolles, G. Synnaeve, H. Jégou, "Going deeper with Image Transformers," in Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), 32–42, 2021.
[24] A.G. Howard, M. Zhu, B. Chen, D. Kalenichenko, W. Wang, T. Weyand, M. Andreetto, H. Adam, "MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications," arXiv, 2017, doi:10.48550/ARXIV.1704.04861.
[25] J.K. Duggal, M. El-Sharkawy, "Shallow squeezenext: An efficient shallow DNN," in 2019 IEEE International Conference on Vehicular Electronics and Safety, ICVES 2019, Institute of Electrical and Electronics Engineers Inc., 2019, doi:10.1109/ICVES.2019.8906416.
[26] J.K. Duggal, M. El-Sharkawy, Shallow SqueezeNext Architecture Implementation on Bluebox2. 0, Advances i, Springer, Cham, 2021.

ASTES

Advances in Science, Technology and Engineering Systems Journal Vol. 7, No. 3, 82-86 (2022)

www.astesj.com

ASTESJ

ISSN: 2415-6698

Hole-Confined Polar Optical Phonon Interaction in $Al_{0.35}Ga_{0.65}As/GaAs/Al_{0.25}Ga_{0.75}As$ Quantum Wells

Mohamed Boumaza[1,*], Yacine Boumaza[2]

[1]*Teacher Education college of Setif Messaoud Zeghar, Department of sciences, El Eulma, 19600, Algeria*

[2]*Batna 1 University, Department of Electrical Engineering, Faculty of Technology, 05000 DZ, Algeria*

ARTICLE INFO

Article history:
Received: 19 March, 2022
Accepted: 15 May, 2022
Online: 25 May, 2022

Keywords:
Quantum Well
Confined Phonon
Semiconductor

ABSTRACT

In $Al_{0.35}Ga_{0.65}As/GaAs/Al_{0.25}Ga_{0.75}As$ quantum wells, the hole-confined polar optical phonon interaction is investigated. To calculate the valence band structure, we use the Luttinger-Kohn Hamiltonian with the k.p method. Within the dielectric continuum model, the hole-confined phonon scattering rates of intrasubband heavy holes in quantum well are calculated. It is found that the scattering rates are governed by an overlap integral and the density of states. Moreover, the scattering rates are reduced under compressive hydrostatic strain for low hole energy. The anisotropic effect on hole-confined phonon interaction is also studied.

1. Introduction

In physics, electron-phonon interaction plays an important role such as spin relaxation [1], superconductivity [2], quantum laser [3], mobility [4], Carrier thermalization [5]. Over the past decade, there has been an increasing interest of $GaAs/Al_xGa_{1-x}As$ heterostructures with a variety of structures such as heterojunction [6], quantum well [7], quantum wire [8], quantum dot [9], multiquantum well [10], superlattice [11]. It is well known that phonons are confined in quantum well which has proven experimentally [12–14]. Furthermore, to describe the optical phonons in quantum well there are several models such as the dielectric continuum model (DCM) in [15], the hydrodynamic model in [16], the microscopic model in [17], and the hybrid model of in [18]. In the case of other structures such as periodic soliton we use the new generalized (G'/G)-expansion method [19].

In this paper, we calculate the valence band structure using the 6x6 Luttinger-Kohn Hamiltonian, taking into account the warping in $Al_{0.35}Ga_{0.65}As/GaAs/Al_{0.25}Ga_{0.75}As$ asymmetric quantum well [20,21] within the k.p method due to its simplicity and accuracy [22], whereas to describe the phonons in the quantum well, we use the dielectric continuum model which has been used by several authors [23–27], and given excellent results compared to the experimental results [28, 29]. In addition, we investigate hole confined phonon scattering rates for different quantum well widths and the results are discussed. We also study the scattering rates under compressive hydrostatic strain using the theory of Luttinger–Kohn and Bir–Pikus [30].

2. Theory

In our work, we consider an asymmetric quantum well grown along the z direction. The 6x6 Luttinger Hamiltonian is transformed into two 3×3 matrixes [31,32], we calculate the hole band structure, by solving the Schrödinger equation including the heavy hole (HH), light hole (LH), and spin-orbit split-off subbands.

With the dielectric continuum model, the Frohlich Hamiltonian is written as [25]

$$H_{h-ph} = \sum_q \sum_m e\Phi_m(z)\,(a_m(q) + a_m^+(-q))\,e^{i\boldsymbol{q}.\boldsymbol{r}} \quad (1)$$

here $a_m^+(-q)$ is the phonon creation operator, $a_m(q)$ is the phonon annihilation operator, $\Phi_m(z)$ is the normalized phonon potential, $\boldsymbol{r}$ is the position vector in the xy plane, m denotes the LO mode order index and $\boldsymbol{q}$ is the in-plane phonon wave vector.

Using the Fermi's golden rule, the hole-confined phonon scattering rates from the initial hole state with the wave vector $\boldsymbol{k_i}$ within subband i to the final hole states in subband f with wave vector $\boldsymbol{k_f}$ are calculated as [33]

*Corresponding Author: Mohamed Boumaza, mboumazaphy1@gmail.com

$$\gamma_{i\to j} = \frac{2\pi}{\hbar}\int |M(f,i)|^2 \delta(E_f \pm \hbar\omega - E_i)\, dN_f \quad (2)$$

here, N_f is the number of final states, E_i is the initial hole state energie, E_f is the final hole state energie, $M(f,i)$ is the function connecting between the initial and the final hole states. Equation **(2)** yields

$$\gamma_{i\to j} = \frac{1}{2\pi\hbar}\frac{k_f}{\frac{\partial E_f}{\partial k}}\left(N_q + \frac{1}{2} \pm \frac{1}{2}\right)\Gamma_{if} \quad (3)$$

where N_q is the phonon occupation number, Γ_{if} is the function of the hole wave function and the phonon potential, which is written as

$$\Gamma_{if} = \sum_m \int \left| \int \left\langle f, k_f \left| e\Phi_m(z) \right| i, k_i \right\rangle dz \right|^2 d\theta \quad (4)$$

In Equation **(4)** the integration is done numerically where momentum and energy are conserved. The phonon potential $\Phi_m(z)$ is provided by [25]

$$\Phi_m(z) = A_C \cos\left(\frac{m\pi z}{L}\right) \qquad m=1,3,5,.. \quad (5)$$

$$\Phi_m(z) = A_C \sin\left(\frac{m\pi z}{L}\right) \qquad m=2,4,6,.. \quad (6)$$

here L is quantum well width, A_C is the normalization constant [25].

3. Results and discussions

In our work, we use the material parameters listed in Table [34–36]

Table 1: Parameters used in our work

Parameter	Unit	GaAs	AlAs
γ_1		6.85	3.69
γ_2		2.1	0.79
γ_3		2.9	1.4
Δ	eV	0.341	0.28
E_g	eV	1.424	2.671
a_v	eV	-1.16	-2.47
$\hbar\omega_{LO}$	eV	0.03625	0.05009

Figure 1 shows the valence band structure of a 25 Å $Al_{0.35}Ga_{0.65}As/GaAs/Al_{0.25}Ga_{0.75}As$ asymmetric quantum well in the k_x - k_y plane showing a great nonparabolicity with lifted spin degeneracy. Because of the coupling between the heavy hole and light hole subbands, our results exhibit a strong anisotropy along the [10] and [11] directions. We note here that the heavy hole subband is more anisotropic than the light hole subband in particular for high energies.

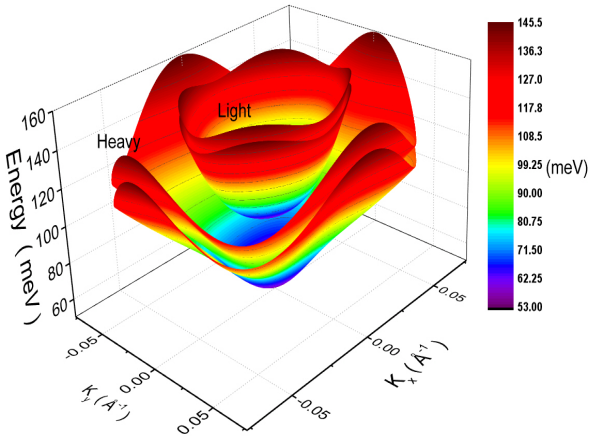


Figure 1: Heavy hole subband and light hole subband structures as a function of wave vector k in the $k_x - k_y$ plane for L = 25 Å and for clarity, the split off subband is not shown.

Figure 2.a shows the intrasubband heavy holes scattering rates $\gamma_{H\to H}$ of the confined optical phonon absorption as function of initial hole energy for different well widths, whereas the function Γ_{if} is shown in **(b).**

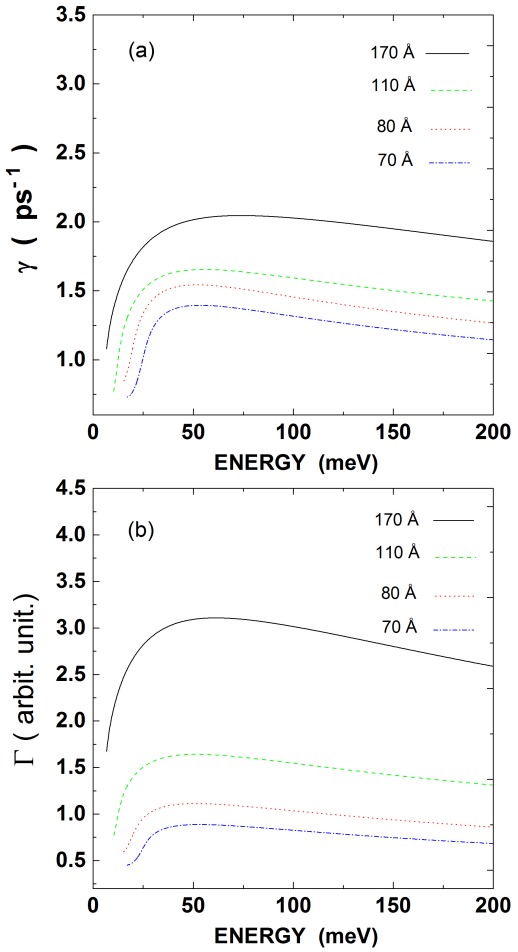


Figure 2 : **a)** Intrasubband scattering rates of heavy hole for *different well widths* respectively 70 Å, 80 Å , 110 Å, 170 Å **b)** The function Γ_{if}

The scattering rates depend on the density of states and the function $\Gamma_{i\,f.}$ Therefore, to understand our results, we plot in Figure 3 the dependence of the density of states on the hole energies.

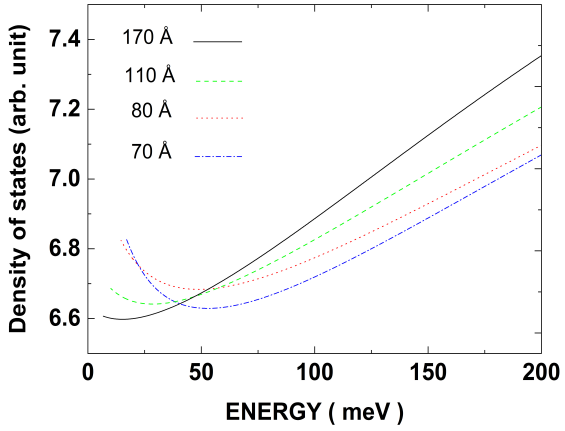


Figure 3: Density of states D(E) of heavy hole as function of hole energy and for different well widths.

One can see that for low hole energy scattering rates increase rapidly with increasing hole energy to reach its maximum value. However, for high hole energies although the density of states increases with increasing hole energy, scattering rates show a very slight decrease. This indicates that the scattering rate shows only a weak dependence on the density of states. The maximum value of the scattering rate is 1.39×10^{12} s^{-1} for L = 70 Å while the scattering rate reaches its highest value 2.045×10^{12} s^{-1} for L = 170 Å

It is well known that under hydrostatic strain valence band structure is altered, which leads to a significant change in the scattering rate. In order to study the scattering rates under strain, we show in figure 4 a) Heavy hole scattering rates for the confined optical phonon absorption as function of the initial hole energy for two quantum well widths 70 Å and 170 Å b) The overlap integral $\Gamma_{i\,f.}$ The dashed lines depict the scattering rates for 2 % of compressive hydrostatic pressure, whereas the solid lines for the results in the absence of strain. For low hole energy and for quantum well width L = 70 Å (L = 170 Å) scattering rates under strain are reduced by about 46.5 % (64 %), on the other hand, for high hole energy and for L = 70 Å (L = 170 Å) scattering rates are increased by about 2.8 % (2.3 %). This behavior is similar to the function $\Gamma_{i\,f.}$

Figure 5 shows the scattering rate for the confined optical phonon as a function of the initial two dimensional wave vector k with including the warping in the valence subband structure. For clarity, we also plot in Figure 6 and 7 the scattering rates for confined phonon absorption as a function of the initial hole wave vector k in polar coordinates for two well widths L = 25 Å and L = 170 Å respectively. One can see that for the quantum well width L = 25 Å our results exhibit significant anisotropic behavior for high hole energies between the directions [0 1] and [1 1], in which scattering rates increase by 14.7 %. However, for the quantum well width L = 170 Å scattering rates decrease by 54.5 %. This anisotropy is due to the strong valence subband anisotropy.

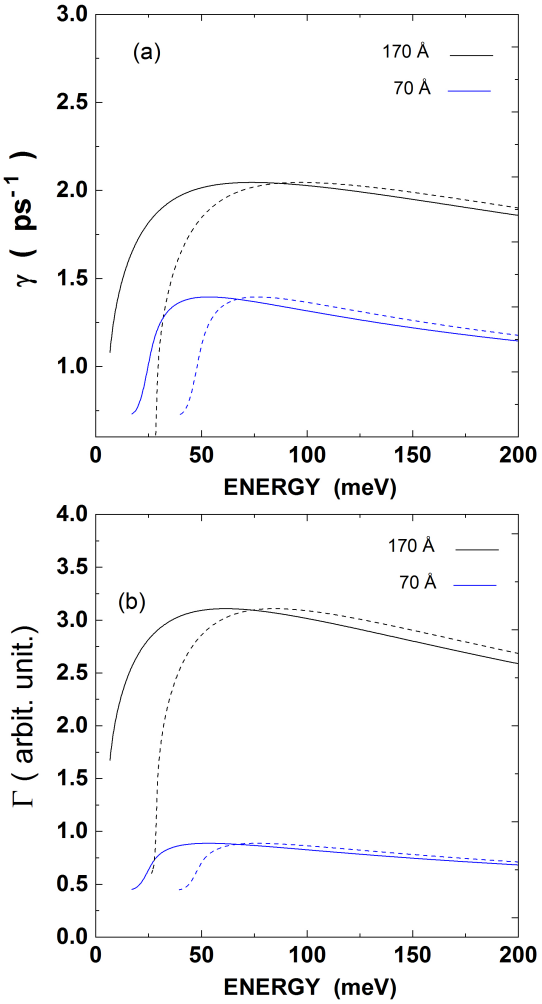


Figure 4: **a** Scattering rates within heavy hole subband for the confined optical phonon absorption and for two different well widths 70 Å and 170 Å b**)** The overlap integral $\Gamma_{i\,f.}$ The solid lines and dashed lines stand for the results without a strain with 2 % compressive hydrostatic pressure respectively.

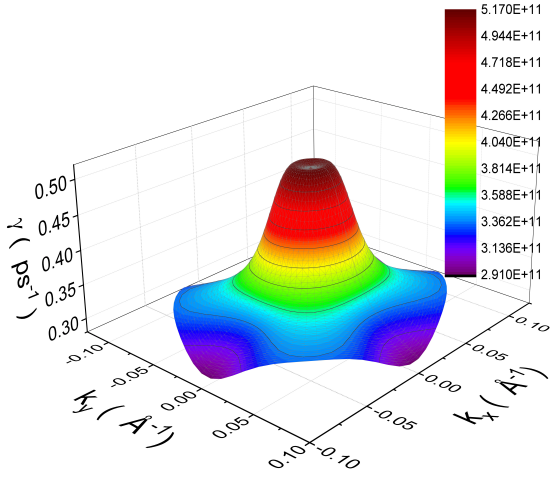


Figure 5: Scattering rates within heavy hole subband for confined optical phonon absorption with including warping as a function of the initial hole wave vector k in the kx-ky plane and for L = 25 Å

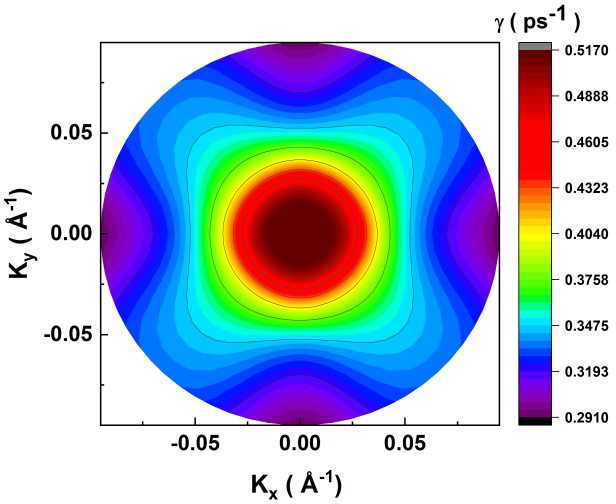


Figure 6: Scattering rates within heavy hole subband for confined optical phonon absorption as a function of the initial hole wave vector k in polar coordinates and for L = 25 Å

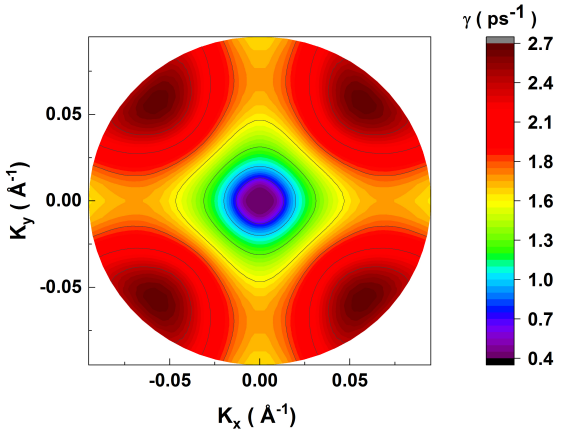


Figure 7: Scattering rates within heavy hole subband for confined optical phonon absorption as a function of the initial hole wave vector k in polar coordinates and for L = 170 Å

4. Conclusion

In summary, with the k.p method, the valence band structure is calculated including spin-orbit split-off subbands effect in the $Al_{0.35}Ga_{0.65}As/GaAs/Al_{0.25}Ga_{0.75}As$ asymmetric quantum wells. Hole-confined polar optical phonon scattering rates are investigated using the dielectric continuum model. It is found that scattering rates increase with increasing quantum well width. Moreover, under compressive hydrostatic strain, the scattering rates are reduced, in particular for low hole energy. In addition, scattering rates follow mostly the behavior of the overlap integral and exhibit a strong anisotropy for high hole energy. This anisotropy increases with increasing quantum well width. Our results show the importance of the band structure engineering quantum well via strain and within the asymmetric quantum well to reduce scattering rates and, consequently, the mobility of carriers can be increased. In the future, we will extend our work to different quantum well growth directions.

Conflict of Interest

The authors declare no conflict of interest.

References

[1] J. Park, J.-J. Zhou, M. Bernardi, "Spin-phonon relaxation times in centrosymmetric materials from first principles," Physical Review B, **101**(4), 045202, 2020, doi:10.1103/PhysRevB.101.045202.

[2] M. Li, Y. Fang, C. Kenney-Benson, L. Wang, "Superconductivity and electron–phonon interaction in SrxBi2Se3 under pressure," New Journal of Physics, **23**(8), 083011, 2021, doi:10.1088/1367-2630/ac14cf.

[3] A. Demić, Z. Ikonić, P. Dean, D. Indjin, "Dual resonance phonon–photon–phonon terahertz quantum-cascade laser: physics of the electron transport and temperature performance optimization," Optics Express, **28**(26), 38788–38812, 2020, doi:10.1364/OE.410014.

[4] T. Nematiaram, D. Padula, A. Landi, A. Troisi, "On the Largest Possible Mobility of Molecular Semiconductors and How to Achieve It," Advanced Functional Materials, **30**(30), 2001906, 2020, doi:10.1002/adfm.202001906.

[5] A. Asgari, L. Faraone, "Thermal broadening of electron mobility distribution in AlGaN/AlN/GaN heterostructures," Journal of Applied Physics, **114**(5), 053702, 2013, doi:10.1063/1.4813866.

[6] A. Sužiedėlis, S. Ašmontas, J. Gradauskas, V. Nargelienė, A. Čerškus, A. Lučun, T. Anbinderis, I. Papsujeva, A. Narkūnas, B. Kundrotas, R. Rinkevičienė, "Peculiarities of Temperature Dependence of Detected Voltage by GaAs/Al0.25Ga0.75As Heterojunction Microwave Diode Near Interwalley Crossover," Materials Science, **20**(2), 138–140, 2014, doi:10.5755/j01.ms.20.2.6319.

[7] T. Kawazu, "Valence Band Mixing in GaAs/AlGaAs Quantum Wells Adjacent to Self-Assembled InAlAs Antidots," Journal of Nanomaterials, **2019**, e5349291, 2019, doi:10.1155/2019/5349291.

[8] K.D. Pham, C.V. Nguyen, N.N. Hieu, H.V. Phuc, B.D. Hoi, B.M.H. Hoa, L.T.T. Phuong, "Theoretical investigation of hot electron cooling process in GaAs/AlAs cylindrical quantum wire under the influence of an intense electromagnetic wave," Optical and Quantum Electronics, **50**(9), 342, 2018, doi:10.1007/s11082-018-1606-x.

[9] S. Germanis, P. Atkinson, R. Hostein, S. Suffit, F. Margaillan, V. Voliotis, B. Eble, "Electrical control of optically pumped electron spin in a single GaAs/AlAs quantum dot fabricated by nanohole infilling," Physical Review B, **102**(3), 035406, 2020, doi:10.1103/PhysRevB.102.035406.

[10] O. Kojima, T. Kita, M.J. Steer, R.A. Hogg, "Modulation of exciton states through resonant excitation by continuous wave lasers in a GaAs/AlAs multiple quantum well," Journal of Physics D: Applied Physics, **54**(33), 2021.

[11] H. Eisele, L. Li, E.H. Linfield, "High-performance GaAs/AlAs superlattice electronic devices in oscillators at frequencies 100–320 GHz," Applied Physics Letters, **112**(17), 172103, 2018, doi:10.1063/1.5020265.

[12] A.K. Sood, J. Menéndez, M. Cardona, K. Ploog, "Interface Vibrational Modes in GaAs-AlAs Superlattices," Physical Review Letters, **54**(19), 2115–2118, 1985, doi:10.1103/PhysRevLett.54.2115.

[13] M.V. Klein, "Phonons in semiconductor superlattices," IEEE Journal of Quantum Electronics, **22**(9), 1760–1770, 1986, doi:10.1109/JQE.1986.1073174.

[14] A.K. Sood, J. Menéndez, M. Cardona, K. Ploog, "Resonance Raman Scattering by Confined LO and TO Phonons in GaAs-AlAs Superlattices," Physical Review Letters, **54**(19), 2111–2114, 1985, doi:10.1103/PhysRevLett.54.2111.

[15] R. Fuchs, K.L. Kliewer, "Optical Modes of Vibration in an Ionic Crystal Slab," Physical Review, **140**(6A), A2076–A2088, 1965, doi:10.1103/PhysRev.140.A2076.

[16] M. Babiker, "Longitudinal polar optical modes in semiconductor quantum wells," Journal of Physics C: Solid State Physics, **19**(5), 683–697, 1986, doi:10.1088/0022-3719/19/5/008.

[17] K. Huang, B. Zhu, "Dielectric continuum model and Frohlich interaction in superlattices," Physical Review B, **38**(18), 13377–13386, 1988, doi:10.1103/PhysRevB.38.13377.

[18] B.K. Ridley, "Electron-hybridon interaction in a quantum well," Physical Review B, **47**(8), 4592–4602, 1993, doi:10.1103/PhysRevB.47.4592.

[19] Md.K. Alam, Md.D. Hossain, M.A. Akbar, K.A. Gepreel, "Determination of the rich structural wave dynamic solutions to the Caudrey–Dodd–Gibbon equation and the Lax equation," Letters in Mathematical Physics, **111**(4), 103, 2021, doi:10.1007/s11005-021-01443-9.

[20] P.C. Klipstein, Y. Livneh, O. Klin, S. Grossman, N. Snapi, A. Glozman, E. Weiss, "A k·p model of InAs/GaSb type II superlattice infrared detectors," Infrared Physics & Technology, **59**, 53–59, 2013, doi:10.1016/j.infrared.2012.12.009.

[21] S. Rodríguez, J.A. López-Villanueva, I. Melchor, J.E. Carceller, "Hole confinement and energy subbands in a silicon inversion layer using the

effective mass theory," Journal of Applied Physics, **86**(1), 438–444, 1999, doi:10.1063/1.370749.

[22] Z. Becer, A. Bennecer, N. Sengouga, "Modeling Energy Bands in Type II Superlattices," Crystals, **9**(12), 629, 2019, doi:10.3390/cryst9120629.

[23] M. Boumaza, S. Lamari, "Anisotropic intrasubband hole scattering by polar optical phonon modes in thin GaAs/AlxGa1−xAs quantum wells," Superlattices and Microstructures, **72**, 156–163, 2014, doi:10.1016/j.spmi.2014.03.049.

[24] W.-D. Huang, G.-D. Chen, Y.-J. Ren, "Effect of ternary mixed crystals on interface optical phonons in wurtizte InxGa1−xN/GaN quantum wells," Journal of Applied Physics, **112**(5), 053704, 2012, doi:10.1063/1.4748173.

[25] N. Mori, T. Ando, "Electron--optical-phonon interaction in single and double heterostructures," Physical Review B, **40**(9), 6175–6188, 1989, doi:10.1103/PhysRevB.40.6175.

[26] J. Požela, K. Požela, V. Jucienė, "Scattering of electrons by confined interface polar optical phonons in a double-barrier heterostructure," Semiconductors, **41**(9), 1074–1079, 2007, doi:10.1134/S1063782607090126.

[27] A.M. Alcalde, G. Weber, "Scattering rates due to electron-phonon interaction in CdS1-xSex quantum dots," Semiconductor Science and Technology, **15**(11), 1082–1086, 2000, doi:10.1088/0268-1242/15/11/312.

[28] X. Zianni, C.D. Simserides, G.P. Triberis, "Electron scattering by optical phonons in AlxGa1−xAs/GaAs/AlxGa1−xAs quantum wells," Physical Review B, **55**(24), 16324–16330, 1997, doi:10.1103/PhysRevB.55.16324.

[29] V.V. Bondarenko, V.V. Zabudskii, F.F. Sizov, "Electron-phonon interaction and electron mobility in quantum-well type-II PbTe/PbS structures," Semiconductors, **32**(6), 665–667, 1998, doi:10.1134/1.1187461.

[30] Y. Kajikawa, "Optical anisotropy of (11l)-oriented strained quantum-wells calculated with the effect of the spin-orbit split-off band," Journal of Applied Physics, **86**(10), 5663–5677, 1999, doi:10.1063/1.371577.

[31] D.A. Broido, L.J. Sham, "Effective masses of holes at GaAs-AlGaAs heterojunctions," Physical Review B, **31**(2), 888–892, 1985, doi:10.1103/PhysRevB.31.888.

[32] C.Y.-P. Chao, S.L. Chuang, "Spin-orbit-coupling effects on the valence-band structure of strained semiconductor quantum wells," Physical Review B, **46**(7), 4110–4122, 1992, doi:10.1103/PhysRevB.46.4110.

[33] B.K. Ridley, Electrons and Phonons in Semiconductor Multilayers, 2nd ed., Cambridge University Press, Cambridge, 2009, doi:10.1017/CBO9780511581496.

[34] S. Boyer-Richard, F. Raouafi, A. Bondi, L. Pédesseau, C. Katan, J.-M. Jancu, J. Even, "30-band k·p method for quantum semiconductor heterostructures," Applied Physics Letters, **98**(25), 251913, 2011, doi:10.1063/1.3600643.

[35] I. Vurgaftman, J.R. Meyer, L.R. Ram-Mohan, "Band parameters for III–V compound semiconductors and their alloys," Journal of Applied Physics, **89**(11), 5815–5875, 2001, doi:10.1063/1.1368156.

[36] GaAs, AlAs, and AlxGa1−xAs: Material parameters for use in research and device applications: Journal of Applied Physics: **58**(3), Mar. 2022.

ASTES

Advances in Science, Technology and Engineering Systems Journal
Vol. 7, No. 3, 87-94 (2022)
www.astesj.com

ASTES Journal
ISSN: 2415-6698

A CMOS On-Chip High-Precision PVTL Detector

Pang-Yen Lou[1], Ying-Xuan Chen[1], Chua-Chin Wang[1,*], Wei-Chih Chang[2]

[1]*National Sun Yat-Sen University, Department of Electrical Engineering, 70 Lian-Hai Rd., Kaohsiung, Taiwan*

[2] *FocalTech Systems Co., Ltd, 6, Du Sing 1st Rd, East District, Hsinchu, Taiwan*

ARTICLE INFO

Article history:
Received: 14 April, 2022
Accepted: 16 May, 2022
Online: 27 May, 2022

Keywords:
High resolution
PVT corner
Leakage detecting
Linearity
Pulse count

ABSTRACT

A novel PVTL (Process, Voltage, Temperature, Leakage) detection circuit consisting of four individual detectors is proposed in the investigation. Voltage Variation Detector is composed of a feedback control block comprising multi-stage delay cells using high Vth devices such that 0.5% of VDD variation can be detected. Temperature Detector based on a current to pulse converter is proved to attain high linearity of temperature sensing. PMOS Variation Detector and NMOS Variation Detector are carried out using threshold voltage sensors and ring oscillators, respectively. Thus, all process corners can be clearly differentiated using pulse counts. Leakage Detector is realized by a single-MOSFET leakage current detector. Most of prior leakage detectors compensate for leakage current instead of detecting the precise amount of the leakage current. The proposed leakage detector, however, can accurately detect the leakage current of CMOS transistors, where a Strobe pulse generator is used as a detection switch. Thus, the detection time is predictable. It elevates the reliability of the detection result. The proposed PVTL detector design is implemented using a typical 180 nm CMOS process to justify the performance. Measurement shows that the proposed design is the best of all prior PVTL detectors in terms of accuracy.

1 Introduction

The transistors shrink in size as CMOS process advances, which benefits the digital circuits, including lower cost per area, lower supply voltage, lower power consumption, as well as higher operating speed. However, in semiconductor manufacturing, a 3-σ rule is required to overcome different doping concentrations on each N-type and P-type transistor, namely process corners, which might severely affect the performance of digital circuits. Meanwhile, as shown in Figure 1 [1], voltage and temperature are also unavoidable variations in any environment, which are needed to be detected to neutralize their effects on transistor operations. Moreover, the leakage issue becomes even more important by advancing of CMOS processes. Figure 2 shows the average gate leakage of different technology nodes, where the leakage increases as the process node evolves to ever-smaller devices [2], [3]. PVTL variations must be considered during chip design especially in digital circuits. To ensure the reliability of the performance, detections as well as auto-adjustments must be included in the system. Many prior PVTL detectors have been reported to address this problem with solutions [3]-[12]. A few other works reported to adjust for leakage problems [13]-[17]. Therefore, according to the challenges mentioned above, a high-precision detector for process, voltage, temperature, and leakage (PVTL) variation could be a solution.

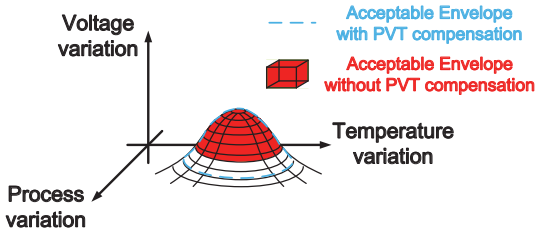


Figure 1: Acceptable envelope between with and without PVT detection [1]

In this investigation, high precision P, V, T, and L detectors are proposed, respectively. The proposed Voltage Variation Detector is featured with a delay line composed of high-Vth MOS-based delay cells. The proposed Temperature Detector consists of a temperature-sensitive current generator, a Charge and Discharge Circuit (CDC), and a Voltage Window Comparator (VWC). Lastly, the proposed process variation detectors comprises separate NMOS and PMOS process variation detectors such that all the possible process corners will be detected. Measurement given that the proposed corner detectors are realized using 0.18 micro meter CMOS technology node show that the voltage detecting resolution is as good as 0.5% of VDD, and temperature detecting resolution is proved to be 3 °C

*Corresponding Author. He is also with Institute of Undersea Technology (IUT) of NSYSU, Taiwan. Tel: 886-7-5252000 ext. 4144; Fax: 886-7-5254199.

in [-40 °C, +80 °C], and all the process variations, (SS, SF, TT, FS, FF), are detectable. Moreover, the leakage of PMOS and NMOS can be exactly detected by Leakage Detector, respectively.

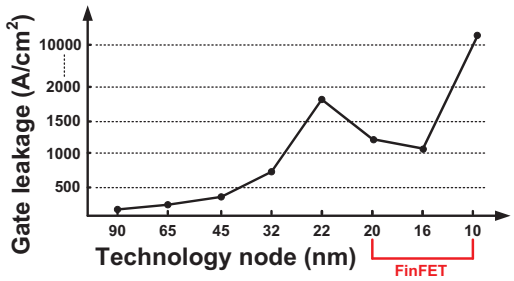


Figure 2: Gate leakage of different technology nodes [2], [3]

2 High-precision PVTL detector

Figure 3 shows the block diagram of the proposed on-chip high-precision PVTL detector contains the PMOS Variation Detector, NMOS Variation Detector, Temperature Detector, Voltage Variation Detector, and Leakage Detector. The details of each sub-circuit are given in the following text.

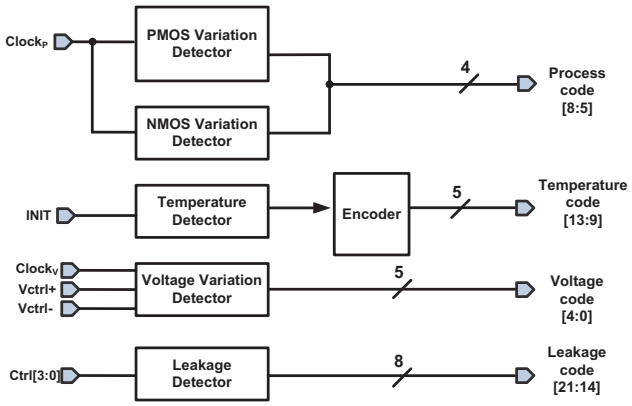


Figure 3: Proposed PVTL detecting system

2.1 Voltage Variation Detector

The schematic of Voltage Variation Detector in Figure 3 is shown in Figure 4, consisting of a buffer delay line (BF1-BF6), DFFs (FF1-FF6), XORs, and a Controllable delay block. Notably, all the cells in this circuit is also driven by the same input voltage (VDD). The operation of this circuit is summarized as follows:

- Due to the variation of input voltage (VDD), the delay generated by the buffer delay line will be varied accordingly.
- The generated codes at each delay cell are registered by corresponding DFFs, which are then the output of the adjacent DFF to generate Voltage code, V[0]-V[4].
- The last output of the buffer delay line, namely D6, is coupled to an input of Controllable delay block to form a feedback system.

Since Controllable delay block is meant to monitor the clock drift caused by voltage variation, the delay of each stage therein shall be auto-tuned by input voltage (VDD) and the final delay generated by the mentioned buffer delay line, e.g., D6. The schematic of the delay stage in Controllable delay block is revealed in Figure 5. It is notably featured with high Vth devices to prevent possible device parameter variation thanks to its thick gate oxide. D6 of buffer delay line is coupled to Vctrl+ and Vctrl- such that both the pull-up and pull-down switches have turned into a voltage controlled resistor.

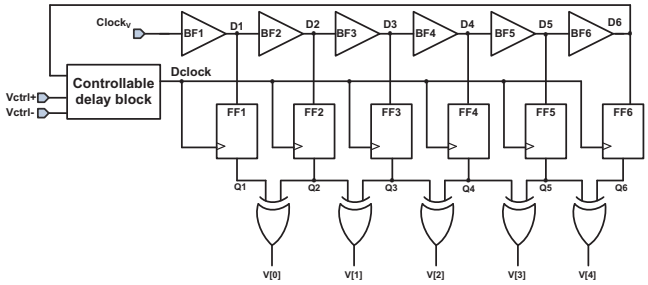


Figure 4: Schematic of the proposed Voltage Variation Detector

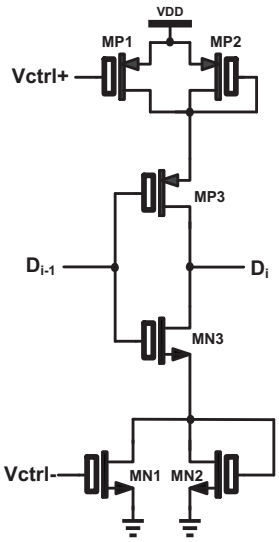


Figure 5: Schematic of the symmetric load delay cell

2.2 Process Variation Detector

The schematics of PMOS Variation Detector and NMOS Variation Detector in Figure 3 are shown in Figure 6 and Figure 7, respectively. Two individual process variation detectors are required for PMOS and NMOS to find out all the possible process corners. Take the N-type process variation detector in Figure 6 as an example. $Clock_P$ is coupled from the system clock source to an NAND gate to trigger the ring oscillator composed of only NMOS delay stages. As shown in Figure 7, different pulse counts (11, 6, 2) are attained

at the counter output given that NMOS are at F (fast), T (typical), S (slow) corners respectively, when $Clock_P$ is high.

As for PMOS process variation detector, because the pulse count generated by PMOS devices in different corners is small, the threshold voltage is used to judge the changes in different process corners. When the $Clock_P$ drops to low, MP17 in Figure 6 will be turned on and discharged to Vth of the PMOS. Similarly, MP18 will be discharged to twice Vth of the PMOS. The comparator is compared with the two bias voltages (vbias1 and vbias2) to determine the corner which it is now.

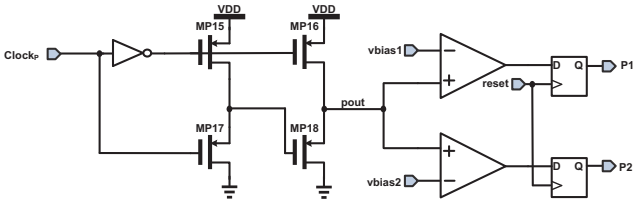


Figure 6: Schematic of the proposed PMOS Variation Detector

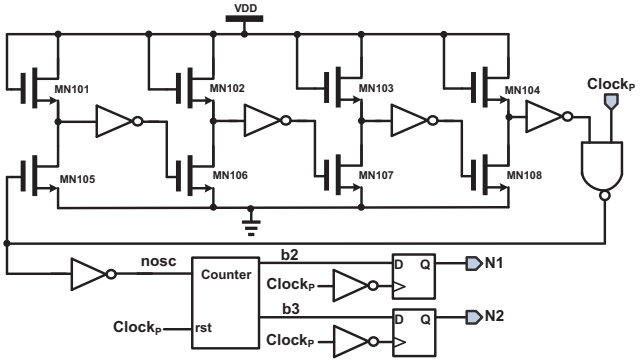


Figure 7: Schematic of the proposed NMOS Variation Detector

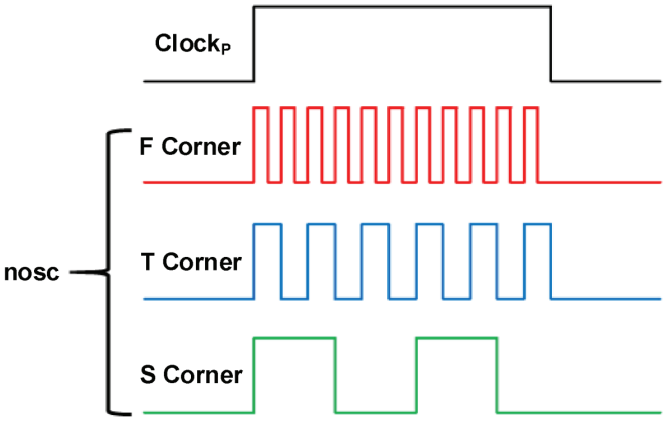


Figure 8: Waveforms of NMOS Variation Detector

2.3 *Temperature Detector*

The block diagram of Temperature Detector in Figure 3 is shown in Figure 9, consisting of a Current Generator (CG), a Charge and Discharge Circuit (CDC), a Voltage Window Comparator (VWC), and an Encoder. The schematic of CG, CDC, and VWC is given in Figure 10. Current Generator is charge of generating a current highly correlated to temperature variation. The operation of CDC and VWC is summarized as follows:

- Charging operation: The switch sw1 is shorted to node a1. Then, the storage capacitor, cap, starts to be charged via saturated MP12.
- Discharging operation: As soon as the voltage of the cap, $V_{cap}(T)$, reaches VH, the output of VWC, $V_{OUT}(T)$, is switched low to short-circuit sw1 to node b1. MN13 is tuned on to be able to sink a current which is twice of the charging current provided by MP13. Thus, cap is discharged. As soon as the $V_{cap}(T)$ is pulled down to VL, $V_{OUT}(T)$ will be turned high to start another cycle of charging-and-discharging operation.

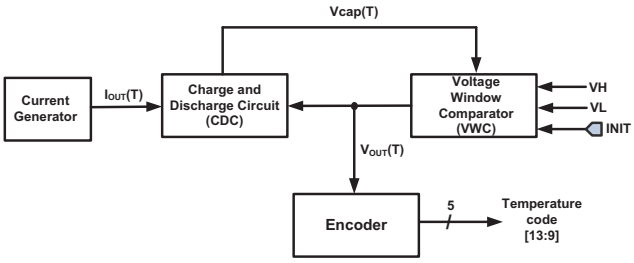


Figure 9: The block diagram of the proposed Temperature Detector

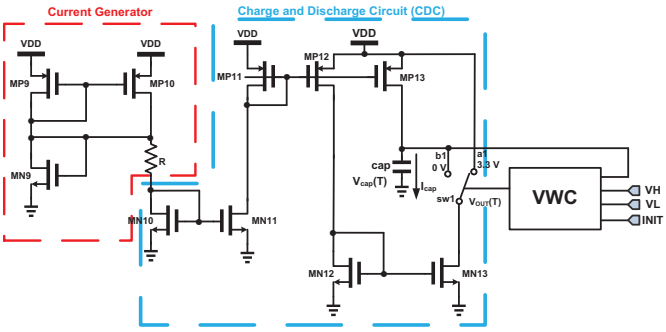


Figure 10: Schematic of the proposed Temperature Detector

$V_{cap}(T)$ is compared with two pre-defined voltages, VH and VL, in VWC. If $V_{cap}(T)$ ¿ VH, $V_{OUT}(T) = 0$. On the contrary, if $V_{cap}(T)$ ¡ VL, $V_{OUT}(T) = VDD$. This result in the temperature is converted into a string of pulses, where higher temperature results in higher pulse count and vice versa. The frequency of the output $F_{OUT}(T) = I_{cap}(T)/2 \cdot C_{cap} \cdot (VH - VL)$. The relationship between $F_{OUT}(T)$ and temperature by simulation is shown in Figure. 11, which temperature would be detected correctly in every °C, and the maximum error is 0.59% at 37°C.

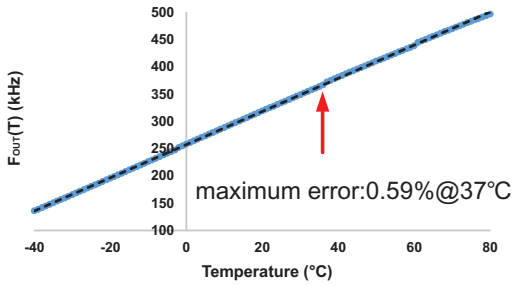


Figure 11: Theoretical function between F_{OUT}(T) and temperature by simulation

2.4 *Leakage Detector*

The schematic of Leakage Detector in Figure 3 is shown in Figure 12, consisting of a Strobe pulse generator, a Switch, two PMOS Leakage detectors, two NMOS Leakage detectors, a Voltage-to-frequency converter (VFC), and a Counter.

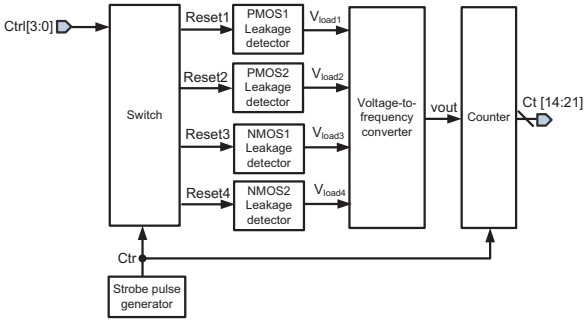


Figure 12: Schematic of the proposed Leakage Detector

The operation of the proposed Leakage Detector is summarized as follows:

- The Strobe pulse generator generates pulses which is the control signal of the four Leakage detectors and the Counter.
- Using the control signal Ctr and Ctrl[3:0] to select PMOS and NMOS with different aspect ratios as the object of leakage detection.
- The selected Leakage detector charges the capacitor with the leakage of a single PMOS or NMOS, and then converts the capacitor voltage output into a periodic square wave by the Voltage-to-frequency converter.
- The number of pulses of the periodic square wave is counted by Counter. If the larger the leakage is, the charging voltage of the capacitor will rise faster. Thus, the number of pulses registered in the counter will be higher.

Figure 13 shows the schematic of a PMOS Leakage detector and the equivalent voltage source model, which is the same as a DC transient circuit for RC charging and discharging. The function of NMOS Leakage detector is similar.

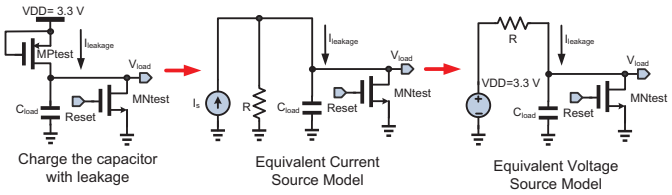


Figure 13: Equivalent model of Leakage Detector

Therefore, according to Eqn. (1) governed by Kirchhoff's voltage law (KVL), the capacitor charging voltage Eqn. (2) and the capacitor charging current Eqn. (3) can be derived,

$$E(t) = V_R(t) + V_C(t) = I(t) \times R + \frac{Q(t)}{C} = I(t) \times R + \frac{1}{C}\int I(t)dt \quad (1)$$

$$V_C(t) = E(t) - V_R(t) = V_S \times (1 - e^{-t/RC}) \quad (2)$$

$$I(t) = \frac{V_S}{R} \times e^{-t/RC} \quad (3)$$

where $V_R(t)$ is the voltage across the resistor, $V_C(t)$ is the capacitor voltage, V_S is the supply voltage (VDD), and e is the natural logarithm. When the circuit current drops from the maximum value to 36.8 %, the time is the product of the resistance and the capacitance, so the time constant (τ) of the RC charging circuit is the product of resistance and capacitance, as shown in Eqn. (4).

$$\tau = R \times C \quad (4)$$

Through the natural logarithmic function calculation, as shown in Eqn. (5) and (6), the time for the capacitor charging to be stable is 5τ, and then the resistance R is derived. Thus, referring to Eqn. (3) to derive the capacitor charging current at any time. In addition, the charging current would decrease by time. The initial PMOS leakage charging current I(0) is as shown in Eqn. (7).

$$t = 5\tau \Rightarrow (1 - e^{-5\tau/RC}) = 0.993 \div 1 \quad (5)$$

$$V_C(t = 5\tau) = V_S \times (1 - e^{-5}) \div 1 \quad (6)$$

$$I(t = 0) = \frac{V_S}{R} = I(0) \quad (7)$$

3 Measurement and Verification

This proposed PVTL detector design is realized using TSMC 180 nm CMOS process. The layout and die photo of the entire PVTL detector are shown in Figure 14 (a) and (b), respectively, where the core area is 1156×1671 μm^2, and the total chip area is 2493×2553 μm^2. The chip measurement setup of the proposed design is shown in Figure 15. The chip is soldered on the PCB to reduce noise interference. The Power Supply Agilent E3631A provides the required voltages to the chip. Arbitrary Waveform Generator Agilent 33522A and Signal Generator provide the $Clock_V$ and $Clock_P$, respectively. The Programmable Compact Temperature & Humidity System is the equipment to define the environment temperature. The oscilloscope WaveRunner610Zi is used to observe waveforms and check the circuit operations.

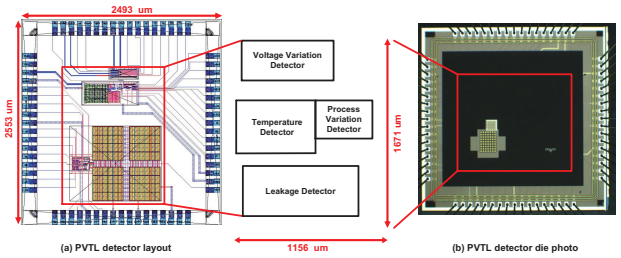


Figure 14: (a) PVTL detector layout; (b) PVTL detector die photo

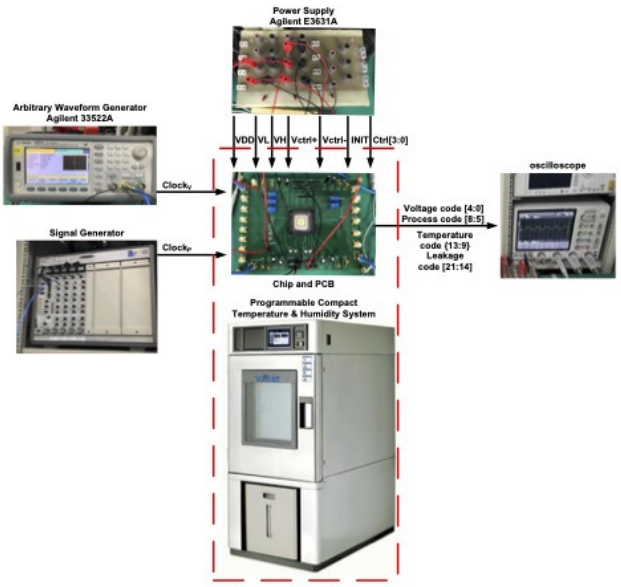


Figure 15: Measurement setup and equipment

A total of six chips are measured 10 times per chip to verify the reliability. The details of each measurement of sub-circuit are given in the following text.

3.1 Voltage Variation Detector

Figure 16 shows the measurement waveforms of the output signals of Voltage Variation Detector. It shows that Voltage code[4:0] is changed when VDD varies. The VDD is drifted by ±1%, ±0.5%, and 0%. Therefore, the measurement result proves the correctness of Voltage Variation Detector.

3.2 Process Variation Detector

Figure 17 shows the measurement waveforms of the output signals of Process Variation Detectors. It shows that the Process code is 0000 by all six chips. Thus, all PMOS and NMOS are made by SS corner. Notably, it is hard to ask foundry to deliberately fabricate the dies/chips at different process corners.

3.3 Temperature Detector

The interval between each measurement is at least an hour to ensure that the temperature of the entire chip and PCB is stable. Table 1 summarizes all the readings in the measurement, where the initial temperature is 20 °C, and the initial F_{OUT}(T) is 206 kHz. By tuning Programmable Compact Temperature & Humidity System, the temperature and F_{OUT}(T) rise continuously. In addition, F_{OUT}(T) in measurement is lower than that given by the simulations due to the loading of the passive components and PCB. However, F_{OUT}(T)-temperature curve is still linear in measurement. The error in Table 1 is calculated with the F_{OUT}(T) by the linear regression. The maximum error is 2.24 % at -5 °C, while the overage error is 0.773 %.

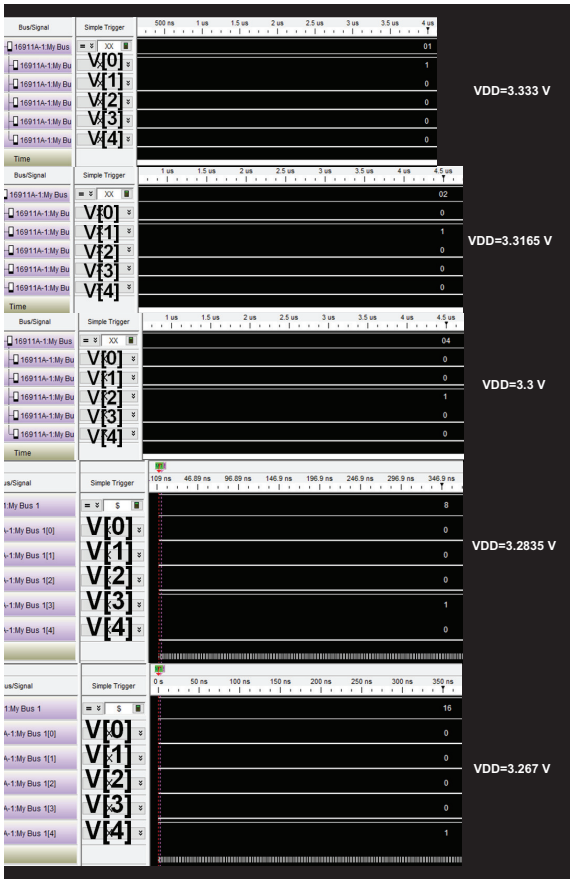


Figure 16: The measurement result of voltage variation

3.4 Leakage Detector

Referring to the analysis of the pulse count and leakage current which is shown in Figure 18 and the measurement waveforms of Leakage Detector shown in Figure 19. Figure 19 (a) and (b) are the PMOS Leakage Detector measurement outcome. The size of the PMOS under test in (a) is 50 μm/0.35 μm, and the pulse count is 31 equivalent to 31 pA. The size of the PMOS under test in (b) is 100 μm/0.35 μm, and the pulse count of (b) is 35 equivalent to 35 pA. Figure 19 (c) and (d) are the NMOS Leakage Detector measurement outcome. The size of the NMOS under test in (c) is 50 μm/0.35 μm, and the pulse count is 30 equivalent to 30 pA. The size of the NMOS under test in (d) is 100 μm/0.35 μm, and the pulse count is 55 equivalent to 55 pA. Therefore, the pulse count of the proposed Leakage Detector is positively correlated with the leakage of PMOS and NMOS.

Table 1: Measurement outcome of the proposed Temperature Detector

F_{OUT}(T) (kHz)	Temperature (°C)	Error (%)	F_{OUT}(T) (kHz)	Temperature (°C)	Error (%)
106	-5	2.24	285	40	0.92
129	0	1.25	304	45	0.53
148	5	0.38	325	50	0.11
163	10	0.27	348	55	0.68
187	15	0.86	371	60	0.64
206	20	0.92	393	65	0.34
224	25	1.16	419	70	0.35
242	30	1.08	441	75	0.3
267	35	1.6	463	80	0.29

Table 2: Performance comparison of Output Buffers

	[18] *TCAS-I 2013*	[19] *TCAS-II 2010*	[20] *TCAS-II 2019*	[21] *TVLSI 2017*	[22] *ICICDT 2019*	This work *2022*
Process (nm)	90	180	40	40	28	180
VDD (V)	1.2	1.8	0.9	0.9	0.9	3.3
VDD Variation (%)	10	10	10	10	10	0.5
Temperature Range (°C)	N/A	0-40	0-75	0-75	0-75	-40-80
PVT corner detected	5 (All)	5 (All)	5 (All)	5 (All)	5 (All)	5 (All)
Leakage detection	N/A	N/A	N/A	Yes	Yes	Yes
Verification	Mea.	Mea.	Mea.	Mea.	Mea.	Mea.
Core Area (mm^2)	-	0.009	0.011	0.011	0.013	1.931
Core Area (Normalization) (mm^2)	-	1.11	6.875	6.875	16.581	59.5
FOM^a	N/A	18.01	2.91	5.45	2.26	1.01
FOM^b	N/A	19.99	20	37.45	37.47	60.9

$$FOM^a = \left(\frac{\text{Temperature Range} \cdot \text{PVT corner detected}}{\text{VDD Variation} \cdot \text{Normalized Chip Area}}\right)$$

$$FOM^b = \left(\frac{\text{Temperature Range} \cdot \text{PVT corner detected}}{\text{VDD Variation}}\right)$$

3.5 Performance comparison

Table 2 tabulates the performance comparison of several recent PVTL detector works. This work attain the best FOM^b and the highest accuracy among all PVTL detector works in 2011-2020. However, due to the fact that four individual detectors are used, the proposed design pay the price of larger area. If the area is a factor be considered, FOM^a, is also given and defined in Table 2. Apparently, it is a trade-off between high performance and large area.

4 Conclusion

A highly accurate on-chip PVTL detector design is demonstrated in this investigation, where four individual detectors are used to assist the quality improvement of chips implemented by CMOS. If the ICs' leakage and PVT corner are correctly estimated, the host controller will be able to perform better.

5 Acknowledgment

This work was partially supported by FocalTech Systems Co., Ltd, Taiwan, under grant NO. N109130. The authors would like to express their deepest gratefulness to TSRI (Taiwan Semiconductor Research Institute), Taiwan, for their chip fabrication service.

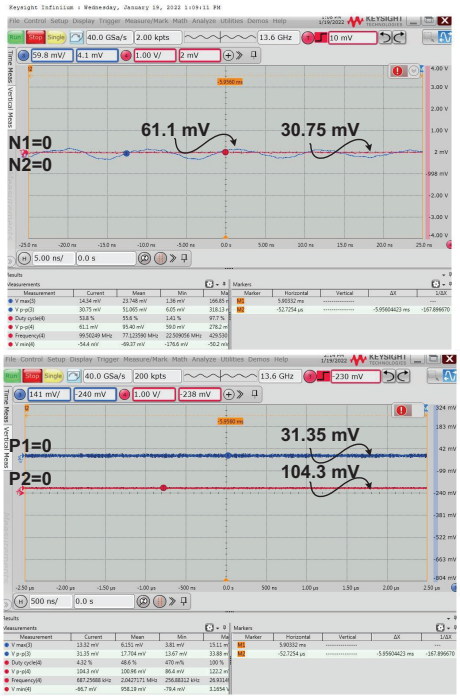


Figure 17: The measurement result of process variation

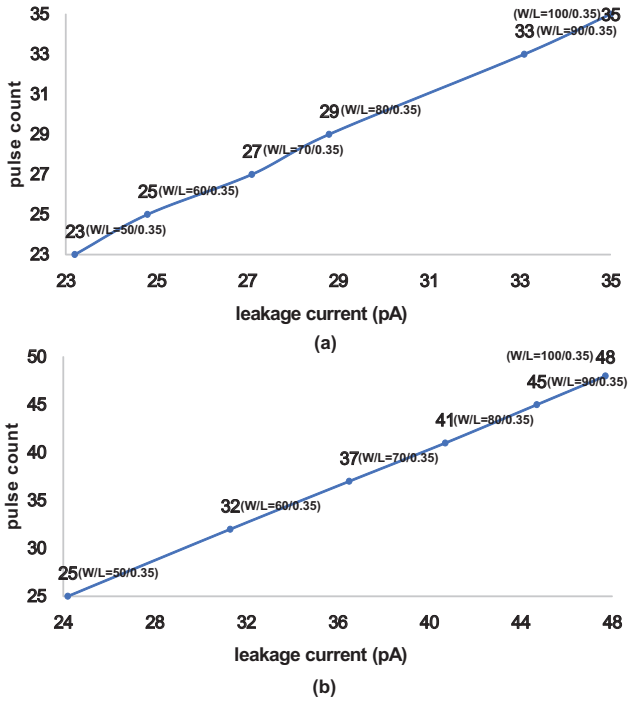


Figure 18: The analysis of the pulse count and leakage current: (a) PMOS leakage vs pulse count; (b) NMOS leakage vs pulse count

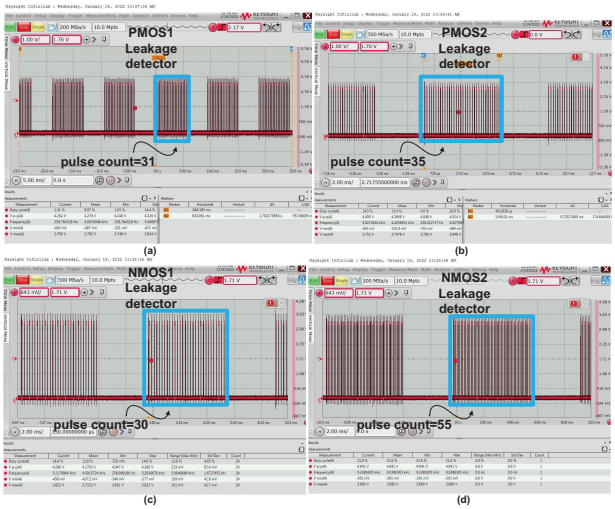


Figure 19: The measurement waveform of Leakage Detector:(a)PMOS (W/L=50/0.35) Leakage Detector; (b)PMOS (W/L=100/0.35) Leakage Detector; (c)NMOS (W/L=50/0.35) Leakage Detector; (d)NMOS (W/L=100/0.35) Leakage Detector

References

[1] C.-C. Wang, W.-J. Lu, K.-W. Juan, W. Lin, H.-Y. Tseng, and C.-Y. Juan, "Process corner detection by skew inverters for 500 MHz 2×VDD output buffer using 40-nm CMOS technology," *Microelectronics Journal,* **46**(1), 1-11, Jan. 2015. doi: 10.1016/j.mejo.2014.09.010

[2] H. F. Dadgour, S. Lin, and K. Banerjee, "A statistical framework for estimation of full-chip leakage-power distribution under parameter variations," *IEEE Trans. on Electron Devices (TED),* **54**(11), 2930-2945, Nov. 2007. DOI: 10.1109/TED.2007.906960

[3] Z. Abbas, A. Zahra, M. Olivieri, and A. Mastrandrea "Geometry scaling impact on leakage currents in FinFET standard cells based on a logic-level leakage estimation technique," *Microelectronics, Electromagnetics and Telecommunications,* **521**(4), 283-294, Jun. 2018. doi: 10.1007/978-981-10-7329-8-29

[4] S.-K. Shin, W. Yu, Y.-H. Jun, J.-W. Kim, B.-S. Kong, and C.-G. Lee, "Slew-rate-controlled output driver having constant transition time over process, voltage, temperature, and output load variations," *IEEE Trans. on Circuits and Systems II: Express Briefs (TCAS-II),* **54**(7), 601-605, Jul. 2007. DOI: 10.1109/TCSII.2007.895314

[5] R.-C. Kuo, H.-Y. Tseng, J.-W. Liu, and C.-C. Wang, "On-chip process and temperature compensation and self-adjusting slew rate control for output buffer," in *Proc. Asia Pacific Conf. on Postgraduate Research in Microelectronics Electronics (PrimeAsia),* 37-40, Oct. 2011.

[6] T.-J. Lee, W. Lin, and C.-C. Wang, "Slew rate improved 2×VDD output buffer using leakage and delay compensation," in *Proc. IEEE Inter. Conf. on Electron Devices and Solid-State Circuits (EDSSC),* pp. 1-2, Jun. 2014.

[7] C.-L. Chen, H.-Y. Tseng, R.-C. Kuo, and C.-C. Wang, "On-chip MOS PVT variation monitor for slew rate self-adjusting 2×VDD output buffers," in *Proc. IEEE Inter. Conf. on IC Design Technology (ICICDT),* 1-4, May 2012.

[8] C.-L. Chen, H.-Y. Tseng, R.-C. Kuo, and C.-C. Wang, "A slew rate self-adjusting 2×VDD output buffer with PVT compensation," in *Proc. IEEE VLSI Design, Automation and Test (VLSI-DAT),* 1-4, Apr. 2012.

[9] C.-C. Wang, C.-L. Chen, R.-C. Kuo, H.-Y. Tseng, J.-W. Liu, C.-Y. Juan, "On-chip process and temperature monitor for self-adjusting slew rate control of 2×VDD output buffers," *IEEE Trans. on Circuits and Systems I: Regular Papers (TCAS-I),* **60**(6),1432-1440, Jun. 2013. DOI: 10.1109/TCAI.2013.895314

[10] T.-J. Lee, K.-W. Ruan, and C.-C. Wang, "32% slew rate and 27% data rate improved 2×VDD output buffer using PVTL compensation," in *Proc. IEEE Inter. Conf. on IC Design Technology (ICICDT),* 1-4, May 2014.

[11] C.-C. Wang, T.-Y. Tsai, T.-J. Lee, and K.-W. Ruan, "2×VDD output buffer with 36.4% slew rate improvement using leakage current compensation," *Electronics Letters (EL),* **53**(2), 62-64, Jan. 2017. doi: 10.1049/el.2016.2351

[12] C.-C. Wang, T.-Y. Tsai, and W. Lin, "A 90-nm CMOS 800 MHz 2×VDD output buffer with leakage detection and output current self-adjustment," *Analog Integrated Circuits and Signal Processing (AICSP),* **97**(2), 343-350, Nov. 2018. DOI: 10.1109/AICSP.2018.895314

[13] S. A. Tawfik, and V. Kursun, "Multi-Vth level conversion circuits for multi-VDD systems," in *Proc. Inter. Symp. on Circuits and Systems (ISCAS),* 1397-1400, May 2007.

[14] W.-P. Tu, S.-W. Wu, S.-H. Huang, and M.-C. Chi, "NBTI-aware dual threshold voltage assignment for leakage power reduction," in *Proc. Inter. Symp. on Circuits and Systems (ISCAS),* 349-352, May 2012.

[15] M. Liu, W.-S. Wang, and M. Orshansky, "Leakage power reduction by dual-Vth designs under probabilistic analysis of Vth variation," in *Proc. Inter. Symp. on Low Power Electronics and Design (ISLPED),* 2-7, Aug. 2004.

[16] T. Shirai, and K. Usami, "Hybrid design of dual Vth and power gating to reduce leakage power under Vth variations," in *Proc. Inter. SoC Design Conf. (ISOCC),* 310-313, Nov. 2008. DOI: 10.1109/ISOCC.2008.2736782

[17] C. S. Nagarajan, L. Yuan, G. Qu, and B. G. Stamps, "Leakage optimization using transistor-level dual threshold voltage cell library," in *Proc. Inter. Symp. on Quality Electronic Design (ISQED),* 62-67, Mar. 2009.

[18] M.-D. Ker, and P.-Y. Chiu, "Design of 2×VDD-tolerant I/O buffer with PVT compensation realized by only 1×VDD thin-oxide devices," *IEEE Trans. on Circuits and Systems I: Regular Papers (TCAS-I),* **60**(10), 2549-2560, Oct. 2013.

[19] Y.-H. Kwak, I. Jung, and C. Kim, "A Gb/s+ slew-rate/impedance-controlled output driver with single-cycle compensation time," *IEEE Trans. on Circuits and Systems II: Express Briefs (TCAS-II),* **57**(2), 120-125, Feb. 2010.

[20] T.-J. Lee, T.-Y. Tsai, W. Lin, U.-F. Chio, and C.-C. Wang, "A slew rate variation compensated 2×VDD I/O buffer using deterministic P/N-PVT variation detection method," *IEEE Trans. on Circuits and Systems II: Express Briefs (TCAS-II),* **66**(1), 116-120, Jan. 2019.

[21] T.-J. Lee, T.-Y. Tsai, W. Lin, U.-F. Chio, and C.-C. Wang, "A dynamic leakage and slew rate compensation circuit for 40-nm CMOS mixed-voltage output buffer," *IEEE Trans. on Very Large Scale Integration (VLSI) Systems (TVLSI),* **25**(11), 3166-3174, Nov. 2017. DOI: 10.1109/TVLSI.2017.2736782

[22] T.-Y. Tsai, Y.-Y. Chou, and C.-C. Wang, "A method of leakage reduction and slew-rate adjustment in 2×VDD output buffer for 28 nm CMOS technology and above," in *Proc. Inter. Conf. on IC Design and Technology (ICICDT),* 1-4, Jun. 2016.

ASTES

Advances in Science, Technology and Engineering Systems Journal Vol. 7, No. 3, 95-105 (2022)

www.astesj.com

Special Issue on Multidisciplinary Sciences and Engineering

ASTESJ

ISSN: 2415-6698

A Secure Trust Aware ACO-Based WSN Routing Protocol for IoT

Afsah Sharmin*, Farhat Anwar, S M A Motakabber, Aisha Hassan Abdalla Hashim

Faculty of Engineering, Department of Electrical and Computer Engineering, International Islamic University Malaysia, Kuala Lumpur, 53100, Malaysia

ARTICLE INFO

Article history:
Received: 14 April, 2022
Accepted: 16 May, 2022
Online: 27 May, 2022

Keywords:
Internet of Things
Wireless Sensor Network
Routing algorithms
Ant Colony Optimization
Security
Trust Value
Energy consumption

ABSTRACT

The Internet of Things (IoT) is the evolving paradigm of interconnectedness of objects with varied architectures and resources to provide ubiquitous and desired services. The popularization of IoT-connected devices facilitating evolution of IoT applications does come with security challenges. The IoT with the integration of wireless sensor networks possess a number of unique characteristics, so the implementation of security in such a restrictive environment is a challenging task. Due to the perception that security is expensive in terms of computation, power and user-interface components, and as sensor nodes or low-power IoT objects have limited resources, it is desired to design security mechanisms especially routing protocols that are light weighted. Bio-inspired mechanisms are shown to be adaptive to environmental variations, robust and scalable, and require less computational and energy resources for designing secure routing algorithms for distributed optimization. In IoT network, the malicious intruders can exploit the routing system of the standardized routing protocol, e.g., RPL (The Routing Protocol for Low-Power and Lossy Networks), that does not observe the node's routing behavior prior to data forwarding, and can launch various forms of routing attacks. To secure IoT networks from routing attacks, a secure trust aware ACO-based WSN routing protocol for IoT is proposed here that establishes secure routing with trustworthy nodes. The trust evaluation system, is enhanced to evaluate the node trust value, identify sensor node misbehavior, and maximize energy conservation. The performance of the proposed routing algorithm is demonstrated through MATLAB. Based on the proposed system, to find the secure and optimal path while aiming at providing trust in IoT environment, the average energy consumption is minimized by nearly 50% even as the number of nodes has increased, as compared with the conventional ACO algorithm, a current ant-based routing algorithm for IoT-communication, and a present routing protocol RPL for IoT.

1. Introduction

The IoT (Internet of Things) is an evolving technology that performs a significant role in interconnecting intelligent devices or objects that surround us into a network. Integration of wireless sensor networks (WSNs) and IoT, offer a wide variety of applications domains that contour human life and also have influence on economic benefits. The IoT applications have touched its presence in many spaces, such as smart homes, smart cities, smart grid systems, banking, healthcare, environmental monitoring, transportations, data management and analysis and agriculture etc. The evolution of novel applications, systems, and technologies are intensifying attention from the research perception as well. Fueled by the extensive use of systems of interconnected intelligent objects or things enabled by wireless technology such as radio frequency identification (RFID), Wi-Fi, Bluetooth, embedded sensor and actuator nodes, cell phones, IoT is transforming into a fully integrated future internet from the static internet that would provide autonomous, smart behavior and pervasive communication networks for smart connectivity and context-aware computation. This paper is an extension of work originally presented in ICCCE'21 [1].

The present network protocols for wireless communications become inadequate when it comes to the IoT because of the large upsurge of IoT objects, diversity of continually emerging devices and possessions, and heterogeneity among objects' architectures.

*Corresponding Author: Afsah Sharmin, Faculty of Engineering, Department of Electrical and Computer Engineering, International Islamic University Malaysia, 53100 Kuala Lumpur, Malaysia Tel: +60142217380
Email: afsahsharmin@gmail.com

Especially, when the objects or nodes in the IoT system have limited resources in term of energy, memory and processor and the topology changes due to the mobility of the nodes. There are numerous applications, systems, and services from diverse manufacturers, as well as a wide range of hardware and software requirements, making a comprehensive compliance process for efficient and secure IoT-communication difficult to achieve. Also, the adaptation process of conventional communication protocol to take into account the structural and logical characteristics of the IoT modules is also very challenging.

During IoT routing, which influences and aids the interconnectivity of devices, a crucial deliberation focuses on energy efficiency, secure communication, scalability, computational complexity, autonomy, changed environmental issues, node mobility, resource constraints, and QoS (quality of service) requirements for specific applications. The sensors deployed in an IoT system, are energy-constrained and characterized by their self-organization; they sense, monitor and collect data, and perform computational functions while communicating over wireless networks and lossy channels. Because of the distinctive features of IoT networks, the system is subject to a variety of attacks, and many IoT devices are low-powered and computationally weak, and are not built to address security and privacy issues, a security breach could occur in such a system. Despite the fact that various IoT-specific routing protocols have been designed for providing routing decisions, within satisfying resource consumption, they have not been exhaustively verified for trustworthiness [2]. A secure routing protocol is essential for secure exchange of data with the intended parties rather an attacker and a mechanism is required for the stipulation of predetermined participants or discovering trustworthy nodes to collaborate with. Wireless sensor network which is constructed on autonomous nodes collaboration, plays a vital role in providing ubiquitous computing facilities to the diversification of IoT. There are numerous threats and challenges in the area of communication security, and wireless communication is particularly vulnerable to data exposure. The importance of route security is likewise high because the nodes are spatially scattered over a large area and the base station may be located distant from the information-carrying sensor node or device, requiring multi-hop communication to cooperatively send data over the network to a main location or the sink node for which routing path is necessary [3].

To protect the information in IoT devices, a significant range of secure routing algorithms and security mechanisms, including cryptographic techniques for message integrity, have been proposed by the scientific community. When malicious nodes or internal adversarial nodes or internal compromised nodes are present, the keys exchanged for interactions with the other nodes in the network are compromised as well. Most of the secret keys distribution algorithms are computationally expensive and take additional resources such as large memory space and CPU cycles, and that would result in performance degradation, while making it difficult to distinguish between malicious and non-malicious nodes using solely cryptographic measures. As a result, they are inappropriate for resource-constrained network systems. The notion of providing security is pricy with regard to compute, electrical energy, and user-interface components due to low-powered IoT objects, sensor and actuator nodes. If the encryption keys are accessed by the attackers, the whole network's data could be susceptible to exposure. If the protocol does not take the node's behavior into account throughout the routing process, security attacks like Rank attacks and Sybil attacks can be carried out without difficulty, paving the way for further insider attackers. These attacks can be mitigated by employing trust-aware secure routing protocols.

Existing WSN and IoT routing protocols are unable to adequately set of scales security and energy consumption, resulting in routes that are not globally optimum and might fail to function in the face of malicious attacks, threats and vulnerabilities. Bio-inspired processes offer low-cost options for developing secure routing algorithms that find the optimal path. Furthermore, finding trustworthy neighbors is a critical responsibility. Thus, an accompanying security solution known as trust management has been applied and enhanced [4]. In order to manage the network's highly dynamic topology while preserving energy efficiency during data transfer, various intelligent systems and biological systems, as well as the techniques by which they solve their everyday challenges, are used in the construction of secure routing algorithms. The ant colony optimization (ACO) system is a bio-inspired algorithm that uses the notion of self-organization to aid ants' coordination for solving problems. This technique is notably inspiring for addressing security issues in IoT network routing, as ants create paths that satisfy precise constraints in a graph. Bio-inspired processes are robust, adaptive, and scalable, and they aid in the design of optimal algorithms and distributed systems. The probability formula is utilized for route selection in ACO, which is a probabilistic process, while the pheromone update formula is used for pheromone trail updating [5].

EICAntS (Efficient IoT communications based on ant system) is an ant-inspired routing strategy for optimizing IoT communications that was proposed in [6]. The energy parameter is used in the calculation of the global efficiency factor, which represents the ant colony system's pheromone estimations. This approach extends network lifetime while reducing energy usage. The energy impact concentrates the data class that the node manipulates. The various difficulties afflicting the energy factor, for instance small-scale multi-path fading and large-scale fading, free-space path loss in wireless communications are not indicated here. Furthermore, no precise details for calculating the energy level of the nodes are provided. Three routing metrics, ETX (Expected Transmission Count), load or content, and residual energy, are utilized separately and in combination in [7] to enhance the design of the proactive routing protocol RPL (Routing Protocol for Low Power Lossy Network) objective function (OF), that is used to automate the route development method, for IoT applications. Residual energy (RE) in conjunction with ETX (EE) and an upgraded timer setup is effective for energy consumption. On the other hand, unlike the ant colony based approach, which employs the mechanism at work in ant-colony foraging, there is no optimization model used. Using the principles of rank threshold limitations and hash chain authentication, a secure-RPL (SRPL) protocol is suggested in [8] to minimize the influence of rank manipulation. This technique is seemed to be computationally expensive as it combines cryptography with hash chain authentication. In addition, nodes are vulnerable to insider attacks. In [9], the authors suggested a trust-based threshold method for the selection of a parent node to provide security countermeasures to Rank attacks amid RPL routing. The scheme's benefit is that the attacking node is recognized in the course of selection process of

the parent node, which mitigates Rank attacks. The scheme's downside is that additional susceptible attacks, such as blackhole and Sybil attacks, cannot be identified and alleviated well. The authors proposed approaches for the detection and mitigation of Rank attacks that are inconsistent with RPL-supported IoT in [10]. The node's trustworthiness is not considered by the technique, which leads to further security issues, targeting network traffic and resources.

In RFSN [4] framework, the sensor nodes keep reputation about other nodes in the system. Within this framework, a beta reputation system that uses Bayesian formulation has been employed. Using a watchdog mechanism, a node observers the behavior of other nodes. In this way, their reputation is built over a period which help to evaluate their trustworthiness to collaborate. Also, their future behavior is predicted. Direct and indirect reputation are built up using direct observations and second hand information respectively. The statistical expectation of the probability distribution signifies the reputation, which is used to calculate trust. However, this schema does not include a provision for distributing information about a bad reputation. As a result, it is unable to cope with uncertainty. Secure alternate path routing in sensor network (SeRINS) detects and isolates the compromised nodes by providing key management system along with the neighbor report system where the inconsistent routing information have been injected by those nodes [11]. Here, the compromised node is found out using neighbor report technique. The base station then broadcasts the compromised node's ID, key ring to the entire network and that malicious node is excluded using revocation of its cryptographic keys network-wide. However, the proposed technique needs huge changes to apply in the network as it is majorly embedded in the routing arrangement and neighboring nodes can eavesdrop.

A hybrid tree-based search approach called ANT-BFS is presented in [12] to determine the best and shortest information transmission route in order to enhance network performance. ACO is used in conjunction with breadth first search algorithm to investigate the neighbors so as to identify the solution or the requisite node. The amount of energy used is reduced with this strategy. The execution of BFS in ACO, on the other hand, necessitates more memory and computation time. In [13], quantum computation method is introduced and a new WSN routing algorithm, named the Quantum Ant Colony Multi-Objective Routing (QACMOR), is proposed to monitor in complex manufacturing environments. The node pheromone is characterized by quantum bits and to update the pheromone concentration of the path, the quantum gates are rotated. This method improves convergence performance and saves energy consumption. However, the computational complexity of the algorithms and effects on QoS are not addressed in this technique and need to be considered as QoS is posed by real-time applications. In [14], a routing protocol REL for IoT based on residual energy and wireless link quality estimate is suggested to improve reliability and energy efficiency. It enhances the quality of service (QoS) of IoT applications. To improve protocol reliability, received signal strength indication (RSSI) and signal to noise ratio (SNR) are used to generate the link quality estimate for wireless links. To reduce protocol overhead, an opportunistic piggyback technique is implemented, and the residual energy is transmitted to adjacent nodes to increase energy consumption. Despite this, no better approach is used, as opposed to the ant-inspired routing algorithm, which makes use of the ant-based system.

The proposed system of ours [15] has been analyzed more here to efficiently balance security and energy consumption. The proposed routing algorithm has considered important communication parameters for data transmission in an IoT network, such as mobility and energy parameters. This paper extends the work reported in [1], where a secure bio-inspired routing protocol based on ant colony optimization (ACO) systems is proposed, with the intension of providing trust in WSNs while improvising efficient IoT communications.

The relevant work is introduced in Section 1 and the rest of this paper is organized as follows. The system model and energy consumption model are discussed in Section 2. Section 3 depicts the proposed scheme in detail, including the proposed secure ACO algorithm and its design concept, the trust model used as a security mechanism, and trust assessment. The performance of the proposed system is evaluated in Section 4. Finally, some concluding remarks are provided in Section 5.

2. System Model and the Energy consumption model

The chosen network system is based on an IoT sensor organization in which the deployed nodes, sensors and actuators M_i are dispersed throughout the monitored area at random. The graph G linked with the nodes and symmetrical communication links makes up the system model for an IoT communication network. The energy and computational resources available to the sensor nodes in the region are the same. The received signal strength indication (RSSI) can be used by the nodes to calculate the estimated distance of the transmitters, where the transmission power must be acknowledged. The nodes can adjust transmission power and keep records of their neighbors' information updates. The presented system uses the radio energy model of wireless communications [16] and is implemented utilizing (1) to analyze the energy consumption. The quantity of energy consumed is determined by the distance between transmitting and receiving nodes, the size of the packets, and a distance-threshold value, d_0. The two types of energy consumption models applied here are free-space (the transmission power attenuates inversely proportional to d^2) and multi-path fading (the received power is falling off inversely with d^4) models. The following equations are used to calculate the energy consumption (E_{tran}) by the sensors during the transmission of an m-bit data:

$$E_{tran}(m,d) = \begin{cases} mE_{elec} + m\epsilon_{fs}d^2 & if\ d < d_0 \\ mE_{elec} + m\epsilon_{mp}d^4 & if\ d \geq d_0 \end{cases} \quad (1)$$

where ϵ_{fs} and ϵ_{mp} are the amplifying radio's energy consumption in the free-space and multi-path fading models, respectively. The distance is denoted by d, while the threshold value for the distance is denoted by d_0. Electronic devices' circuitry is powered by energy dissipation, E_{elec}. Now, $E_{Rx}(m)$, the reception energy for an m-bit data for a node, can be calculated as follows:

$$E_{Rx}(m) = mE_{elec} \quad (2)$$

The following equation is used to compute the residual energy (E_{res_i}) of a node n_i :

$$E_{res_i} = E_{tot_i} - E_{tran_i} \quad (3)$$

where E_{res_i} denotes the residual energy, E_{tot_i} represents the total initial energy and E_{tran_i} represents the transmission energy.

3. Proposed Secure Routing Protocol based on ACO

3.1. Proposed improved ACO Algorithm

a) The state-transition formula:

By examining the nodes' stable energy consumption while keeping in consideration the security issues for next-hop routing to determine the most trustable route getting to the node that delivers the certain required provision, an enhanced network routing algorithm based on ACO is proposed here. Hence, the next hop selection by the ants depends upon residual energy level of the neighbor nodes, i.e., the node with greater energy level possessing more probability of being selected higher, and their trust value, i.e., regarding the nodes' high trust value as probabilistic next-hop for routing. Assume if an ant *m* is located at node *i* at time *t*, it will comply to the following probability formula to choose the subsequent node *j* as the information forwarding node of the ensuing route for the proposed improvement of our ant colony optimization based routing algorithm [15]:

$$P_{ij}^{m} = \begin{cases} \frac{[\tau_{ij}(t)]^{\alpha}[\eta_{ij}(t)]^{\beta}[\vartheta_{ij}(t)]^{\gamma}[T_{ij}(t)]^{\psi}E_j}{\sum_{s\subset allowed_m}[\tau_{is}(t)]^{\alpha}[\eta_{is}(t)]^{\beta}[\vartheta_{is}(t)]^{\gamma}[T_{is}(t)]^{\psi}E_s} & , j \subset allowed_m \\ 0, \quad others & \end{cases} \quad (4)$$

$$\text{here,} \quad \eta_{ij}(t) = \frac{1}{d_{ij}} \quad (5)$$

where $\tau_{ij}(t)$ is the amount of pheromone deposited on edge (i, j) and $\eta_{ij}(t)$ is the state transition desirability of edge (i, j). A priori knowledge, typically the heuristic value $\eta_{ij}(t)$ is $1/d_{ij}$ and d_{ij} is the distance between *i* and *j*. There are two impact factors, α and β, that control the influence of the pheromone intensity and heuristic value respectively. In accordance with the average node mobility or speed, the stability factor, $\vartheta_{ij}(t)$, is determined, where γ is the mobility constant. $T_{ij}(t)$ is the high trust value of nodes at time *t* or the trust metric and ψ is the impact factor that control the influence of trust level among the nodes to further communication. The calculation of the trust metric will be provided below. E_j represents the node residual energy that ant *m* would visit.

a) The local update:

After an ant finish mapping a node *i* to node *j*, the corresponding pheromone intensity $[\tau_{ij}(t)]$ is updated by a local pheromone updating rule according to (6). Besides enhancing diversity of the algorithm, local pheromone update is augmented here in case a large quantity of pheromone value is accumulated down the pathways, while preventing faster local convergence. As a result, the pheromone measure is restored and controlled through the use of a threshold rating, and the updating rule which is assessed by following equation:

$$\tau_{ij}(t+1) = \begin{cases} T, & \tau_{ij}(t+1) > T \\ (1\text{-}\rho)\tau_{ij}(t) + \Delta\tau_{ij}(t) & else \end{cases} \quad (6)$$

$$\Delta\tau_{ij}(t) = \sum_{k=1}^{m} \Delta\tau_{ij}^{k} \quad (7)$$

where ρ signifies the local pheromone decay parameter, $\rho \in (0,1)$, a threshold value T is provided to restrict excessive pheromone accumulation, $\Delta\tau_{ij}(t)$ is the appended pheromone deposition of link (i, j), which is typically specified as below:

$$\Delta\tau_{ij}^{k} = \begin{cases} \frac{S}{L_k} & if\ kth\ ant\ travels\ on\ the\ edge\ (i,j) \\ 0 & otherwise \end{cases} \quad (8)$$

where *S* is a constant represents the strength of pheromone, L_k is the cost of the k^{th} ant's tour known as length, and *m* is the number of ants.

3.2. Trust Model used as a Security Mechanism

Trust framework involves two participants: trustors and trustees, who work associatively to accomplish a particular job based on a node's estimated trust value to determine its trustworthiness. A weighed value index, denoted as the node's trust rating, is assessed depending on the judgement of a node's prior behaviour, which also establishes the node's reputation [4]. The trust management approach identifies malicious and other attackers and compromised nodes in the communication network by evaluating their trust value in order to deal with unpredictability about the nodes' future activities. This reflects a node's trustworthiness while engaging with its neighbors, either directly or indirectly, to complete a set of specified activities. Consequently, the behaviour of a node is observed and that defines the positive or negative interactions over time, which is demonstrated as expressions to represent the calculable range of values, such as the trust value rating, while also indicating a node's reputation metric. For secure routing, neighbors or adjacent nodes with greater trust metric values are chosen, whereas nodes having lower values of trust or if trust values of these suspect nodes do not increment with time, then these nodes are identified as malicious. The security mechanism described in [4], manipulating a watchdog mechanism, is utilized and expanded upon here, where a beta reputation system based on Bayesian formulation is applied via direct/indirect observations to signify reputation metric.

The direct trust (DT_{ij}) is the rating of the latest activities of node *j* by node *i*. and it is calculated through the expected value of the probability distribution function. While allowing for consideration of the beta distribution and beta function, as a previous distribution property in the communications between the nodes, yields the direct trust value that node *i* has on node *j*. The direct trust is represented as follows:

$$DT_{ij} = \frac{\acute{\alpha}_j + 1}{\acute{\alpha}_j + \acute{\beta}_j + 2} \quad (9)$$

where $\acute{\alpha}_j$ indicates the successful or cooperative interactions and $\acute{\beta}_j$ indicates the unsuccessful or non-cooperative interactions or interactive behaviors between node *i* and *j* accordingly from the perspective of node *i*.

While an entity can make a precise direct trust judgement based on direct observation without a third party involvement for its adjacent nodes, it relies on the recommendations of trusted nodes to assess trust for packet transmission to nodes that are not directly connected. In case of uncertainty, presume that the evaluating node *i* needs the recommendation from a third entity and acquires reputation rating of node *j*, via their commonly adjoining nodes *k*. According to the principle of trust transfer decline, the recommended trust metric is calculated by following equation:

$$RT_{ij}^{k} = DT_{ik} * DT_{kj} \tag{10}$$

here, RT_{ij}^{k} gives the recommended trust value that node *i* possesses about node *j* offered by the common neighbor nodes *k*, and is derived by the product of the direct trust values, DT_{ik} and DT_{kj}. Accordingly, DT_{ik} represents direct trust rating between nodes *i* and *k*, and DT_{kj} represents direct trust rating between nodes *k* and *j*.

The trust metric $T_{ij} \in [0,1]$ of node *i* holds for *j* is the operational trust rating that is computed by collecting interactive records from third parties through direct observation or indirect observation. The weighted average, an associatory trust aggregation function, is computed by combining the estimates RT_{ij}^{k}, where distinct recommenders' provisions are brought into consideration for computing RT_{ij}^{k} particularly from each trusted edge. The trust metric is given as follows:

$$T_{ij} = \sum_{k \in N_i} \left(RT_{ij}^{k} * w_k \right) \tag{11}$$

$$\text{here, } w_k = \frac{DT_{ik}}{\sum_{k \in N_i} DT_{ik}}, \; k = 1,2, \ldots, N_i. \tag{12}$$

$$DT_{ik} = \frac{\acute{\alpha}_k + 1}{\acute{\alpha}_k + \acute{\theta}_k + 2} \tag{13}$$

where the weight w_k is assigned depending on recommenders' trust levels to lessen the influence of personal choice. In the above equation, $w_k (0 \le w_k \le 1, \sum_{k=1}^{N_i} w_k = 1)$ is the weight of RT_{ij}^{k}. The direct or indirect recommendations for node *j* received by node *i* from a set of trusted nodes denoted as N_i. It also indicates the number of received recommendations that is utilized. DT_{ik} represents the direct trust values between nodes *i* and *k*, while $\acute{\alpha}_k$ and $\acute{\theta}_k$ represent the prior recommendation or reputation metric, successful and unsuccessful interactive records accordingly that node *i* already possesses about node *k*.

In the proposed system, every sensor node manages and controls its own pheromone traces while not adding too much overload to the network, and maintaining the lightness of the model. Moreover, to gather ratings, there is not any central or supervising entity and each transmitted ant carries the sensors' identifications along with the pheromone traces.

On the contrary, a reputation rating depending upon pheromone value, τ_{ij}, of a communication route can be established here where the higher is the pheromone trace, the quality of the path, the higher is the security. Every node saves its own pheromone traces and the pheromone traces for its neighbors. In this sense, a more secured route is with more pheromone deposits, implying that a linked node holds greater packet forwarding or collaborative capabilities. The deterministic factor as well as this pheromone measure, $\tau_{ij} \in [0, 1]$, will determine the probability of ants selecting one path or another and the trust value in association with reputation metric provided by an entity to another node specifies the deterministic factor. If the reputation of a node at time *t* is denoted by $\varphi_{ij}(t)$, then the following equation can be applied for the detection of a malicious node:

$$\tau_{minimum} = \frac{\sum_{i=1}^{n_k} \tau_{ij}(t)}{n_k} \tag{14}$$

where τ_{ij} represents the pheromone quantity in between nodes *i* and *j*, and the number of *i*'s neighbors is n_k. If $\varphi_{ij}(t) < \tau_{minimum}$, which indicates the node's reputation falls below the minimum reputation conditions, $\tau_{minimum}$, then security threat or node's misbehavior is detected, and this node is identified for its malicious tasks, and will have fewer forwarding capabilities.

3.3. Trust Assessment

High-trust level nodes are used for routing decisions or secure communications by the proposed method. During the trust calculation process when the trust values have been determined, a trust assessment system is further adopted for ranking the highest to the lowest trust values, T ([0, 1]). This will further help to detect and eliminate the misbehaving node, where nodes with lower trust values are categorized as malicious. The membership degree and fuzzy classification of nodes' trust are implemented here. Three grades or level of trust have been provided for trust evaluation of a node by using fuzzy judgment as: distrust, uncertain and completely trust level or state which is represented in Table 1. Three fuzzy subsets T1, T2 and T3, as shown in Fig. 1, and the corresponding membership functions are defined as m1 (t), m2 (t) and m3 (t) and m1 (t) + m2 (t) + m3 (t) = 1.

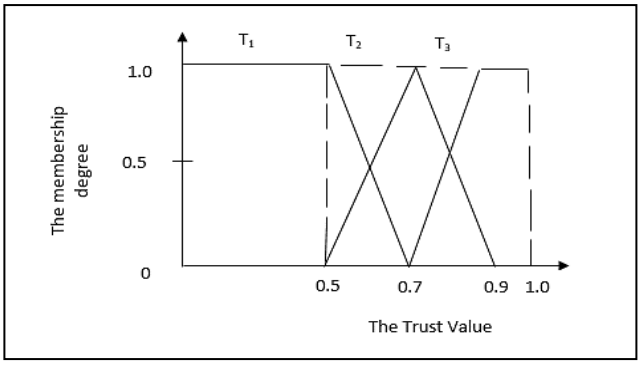


Figure 1: The membership function of node's trust

Table 1: Trust States

Three fuzzy subsets (T)	Trust level
$\boldsymbol{T_1}$	Distrust

T_2	Uncertain
T_3	Completely trust

3.4. The global update and the fitness-function

The global updating of pheromone concentration is worked out after all of the ants have constructed their solutions, finishing their search and have appeared at the target node. In addition, after the search is completed, each ant corresponds to a routing path. To begin, the path estimation rating, which is a route assessment function, is provided as (15) using the existing node energy, route length [m^{th} ant's route length L_m^k in k^{th} iteration], and trust metric. Some nodes will die prematurely if the residual energy [E_{res_i} for a sensor node n_i] is not analyzed as it is in the typical ACO method, reducing the network's overall lifespan. The path's fitness value can be calculated as follows:

$$f_{(fitness)_m^k} = \frac{E_{res_i}}{L_m^k} * T_{ij} \quad (15)$$

here T_{ij} is the trust metric of node i holds for j. Then the global pheromone updating applies on the optimum path which is the best-so-far solution, offering the largest fitness value. The pheromone intensity is updated globally according to the following equation:

$$\tau_{ij}(t+1)=(1-\delta)\tau_{ij}(t)+\sum_{m=1}^{n}\Delta\tau_{ij}^m \quad (16)$$

$$\Delta\tau_{ij}^m = \begin{cases} R * f_{(fitness_{best})_m^k}, & if\ mth\ ant\ visits\ the\ edge\ i,j \\ 0, & otherwise \end{cases} \quad (17)$$

where $0< \delta <1$ is the global pheromone decay parameter, R is the constant for recompensing the pheromone, n denotes the total number of ants, and $\Delta\tau_{ij}^m$ is the increase of pheromone concentration of the edge (i,j) utilized by m^{th} ant, which is proportionate to the maximal cost of fitness equation, $f_{(fitness_{best})_m^k}$, if edge (i,j) is associated with the global best route.

3.5. Proposed Improvement

Clustering is contemplated on attaining scalability while maintaining security. As a result, the routing protocol presented in [15] can be used in conjunction with a clustering based routing approach, such as LEACH [16], a hierarchical clustering protocol. It takes into account the data forwarding probability, nodes' current residual energy, the trust metric, and nodes distance from the base station (BS) and improves the optimal cluster head (CH) selection technique. The flow diagram representing the proposed enhancement is shown in Figure 2.

The probability ($Prob_i$) of a node being selected as a cluster head, an ant m can apply the probability calculation equation given in (18). Node i is presumed to be the present cluster head node then the next node j to be selected as the subsequent cluster head, where the trust metric (T_{ij}), P_{ij}^m , the node distance ($dist_i$), as well as two control parameters (α and β) are used, and the following probability equation is applied:

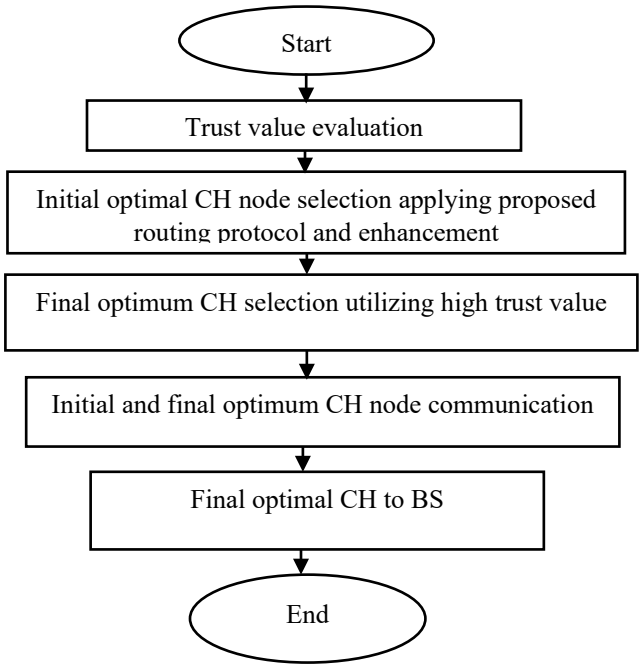


Figure 2: The flow diagram of the proposed improvement

$$Prob_i(t)=\frac{dist_i*\alpha+[P_{ij}^m(t)]*\beta}{\sum_1^{N_i} dist_i*\alpha+[P_{ij}^m(t)]*\beta} * T_{ij} \quad (18)$$

here T_{ij} signifies the trust metric, P_{ij}^m is computed from (4), and N_i is the set of nodes in the cluster.

3.6. Trusted Parent Selection

Algorithm 1: Trust Calculation and selection of trusted parent

Let $M_1 \leftarrow$ any obtainable entity in the Neighbour_List[ ]
Let $M_2 \leftarrow$ another entity next to M_1 in the Neighbour_List[ ]
Calculate

$$T_{ij} = \sum_{k \in M_i} \left(RT_{ij}^k * w_k\right)$$

while node is not found in Malicious_Class_List do
If (M_1.*ETX_metric*<= *ETX_metric-limit*) & (M_2.*ETX_metric*<= *ETX_metric-limit*)
If (M_1.*Rank* <= *Self_Rank*) & (M_2.*Rank* <= *Self_Rank*)
Selected_Parent = $M_1.T_{ij} > M_2.T_{ij}$? M_1:M_2;
else
if (M_1.*Rank* <= *Self_Rank*) || (M_2.*Rank* <= *Self_Rank*)
Selected_Parent = M_1.*Rank* < M_2.*Rank* ? M_1: M_2
else
Selected_Parent = NULL;

end if
else
If (M_1.*ETX_metric* <= *ETX_metric-limit*) ||
(M_2.*ETX_metric* <= *ETX_metric-limit*)
Selected_Parent = M_1.*ETX_metric* <= M_2.*ETX_metric* ?
M_1: M_2;
else
Selected_Parent = NULL;
end if
end while
return Selected_Parent
End. //of program.

The algorithmic procedure implemented here has been given above for selecting the trusted-parents. It includes calculation of the trust values of the nodes and a trust-based method for the selection of parents. The algorithm utilizes the ETX metric as specified in [17]. For the initiation for the optimum parent swap, the minimum required variation of the computed trust value for a node is denoted as $M_1.T_{ij}$. The node having the maximal trust rating along the node's routing path is searched for by the algorithm among all the routes, while the path would also have minimum ETX values, given in (19). The ETX limit represents the maximum ETX rating assessed to be the optimal prospective parent, whereas a node will not select its neighbours that have superior rank as its possible chosen parents. It will also ensure that there is no loop. The trust threshold (Trust assessment Table 1) is utilized for a trusted parent preference, during trust calculation for selecting the node as the chosen parent. Moreover, the rank order is maintained as specified in [18]. Upon identification of a malicious node as a parent, the child node reassigns itself with a different parent from the offered list for selecting a parent node.

The ETX metric, or expected transmission count is calculated as:

$$ETX_{(i,j)} = \frac{1}{D_f * D_r} \tag{19}$$

where D_f defines the forward data delivery and D_r is the reverse data delivery or acknowledgement from the receiver.

4. Result and Discussion

MATLAB is used to accomplish the performance evaluation and simulation. The routing protocol proposed here is compared to the benchmark protocols, where the conventional ACO algorithm, EICAntS algorithm [6], a current ant-based routing method for IoT communication, and a present proactive routing protocol for low power lossy network (RPL) [7] for IoT have been considered as benchmark protocols. There are 100 nodes dispersed in a $100m \times 100m$ area. Some malicious nodes are also deployed across the network at random. The initial trust value is calculated which is set as 0.6, observing the number of interactions, and for that taking reasonable value is crucial. The simulation parameters are set as: $\alpha = 1, \beta = 1, \gamma = 1, \rho = 0.05, \delta = 0.05$. More parameters are presented in the Table 2.

Table 2: Simulation Parameters

Parameters	Values
T	100
Initial trust value	0.6
Initial energy per node	0.5 joule
Node-speed	2 m/s ~ 5 m/s
Transmitted message bits	4000 bits
Distance of transmission	50 m

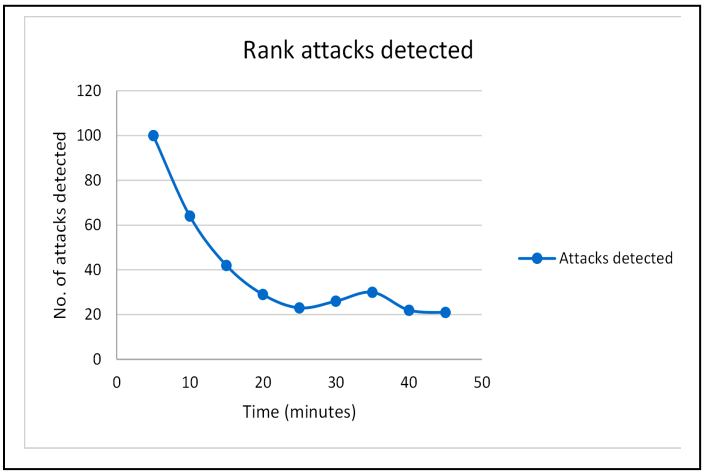


Figure 3: Rank attack detected by proposed secure ACO algorithm

In Trust calculation process, by using the computed trust value, a trustor node assesses a trustee node. It employs the trust metric value to evaluate whether the trustee node is adequately reliable enough of fulfilling an allotted task, and the trust threshold system (Trust assessment Table 1) is utilized for the assessment. Each node collects the direct trust value of directly connected neighbours and recommended trust value of indirectly connected neighbours. The focus of this research is on detecting and isolating internal attacks, particularly Rank and Sybil attacks. A malicious node modifies its rank in a rank attack for disrupting the network route topology, whereas a Sybil node, by using fake identities, tries to subvert the network process. The malicious nodes, taking part in internal attacks of the network, are more challenging to detect as they are aware of the system information of the network. By using node overhearing and monitoring methods, this secure trust-based system perceives unusual route transmission towards a node, and that might be an indication of a rank attack. By assigning a greater weight to a node's existing trust value, a Sybil attack node is detected and isolated. It is also needed not to attribute its observed prior behavior too much weight while defending against a Sybil node as its initial behaviour might be well. So, if it does not have any worthy packet sending behavior that can be observed, its trust value will remain below the threshold, which is necessary for secured communication.

For the simulation study, in the phase of implementing Rank attack, a malicious node initially keeps up with a fine prior behaviour for roughly 5 to 10 seconds. After that during every cycle, it broadcasts spuriously low Rank values and initiates its attack.

From Figure 3, it can be seen that the proposed secure routing protocol is effective at detecting and isolating the Rank attacks. In the course of routing operations, about 100 attacks have been

detected in the first five minutes. Although with the simulation progress, the number of attacks detected has steadily decreased.

In RPL routing operation, a node examines potential parents that have lower rank values than itself and then selects as its chosen parent. In this way, the rank of a node changes and realignment takes place for a child node to another selected parent node that has a smaller rank value. A Rank attack proceeds where the attacker takes advantage of this attribute in RPL routing. It presents itself with a superior rank value to its adjacent nodes and the neighbours are attracted and deceived by this.

The frequency of node rank changes is shown in Figure 4. From the comparison in between MRHOF-RPL (Minimum Rank with Hysteresis Objective Function-RPL) [18] and the secure system presented here, it is observed that the benchmark protocol has notably higher vulnerability to node rank changes than the proposed algorithm, demonstrating a vulnerability to Rank attacks. However, this proposed scheme, persistently has maintained low frequency of node rank changes during all of the simulation period.

From Figure 5, it can be observed that the proposed secure routing protocol is effective at detecting and isolating the Sybil attacks. During the routing procedures, about 272 attacks have been detected in the first five minutes but the number of attacks detected have decreased with time.

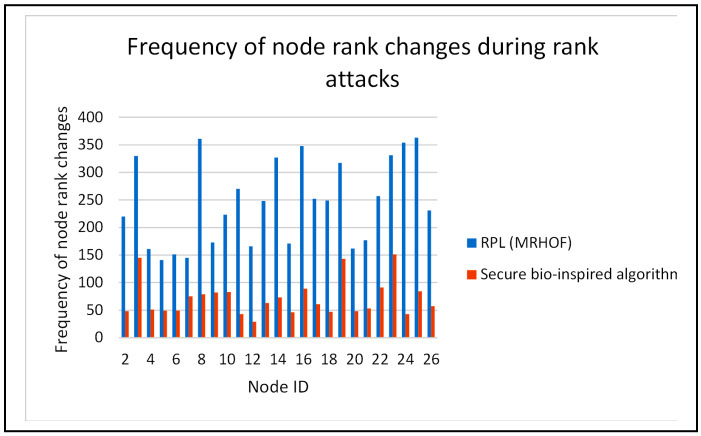


Figure 4: Frequency of node rank changes comparison

The offered ant colony metaheuristic-based routing method is used to discover the optimum pathway for routing packets from the originating node to the target node with the least amount of energy consumption and the highest level of security. The trustworthy nodes are chosen for data transfer in order to build a secure routing path. The graph in Figure 6 shows the relationship between the detection times of a malicious node and the number of nodes in the network. The detection time is defined as the number of malicious nodes found in relation to the simulation time. The percentage of malicious nodes has been retained fixed in this graph, and the value 1 for detection time indicates that no malicious nodes have been found.

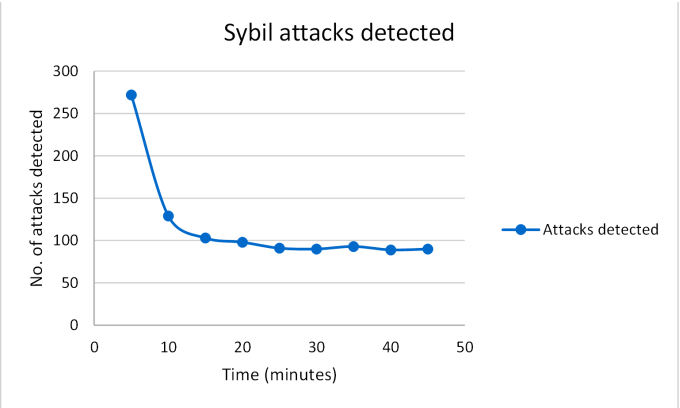


Figure 5: Sybil attack detected by proposed secure ACO algorithm

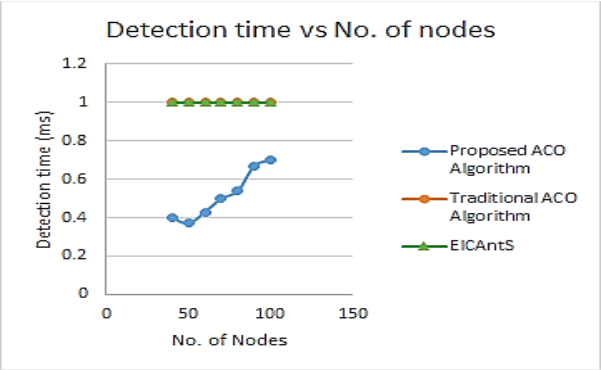


Figure 6: Detection times comparison of a malicious node to the No. of nodes.

The detection of malicious nodes in the network is not considered by the traditional ACO algorithm and EICAntS protocol. As a result, the systems fail to detect any malicious nodes, resulting in lower security and performance. However, the proposed scheme not only converges into the best-so-far path but also the most secure route by taking into account essential transmission factors along with the use of trust to improve security. With the detection and isolation of malicious nodes, it outperforms benchmark protocols while discovering and collaborating with trustworthy nodes via utilizing a trust assessment system. Figure 7 shows the comparison of a malicious node's detection times represented on the ordinate to the percentage of malicious nodes represented on the abscissa accordingly.

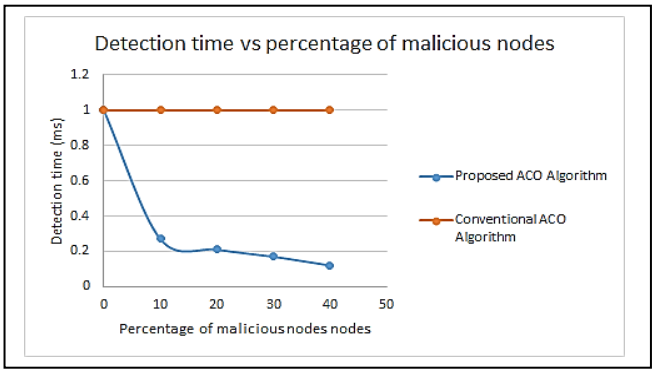


Figure 7: Detection times comparison of a malicious node to the percentage of malicious nodes.

Figure 8 shows a contrast of the consumed energy per transmission of the nodes for the protocol proposed here, the ant-based routing method for IoT communication, i.e., EICAntS, the standard ACO algorithm, and the RPL protocol, demonstrating that the suggested ACO algorithm is better in terms of consuming a lesser amount of energy. The cumulative energy consumption of each node is displayed here every transmission for each individual search operation. In comparison to the previous benchmark protocols, the suggested calculation has clearly achieved refinement, resulting in a substantially lower energy consumption, nearly 50% less for the majority of nodes.

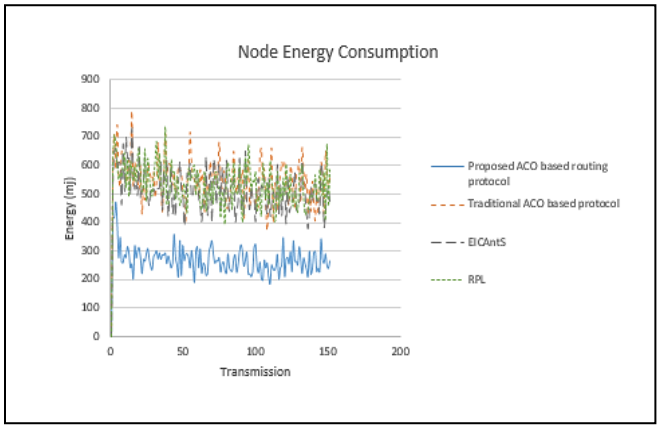


Figure 8: The energy consumption per transmission of the nodes comparison

The presented routing technique based on ACO consumes less energy compared to the traditional ant colony optimization metaheuristic algorithm, proactive routing protocol for low power lossy network (RPL), and efficient IoT communications based on ant system (EICAntS) routing protocol, even in the situations where the number of nodes increases, as shown in Figure 9, where the average energy consumption is lower. When contrasted to the benchmark routing techniques, it is clear that using the suggested ACO-based routing algorithm as an explication reduces average energy consumption, by approximately 50% less and makes the algorithm lightweight. Because the proposed approach enables the optimum packet forwarding path for transmission to be determined, and retransmissions are avoided, providing reliable communication. This method likewise reduces the number of updating phases while optimizing the route selection strategy. According to the outcomes, the more nodes there are, the higher the energy consumption. Another result is that the more malicious nodes there are, the more energy is consumed. The presented framework retains scalability by using less power than the standard protocols taken as the benchmark, even as the number of nodes in the network expands.

As demonstrated in Figure 10, the average End-to-end delay performance metric rises when the number of nodes grows. The proposed routing protocol lessens the repetition issue while also enhancing the procedure for selecting a route because numerous packets have to be sent again to the intended destination if the optimal path is not found and utilized to deliver the packets. Compared to the mentioned benchmark algorithms here, the suggested technique performed well with regard to average End-to-end delay, achieving a nearly 40% decrease in end-to-end delay. Figure 11 shows the throughput results. The network throughput is measured by calculating the total number of packets sent over the complete simulation time, or the measure of digital data transmitted per time unit via a communication link. It is usually expressed in bits per second (bps), although it can also be expressed as data packets delivered per-second or per-time-slot. From the contrast, it is clearly shown that the results obtained by the method proposed here are better than the results recorded by the other network.

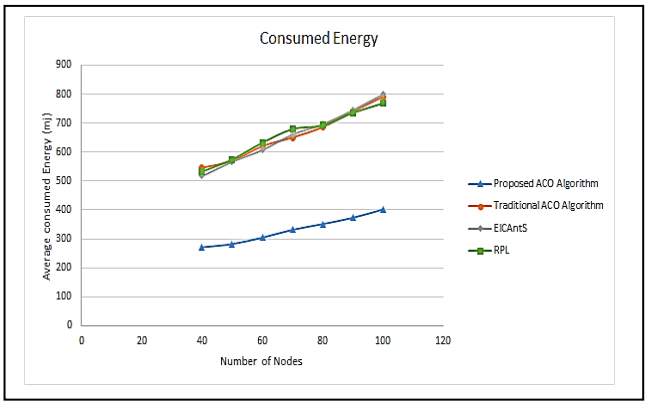


Figure 9: The average energy consumption comparison.

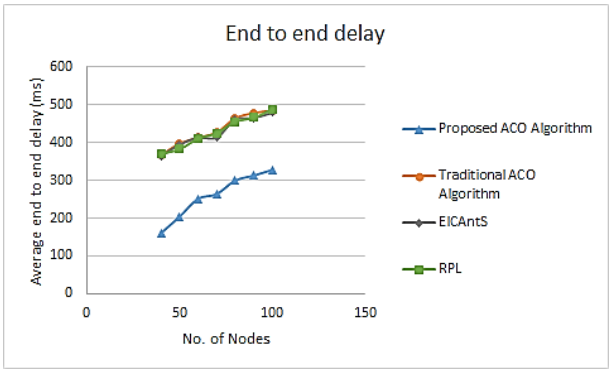


Figure 10: Comparison of the average End-to-end delay with regard to the number of nodes using fitness function

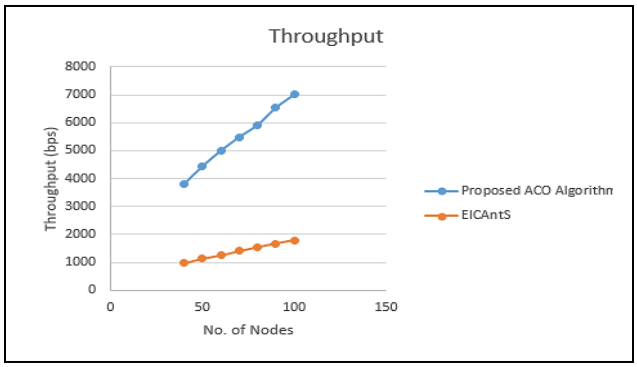


Figure 11: Comparison of Throughput

The findings of the proposed algorithm's calculation of packet delivery rates are shown in Figure 12. Since the protocol devises the ability to deliver numerous packets shortly while also reaching the destination, the proposed approach achieved satisfactory outcomes despite the crucial node quantity. It offers a system for determining the most secure information-transmission path among the network's several routes. Many packets are diverted or

dropped out when there are malicious nodes utilizing other strategies. However, due to security mechanisms, the proposed technique is used to deliver most of the data.

The packet loss ratio is also seen in Figure 13 when malicious nodes are present. The packet delivery ratio diminishes when the percentage of nodes that are malicious rises or as other attackers and compromised nodes exist in the pathways amid communication nodes. When attackers and compromised nodes cannot be detected and packets outreaching the target node successfully decline, the ratio of packet loss or misdirection is significant in the case of the specified benchmark methodologies.

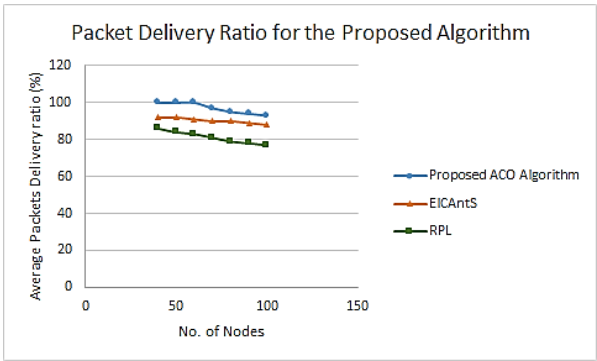


Figure 12: Packet Delivery Ratio (PDR) with the number of nodes

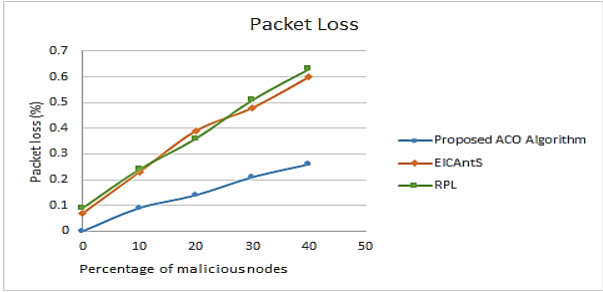


Figure 13: Packet Loss in the existence of malicious nodes

5. Conclusions

Though the IoT technology would be evolving in the coming decade, its multiple and complex aspects need to be considered in the process of developing effective communication protocols. An ACO-based WSN routing algorithm for IoT is proposed in this paper, which uses a trust-based security system while considering the limited resources restraints in sensors or low-power IoT objects, as well as the special necessity of security in the data forwarding process. The trust value is worked out to determine a node's trustworthiness for packet transfer. In order to evaluate route performance, the proposed enhancement and the route assessing function include the trust metric, as well as the existing energy of the nodes and the route length. The energy factor, the trust metric, and the average mobility of the nodes are all included in the ACO algorithm's probability formula as well. When compared to the benchmark methods, the presented ACO-based routing algorithm lowered energy consumption by almost 50% even as the number of nodes rose, making the algorithm lightweight and scalable. It also showed a nearly 40% reduction in end-to-end delay. The routing protocol generates a secure and globally optimal route based on the related information, which includes the neighboring nodes' trust value and residual energy, as well as the path cost from the adjacent node to the sink node. The proposed technique can retain a higher packet delivery ratio due to the security mechanism, which ensures the system's efficacy in addition to the global optimization. Furthermore, by providing trustworthy routing paths, the proposed routing protocol can efficiently balance energy consumption and security.

As future work, presented secure routing protocol would be improved to implement in a real-world setting to estimate the algorithm's performance. Moreover, it will be elaborated to deal with additional conspiring attacks such as a Rank attacking node colluding with Selective Forwarding attacks or having collusion with a Blackhole or a Sybil attack. Finally, previously trusted nodes will be re-assimilated based on their trust levels after having recouped their battery power. These nodes will be deployed administratively to ensure network's balanced secure communication.

References

[1] A. Sharmin, F. Anwar, S.M.A. Motakabber, A.H.A. Hashim, "Secure ACO-Based Wireless Sensor Network Routing Algorithm for IoT," in Proceedings of the 8th International Conference on Computer and Communication Engineering, ICCCE 2021, 190-195, 2021, doi:10.1109/ICCCE50029.2021.9467223.

[2] J. Granjal, E. Monteiro, J. Sa Silva, "Security for the internet of things: A survey of existing protocols and open research issues," IEEE Communications Surveys and Tutorials, **17**(3), 1294– 1312, 2015, doi:10.1109/COMST.2015.2388550.

[3] W. Dargie, C. Poellabauer, Fundamentals of Wireless Sensor Networks: Theory and Practice, 2011, doi:10.1002/9780470666388.

[4] S. Ganeriwal, M.B. Srivastava, "Reputation-based framework for high integrity sensor networks," in Proceedings of the 2004 ACM Workshop on Security of Ad Hoc and Sensor Networks, SASN'04, 66–77, 2004, doi:10.1145/1029102.1029115.

[5] L. Bianchi, M. Dorigo, L.M. Gambardella, W.J. Gutjahr, "A survey on metaheuristics for stochastic combinatorial optimization," Natural Computing, 8(2), 239-287, 2009, doi:10.1007/s11047-008-9098-4.

[6] S. Hamrioui, P. Lorenz, "Bio inspired routing algorithm and efficient communications within IoT," IEEE Network, **31**(5), 74-79, 2017, doi:10.1109/MNET.2017.1600282.

[7] S.S. Solapure, H.H. Kenchannavar, "Design and analysis of RPL objective functions using variant routing metrics for IoT applications," Wireless Networks, **26**(6), 4637–4656, 2020, doi:10.1007/s11276-020-02348-6.

[8] G. Glissa, A. Rachedi, A. Meddeb, "A secure routing protocol based on RPL for internet of things," in 2016 IEEE Global Communications Conference, GLOBECOM 2016-Proceedings, 1-7, 2016, doi:10.1109/GLOCOM.2016.7841543.

[9] I. Kenji, T. Matsunaga, K. Toyoda, I. Sasase, "Secure parent node selection scheme in route construction to exclude attacking nodes from RPL network," IEICE Communications Express, 299–303, 2015, doi:10.1587/comex.4.340.

[10] R. Stephen, L. Arockiam, "E2V: Techniques for Detecting and Mitigating Rank Inconsistency Attack (RInA) in RPL based Internet of Things," in Journal of Physics: Conference Series, **1142**(1), 012009, 2018, doi:10.1088/1742-6596/1142/1/012009.

[11] S.B. Lee, Y.H. Choi, "A secure alternate path routing in sensor networks," Computer Communications, **30**(1), 153–165, 2006, doi:10.1016/j.comcom.2006.08.006.

[12] R. Khoshkangini, S. Zaboli, "Efficient Routing Protocol via Ant Colony Optimization (ACO) and Breadth First Search (BFS)," International Conference on Internet of Things (IThings 2014), (March), 375–381, 2014, doi:10.1109/iThings.2014.69.

[13] F. Li, M. Liu, G. Xu, "A quantum ant colony multi-objective routing algorithm in WSN and its application in a manufacturing environment," Sensors (Switzerland), **19**(15), 3334, 2019, doi:10.3390/s19153334.

[14] K. Machado, D. Rosário, E. Cerqueira, A.A.F. Loureiro, A. Neto, J.N. de Souza, "A routing protocol based on energy and link quality for internet of things applications," Sensors (Switzerland), **13**(2), 1942–1964, 2013, doi:10.3390/s130201942.

[15] A. Sharmin, F. Anwar, S.M.A. Motakabber, "Energy-Efficient Scalable Routing Protocol Based on ACO for WSNs," 2019 7th International

Conference on Mechatronics Engineering, ICOM 2019, 1-6, 2019, doi:10.1109/ICOM47790.2019.8952053.

[16] W.B. Heinzelman, A.P. Chandrakasan, H. Balakrishnan, "An application-specific protocol architecture for wireless microsensor networks," IEEE Transactions on Wireless Communications, **1**(4), 660-670, 2002, doi:10.1109/TWC.2002.804190.

[17] J. Vasseur, M. Kim, K. Pister, N. Dejean, D. Barthel, "Routing metrics used for path calculation in low power and lossy networks," Draft-Ietf-Roll-Routing-Metrics, 2011.

[18] T. Winter and P. Thubert "RPL: IPv6 Routing Protocol for Low power and Lossy Networks," IETF Internet-Draft, 2010.

ASTES

Advances in Science, Technology and Engineering Systems Journal
Vol. 7, No. 3, 106-115 (2022)
www.astesj.com
Special Issue on Innovation in Computing, Engineering Science & Technology

ASTES Journal
ISSN: 2415-6698

Secured Multi-Layer Blockchain Framework for IoT Aggregate Verification

Ming Fong Sie, Jingze Wu, Seth Austin Harding, Chien-Lung Lin, San-Tai Wang, Shih-wei Liao*

Department of Computer Science and Information Engineering, National Taiwan University, Taipei, 106, Taiwan

ARTICLE INFO

Article history:
Received: 27 March, 2022
Accepted: 25 May, 2022
Online: 04 June, 2022

Keywords:
IoT
Blockchain
ID-based Signature
Aggregate Verification
Digital Provenance
Supply Chain Management

ABSTRACT

Technologies designed for digital provenance, especially the Internet of Things (IoT) and blockchain, may allow for security, transparency, and traceability in the global supply chain. However, upstream nodes in the supply chain that work for large-scale production suppliers are not considered. In addition, most IoT blockchain systems adopt an ID-based signature scheme that may affect the efficiency of IoT devices. We propose using aggregate verification to improve the security and efficiency of ID-based verification, reduce network traffic on the blockchain, and transfer computing overhead to aggregator nodes. This paper implements a multi-layer blockchain for Agriculture 4.0 supply chain management that has higher efficiency, effectiveness, and security in comparison to conventional blockchains. We design a Multi-Layer Aggregate Verification (MLAV) solution to improve supply chain management with IoT Blockchain for Agriculture 4.0 through the following methods. First, we use a multi-layer IoT blockchain system to reduce Ethereum gas fee. Second, we design an ID-based Aggregate Verification scheme, thereby eliminating the certificate management cost in the traditional Public Key Infrastructure (PKI) and reducing bandwidth and computation time requirements. Third, we implement a three-layer blockchain infrastructure. In Layer 1, IoT devices sense and upload data to the system's database; in Layer 2, smart contracts execute aggregate ID-based signature verification from IoT devices and upload the transactions to the private blockchain; in Layer 3, a batch converts the layer 2 data and uploads its Merkle root to Ethereum, thereby reducing the required gas fee.

1 Introduction

This paper is an extension of a work originally presented in BRAINS 2021 [1] that uses a multi-layer architecture designed to facilitate smallholders in joining an agricultural blockchain infrastructure. We use aggregate verification to solve the efficiency bottleneck of ID-based signature verification. We also lay out the framework for distributed supply chain management for access control with smart contracts to allow the smallholder to gain access to loans.

With the ushering in of Agriculture 4.0, in recent years there has been widespread adoption of technologies such as the Internet of Things, big data, artificial intelligence, cloud computing, and remote sensing [2]. Agriculture 4.0, also known as digital farming or smart farming, has been brought about by combining telematics and data management with known precision agriculture concepts. These changes have improved the accuracy and practicality of farming operations [3].

However, regarding data management, food safety, and quality monitoring of the agricultural supply chain, Agriculture 4.0 still has significant shortcomings. Particularly in the COVID-19 pandemic, it is essential to prevent cross-contamination and food pathogen outbreaks. The agricultural supply chain should have data transparency, and there should exist a high level of traceability from source to consumer. The World Government Summit in 2018 pointed out that Agriculture 4.0 will need to focus on both the demand side and the value chain (supply) side of the food equation to use technology to meet the real needs of the consumer and to re-engineer the value chain [4]. The tamper-proof property and transparency of blockchain technology allow it to meet these requirements effectively. IoT Blockchain is highly suited for ensuring traceability and consistency of information generated by IoT devices.

Large companies have already begun to adopt distributed supply chain management systems with blockchain technology. In 2017, Walmart established the Walmart Food Safety Center in Beijing and invested US$25 million to use IBM's blockchain solutions to build a global food safety system [5] already tracking 1,500 items on the supply chain blockchain in 2021. Proof of Concept (PoC) and blockchain pilot projects have been established in the United

*Corresponding Author: Shih-wei Liao, Email: liao@csie.ntu.edu.tw

States and China for two products: mango slices and fresh-cut pork products [6]. The cost of supply chain management systems is very high, and there are restrictions on their scalability; therefore, IoT blockchain systems are only suited for large companies. We believe that the benefits brought about by IoT Blockchain should serve not only high-level large-scale food and agricultural suppliers but also grassroots farmers and food producers in providing customized platforms and advice. Blockchain is becoming ever-more democratized and is decentralized in nature; it is therefore highly suited to providing equal opportunity for traditionally disadvantaged entities.

Therefore, the contributions of this paper can be outlined as follows. An Agricultural Supply Chain Finance operational procedure is created with smart contracts that define the rules and functions of three types of nodes: farmer, distribution channel, and financial institution. A corresponding IoT blockchain system is then designed and implemented to record valuable data on both production activity and order history of farmers. There are two types of IoT devices in this system: farming sensors and mobile phones. The farming sensor is a custom-designed piece of hardware that is used in an agricultural setting; for mobile phones, a custom-designed Android app is developed that connects to our IoT blockchain system. Batch verification is leveraged by our IoT Blockchain system to increase ID-based signature verification.

The remainder of this paper is structured as follows: In Section II, we summarize related works. In Section III, we discuss the benefits of using our multi-layer blockchain. Section IV defines and details the aggregate verification algorithm we use. In Section V we describe the blockchain management framework and system architecture. In Section VI, we propose the implementation and evaluation. In Section VII, we have our conclusion.

2 Related Works

2.1 Distributed Ledger Technology

Distributed ledger technology (DLT) is a shared transaction ledger technology that can store, distribute, and exchange certifications publicly or privately between peer entities. In the context of the blockchain, a distributed ledger records transactions between participants and nodes. This data can be duplicated and synchronized across decentralized peer-to-peer networks with consensus algorithms. DLT may then facilitate the flow of information between nodes and help to resolve inefficiencies relating to information asymmetry[7] [8].

Blockchain technology is a specific type of DLT that was developed in 2008 for the implementation of the cryptocurrency Bitcoin [9]. In the Bitcoin network, by leveraging the Proof of Work (PoW) consensus mechanism, blocks are added to a linearly growing, chronologically ordered blockchain. Each block contains the timestamp, transaction data, and hash value of the previous block.

2.2 Ethereum Layer 2 and Smart Contract

Our platform uses the Ethereum blockchain, an open-source, public blockchain structured around a decentralized Ethereum Virtual Machine (EVM) that may process smart contracts [10]. Data agreements are reached via the Proof of Work consensus algorithm. Ethereum 2.0 is expected to switch to the Proof of Stake (PoS) consensus algorithm, which may significantly reduce computational resources wasted during mining and prevent attacks from application-specific integrated circuits (ASIC) [11]. In addition, Ethereum layer-2 technologies such as Arbitrum could collect transactions off-chain and batch it on-chain, scaling up transaction speed, improving privacy, and holding EVM compatibility by using Arbitrum Virtual Machine (AVM), all while still benefiting from layer-1 security [12].

The concept of smart contracts was first proposed in [13]. Its original idea was centered around a ”computerized transaction protocol that executes the terms of a contract.” Developers began integrating this innovative concept into the blockchain environment. The idea of a smart contract has become ”executable code that runs on top of the blockchain to facilitate, execute, and enforce an agreement between untrusted parties without the involvement of a trusted third party” [14].

2.3 IoT Blockchain in Agriculture 4.0

IoT Blockchain is a new technology that integrates the Internet of Things with blockchain technology. It features the following advantages [15] in Agriculture 4.0:

- Transparency for participating companies [16]: New levels of transparency and visibility are essential for improving product traceability and ensuring product authenticity and legality [17, 18].
- Food safety and quality monitoring [19]: A real-time food tracking system built on blockchain technology provides an information platform that enables all supply chain members to access all information, thereby providing openness, neutrality, and reliability for the food supply chain[20].
- Promoting the digitization and disintermediation of the supply chain: Blockchain reduces verification and transaction costs by eliminating intermediaries [21]. By replacing trade financing (banks acting as financial intermediaries) with a blockchain platform, processing time may be reduced from between 7 and 10 days to between 1 and 4 hours [22].
- Improving data security of information sharing: Centralized databases may be prone to data loss or have data that is difficult to retrieve [23]. All data in the blockchain is immutable because the order of transactions is stored in chronological blocks and broadcast to all nodes [24]. The stored data is tamper-proof because updating and deleting of transactions is determined by the consensus mechanism[25, 26].
- Agricultural Finance: Blockchain technology can implement fast, real-time payments for agricultural financial services that increase cash flow and working capital while reducing transaction costs and risks [27].

Ethereum is the first Turing-complete blockchain framework that enables smart contract integration [10]; given enough time and memory as well as the necessary instructions, smart contracts can solve any computational problem no matter the complexity. There

are numerous advantages of smart contracts on the blockchain over traditional contracts.

- The content of the contract is open-source, transparent, and tamper-proof. The immutable smart contract code is guaranteed to execute, which reduces the occurrence of fraud.
- Higher efficiency: using a programming language, there are almost no misunderstandings or disputes, and consensus may easily be reached.
- No third-party arbitration is required. The system automatically executes according to the smart contract, thereby reducing time and verification costs.

2.4 IoT Blockchain

Massive amounts of data are generated from IoT devices and stored in the cloud. It is also an emerging security concern. Blockchain-based IoT systems might solve this security problem in distributed infrastructure [28]. The user could integrate with PKI to access data from IoT devices and interact with blockchain miner nodes to keep track of every transaction on chain [29]. In [30], the author introduce multi-layered network architecture by defining the IoT device layer, router layer, cloud compute layer, offers an authentication framework, and reducing IoT network burden also improves transaction throughput and security.

2.5 Supply Chain Finance

This research focuses on the agricultural supply chain, which is highly competitive. Large-scale downstream enterprises hold an advantageous position. They place heavy requirements on upstream suppliers for purchases, prices, and payment conditions, causing massive pressure on upstream suppliers. Most of these upstream suppliers are small-scale enterprises or farmers; it is difficult for suppliers to obtain loans from financial institutions, resulting in limited funding for upstream entities. A new set of modes has been developed to solve this problem: Agricultural Supply Chain Finance. Agricultural Supply Chain Finance is a proposed framework for financial support for upstream suppliers that promotes the establishment of long-term strategic, synergistic relationships between upstream suppliers and major distributors and improves the competitiveness of the entire agricultural supply chain.

The lack of mutual trust in the agricultural supply chain stems from one primary source: agricultural production suppliers rarely keep financial records. Since most agricultural operations are tax-exempt, farmers neither pay taxes nor report taxes. Therefore, most farms in Taiwan do not have records of either production or sales. As a result, when applying for loans from financial institutions, it is difficult for these agricultural production suppliers to provide sufficient financial data; coupling this with the lack of sufficient existing credit information, financial institutions are often unwilling to lend to these suppliers.

2.6 Aggregate Verification

Aggregate verification is a computationally efficient method for verifying a large number of digital signatures quickly. It is more computationally efficient than individual verification of each signature. With large quantities of data rapidly generated by a variable number of IoT devices, we adopt an aggregate verification method that may be used even in cases of high traffic in the network.

Among the most important considerations in implementing blockchain technology is the measurement of time between uploading data and attaining immutability (confirmation time). For systems with a large number of nodes frequently generating new data, aggregate verification may be implemented to efficiently verify many signatures simultaneously in one action [31]. One such example of this form of implementation may be in IoT devices which require periodic firmware and software updates. In this case, aggregate verification may be implemented for future-proofing the devices by evolving to internet threats, fixing functionality by releasing firmware updates [32]. Upon release of updated firmware, a group of distributed IoT devices may have an update securely and remotely installed automatically rather than being individually, manually installed by the owner of each device.

In the case of immediate verification of items uploaded to the blockchain, some blockchains may choose to require fewer block confirmations for signature verification, thereby compromising security [33]. Therefore, in addition to the aggregate verification case described above, a new aggregate verification application has been proposed in [31] in which a system may need to verify a smaller number of signatures with high efficiency and minimized confirmation time. For example, an IP camera surveillance system may have cameras with a framerate of 15 FPS, with each device generating approximately 1.25 million images in one day. Digital signatures may be used to quickly verify large amounts of pictures and uploaded to the blockchain in large clusters.

3 Benefits of Multi-Layer Blockchain

This section describes the advantages of using our multi-layer blockchain on the system level and on the framework level.

3.1 Multi-Layer Blockchain System

With the application of blockchain technology in the supply chain, corporations and large upstream manufacturers simplify collecting information and production-related data for all transactions in the supply chain (e.g. IBM's global food safety system), assisting institutions in strengthening risk management and control. However, the disadvantaged small upstream producers do not benefit from blockchain technology. Low-income smallholders who operate without a working contract are among the highest upstream entities in the agricultural supply chain; they and the information related to their production activities are essentially nonexistent or anonymous on the supply chain. As participants in the supply chain, they have limited or no access to the information in the blockchain system.

Following our previous work [1], we create a platform to facilitate these small upstream entities in joining the blockchain system; the overview of our platform is shown in Fig. 1. We propose the Certificateless verification process on a client node based on aggregate verification, identity-based cryptography, and zero-knowledge cryptography, which resolves the critical escrow problem and secures the

proof. The Key Generation Center (KGC) constructs only the partial private key of the IoT device. The IoT device generates the entire private key by choosing a piece of secret information and combining it with the partial private key constructed by the KGC. The system parameters published by KGC compute the corresponding public key of the IoT device, and the IoT's secret information is chosen by itself. In addition to our multi-layer platform, we also develop an Android app. The outermost layer of our platform is the Database layer which allows app users to record and read production-related data. The Database layer also provides storage for IoT-generated data for large manufacturers. In order to ensure the tamper-proof property and connectivity of all product information, the Smallholder Node and the Aggregator Node upload information to the Quorum chain in the form of a hash. As Quorum is a permissioned blockchain and lacks information transparency, the data is then uploaded to Ethereum for public access [34]. In order to reduce the gas fee and increase the blockchain recording rate in Ethereum, we convert the data for each product unit into Merkle tree format and only upload the Merkle root on Ethereum. The detailed technical process description may be found in Section V.

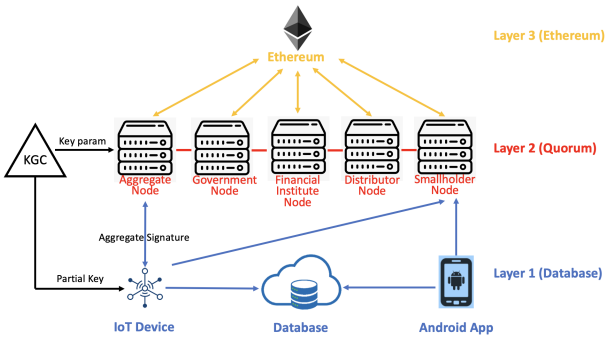


Figure 1: Multi-layer Platform: Blockchain View

3.2 *Supply Chain Management Framework Based On Smart Contracts*

When operating blockchain-based supply chains, the blockchain system must define a set of rules to determine the roles, responsibilities and read-write permissions of participants, which involves maintaining a distributed, authenticated, and synchronized ledger of transactions [35, 36]. Traceable transactions with roles could ensure the stability of the supply chain system [36].

Our supply chain framework is designed to benefit smallholders equally; a blockchain system with smart contracts clearly defines the different rules, responsibilities, and relations of the three types of nodes: Upstream Producer, Financial Institution, and Distribution Channel. The detailed operation process description may be found in "Operation Procedure" in Section V.

With the multi-layer blockchain framework, small upstream entities may upload to the blockchain and access their production activity records, contracts, and transaction records while also solving the largest hurdle: financing. In the past, due to factors such as low fixed assets, low equivalent collateral, and no access to personal financial information, it has proven difficult for smallholders to gain the trust of financial institutions or to obtain any form of funding channel. Through the application of IoT Blockchain, the upstream producer's production activity data and historical order information may be provided to third-party entities for evaluation. This ultimately provides a means by which they may gain trust and become eligible for loans from financial institutions.

Blockchain is increasingly being viewed as a viable solution for alleviating the complexity of global supply chains [20, 37]. The solutions we provide may also be applied to international trade and global supply chains. Addressing the eight technical and non-technical challenges summarized by the OECD [38], our implementation framework and blockchain system may realize the goal of both Inclusion of Informal Actors and Governance.

4 Verification Process

This section presents the details of our verification process on a client node (farmer node or aggregator node) based on aggregate verification, identity-based cryptography, and zero-knowledge cryptography. The aggregator node scheme includes system setup and registration phases, data uploading procedure, aggregate verification, and performance simulation.

4.1 *Aggregator Node*

Smallholders only need to upload a few pictures and messages to the chain; there is no real-time requirement, and they do not need to consider its writing efficiency or computational cost. Smallholders may register as a client node and perform write/read operations to the blockchain through our Android app.

For large suppliers (such as central kitchen) nodes, one node may have hundreds of IoT devices operating simultaneously, with huge amounts and multiple types of complex data simultaneously uploading to the chain; special consideration must be made for throughput, security, and the tamper-proof property.

Therefore, we design an aggregator node; IoT devices are not directly connected to the blockchain but send messages and signatures to an aggregator node which is a relay device that holds powerful computing resources for the network. After executing aggregate verification, the aggregator node then uploads the data to the chain. The advantages of using an aggregator node include:

- Reducing computational cost: IoT devices have limited computing resources, and many devices that use batteries have power restrictions. Therefore, we place energy-consuming operations (uploading procedure) on the aggregator node after centralizing the signature and message to ensure durability in IoT device operations.
- Meet real-time requirements: The message may be quickly uploaded to the chain after aggregate verification, avoiding the possibility of human tampering.

4.2 *System Setup and Registration*

When an IoT device or a client node (aggregator or farmer) first enters the system, the System Setup and Registration processes are executed. According to the aforementioned multi-layer architecture,

data uploading to the chain takes place on Layer 2: Quorum. A Quorum chain is a consortium blockchain; its implementation includes an identity-based signature scheme (IBS). A short public identifier such as an IoT device's MAC address may be used as a verification key. The following system parameters are defined in a tuple: (q, P, G1, G2, e) [39]. The notations and parameters throughout this paper are listed in Table 1.

Table 1: Notation and Parameters

Notation	Description
RID	Real Identity of the device
q	Prime order
P	Generator of G1
G1	Additive group of q
G2	Multiplicative group with same order as G1
e	G1 x G1 → G2, a bilinear mapping
h()	One-way hash function like SHA3-256
H1()	MapToPoint hash function $H1 : \{0,1\}^* \rightarrow G_1$
⊕	Exclusive or
DM	IoT device
MT	Model type
Chk	Checksum generated by SHA3-256
BT_seed	BitTorrent seed of the data

Next, the IBS selects two randomly generated numbers c_1 and c_2 as a pair of secret master keys and generates their corresponding public keys PK_{MAC1} and PK_{MAC2} as follows:

$$\begin{aligned} &c1, c2 \in Z_q^* \\ &PK_{MAC1} = c1 \times P \\ &PK_{MAC2} = c2 \times P \end{aligned} \tag{1}$$

The following hash functions are formed [40]:

$$\begin{aligned} &h : \{0,1\}^* \rightarrow \{0,1\}^f \\ &H1 : \{0,1\}^* \rightarrow G_1 \end{aligned} \tag{2}$$

The system publishes <q, P, G1, G2, e, h, H1, PK_{MAC1}, PK_{MAC2}>.

Upon entry of an IoT device or a client node, our platform must also execute a key generation procedure. The Hardware Security Module (HSM) stores system parameters c_1, c_2 and generates the pseudo-identity and its corresponding public and private keys for the client node by executing the following three procedures.

- Identity Authentication: Each device must first input <RID, fp, or pw> where RID refers to the real ID of the farmer, fp refers to fingerprint, and pw refers to password. This step is necessary for activating the device's HSM.
- Sub-Identity Calculation: HSM selects $t \in Z_q^*$ and computes the first sub-identity $SID1 = tP \in G1$ as well as the second sub-identity $SID2 = RID \oplus H1(t \times c1 \times PK_{MAC2})$.
- Key Calculation: To finish the procedure, HSM calculates the corresponding private key $PrK1 = c1 \times SID1$, $PrK2 = c2 \times H1(SID1 \oplus SID2)$. After Identity Authentication, Sub-Identity Calculation, and Key Calculation procedures have been executed, HSM outputs <ID = (SID1, SID2), PrK = (PrK1, PrK2)>.

After Identity Authentication, Sub-Identity Calculation, and Key Calculation procedures have been executed, HSM outputs <ID = (SID1, SID2), PrK = (PrK1, PrK2)>.

4.3 Data Uploading Procedure

IoT device D uses off-chain programs to download device information DM from BT_seed, and computes DM's checksum. IoT device D uploads new data to an aggregator node according to the following procedure.

- D prepares device DM and retrieves model type MT
- DM collects data, recording TimeStamp and checksum Chk generated by h(DM).
- D stores data into a peer-to-peer file sharing system such as BitTorrent or IPFS while also acquiring the BT_seed of DM.
- D provides a signature as follows:

$$\begin{aligned} v &\leftarrow h(ID \parallel MT \parallel TimeStamp) \\ w &\leftarrow h(ID \parallel Chk \parallel BT\_seed) \\ \sigma_D &= (v \times PrK1) + (w \times PrK2) \end{aligned} \tag{3}$$

Finally, D sends <ID, MT, TimeStamp, Chk, BT_seed, σ_D > to the aggregator node. The aggregator node then performs aggregate verification. The smart contract checks the validity of the aggregator's signature; if correct, the smart contract records (MT, TimeStamp, Chk, BT_seed) onto the blockchain.

4.4 Aggregate Verification

In our implementation of aggregate verification, we apply a new signature verification method that allows for superior efficiency and zero-knowledge proof. The algorithm for verifying the signature is proposed in Algorithm 1. An aggregator node receives a new transaction sent from D in the following format: <ID, MT, TimeStamp, Chk, BT_seed, σ_D >, and the aggregator node can then verify the signature σ_D with the following equation [32]: $e(\sigma_D, P) \stackrel{?}{=} e(v \cdot SID1, PK_{MAC1}) \cdot e(w \cdot H1(SID1 \oplus SID2), PK_{MAC2})$, where e is the bilinear mapping function [40] as in Table 1. The verification procedure is shown below:

$$\begin{aligned} &e(\sigma_D, P) \\ &= e(v \cdot PrK_1 + w \cdot PrK_2, P) \\ &= e(v \cdot PrK_1, P) \cdot e(w \cdot PrK_2, P) \\ &= e(v \cdot c_1 \cdot SID_1, P) \cdot e(w \cdot c_2 \cdot H_1(SID_1 \oplus SID_2), P) \\ &= e(v \cdot SID_1, c_1 \cdot P) \cdot e(w \cdot H_1(SID_1 \oplus SID_2), c_2 \cdot P) \\ &= e(v \cdot SID_1, PK_{MAC1}) \cdot e(w \cdot H_1(SID_1 \oplus SID_2), PK_{MAC2}) \end{aligned} \tag{4}$$

Algorithm 1: Verify the signature D

Data: <ID, MT, TimeStamp, Chk, BT_seed, σ_D >
Result: *true* or $false$
if $e(\sigma_D, P) == e(v \cdot SID1, PK_{MAC1}) \cdot e(w \cdot H1(SID1 \oplus SID2), PK_{MAC2})$ **then**
  if *Chk of DM is correct* **then**
    upload hash of BT_seed on the blockchain;
    return *true*;
  else
    return $false$;
else
  return $false$;

In the case that an aggregator node receives a large quantity of new data in a short time period, the aggregator node must verify multiple or numerous signatures. Under these circumstances, aggregate verification is applied as shown below. Let us suppose that n distinct transactions must be verified. Each transaction is denoted as < MT^n, TimeStamp^n, Chk^n, BT_seed^n, σ_D^n > . The aggregator node can perform aggregate verification on all signatures with the following calculation: $e(\sum_{i=1}^{n} \sigma^i, P) \stackrel{?}{=} e(\sum_{i=1}^{n} v_i \cdot SID_1^i, PK_{MAC1}) \cdot e(\sum_{i=1}^{n} w_i \cdot H1(SID_1^i \oplus SIDi_2^i), PK_{MAC2})$. The verification procedure is shown below:

$$
\begin{aligned}
& e(\sum_{i=1}^{n} \sigma^i, P) \\
&= e(\sum_{i=1}^{n} (v^i \cdot PrK_1^i + w^i \cdot PrK_2^i), P) \\
&= e(\sum_{i=1}^{n} (v^i \cdot Prk_1^i, P) \cdot e(\sum_{i=1}^{n} (w^i \cdot PrK_2^i), P) \\
&= e(\sum_{i=1}^{n} (v^i \cdot c_1 \cdot SID_1^i, P) \cdot e(\sum_{i=1}^{n} w^i \cdot c_2 \cdot H_1(SID_1^i \oplus SID_2^i), P)) \\
&= e(\sum_{i=1}^{n} v^i \cdot SID_1^i, c_1 \cdot P) \cdot e(\sum_{i=1}^{n} w^i \cdot H_1(SID_1^i \oplus SID_2^i), c_2 \cdot P) \\
&= e(\sum_{i=1}^{n} v^i \cdot SID_1^i, PK_{MAC1}) \cdot e(\sum_{i=1}^{n} w^i \cdot H_1(SID_1^i \oplus SID_2^i), PK_{MAC2})
\end{aligned}
\tag{5}
$$

4.5 Performance Simulation

We evaluate the performance of aggregate verification and other cases presented [41]. Cases 1-5 consist of two types of operations: the first is response from a verification node to a requesting node (R1). The second is the response from a response node to a requesting node (R2). R1 is then divided into two sub-cases (Case 1 and Case 2), while R2 is similarly separated into Case 3, Case 4, and Case 5. Each case depends on the type of response node which receives the version-check request message sent from the requesting node. Their proposed blockchain scheme has five different operation cases:

- Case 1: an IoT device sends a request to a verification node that has the newest data to verify whether its data is the newest.
- Case 2: an IoT device sends a request to a verification node that has the newest data to confirm its data is old and then downloads the newest data.
- Case 3: an IoT device first asks its neighbor IoT device to verify the data version and then verifies whether it has the newest data.
- Case 4: an IoT device first asks its neighbor IoT device to verify whether the data has a newer version and sends a download request to a verification node if it does not have the newest data.
- Case 5: an IoT device broadcasts a join verification message to a blockchain network to ask other nodes to join the verification process to check whether the data version of a device is equal to that of its neighbor node.

We define the cost of cryptographic operations required for each verification procedure. The parameters in Table 1 let S be the time of scalar multiplication in G1, P the time of bilinear pairing operation, and the time of MapToPoint hash operation is H. The pairing operation is the most time-consuming of those operations. We only consider these operations, which determine the speed of signature verification, omitting all other procedures such as one-way hash and point addition.

We use parameter size selection for the elliptic curve cryptography scheme [42] to ensure a security level of the 1024-bit RSA algorithm as a benchmark. G is an additive cyclic group of order q (160-bit prime number) on the elliptic curve, and P is the generator of G. The processing time of Tate pairing on an MNT curve with an 80-bit security level, 160-bit q, and embedding degree k=6 running on an Intel i7 3.07 GHz machine in experiment [43]. Table 2 shows the symbol and execution time in milliseconds.

Table 2: Symbols and Execution Time

Symbol	Description	Execution Time
P	Bilinear pairing operation	3.25
S	Scalar multiplication	0.41
H	MapToPoint hash operation	0.13

Table 3: The Comparison of Operations

Case	Operations
Case1	$2S + 1P$
Case2	$2S + 1P$
Case3	$5S + 3P$
Case4	$5S + 3P$
Case5	$6H + 3S + 2P$
Our scheme	$2H + 3S$

Our computational cost-performance comparison and overhead are shown in Fig. 2. In our scheme, there are five components (RID, q, P, G1, and e) that are passed into the smart contract for updating the data. Next, we explain the update message calculation of Case 1 to Case 5 based on [41], which requires more operations than in Case 1; Case 2 includes two decryption operations and one verification operation. The time in Case 3 and Case 4 to process n messages

in five decryption and three verification processes. In Case 5, the data is considered verified if the original verification is confirmed by other nodes; this is done via a six-block confirmation with PoW consensus. The request nodes need to receive six verification messages from other blockchain nodes to confirm the integrity of the data. As a result, the total computational cost is three decryption operations, two verification operations, and six ECDSA verifications.

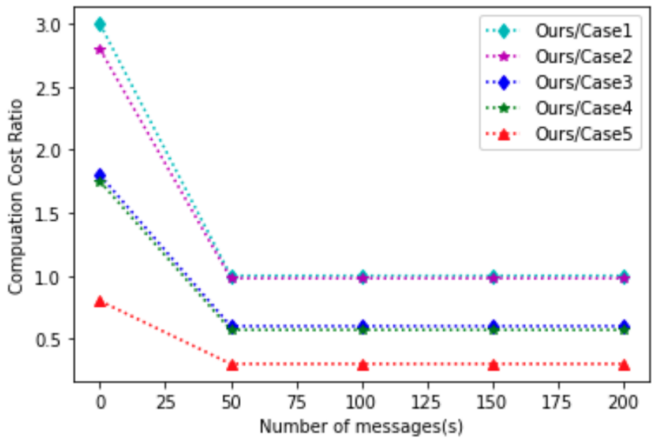


Figure 2: Computation Ratio

5 Management Framework and System Architecture

This section will focus on our blockchain management framework operation procedure and the details of our system architecture. Traditional paper contracts are replaced with smart contracts in the practical process of smallholder agricultural supply chain finance. We digitize the loan application and the information that may prove an upstream producer can reimburse the lender (operational plans, historical transaction records, agricultural production status, current delivery schedule, etc.) and upload this information to the blockchain system. We describe the relationship of the three layers that compose the system architecture.

5.1 Objects and Nodes

There are three main roles: 1. Smallholder (loan applicant) 2. Agricultural financial institution (lending institution) 3. Distribution channel (supermarket or hypermarket)

There are two separate procedures for signing the transaction contract and for the payment verification by the distribution channel: I. The buyer (distribution channel) and the seller (smallholder) sign a smart contract and upload it to Ethereum. II. Smallholders supply to large distributors. After the distribution channels confirm acceptance, they upload the acceptance records to the blockchain. The distribution channel regularly settles the sales volume and pays the smallholders, and at the same time, the sales volume data and payment information are uploaded to the blockchain. An order is not complete until the final payment has been processed.

5.2 Operation Procedure

The operation structure is shown in Fig. 3. The distribution channel (DC) first signs a contract with the smallholder through a smart contract and uploads it to the blockchain after a certain number of parties confirm its correctness. The smallholder must provide other information to financial institutions that may prove their ability to repay the loan, such as business plan books, historical transaction records, and agricultural production conditions to the agricultural financial institution (AFI). After the AFI confirms eligibility and approves the loan, the smallholder begins production. If the DC has paid the deposit, the deposit is then transferred to the smallholder's account in the AFI.

While the smallholder grows crops, production status is regularly uploaded to the blockchain through IoT devices for inspection by AFI and DC. For example, at the i'th production checkpoint, the IoT device uploads photos of the production progress to the blockchain via aggregate verification. Therefore, AFI can make phase i payment to the smallholder based on the production progress ledger. If the photos of the production status do not meet the mark and the smallholder cannot complete production on time, the AFI will recover the funds.

After the smallholder has completed production, they will upload the supply orders to the system when they deliver the produce to the DC. After the DC accepts the produce, the supply orders are uploaded to the blockchain to complete the acceptance of produce. Finally, the AFI makes the final payment to the smallholder, deducting interest and handling fees of the loan according to the previous production progress account book, and settles the remaining receivables.

On our agricultural supply chain blockchain financing platform, all kinds of transaction contracts, supply documents, and production activities are recorded in the blockchain system to allow for agricultural production and sales activities to be transparent and immutable. Relevant information from farmers stored in the blockchain system also proves their repayment ability, providing them with credentials for loans.

5.3 System Architecture

Fig. 4 shows the architecture of the blockchain system for the proposed financing platform. We choose Ethereum as our blockchain system because smart contract functionality is built-in. As shown in the user group located at the bottom of Fig. 4, the IoT device uploads new photos over a set interval. The original data is stored in the database while the following hash value is uploaded to the blockchain: <ID, CT, TimeStamp, Chk, BT_seed, σF >. Other data such as transaction contracts, supply records, and sales records are uploaded to the blockchain in smart contracts. After applying for permission, distribution channels may confirm contract information through the Inspection API. Agricultural financial institutions and agricultural loan reviewing entities may also view the field of agricultural product production materials and inspection documents.

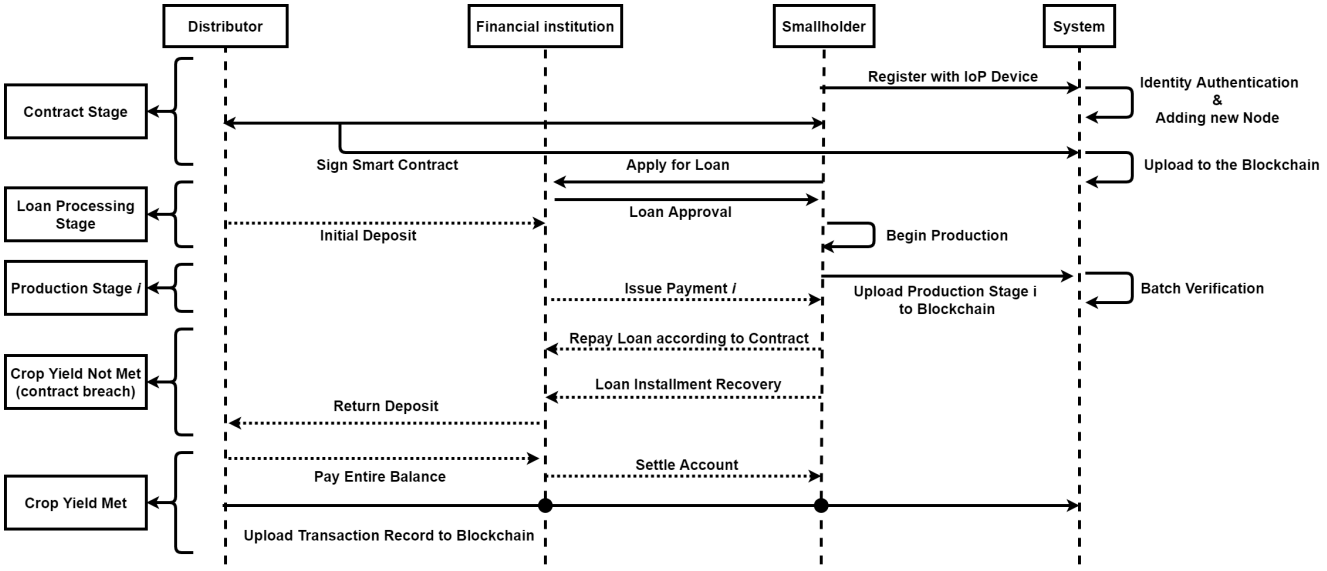


Figure 3: Operation Procedure

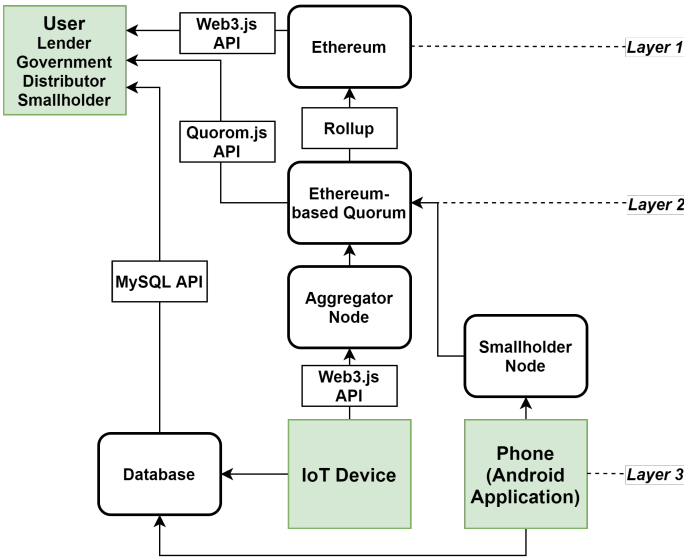


Figure 4: System Architecture

6 Implementation

Layer 2 is built on Quorum (version 2.7.0) with a modified version of the RAFT consensus mechanism [44]. It was initially built on three nodes in different data centers, where each node has 2GB of memory. In addition to fast handling and verification of transactions, this blockchain system must also support frequent on-chain data and contract status queries. Therefore, according to our evaluation results, the on-chain data must be stored in an SSD with at least 5000 read and write IOPS to handle more than 1,000 transactions and queries per second.

In the current situation, each node requires about 10 GB of disk space; it may also be stored in RAM when there is a backup power supply, but the cost is higher. Regarding the parameter settings of the blockchain, to avoid frequent retransmission of transactions between nodes, we set both global slots and global queue block numbers to 76798.

The need for a cheap and easy-to-manufacture device to monitor crops from a close range has arisen from a lack of availability of sensors suited for camera monitoring that have environmental resistance. Most smart farming IoT sensors do not employ the use of cameras. We create a new smart farming IoT device with a simple but robust design that may allow for crop monitoring using a mini-spherical camera: Hi3516CV300 with auto-zoom lens (3x optical zoom) *PTZ 355 degrees left and right, 90 degrees up and down. The camera is mounted to the top of a 1m x 1m x 1m transparent plastic box that crops can grow in.

Our IoT devices include (1) Wireless temperature and humidity sensing devices: measure temperatures ranging from -50C to 120C, and Humidity measurements ranging from 0 percent to 100 percent.; (2) 2.4GHz wireless sensing control device; (3) Weighing transducer: the force or mass is converted into a measurable electrical signal, and the detected weight data is uploaded to the cloud platform through the communication device; (4) Carbon monoxide temperature and humidity sensor (with routing function): Carbon dioxide sensing range: 400 10,000ppm; (5) The above-mentioned camera mount.

The number of transactions per month is shown in Fig. 5. In the case of high transaction volume in the future, it is necessary to adjust the data ratio of each layer of the LevelDB database and use batch parallel reading to increase the access efficiency of the SSD. In addition, to accelerate smart contracts, we also need to rewrite and compile Quorum to use evmos.

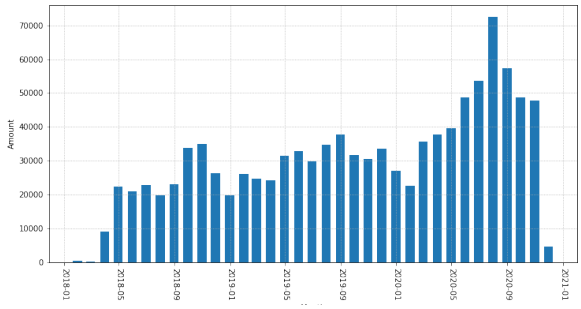


Figure 5: 2018/01 - 2020/12 Number of Transactions

7 Conclusion

In this paper, we propose secured MLAV for an IoT Blockchain architecture. We implement a high-efficiency and high-security agriculture 4.0 framework to allow for smallholders to join the agricultural supply chain blockchain. First, we create the Agricultural Supply Chain Finance operational structure and corresponding smart contracts which define the rules and stipulate the responsibilities and read and write permissions of the three types of nodes in the network (farmers, distribution channels, and financial institutions); second, we design an IoT blockchain system and an IoT camera sensor that allows for upstream producers' production activity data and historical order information to be uploaded to the blockchain. This data will provide guarantees to help them gain trust and qualify for loans from financial institutions; third, through aggregate verification technology, we successfully allow for multiple IoT devices to upload large amounts of data to the blockchain system in a short time. To guarantee security, smallholders can join the blockchain network with high efficiency, high transaction speed, and ID-based signature verification. Through the application of IoT Blockchain, the upstream producer's production activity data and historical order information may be provided to third party entities for evaluation. This ultimately provides a means by which they may gain trust and become eligible for loans from financial institutions.

7.1 Future Work

This paper provides a framework for securely and efficiently generating and storing data on a novel IoT blockchain system in an agricultural supply chain setting. Future work on this research may detail different application scenarios of our IoT blockchain framework in both agricultural and non-agricultural settings. The supply chain considerations, the corresponding smart contracts, and the Operation Procedure in Fig. 3 would need to be changed to fit the application scenario accordingly; however, the IoT blockchain infrastructure and batch verification algorithm may require little to no altering. In addition, future work could focus further on blockchain oracle implementations that allow for guided decisions based on automatic data analysis. An example application scenario in an agricultural setting could be in using our IoT sensor for monitoring of insects on crops. This may hold valuable implications in defining a crop's organic status because in many regions, in order to attain organic status, it must be ensured that insecticides were not used. The number of insects and the type of insects on crops may indicate whether insecticides were used, and therefore confirm the organic status of the crop.

Conflict of Interest The authors declare no conflict of interest.

Acknowledgment Ming-Fong Sie, Jingze Wu, and Seth Austin Harding contributed equally to this paper.

References

[1] J. Wu, M.-F. Sie, S. A. Harding, C.-L. Lin, S.-T. Wang, S.-w. Liao, "Multi-Layer Aggregate Verification for IoT Blockchain," in 2021 3rd Conference on Blockchain Research & Applications for Innovative Networks and Services (BRAINS), 43–44, IEEE, 2021, doi:10.1109/BRAINS52497.2021.9569817.

[2] Z. Zhai, J. F. Martínez, V. Beltran, N. L. Martínez, "Decision support systems for agriculture 4.0: Survey and challenges," Computers and Electronics in Agriculture, **170**, 105256, 2020, doi:10.1016/j.compag.2020.105256.

[3] V. Saiz-Rubio, F. Rovira-Más, "From smart farming towards agriculture 5.0: a review on crop data management," Agronomy, **10**(2), 207, 2020, doi:10.3390/agronomy10020207.

[4] M. De Clercq, A. Vats, A. Biel, "Agriculture 4.0: The future of farming technology," Proceedings of the World Government Summit, Dubai, UAE, 11–13, 2018.

[5] R. Kamath, "Food traceability on blockchain: Walmart's pork and mango pilots with IBM," The Journal of the British Blockchain Association, **1**(1), 3712, 2018.

[6] H. Haswell, M. Storgaard, "Maersk and IBM unveil first industry-wide cross-border supply chain solution on blockchain," IBM, http://www-03. ibm. com/press/us/en/pressrelease/51712. wss, 2017.

[7] J. Nurgazina, U. Pakdeetrakulwong, T. Moser, G. Reiner, "Distributed ledger technology applications in food supply chains: a review of challenges and future research directions," Sustainability, **13**(8), 4206, 2021, doi:10.3390/su13084206.

[8] A. Ellebrecht, "Chain of Custody and Transparency in Global Supply Chains," in Sustainable Global Value Chains, 227–237, Springer, 2019, doi:10.1007/978-3-319-14877-9_13.

[9] S. Nakamoto, A. Bitcoin, "A peer-to-peer electronic cash system," Bitcoin.–URL: https://bitcoin. org/bitcoin. pdf, 2008.

[10] G. Wood, et al., "Ethereum: A secure decentralised generalised transaction ledger," Ethereum project yellow paper, **151**(2014), 1–32, 2014.

[11] V. Buterin, "Ethereum: Platform Review," Opportunities and Challenges for Private and Consortium Blockchains, 2016.

[12] H. Kalodner, S. Goldfeder, X. Chen, S. M. Weinberg, E. W. Felten, "Arbitrum: Scalable, private smart contracts," in 27th USENIX Security Symposium (USENIX Security 18), 1353–1370, 2018.

[13] N. Szabo, "Formalizing and securing relationships on public networks," First Monday, **2**(9), 1997.

[14] M. Alharby, A. Van Moorsel, "Blockchain-based smart contracts: A systematic mapping study," arXiv preprint arXiv:1710.06372, 2017.

[15] Y. Wang, J. H. Han, P. Beynon-Davies, "Understanding blockchain technology for future supply chains: a systematic literature review and research agenda," Supply Chain Management: An International Journal, 2019.

[16] S. Mansfield-Devine, "Beyond Bitcoin: using blockchain technology to provide assurance in the commercial world," Computer Fraud & Security, **2017**(5), 14–18, 2017, doi:10.1016/S1361-3723(17)30042-8.

[17] X. Lin, S.-C. Chang, T.-H. Chou, S.-C. Chen, A. Ruangkanjanases, "Consumers' intention to adopt blockchain food traceability technology towards organic food products," International Journal of Environmental Research and Public Health, **18**(3), 912, 2021, doi:10.3390/ijerph18030912.

[18] F. Casino, V. Kanakaris, T. K. Dasaklis, S. Moschuris, S. Stachtiaris, M. Pagoni, N. P. Rachaniotis, "Blockchain-based food supply chain traceability: a case study in the dairy sector," International Journal of Production Research, **59**(19), 5758–5770, 2021, doi:10.1080/00207543.2020.1789238.

[19] J. Xu, S. Guo, D. Xie, Y. Yan, "Blockchain: A new safeguard for agri-foods," Artificial Intelligence in Agriculture, **4**, 153–161, 2020, doi:10.1016/j.aiia.2020.08.002.

[20] A. Park, H. Li, "The effect of blockchain technology on supply chain sustainability performances," Sustainability, **13**(4), 1726, 2021, doi:10.3390/su13041726.

[21] H. Chen, Z. Chen, Y. Cheng, X. Deng, W. Huang, J. Li, H. Ling, M. Zhang, "Poster: An Efficient Permissioned Blockchain with Provable Reputation Mechanism," in 2021 IEEE 41st International Conference on Distributed Computing Systems (ICDCS), 1134–1135, IEEE, 2021, doi:10.1109/ICDCS51616.2021.00123.

[22] Y. Guo, C. Liang, "Blockchain application and outlook in the banking industry," Financial Innovation, **2**(1), 24, 2016, doi:10.1186/s40854-016-0034-9.

[23] R. Kumar, R. Tripathi, N. Marchang, G. Srivastava, T. R. Gadekallu, N. N. Xiong, "A secured distributed detection system based on IPFS and blockchain for industrial image and video data security," Journal of Parallel and Distributed Computing, **152**, 128–143, 2021, doi:10.1016/j.jpdc.2021.02.022.

[24] A. S. Rajawat, R. Rawat, K. Barhanpurkar, R. N. Shaw, A. Ghosh, "Blockchain-based model for expanding IoT device data security," in Advances in Applications of Data-Driven Computing, 61–71, Springer, 2021, doi:10.1007/978-981-33-6919-1_5.

[25] J. Lian, S. Wang, Y. Xie, "Tdrb: An efficient tamper-proof detection middleware for relational database based on blockchain technology," IEEE Access, **9**, 66707–66722, 2021, doi:10.1109/ACCESS.2021.3076235.

[26] X. Fu, H. Wang, P. Shi, "A survey of Blockchain consensus algorithms: mechanism, design and applications," Science China Information Sciences, **64**(2), 1–15, 2021, doi:10.1007/s11432-019-2790-1.

[27] M. Tripoli, J. Schmidhuber, "Emerging Opportunities for the Application of Blockchain in the Agri-food Industry," FAO and ICTSD: Rome and Geneva. Licence: CC BY-NC-SA, **3**, 2018.

[28] N. Kshetri, "Can blockchain strengthen the internet of things?" IT professional, **19**(4), 68–72, 2017, doi:10.1109/MITP.2017.3051335.

[29] A. Dorri, S. S. Kanhere, R. Jurdak, P. Gauravaram, "Blockchain for IoT security and privacy: The case study of a smart home," in 2017 IEEE international conference on pervasive computing and communications workshops (PerCom workshops), 618–623, IEEE, 2017, doi:10.1109/PERCOMW.2017.7917634.

[30] S. Varshney, P. Vats, S. Choudhary, D. Singh, "A Blockchain-based Framework for IoT based Secure Identity Management," in 2022 2nd International Conference on Innovative Practices in Technology and Management (ICIPTM), volume 2, 227–234, IEEE, 2022, doi:10.1109/ICIPTM54933.2022.9753887.

[31] K. Hakuta, Y. Katoh, H. Sato, T. Takagi, "Batch verification suitable for efficiently verifying a limited number of signatures," in International Conference on Information Security and Cryptology, 425–440, Springer, 2012, doi:10.1007/978-3-642-37682-5_30.

[32] J.-W. Hu, L.-Y. Yeh, S.-W. Liao, C.-S. Yang, "Autonomous and malware-proof blockchain-based firmware update platform with efficient batch verification for Internet of Things devices," Computers & Security, **86**, 238–252, 2019, doi:10.1016/j.cose.2019.06.008.

[33] E. K. Kogias, P. Jovanovic, N. Gailly, I. Khoffi, L. Gasser, B. Ford, "Enhancing bitcoin security and performance with strong consistency via collective signing," in 25th {usenix} security symposium ({usenix} security 16), 279–296, 2016.

[34] J. Polge, J. Robert, Y. Le Traon, "Permissioned blockchain frameworks in the industry: A comparison," Ict Express, **7**(2), 229–233, 2021, doi:10.1016/j.icte.2020.09.002.

[35] A. Shwetha, C. Prabodh, "A Comprehensive Review of Blockchain Based Solutions in Food Supply Chain Management," in 2021 5th International Conference on Computing Methodologies and Communication (ICCMC), 519–525, IEEE, 2021, doi:10.1109/ICCMC51019.2021.9418305.

[36] Q. Song, Y. Chen, Y. Zhong, K. Lan, S. Fong, R. Tang, "A supply-chain system framework based on internet of things using blockchain technology," ACM Transactions on Internet Technology (TOIT), **21**(1), 1–24, 2021, doi:10.1145/3409798.

[37] J. Liu, H. Zhang, L. Zhen, "Blockchain technology in maritime supply chains: applications, architecture and challenges," International Journal of Production Research, 1–17, 2021, doi:10.1080/00207543.2021.1930239.

[38] J. Tholen, D. de Vries, A. Daluz, C.-C. Antonovici, W. V. Brug, R. Abelson, D. Lovell, "Is there a role for blockchain in responsible supply chains?" in OECD Global Blockchain Policy Forum, 2019.

[39] J. Kurose, K. Ross, "Computer Networking: A Top Down Approach, 2012," .

[40] D. Hankerson, A. J. Menezes, S. Vanstone, Guide to elliptic curve cryptography, Springer Science & Business Media, 2006.

[41] B. Lee, J.-H. Lee, "Blockchain-based secure firmware update for embedded devices in an Internet of Things environment," The Journal of Supercomputing, **73**(3), 1152–1167, 2017, doi:10.1007/s11227-016-1870-0.

[42] E. Barker, A. Roginsky, et al., "Transitions: Recommendation for transitioning the use of cryptographic algorithms and key lengths," NIST Special Publication, **800**, 131A, 2011.

[43] N.-W. Lo, J.-L. Tsai, "An efficient conditional privacy-preserving authentication scheme for vehicular sensor networks without pairings," IEEE Transactions on Intelligent Transportation Systems, **17**(5), 1319–1328, 2015, doi:10.1109/TITS.2015.2502322.

[44] D. Ongaro, J. Ousterhout, "In search of an understandable consensus algorithm," in 2014 USENIX Annual Technical Conference (Usenix ATC 14), 305–319, 2014.

ASTES

Advances in Science, Technology and Engineering Systems Journal Vol. 7, No. 3, 116-122(2022)

www.astesj.com

ASTESJ

ISSN: 2415-6698

Physical and Emission Properties of Blended Bio-Coal Briquettes Derived from Agro-Wastes in Nigeria

Cosmas NgozichukwuAnyanwu[1,2], ChinazomJanefrancesAnimoke[1], Bonaventure Ugo Agu[1], Izuchukwu Francis Okafor[3], Nneka Juliana Ogbuagu[4,*], Samuel Bentson[5], OnyekwereOjike[1]

[1]*Department of Agricultural and Bioresources Engineering, University of Nigeria Nsukka, Enugu state, 410001, Nigeria*

[2]*Africa Centre of Excellence for Sustainable Power and Energy Development (ACE-SPED), University of Nigeria Nsukka, Enugu State, 41000, Nigeria*

[3]*National Centre for Energy Research and Development, University of Nigeria Nsukka, 410001, Nigeria*

[4]*Department of Agricultural and Bioresources Engineering, Nnamdi Azikiwe University.Awka Anambra State, 420007, Nigeria*

[5]*Aprovecho Research Centre, Cottage Grove, Oregon, 97424, USA*

ARTICLE INFO

Article history:
Received: 24 February, 2022
Accepted: 27April, 2022
Online: 04 June, 2022

Keywords:
Firewood
Briquettes
Corn cob
Palm mesocarp fibre
Emission
Thermal efficiency

ABSTRACT

Nigeria has one of the highest deforestation rates in the world, due mainly to felling of trees for fuelwood and charcoal production. This challenge could be managed if agricultural waste briquettes are used to augment the fuelwood demand for cooking energy provisioning. Energy density of biomass fuels can be raised by blending with coal, but particulate matter emissions are also increased in the process. Therefore, some physical and fuel properties of briquettes produced from blends of corn cob and palm mesocarp fibre (PMF) with Sub-bituminous coal from Onyeama mine, using cassava starch binder were studied. Blending ratios were varied from 0% to 100% biomass. After briquetting in a hydraulic press, drying and characterization, Water Boiling Test (WBT) of the briquette samples was performed using Laboratory Emissions Measuring System (LEMS). Results showed that average High-Power Thermal Efficiency was 28.8% for corn cob/coal and 26.0% for PMF/coal briquettes, but High-Power CO emissions decreased from 10.8 mg/MJd to 8.9 mg/MJd (8.31-6.90 ppm) as the composition of corn cob increased from 0% to 100%. Corn cob/coal briquettes produced lower PM emissions than pmf/coal, although both were above the WHO recommended limits.

1. Introduction

Millions of households in Nigeria rely on the traditional use of firewood for their daily cooking needs. This practice has persisted in spite of the fact that smoke from cooking fire causes over 95,000 deaths in Nigeria annually [1]. It is the third biggest killer after malaria and HIV/AIDS in Nigeria [2]. Majority of poor families using three-stone fire spend much of their food budgets on wood and charcoal; others spend hours collecting wood. Inefficiency in the combustion of wood in open three-stone fires raises the cost of cooking for the poor and contributes to high level of deforestation. Nigeria lost about 2.3% of her forest reserves between 1990 and 2010 [3].

The use of bio-coal briquettes as cooking fuels, especially in improved cookstoves, could provide a solution to the afore-mentioned health and deforestation problems [4] and help cut down on the consumption of fossil fuel [5]. Research has shown that blending coal with biomass to produce briquettes helps to ameliorate the emission problems thus result in anenvironmentally friendly briquette with better combustion and physical characteristics [6].

Nigeria has a large deposit of coals, which are mostly sub-bituminous and lignitic formations. With the exception of the

*Corresponding Author: Ogbuagu Nneka Juliana, Department of Agricultural and Bioresource Engineering, Nnamdi Azikiwe University Awka , 234 806 7699 760 & njogbuagu@unizik.edu.ng

Lafia coal, which is partially coking [7], these coals are non-coking and are not very useful in the metallurgical industry. Nigeria produces about 10.8 MMT of corn annually and about 20.9 million tonnes of Mesocarp Fibre annually [8]. The disposal of these agro-wastes poses a serious challenge for municipal authorities. Blending biomass with coal helps to reduce the emission of particulate matter from coal, while increasing the energy density of the biomass. The overall effect is a reduction in deforestation and enhanced management of agro-wastes.

Bio-coal briquettes are prepared by blending coal, biomass and binders; a sulphur fixation agent may, sometimes be incorporated. A briquette is a compressed block of coal dust or other combustiblebiomass material such as charcoal, sawdust, wood chips, peat, or paper used for fuel and kindling to start a fire[9]. Biomass briquettes, mostly made of green waste and other organic materials, are commonly used for electricity generation, heating, and cooking. These compressed compounds contain various organic materials, including rice husk, bagasse, ground nut shells, municipal solid waste, and agricultural waste. The use of organic briquettes (biomass briquettes) started more recently compared to coal briquettes, which dates back to eighteenth century [10]. Corn cobs and palm waste, such as mesocarp fibre are not only in abundant supply, but constitute all-year-round environmental waste in Nigeria. Industrial starch produced mainly from cassava is a common energy resource in Nigeria, since the country produces over 62.0 MMT of cassava annually, out of which about 53.0 MMT is demanded as food [11]. Starch is usually produced from cassava wastes, especially the peels.

The present work reports the production of bio-coal briquettes, composed of sub-bituminous coal and palm mesocarp fibre or corn cob and the characterization of their physical and fuel properties as cooking fuel using the Laboratory Emissions Measuring System (LEMS).

2. Materials and Methodology

2.1. Materials

The materials used in the work were: Sub-bituminous coal from Onyeama mine, Enugu state Nigeria, palm mesocarp fibre (PMF) obtained from a local oil mill in Nsukka, corn cob obtained from various farms in Nsukka, Enugu state, bowl and water.

The equipment used were locally made attrition mill, which was used to crush the coal, electronic weighing balance (Citizen Instruments Inc.), hydraulic press briquetting machine (locally made), improved briquette stove (locally made) and iKA bomb calorimeter. Hydraulic press briquetting machine was used to compound the materials into blocks. It is a manually operated briquetting machine, designed and fabricated at the National Center for Energy Research and Development (NCERD), University of Nigeria Nsukka. It uses vertically mounted 200 kN capacity hydraulic press to its maximum compaction pressure of 160 N/m^2 and has 12 cavities of 70 mm/100 mm cross-sectional area, each where the biomass, coal and binder are loaded for compaction. The press was designed to handle a maximum load of 7.0 kg of biomass. At full capacity, it can produce about 2.80 tons of briquettes per day.

2.2. Methodology

The coal was first reduced to fine particles of less than 0.5 mm diameter using the attrition mill. The mesocarp fibre is a fluffy material. It was reduced to particles of about 2-3 mm, while the corn cobs were milled into small particles of about 0.5-1.0 mm size. The coal, biomass (i.e. palm mesocarp fibre or corn cob) and starch were weighed using the weighing balance into pre-determined proportions, and tied in nylon bags. The samples were then put into a bowl according to a pre-determined ratio for mixing. Binder (starch) content was varied between 10% and 30% at 10% intervals, giving a total of three treatment levels. Each mixed sample was put in a cavity in the hydraulic press briquetting machine and manually pressed into briquettes. The briquettes were then left to dry for about three days in a cabinet solar dryer. After solar- drying the briquettes, the heating (calorific) values were determined using the bomb calorimeter. The HHV of the blended briquettes were then calculated using Equation. 1 as follows:

$$HHV = \frac{\sum_{1}^{n} C_i hv_i}{100} \quad (1)$$

where,
C_i = Mass composition of i^{th} component (coal, biomass or starch, %);hv_i = Heating value of i^{th} component.

The percentage compositions of the bio-coal briquettes, from where their mass compositions were worked out using a typical briquette final mass of approximately 115 g ranged between 0 and 100 % (mass) for the biomass, coal and starch. The hydraulic press loaded with bio-coal blend for briquetting is shown in Figure 1. Appendix 1 is the photograph of the briquetting machine used for the research.

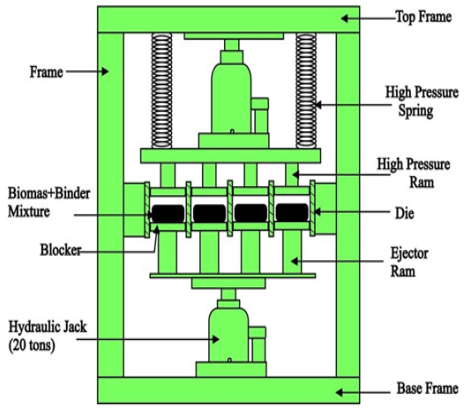


Figure 1: Schematic of the Hydraulic Press Briquetting Machine

The Laboratory Emissions Measuring System (LEMS)™ was used for characterization of the thermal efficiency and

emissions released from the briquettes during the WBT. It is equipped with sensors for CO, CO_2 and Particulate Matter (PM) emissions. The CO sensor is an electrochemical cell [12], which takes advantage of the fact that conductivity between two electrodes in the cell is proportional to the concentration of CO present. This cell has a reference terminal as well and requires a potentiostatic controller. The CO_2 sensor uses non-dispersive infrared (NDIR) to measure CO_2 concentration and outputs voltage. It is self- calibrating, with pure Nitrogen gas used for zero reference. The LEMS has two PM sensors, namely the regular photometer sensor and the gravimetric system. The scattering photometer has both a laser and a light receiver. When smoke enters the sensing chamber, particles of smoke scatter the laser light into the receiver. The amount of scattered light is calibrated with a laboratory-standard nephelometer. The gravimetric system gives a direct measurement of total PM via filter-based sampling. A vacuum pump pulls the exhaust gases through the sample line and the critical orifice, which holds the flow at a steady 16.7 L/min [12]. A cyclone particle separator is used so that all $PM_{2.5}$ (particulate matter of aerodynamic diameter equal to or less than 2.5 μm) is collected on a glass fibre filter while the pump is on. The filter was pre- and post-weighed to calculate the total mass of $PM_{2.5}$ deposited on it.

Flow rate was measured by an orifice meter pressure transducer which outputs a signal based on the pressure drop measured across the flow grid. The temperature of the exhaust gas was measured using a K-type thermocouple sensor in real-time. These data were required to calculate the density of exhaust air in order to determine the mass flow rate of emissions. A thermocouple (TC) temperature sensor was used to record the temperature of the water in the pot. The LEMS sensor box outputs data through an RS-232 serial port of a connected computer.

2.2.1. Water Boiling Test using LEMS

Water Boiling Test (WBT) was conducted using an improved briquette stove and the LEMS equipment in the National Stove Eligibility laboratory, NCERD, University of Nigeria, Nsukka. The stove used for the WBT, the LEMS accessories and the whole experimental set-up are shown in Figure 2 (See appendix 2). The LEMS is suitable for real-time logging of total emissions data emanating from biomass-fired stoves during WBT. It is equipped with sensors for CO_2, CO and Particulate matter (PM), as well as a gravimetric system for more accurate determination of PM emissions. The LEMS sensor box and other accessories were run for at least 1 minute with the blower off in order to capture a zero flow reading. The Magnehelic™ flow meter was adjusted using a small screw to ensure it was reading zero. The blower was turned on and the background period was observed starting from 4 minutes after the LEMS began logging to allow time for the sensors to warm up. The room temperature and other necessary inputs were recorded, while other test materials such as fuel and water were weighed and loaded.

Each WBT test comprised of a Hot start phase (during which 3L of water was brought to boiling point with heat generated from the briquette sample), and Simmer phase (during which the boiling water was kept at a temperature not less than 97°C for 45 minutes), following IWA Protocol and NIS 1000:2018 Part 1, Standard for Clean Cookstoves - Solid Biomass [13]. At the end of the test, the stove and briquettes were removed from under the hood and the system was allowed to run for another 10 minutes to clear out the gases in the sensors. The logged data was then processed using the LEMS software to obtain the thermal efficiency, specific fuel consumption, average CO, CO_2 and PM emissions etc, which are reported according to the ISO International Working Agreement (IWA) format.

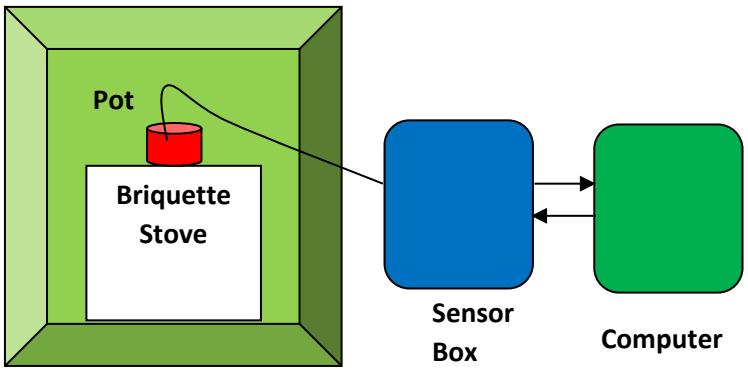


Figure 2: Experimental Set-up for the WBT using LEMS

2.3. Experimental Uncertainty

Uncertainties in experimental studies occur often owing to improper instrument selection, environmental conditions, instrument inaccuracy, readability and human errors [14]. It is very necessary for the investigator to estimate the maximum possible uncertainties in the independent variables (in this case, water temperature, mass of charcoal, mass of water, and exit speed of flue gases) as well as the calculated parameters ($PM_{2.5}$ emissions, CO emissions, CO_2 emissions, thermal efficiency, specific fuel consumption, etc). The uncertainties of the variables are temperature measurement (w_T) ± 0.1°C, relative humidity (w_{RH}) ± 3%, weight of charcoal and water (wm) ± 0.0001kg. The result R is a given function in terms of the independent variables x1, x2, x3, ….., x_n. Let wR be the uncertainty in the result and w1, w2,……,wn be the uncertainties in the independent variables. Therefore, the fractional uncertainty in the thermal efficiency (TE) can be calculated using eqn. 2 as follows:

$$\frac{\partial TE_{eff}}{TE_{eff}} = \sqrt{\left(\frac{\partial \dot{E}_{out}}{\dot{E}_{out}}\right)^2 + \left(\frac{\partial \dot{E}_{in}}{\dot{E}_{in}}\right)^2} \quad (2)$$

Since.,

$$TE = \frac{M_w C_{pw} \Delta T}{M_{ch} HV_{ch}} \quad (3)$$

It follows that,

$$\frac{\partial \mathrm{TE}_{\mathrm{eff}}}{\mathrm{TE}_{\mathrm{eff}}} = \sqrt{\left(\frac{\partial M_w}{M_w}\right)^2 + \left(\frac{\partial M_{ch}}{M_{ch}}\right)^2} \quad (4)$$

where;

C_{Pw} = Specific heat capacity of water (4,190 kJ/kg);

HV_{ch} = Heating value of charcoal (29,400 kJ/kg);
$E_{in(out)}$ = Energy Input (output) from the stove, kJ;
Mw = Mass of water, kg;
Mch = Mass of charcoal, kg;
ΔT = Temperature difference, K.

Solving eqn. 4 leads to fractional efficiency of 2.71% for the system.

2.3. Practical application

In practical terms, this work could help enviro-preneurs wishing to venture into the production of biocoal briquettes as alternative to firewood for domestic and cottage industrial application (e.g. in bakeries etc).

3. Results and Discussion

3.1 Results of bio-coal briquetting

The produced pure sample (coal, corn cob and palm mesocarp fibre) briquettes are shown in Figure 3, while some blended briquettes are shown in Figure 4.

Figure. 3: Pure briquettes of coal, palm mesocarp fibre (pmf) and corn cob.

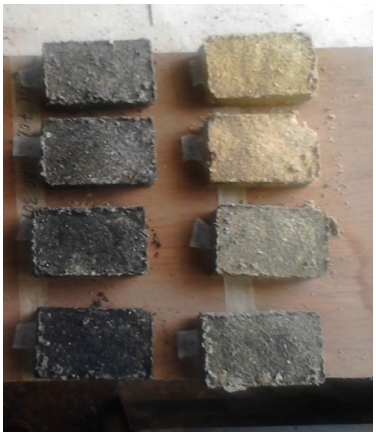


Figure 4: Some bio-coal briquettes.

3.2. Results of briquettes characterization

Table 1 shows the mass, volume, bulk density and HHVs of the coal/corn cob briquettes produced in accordance with their compositions, while the characteristics of the coal/pmf briquettes are presented in Table 2.

Table. 1: Mass, Volume, Bulk Densities and HHV of the Coal/Corn Cob Briquettes.

S/N	Coal %	Corn cob %	Starch %	Mass (g)	Volume (cm^3)	Bulk Density (g/cm^3)	HHV (kJ/kg)
1	100	0	0	119.2	156.24	0.7629	24525
2	90	0	10	115.6	148.59	0.7779	23732.5
3	80	10	10	108.2	164.43	0.6580	22780
4	70	20	10	102.7	160.16	0.6412	21827.5
5	60	30	10	97.2	180.22	0.5393	20875
6	50	40	10	92.7	200.45	0.4625	19922.5
7	40	50	10	82.8	194.18	0.4264	18970
8	30	60	10	81.7	188.79	0.4328	18017.5
9	20	70	10	75.2	204.97	0.3669	17065
10	10	80	10	71.6	218.24	0.3281	16112.5
11	0	90	10	65.9	229.15	0.2876	15160
12	0	100	0	75.9	290.16	0.2616	15000
13	80	0	20	105.1	122.55	0.7629	22940
14	70	10	20	98.0	137.95	0.7779	21987.5
15	60	20	20	92.9	149.69	0.6580	21035
16	50	30	20	89.9	153.79	0.6412	20082.5
17	40	40	20	83.7	175.50	0.5393	19130
18	30	50	20	78.2	187.20	0.4625	18177.5
19	20	60	20	70.1	207.90	0.4264	17225
20	10	70	20	64.3	191.36	0.4328	16272.5
21	0	80	20	59.2	239.20	0.3669	15320
22	70	0	30	98.9	143.00	0.3281	22147.5
23	60	10	30	92.9	154.44	0.2876	21195
24	50	20	30	85.7	154.44	0.2616	20242.5
25	40	30	30	81.1	168.96	0.4799	19290
26	30	40	30	77.0	166.32	0.4629	18337.5

27	20	50	30	73.8	172.80	0.4271	17385
28	10	60	30	66.3	175.50	0.3778	16432.5
29	0	70	30	65.1	187.20	0.3478	15480
30	80	20	0	113.0	167.04	0.6765	22620
31	90	10	0	119.7	160.16	0.7474	23572.5
32	70	30	0	104.9	175.50	0.5977	21667.5

Table. 2: Mass, Volume, Bulk Densities and HHV of the Coal/PMF Briquettes

S/N	Coal %	PMF %	Starch %	Mass (g)	Volume (cm^3)	Bulk Density (g/cm^3)	HHV (kJ/kg)
1	100	0	0	111.1	138.591	0.801639	24525
2	90	0	10	118.8	138.288	0.859077	23732.5
3	80	10	10	114.5	161.04	0.711003	23130
4	70	20	10	116.2	165.088	0.703867	22527.5
5	60	30	10	106.9	168.299	0.635179	21925
6	50	40	10	99.1	187.968	0.527217	21322.5
7	40	50	10	98.9	179.18	0.551959	20720
8	30	60	10	88.7	197.64	0.448796	20117.5
9	20	70	10	78.8	248.31	0.317345	19515
10	10	80	10	77.4	249.744	0.309917	18912.5
11	0	90	10	74.7	246.906	0.302544	18310
12	0	100	0	80	321.9	0.248524	18500
13	80	0	20	102.9	133.5	0.770787	22940
14	70	10	20	102.6	140.94	0.727969	22337.5
15	60	20	20	97.5	133.056	0.732774	21735
16	50	30	20	92.6	146.432	0.632375	21132.5
17	40	40	20	91.1	150.15	0.606727	20530
18	30	50	20	85.6	132.526	0.645911	19927.5
19	20	60	20	73.2	194.568	0.376218	19325
20	10	70	20	71	190.08	0.373527	18722.5
21	0	80	20	72.5	216.448	0.334953	18120
22	70	0	30	89.6	133.055	0.673406	22147.5
23	60	10	30	89.7	133.584	0.671488	21545
24	50	20	30	86.8	137.28	0.632284	20942.5
25	40	30	30	79.7	161.001	0.495028	20340
26	30	40	30	79.1	135.168	0.585198	19737.5
27	20	50	30	74.2	170.346	0.435584	19135
28	10	60	30	64.8	165.184	0.39229	18532.5
29	0	70	30	65.9	170.716	0.386021	17930
30	80	20	0	113.2	148.096	0.764369	23320
31	90	10	0	106.2	144.625	0.734313	23922.5
32	70	30	0	101.2	161.28	0.62748	22717.5

Although PMF is a fluffy material, the bulk densities of the coal/corn cob briquettes were generally lower, due to the smaller aggregate size of the crushed corn cobs when compared with the palm mesocarp fibre.

HHVs of the pure samples were 24,525 kJ/kg for Coal, 16,600 kJ/kg for cassava starch, 15,000 kJ/kg for Corn Cob, and 18,500 kJ/kg for palm mesocarp fibre, respectively. Results of HHV determination indicate that the higher the percentage composition of coal in the coal/corn cob briquettes, the greater the higher heating value, whereas the reverse is the case for coal/pmf briquettes. This is easily explained from the values of the HHV of the pure components used in the study.

3.3. Results of emissions from WBT using LEMS

The major indicators (metrics) determined were High Power Thermal Efficiency (%), Low Power Specific Consumption (MJ/min/L), High Power CO (g/MJd), Low Power CO (g/min/L), High Power PM (mg/MJd), Low Power PM (mg/min/L), Indoor Emissions CO (g/min), Indoor Emissions PM (mg/min).These parameters were analyzed by categorizing the results into five Tiers as described in appendix 3.

The thermal efficiency of a stove varies from about 10% for traditional open fires (regarded as Tier 0 stoves) to above 55% for the most efficient gasifier stoves (Tier 5). It is a very useful indicator that shows how efficiently heat from the fuel is converted to useful energy by the stove, and depends on the stove-fuel configuration. As such, when using the same stove, TE can vary between fuels, but when using the same fuel, it can also vary from stove to stove. The HP Thermal efficiency recorded in the

present work ranged between 27.6% and 29.3% (typically Tier 2), averaging around 28.8% for all the briquettes.

The Low power specific consumption varied between 0.030 and 0.036 MJ/min/L for all the briquettes, which corresponds to Tier 2 performance. Briquettes with coal composition above 60% produced High Power CO in the range of 12.0 g/MJd (9.2 ppm, Tier 1), while for the briquettes with lower coal content, HPCO hovered around 10.5g/MJd (8.1 ppm, Tier 2). Low Power CO varied between 0.6 and 0.7 g/minL (See Fig. 5 and 6).

HPPM emissions were in the range of 302 – 365 mg/MJd (241.3-290.2 µg/m^3), but were generally higher for briquettes with PMF content above 50%, probably due to their oil content. LPPM emissions varied between 1.7 and 2.8 mg/min/L (8.23-9.09 µg/m^3,Tier 3). The High-Power figures are higher than the WHO recommended value of 35 µg/m^3. Indoor emissions of CO were within Tier 2 as they hovered around 0.56g/min, whereas indoor emissions of PM were in the range of 14 mg/min, equivalent to Tier 2. In all, CO emissions were generally higher than the WHO recommended standard of 10 ppm [15], but the HPPM emissions were far higher than the WHO recommended limits.

Many newer biomass cookstoves with chimneys are able tomeet the WHO Targets of 7mg/min. for $PM_{2.5}$ and 1.45 g/min. for CO when tested in the laboratory. WHO permissible exposure to CO emissions for ambient air is 10 ppm. The WHO vented stove Emission Rate Targets are based on 75% of the smoke and gases being removed up the chimney and out of the house. In their review of field studies, an average of 25% of the smoke and gas remained in the kitchen. Almost none of the residential biomass heating stoves in the United States meet the WHO Targets for $PM_{2.5}$ but the chimney transports the smoke outside where it is diluted by clean outdoor air to safe levels of concentration [16]. For a five-minutes sampling period the permissible CO emission level is 200 ppm [17]. The highest CO emissions recorded in this study was 9.2 ppm. This is within permissible limits even for a five-minutes sampling period.

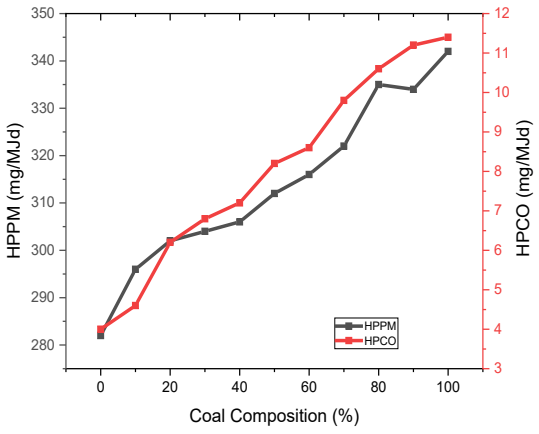


Figure 5: Graph of HPPM and HPCO versus Coal Composition for PMF Briquettes/Coal

The emissions values were generally lower than those obtained from pure coal. In [18], they reported that switching to semi-coke briquettes for household cooking can reduce average emission factors of primary $PM_{2.5}$, elemental carbon, organic carbon, and carbon monoxide by about 92%, 98%, 91%, and 34%, respectively. Although conventional coal devolatilization was not carried out in this study but from the studyon the emissions produced from the combustion of eco-fuel briquettes for domestic applications, blending coal with biomass was found to drastically reduce PM and CO emissions[19]. The CO and $PM_{2.5}$ emissions of the blended bio-coal briquettes are presented in Figures 5 and 6.

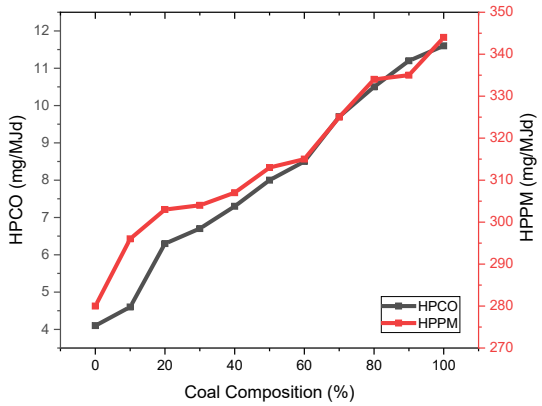


Figure 6: Graph of HPPM and HPCO versus Coal Composition for Corn Cob/Coal Briquettes

4. Findings

The work has shown that although bio-coal briquettes are suitable replacement for firewood as a climate change mitigation measure, the particulate matter (PM2.5) emissions derived from Palm Mesocarp Fibre briquettes are far above the WHO limits probably due to its residual oil content.

5. Conclusion

Bio-coal briquettes were produced from blends of Nigerian sub-bituminous coal and corn cob/palm mesocarp fibre with cassava starch as binding agent, using a purpose-built hydraulic press. Computer-assisted WBT was carried out on the fuel samples using the Laboratory Emissions Measuring System (LEMS). Results of the tests showed that average High Power Thermal Efficiency was 28.8% for Corn cob/coal and 29.2 % for PMF/coal briquettes. High Power CO emissions decreased marginally from 10.8 mg/MJd to 10.5 mg/MJd as the composition of corn cob increased from 0% to 100%. Although the bio-coal briquettes showed improved emissions characteristics over pure coal, their PM emissions were generally above WHO recommended allowable limits.

6. Acknowledgement

The authors wish to acknowledge funding support received from the World Bank Africa Centre of Excellence for Sustainable

Power and Energy Development (ACE-SPED), University of Nigeria, Nsukka.

References

[1] O, Bede-Ojimadu, O, E Orisakwe "Exposure to Wood Smoke and Associated Health Effects in Sub-Saharan Africa" A Systematic Review. Annals of Global Health. 2020; **86**(1): 32, 1–27. DOI: https://doi.org/10.5334/aogh.2725

[2] GACC "The Truth about Cooking Landscape Analysis" Global Alliance for Clean Cookstoves, October 2016.

[3] Mongabay "Forest Cover, Forest types, Breakdown of forest types, Change in Forest Cover, Primary forests, Forest designation, Disturbances affecting forest land, Value of forests, Production, trade and consumption of forest products" The Food And Agriculture Organization of the United Nations' Global Forest Resources Assessment (2005 & 2010) and the State of the World's Forests (2009, 2007, 2005, 2003, 2001), Accessed from http://rainforests.mongabay.com/deforestation on 23rd February, 2017.

[4] C. N. Ibeto, J. N. Oyedele, and C. N. Anyanwu "Evaluation of Pollution Potentials and Fuel Properties of Nigerian Sub-Bituminous Coal and its blends with Biomass" Journal of Mater. Environ. Sci. **7** (8), 2929-2937, 2016.

[5] Yuan, Ye, Yong, He, Jiaxin, Tan, Yongmeng Wang, Sunel Kumar and Zhihua Wang "Co-combustion characteristics of typical biomass and coal blends by thermogravimetric analysis" Frontiers in Energy research, **9**, 1-11, 2021, doi.org/10.3389/fenrg.2021.753622

[6] L.M Esriaty, D. Fatimah and Widodo "Bio-coal briquettes using low-grade coal" in 2018, IOP Conf. Series: Earth and Environmental Science **118**, 2018, doi:10.1088/1755-1315/118/1/012066

[7] D. Mathias "Effect of Locally Available Bitumen on the Coking Properties of Lafia Coal." Thesis, Ahmadu Bello University, Zaria, 2008.

[8] K. E. Anyaoha, R. Sakrabani, K. Patchigolla, and A. M. Mouazen "Evaluating oil palm fresh fruit bunch processing in Nigeria" Waste Management & Research, **36**(3) 236–246, 2018.

[9] Anon., 2016 https://en.wikipedia.org/wiki/Biomass_briquettes last modified on 14 March 2016, sourced 3rd December, 2018.

[10] S. Choudhurl "Briquetting of Organic Residue" proceedings from Bio-Energy Society Convention and Symposium, Delhi, **4**(2) 22-28, 1983.

[11] C. N. Anyanwu, C. N. Ibeto, S. L. Ezeoha, and N. J. Ogbuagu, Sustainability of cassava (Manihot esculenta Crantz) as industrial feedstock, energy and food crop in Nigeria**.** Renewable Energy 81:745-752, 2015, DOI:10.1016/j.renene.2015.03.075

[12] ARC "Instructions for use of the Laboratory Emissions Measuring System (LEMS)" Aprovecho, Research Center, Cottage Grove Oregon, USA, 2013

[13] Standards Organisation of Nigeria "Nigerian Industrial Standards NIS 1000:2018 Part 1, Standard for Clean Cookstoves- Solid Biomass", 2018.

[14] S.Vijayan, T.V.Arjunan, and A.Kumar "Exergo-environmental analysis of an indirect forced convection solar dryer for drying bitter gourd slices" Renewable Energy 146, 2210 -2223, 2020.

[15] WHO "World Health Organisation Air Quality guidelines for Particulate Matter", Ozone, Nitrogen dioxide and Sulfur dioxide. Global update 2005, Geneva 2006.

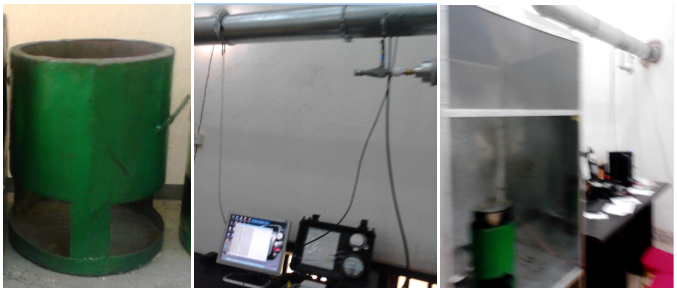
[16] ARC, "Meeting World Health Organization Emission Rate Targets" Newsletter of the Aprovecho Research Center, Cottage Grove Oregon, USA, 2021.

[17] WHO "Carbon Monoxide", Ch. 5 In: Air Quality Guidelines, 2nd Ed. World Health Organisation Regional Office for Europe, Copenhagen, 2010.

[18] Q. Li, X. Li, J. Jiang, L. Duan, S. Ge, Q. Zhang, J. Deng, S. Wang, and J. Hao "Semi-coke briquettes: towards reducing emissions of primary PM2.5, particulate carbon, and carbon monoxide from household coal combustion in China" Scientific Reports **6**, 19306, 2016. DOI: 10.1038/srep19306.

[19] T. J. Pilusa, R. Huberts and E. Muzeda "Emissions Analysis from Combustion of Eco-fuel Briquettes for Domestic Applications". Journal of Energy in Southern Africa, **24**(4), 30-36, 2013.

Appendix 1: Hydraulic Press Briquetting Machine.



Appendix 2: Photo of the Experimental Set-up

		Metric	Value	Unit	Sub-Tier
Efficiency/Fuel Use					
Tier	**3**	High power Thermal Efficiency		%	3
		Low power Specific Consumption		MJ/min/l	4
Emissions					
Tier	**2**	High power CO		g/MJ$_d$	2
		Low power CO		g/min/l	3
		High power PM 2.5		mg/MJ$_d$	4
		Low power PM 2.5		mg/min/l	3
Indoor emissions					
Tier	**3**	High power Indoor emissions CO		g/min	3
		Low power Indoor emissions CO		g/min	4
		High power Indoor emissions PM 2.5		Mg/min	3
		Low power Indoor emissions PM 2.5		Mg/min	3
Safety					
Tier	**4**	Points from 10 weighted safety parameters		points	

Tier 0 → Improving Importance → Tier 4

Appendix 3: Description of Tiers

ASTES

Advances in Science, Technology and Engineering Systems Journal Vol. 7, No. 3, 123-128 (2022)

www.astesj.com

Special Issue on Innovation in Computing, Engineering Science & Technology

ASTESJ

ISSN: 2415-6698

Antenna System Design To Increase Power Transfer Efficiency with NFC Wireless Charging Technology

Jérémy Quignon[1,2,*], Anthony Tornambe[1], Thibaut Deleruyelle[3], Philippe Pannier[2]

[1] *STMicroelectronics, Rousset, 13106, France*

[2] *Université Aix-Marseille, IM2NP (UMR-CNRS 7334), Marseille, 13013, France*

[3] *Yncrea Mediterranée, IM2NP (UMR-CNRS 7334), Toulon, 83000, France*

ARTICLE INFO

Article history:
Received: 18 March 2022
Accepted: 25 May, 2022
Online: 13 June, 2022

Keywords:
NFC technology
Wireless Charging
Power Transfer Efficiency
IoT devices
Optimized Antennas

ABSTRACT

The NFC wireless charging feature is an extension of the NFC technology that can be implanted on wearables. The purpose of this paper is to show how to increase power transfer efficiency on both transmitter and receiver antenna systems. To demonstrate this problematic, firstly this paper gives an overview of how this NFC feature is implemented (architecture, power transfer, carrier frequency, communication bandwidth…), and can be complementary to Qi technology. Then, it provides a study on how to improve the power transfer efficiency on the antennas. To perform this result, the designer can adapt some antenna parameters as coupling coefficient, quality factor, matching method, and the antenna size. If these recommendations are respected, the power transfer efficiency between the antennas could reach between 70% and 80% with the NFC charging technology.

1. NFC Technology Overview

The NFC technology is initially used for contactless communication to exchange data. It is present in wearable applications such as authentication, payment, ticketing, pairing... However, these wearables are mechanical constraints because they embed a lot of components in a limited space. Moreover, for the wearables to be waterproof, and to avoid overheating issues, mechanical links (such as buttons or charger ports) must be removed, which leads to the use of wireless charging technology for batteries. To answer these problematics, a first paper was published [1], which described this new wireless charging technology as a solution.

The NFC wireless charging feature can use the same architecture (antenna, EMI filter, matching circuit, and NFC chip) as the NFC communication feature in order to operate [2]. So, a designer can use both NFC features on the same product [3].

Furthermore, the NFC antenna is a benefit to use NFC wireless charging in targeted devices. Indeed, since antennas are not standardized, they can take every possible size as shown on Figure 1. On the other hand, they do not have to be used with ferrite. So, the wearables that cannot use the Qi charging due to its features, like these standardized antennas with its ferrite, they can use the NFC wireless charging feature. Hence, all products can use wireless charging technology.

Nevertheless, the NFC technology has benefits and drawbacks. The issues that exist in NFC communication feature are also present in NFC charging feature. The main challenge for an NFC product is to be interoperable with other NFC devices available on the market because of their very characteristics, such as the antenna shape, waveform signal, or power signal. To avoid compatibility issues, the designer must check whether the new product, with its feature-related requirements, is compliant with the different existing standards: for instance, ISO 14443, NFC Forum, and EMVCo [4].

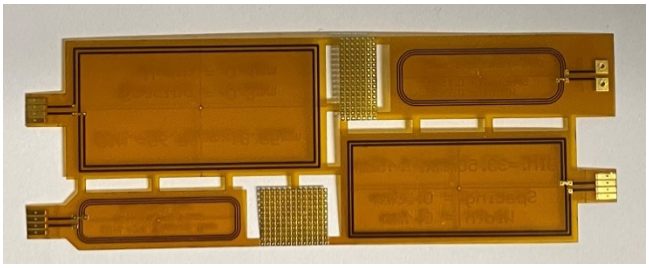


Figure 1: NFC wireless charging antenna with different sizes

*Corresponding Author: Jérémy Quignon, jeremy.quignon@st.com

Another issue is related to the NFC wireless charging, which was standardized only recently. The NFC Forum developed the standard for NFC charging at the end of 2021 to standardize the communication protocol, power supply, and product behavior in relation to issues such as foreign object detection [5]. This means that the products developed and marketed before this standardization operates with a proprietary mode.

The aim of the NFC wireless charging feature is to provide an alternative to Qi charging so that the user can benefit from better power transfer efficiency. This article describes this new wireless charging process and gives some recommendations to optimize the power transfer efficiency between the transmitter and receiver antennas, which is where the highest power losses occur.

2. NFC Wireless Charging Architecture

The NFC wireless charging feature is a radio frequency technology like NFC communication feature. It is based on a transmitter and a receiver, show on Figure 2.

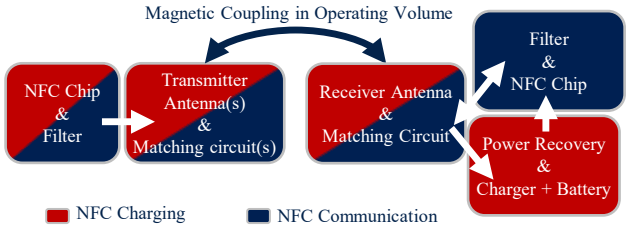


Figure 2: NFC wireless charging architecture

The transmitter uses the same architecture to perform both NFC features. It consists of:

- One NFC chip with its filter circuit, which provides power, and processes the exchanged data.
- One or two antenna(s) with its(their) matching circuit(s).

The difficulty on the transmitter board is on the matching circuit. Indeed, the calibration is different for the two NFC features [6]. So, to increase the power transfer efficiency, the transmitter must be matched in operating mode. This functionality is further described in part III.
On the other hand, to optimize the communication distance, the transmitter must be matched in the open-air mode (there is no metal object and/or antenna in the PCD's operating volume).

In this paper, we chose to use the same antenna and two different matching circuits to address both features.

The receiver, for its part, is composed of one antenna to receive the power and/or data signal with its matching circuit. The matching must be done in open air mode (without metal object in operating volume of the transmitter). Then, this antenna is connected to two different circuits:

- The NFC wireless charging circuit. It consists of a power recovery, in this architecture a diode bridge is used, and a charger chip with a battery.
- The NFC communication circuit. It consists of an EMI filter and an NFC chip to process the exchanged data.

The new NFC wireless charging uses a carrier frequency of fr=13.56MHz, as shown on Figure 3, which also gives the power limit that the device must not exceed. Indeed, the device must not generate a signal with carrier frequency higher than 60dBµA/m. Moreover, the power limit decreases with communication data rate (ΔF).

Furthermore, to enhance the power transfer efficiency, the system quality factor, Q, must be as high as possible. But, to optimize the communication bandwidth, the system quality factor must be limited.

So, if the transmitter and receiver use one antenna, it is important to find the best tradeoff in terms of quality factor to have the best power transfer efficiency sent, according to the resonance frequency fr, and to allow good communication, according to the bandwidth ΔF, between the two boards as shown on (1) and Figure 4.

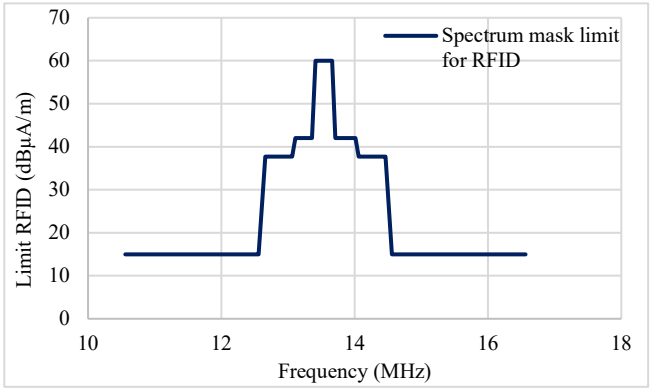


Figure 3: Power limit for RFID transmission

$$Q = \frac{f_r}{\Delta F} \quad (1)$$

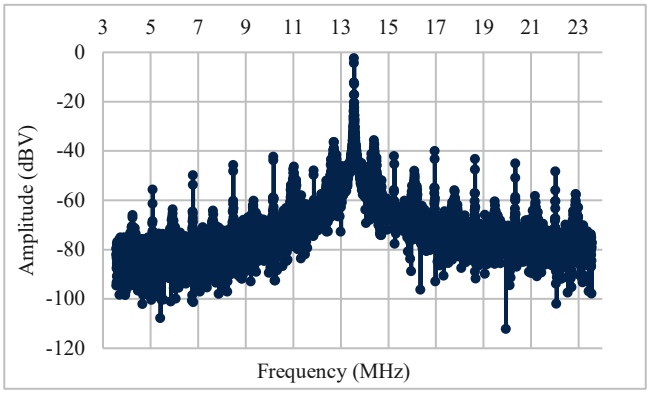


Figure 4: NFC signal with a quality factor allowing both NFC feature

The NFC Forum standardized the power transfer used for NFC charging. They indeed give several power classes as shown in Table 1. This power transfer is guaranteed at the receiver end, and the standard provides the maximum power as an indication. A designer can use this technology at higher power [7]. Moreover, we can see that NFC charging targets a lower power than the Qi power class shown in Table 2.

Table 1: NFC wireless charging power classes given by the NFC Forum.

	Power Classes	Minimum Power
Power Transfer	Class 0	250 mW
	Class 1	500 mW
	Class 2	750 mW
	Class 3	1000 mW

Table 2: Qi wireless charging power classes given by the WPC.

	Power Classes	Power
Power Transfer	BPP	5000 mW
	EPP	15000 mW

3. Optimization of antenna parameters to increase NFC power transfer efficiency

The purpose of the NFC wireless charging feature is to obtain best power transfer efficiency [8]. In order to reach this goal, it is important to mind one critical area, delimited by the transmitter and the receiver antennas themselves and the space between them. The power losses are indeed greatest in this area, and this is where the designer can increase the overall power transfer efficiency.

The Figure 5 shows the used boards to perform the test given on this chapter. The Figure 6 shows a block diagram of NFC wireless charging use to find the result exposed on this chapter.

The first parameter to increase the power transfer efficiency is to put the same antenna size on the transmitter board and the receiver board [9]. Furthermore, since this new technology is used in small devices, the selected antenna must be smaller than the size of the targeted wearable. For the tests described in this article, we used the class 6 antenna as the reference antenna for the transmitter and receiver board, as shown on Figure 7. This is the smallest antenna given by the ISO standard [10].

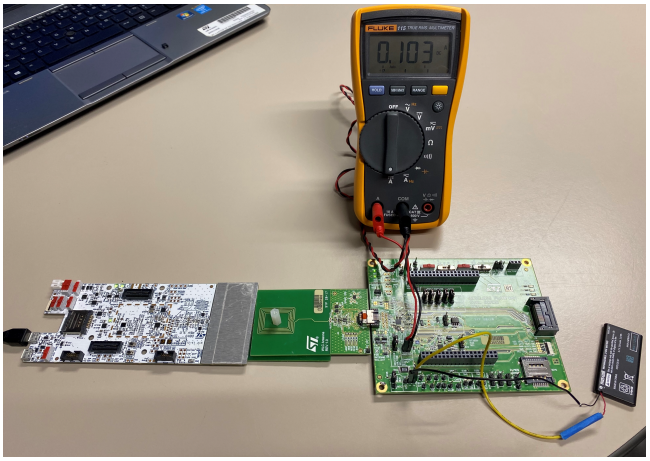


Figure 5: Power transfer efficiency measurement on test board

Therefore, if the transmitter's reference antenna is used in front of different receiver antenna sizes, as shown in Table 3, we can see, on Figure 8, the impact of the antenna design on the power transfer efficiency between the coupled antennas. The results show that the best efficiency is achieved when the PCD and PICC antennas are identical.

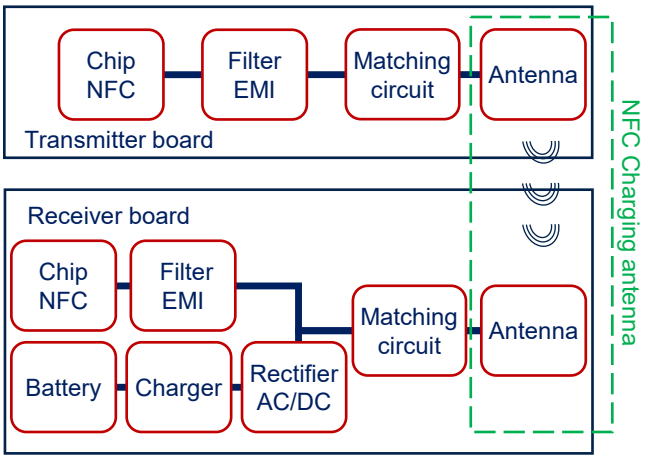


Figure 6: A block diagram of an NFC wireless charging system

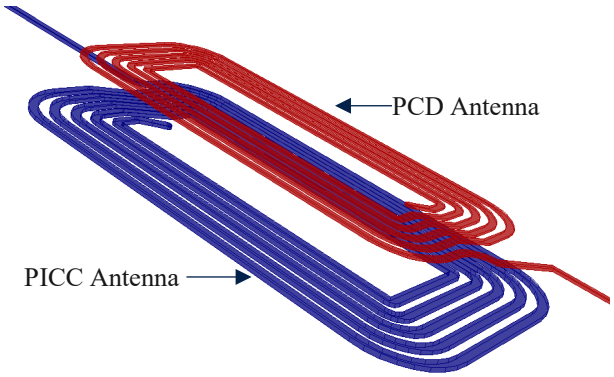


Figure 7: Transmitter and receiver antennas coupled

Table 3: Antenna design used for the different tests.

Receiver antenna size	Length	Width
Class 6 max size	25 mm	20 mm
Bigger than transmitter antenna	49 mm	30 mm
Smaller than transmitter antenna	20mm	20 mm

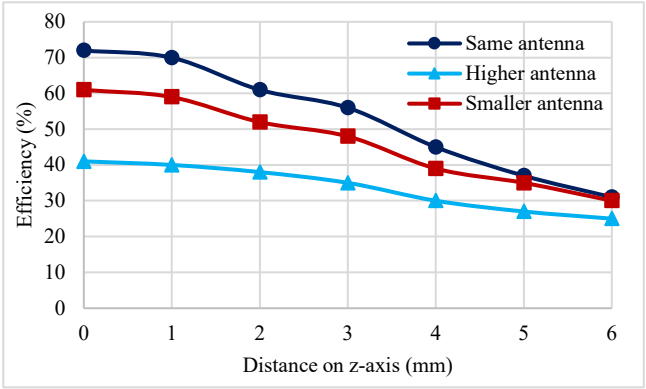


Figure 8: Power transfer efficiency simulated between the PCD and PICC antennas according to the distance along the z-axis

We can split the power losses between both transmitter and receiver antennas into three parts. The first two losses are on the real part of the PCD and PICC antenna impedance. To quantify these P_antenna losses, we measured the RMS voltage V_Antenna across the antennas. We then divided the measured value by the equivalent parallel resistance of the antenna, R_antenna, as shown on (2).

$$P_{Antenna} = \frac{V_{Antenna_{RMS}} \times V_{Antenna_{RMS}}}{R_{Antenna}} \quad (2)$$

The last loss is between the antennas due to the coupling coefficient. To decrease the impact of this parameter, the designer must use the same form factor antenna. Thus, the coupling coefficient k increases as shown on (3) [11]. It uses the equivalent radius of both the receiver (R_Rec) and transmitter R_Trans antenna. The two antennas are centered on a single x-axis.

$$k = \frac{R_{Trans}^2 \times R_{Rec}^2}{\sqrt{R_{Trans} \times R_{Rec}} \left(\sqrt{(R_{Trans}^2 + x^2)}\right)^3} \quad (3)$$

Figure 9 shows that, if we use the three receiver antenna shapes (defined in Table 3) in front of the transmitter with the reference antenna, the power transfer efficiency for a given coupling coefficient is better with the "same antenna" configuration.

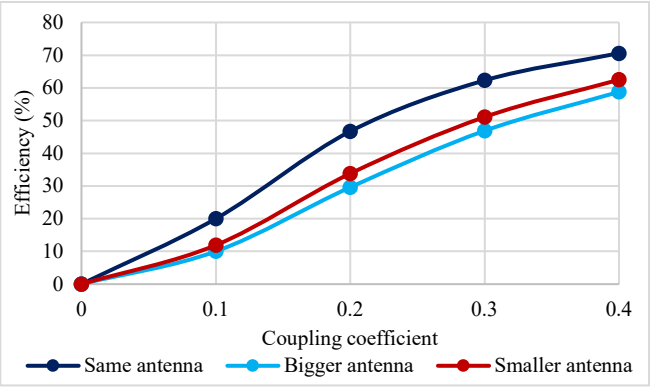


Figure 9: Power transfer efficiency simulated according to the coupling coefficient between the two antennas

The last antenna parameter that a designer can adjust to optimize the power transfer efficiency of the antenna is the antenna quality factor. Indeed, this parameter characterizes the resonance model and allows both the transmitted and received signal amplitudes to be increased [12]. Figure 10 presents the impact of the antenna quality factor at transmission (Tx) and reception (Rx) on power transfer efficiency. However, for the NFC communication use case, the transmitter and receiver quality factor must be adapted to keep a bandwidth large enough not to filter the data signal. The bigger the communication rate is, the smaller the transmitter and receiver quality is. In classical products, the transmitter and receiver quality factor are between 5 and 20.

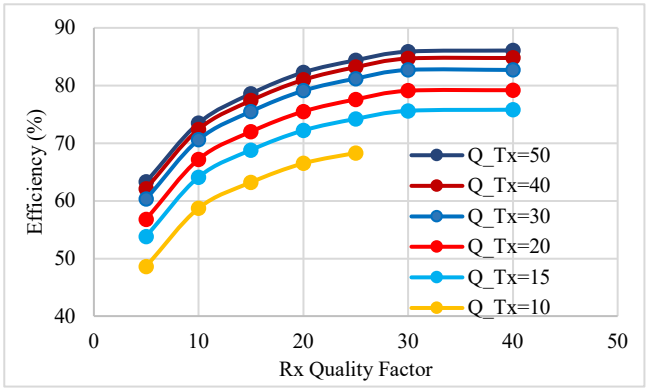


Figure 10: Impact of the quality factor on power transfer efficiency

In addition to the antenna parameters, the designer must keep in mind two other methods to optimize the power transfer efficiency. The first one is the space between both the transmitter and receiver antennas [13]. As announced in part II, both NFC feature use the same architecture. However, the operating volume is different. It is few centimeters for NFC communication whereas it is limited to only a few millimeters for NFC charging. To increase the performance, it is therefore preferable to reduce the operating volume to the location where the receiver receives the most power from the transmitter. Choosing this method also allows the transmitter and receiver boards to be in the matching position, and to increase the coupling coefficient. This again increases the power transfer efficiency.

To perform this test, a custom operating volume is created. Class 6 antennas are used on both transmitter et receiver for this test as shown on Figure 11.

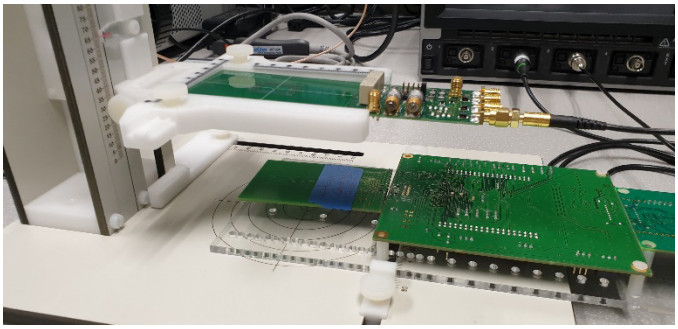


Figure 11. Operating volume test bench

Figure 12 shows power transfer efficiency curves in the custom operating volume. Initially both antennas are centered, and we apply a x offset for different z height. We can see that the power transfer efficiency is spatially limited to just a few millimeters or a single position.

The second method used to increase the power transfer efficiency even more, is to match the transmitter in operating mode. To perform this calibration, the receiver antenna is placed in operating volume of the transmitter antenna, where it receives the most power. Now, the transmitter is set up to send the maximum power when the receiver is near to it. This again shows that maximum power is delivered at the position where the

matching takes place and reduces the operating volume to a static position.

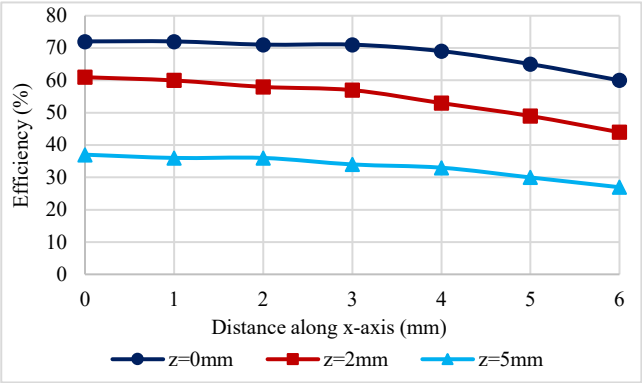


Figure 12: Power transfer efficiency in the operating volume according to the x- and z-axes

With this calibration method, the coupling effect has no more impact. Indeed, when the receiver antenna (L_Rec) is brought closer to the transmitter antenna (L_Trans) which leads to the detuning of the transmitter because the mutual inductance (M) changes as shown in (4). Because of this phenomenon the transmitter is no longer optimized to send maximum power [14]. It is why the transmitter must be matched in operating mode.

$$k = \frac{M}{\sqrt{L_{Trans} \times L_{Rec}}} \quad (4)$$

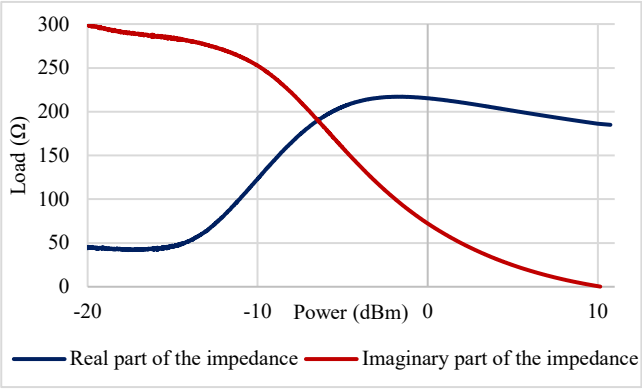


Figure 13: Diode bridge nonlinearity impact on the PICC impedance

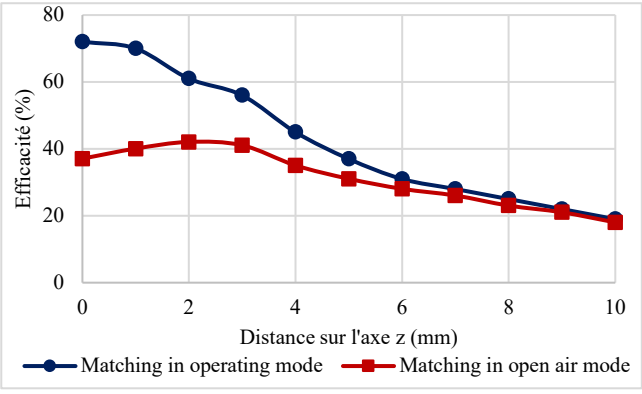


Figure 14: Power transfer efficiency depending on the matching mode

After determining the transmitter matching position, the designer must consider a second parameter to match the transmitter: the diode bridge non-linearity. Indeed, this parameter implies that the receiver impedance varies depending on the power that passes through it, as shown on Figure 13. To consider this nonlinearity, the designer must match the transmitter when it charges the receiver. Figure 14 shows the power transfer efficiency between both transmitter and receiver antennas with the impact of the matching mode used on the transmitter antenna.

4. Conclusion

This paper gives the architecture used by the NFC wireless charging feature. This technology is used to charge the embedded batteries in wearables. The designer can use the NFC products already available on the market to perform NFC charging. It must update the protocol to the new NFC Forum standard on both transmitter and receiver boards.

In addition to this protocol update, this article shows how to reach an optimized power transfer efficiency on the antennas. The results exposed in this article give the antenna parameters (antenna design, coupling coefficient, quality factor) and two additional methods (matching mode and operating volume) that a designer must pay attention to optimize the power transfer efficiency. If the advices are respected, the efficiency on antenna class 6 can reach between 70% and 80%.

Conflict of Interest

The authors declare no conflict of interest.

Acknowledgment

I would like to thank the co-authors who helped me in the writing of this paper with their patience, advice and experience. I would also like to thank the STMicroelectronics proofreading service for improving the English level of this paper.

Finally, I thank the ANRT organization which allowed the feasibility of my PHD in link with the laboratory IM2NP and the company STMicroelectronics.

References

[1] J. Quignon, A. Tornambe, T. Deleruyelle and P. Pannier, "Wireless charging using NFC technology," 6th International Conference on Smart and Sustainable Technologies", 1-4, 2021, doi: 10.23919/SpliTech52315.2021.9566468.

[2] S. Kolev, "Designing a NFC system," 56th International Scientific Conference on Information, Communication and Energy Systems and Technologies, 111-113, 2021, doi: 10.1109/ICEST52640.2021.9483482.

[3] H. Shibuya, T. Tsukuda, H. Suzuky, T. Shimizu, M. Dobashi, S. Nishizono, M.Baba, H. Sasaki, K. Terajima, "A wireless charging and near-field communication combination module for mobile applications," 64th Electronic Components and Technology Conference, 763-768, 2014, doi: 10.1109/ECTC.2014.6897371.

[4] N. S. S. Shobha, K. S. P. Aruna, M. D. P. Bhagyashree and K. S. J. Sarita, "NFC and NFC payments: A review," International Conference on ICT in Business Industry & Government, 1-7, 2016 doi: 10.1109/ICTBIG.2016.7892683.

[5] NFC Forum "Technical specification wireless charging version 2.0," 2021.

[6] L. Li, Z. Gao and Y. Wang, "NFC antenna research and a simple impedance matching method," Proceedings of 2011 International Conference on

Electronic & Mechanical Engineering and Information Technology, 3968-3972, 2011, doi: 10.1109/EMEIT.2011.6023890.

[7] https://www.nucurrent.com/nfc-wireless-charging-explained/ Accessed on January, 2022.

[8] M. Watanabe and K. Akatsu, "A study of high electrical power and high efficiency antenna in 13.56 MHz Wireless Power Transfer," Energy Conversion Congress and Exposition, 1027-1031, 2020, doi: 10.1109/ECCE44975.2020.9235829.

[9] B. Couraud, T. Deleruyelle, R. Vauche, D. Flynn and S. N. Daskalakis, "A Low Complexity Design Framework for NFC-RFID Inductive Coupled Antennas," IEEE Access, **8**, 111074-111088, 2020, doi: 10.1109/ACCESS.2020.3001610.

[10] ISO/IEC, 14443, "Cards and security devices for personal identification _Contactless proximity objects," 2018.

[11] K. Finkenzeller, RFID Handbook: Fundamentals and Applications in Contactless Smart Cards, Radio Frequency Identification and Near-Field Communication, 1999.

[12] P. Baumgartner, O. Biro, W. Renhart, R. Torchio, C. Riener, C. Koger, T. Bauernfeind, "PEEC based design optimization of NFC antennas considering the full system quality factor," Tenth International Conference on Computational Electromagnetics , 1-4, 2019 doi: 10.1049/cp.2019.0123.

[13] W. Lee, J. Kang, J. Lee, Y. -J. Hong, J. Suh and S. J. Kim, "A Compact Base Station System for Biotelemetry and Wireless Charging of Biomedical Implants," Asia-Pacific Microwave Conference, 339-341, 2019, doi: 10.1109/APMC46564.2019.9038396.

[14] M. Meiller, Impact of antenna design tune and match on wireless range, High Frequency Electronics, IEEE Antennas and Propagation Magazine, **62**(5), 95-107. 2015.

ASTES

Advances in Science, Technology and Engineering Systems Journal
Vol. 7, No. 3, 129-138 (2022)
www.astesj.com
Special Issue on Innovation in Computing, Engineering Science & Technology

ASTES Journal
ISSN: 2415-6698

Indoor Position and Movement Direction Estimation System Using DNN on BLE Beacon RSSI Fingerprints

Kaito Echizenya, Kazuhiro Kondo*

Graduate School of Science and Engineering, Yamagata University, Yonezawa, Yamagata 9928510, Japan

ARTICLE INFO

Article history:
Received: 28 March, 2022
Accepted: 05 June, 2022
Online: 24 June, 2022

Keywords:
BLE beacons
RSSI fingerprints
Indoor positioning
Deep learning

ABSTRACT

In this paper, we propose a highly accurate indoor position and direction estimation system using a simple fully connected deep neural network (DNN) model on Bluetooth Low Energy (BLE) Received Signal Strength Indicators (RSSIs). Since the system's ultimate goal is to function as an indoor navigation system, the system estimates the indoor position simultaneously as the direction of movement using BLE RSSI fingerprints recorded indoors. To identify the direction of movement along with the position, we decided to use multiple time instances of RSSI measurements and fed them to a fully-connected DNN. The DNN is configured to output the direction with the location simultaneously. RSSIs are known to be affected by various fluctuating factors in the environment and thus tend to vary widely. To achieve stable positioning, we examine and compare the effects of temporal interpolation and extrapolation as preprocessing of multiple RSSI sequences on the accuracy of the estimated coordinates and direction. We will also examine the number of beacons and their placement patterns required for satisfactory estimation accuracy. These experiments show that the RSSI preprocessing method optimum for practical use is interpolation and that the number and placement of beacons to be installed does affect the estimation accuracy significantly. We showed that there is a minimum number of beacons required to cover the room in which to detect the location if the estimation error is to be minimized, in terms of both location and direction of movement. We were able to achieve location estimation with an estimation error of about 0.33 m, and a movement estimation error of about 10 degrees in our experimental setup, which proves the feasibility of our proposed system. We believe this level of accuracy is one of the highest, even among methods that use RSSI fingerprints.

1 Introduction

In recent years, almost everyone owns a smartphone. Some of the most often used applications on smartphones are those with location and navigation capabilities, such as Google maps. Global Positioning System (GPS) signals received from satellites are used for location detection in many mapping and navigation applications. However, these applications cannot be used indoors since the reception of GPS signals is generally poor due to the building structures. As an alternative to GPS, indoor positioning systems have been developed in recent years using various methods such as WiFi, Indoor Messaging System (IMES), and other wireless LAN standards, as well as dedicated devices such as Light Detection and Ranging (LiDAR) [1–6]. However, there are many issues to be solved for the construction of the system, such as the availability of terminals supporting these standards, and the installation cost.

Many of the indoor localization systems were based on WiFi. For example, [7, 8] use Deep Networks with WiFi RSSI fingerprints. Both use Autoencoders as the DNN. It is shown in [8] that an average error of 1.21 m is possible in an apartment of 14.5 m by 4.5 m, but the number of Access Points (APs) is 59, which is considerably large for a room this size. [9] uses Channel State Information (CSI) fingerprints instead of RSSI fingerprints with an Autoencoder. CSIs provide receive levels of multiple subcarrier signals which can be collectively used to detect the receive level and direction more accurately than a simple RSSI. They were able to achieve an average error of 0.9425 m inside a living room of size 4 m by 7 m with just one AP. However, CSI is not available on all network interface cards.

In this research, we focused on BLE beacons [10, 11]. BLE is one of the short-range wireless communication standards, and it is possible to conduct wireless communication at a lower cost and lower power consumption (often operating on battery for years without the need for replacement), and are easier to install and

*Corresponding Author: Kazuhiro Kondo, 4-3-16 Jonan, Yonezawa, Yamagata 9928510, Japan, kkondo@yz.yamagata-u.ac.jp

maintain compared to other methods, such as WiFi, which generally costs much more and requires a power line. In the real world, BLE technology is quickly spreading in various forms, such as devices for cash-less payment at cash registers and beacons that provide detailed product promotion information, including discount coupons, to nearby mobile devices. In [12], localization performance using BLE beacons is compared to WiFi, and was shown that localization performance using BLE potentially outperforms WiFi. Research using the BLE has been conducted using a variety of approaches [13–19]. Application of such systems to a university campus navigation system has been explored in [20].

The theoretical values of RSSIs received from beacons are found to decay inversely proportional to the square of the distance from the beacons. However, RSSI is known to fluctuate widely in real environments [21]. This is due to the effect of walls, obstacles, and radio interferences from other radio communication devices, among others. Therefore, the Euclidean distance calculation method, which calculates the current position directly from the magnitude of RSSI, generally gives an error of several meters or more. In this research, we propose a method that uses DNNs on the collected RSSI fingerprints [22–24]. This method requires the collection of RSSI data in advance in the environment where we want to perform positioning. It is known that RSSI variations caused by people and obstacles in the environment affect the estimation significantly. In this study, we use DNNs with RSSI fingerprints as training data, and we expect that DNNs can flexibly respond to RSSI variations in position estimation.

One of our target applications for this research is a navigation system that combines an indoor positioning system with augmented reality (AR). One of the intended applications of this system is an auditory navigation system for the visually impaired [25]. In this system, audio sign signals are localized towards the target for navigation. This signal is used as a cue to indicate the direction the user should move next. Localization of the auditory sign signal is achieved by convolving the head-related transfer function (HRTF) corresponding to the direction of the sign signal with the monaural signal. Since this direction is relative to the user's current orientation (the front), it is necessary to detect this orientation at the same time as the position. Therefore, the DNN used in this research estimates the direction of movement (the orientation) as well as the position simultaneously.

There has been some research on using DNNs with BLE RSSI fingerprints (*e.g.*, [26]). However, since our intended application also requires the estimation of the movement of the user, we decided to use multiple time instances of RSSI measurements and fed this to a DNN model to estimate the current position as well as the direction of the movement of the user simultaneously [22–24]. We believe the novelty of our method is this simultaneous estimation, where we employ multiple time instances of RSSI measurements to model the dynamically changing position due to users moving in a path. By explicitly training our DNN model with changing RSSI measurements due to motion, we can detect the direction of the movement from the temporal changes in the RSSI and also estimate the position more accurately when the RSSI measurements are changing due to motion, potentially degrading accuracy with conventional stationary modeling.

There have been attempts to emulate the WiFi RSSI movement data to be fed to an RNN model [1] to estimate the movement. However, their RSSIs are measured in a static state, which does not capture the dynamic change in the RSSIs due to motion. There also have been attempts to track the user position by using both BLE beacons and Pedestrian Dead Reckoning (PDR) [27]. However, the BLE beacons seem to be only used to correct the drift caused by PDR in this work.

This paper is organized as follows. Section 2 describes the proposed method for estimating indoor position and direction of movement using DNN. Section 3 describes the conditions of RSSI preprocessing and the experiments on the placement and number of beacons to be placed. In Section 4, the experimental results are described and discussed. In Section 5, we conclude our work and discuss future challenges.

2 Proposed Method

We propose an indoor position and movement direction estimation system using a DNN that has been previously trained with RSSI fingerprints of a given room. The RSSI fingerprints, *i.e.*, RSSI measurements from all available BLE beacons, are manually collected in the actual environment (test room) to train the DNN. As stated in the introduction, we plan to eventually integrate the proposed method into a navigation system that generates audio sign signals toward the target position. To do so, the direction in which the user is moving (and also facing towards) needs to be detected so that the sign signal can be localized relative to the detected direction. Thus, we decided to use multiple time instances of the RSSI measurements as RSSI fingerprints to detect the direction from the difference between the multiple temporal RSSI measurements. In this paper, we used two time instances, *i.e.*, one set of RSSI measurements at the current time interval, and one measurement from the prior time interval. The explicit integration of motion estimation with position estimation may be what is unique with this method.

We used the RSSI measurements reported in the advertisement packets from the beacons. These advertisement packets are designed to be sent periodically to all nodes in the proximity. However, we found that the time intervals of these packets are often not periodical. Some packets were also lost. Thus, it was necessary to preprocess the RSSI measurements to obtain data in periodic intervals at the exact time instances. In cases where the packets were lost, the RSSI data was estimated from data that was received intact before or after the lost packet. The preprocessing of the RSSI measurements will be described in detail in the following section.

The training configuration of the DNN is shown in Figure 1. The RSSI set for the current time and the prior time interval is fed to the DNN. The DNN estimates the current position and the direction of movement from the given RSSI sets. This is compared to the labeled (true) position and movement direction, and the mean square error (MSE) is calculated, which is fed to the DNN to update the internal weights. This is repeated to minimize the MSE.

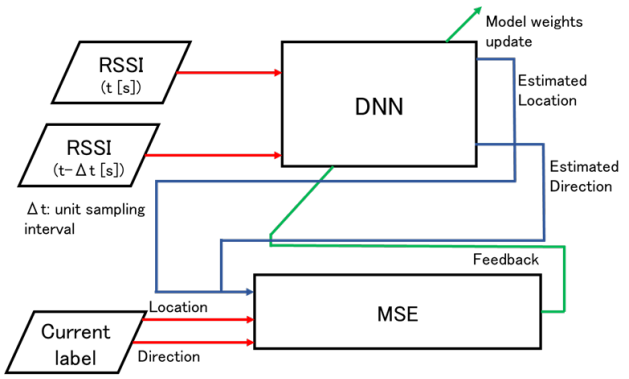


Figure 1: Block diagram of the proposed method. MSE is the Mean Square Error calculation between the estimated and the labeled position and movement direction values.

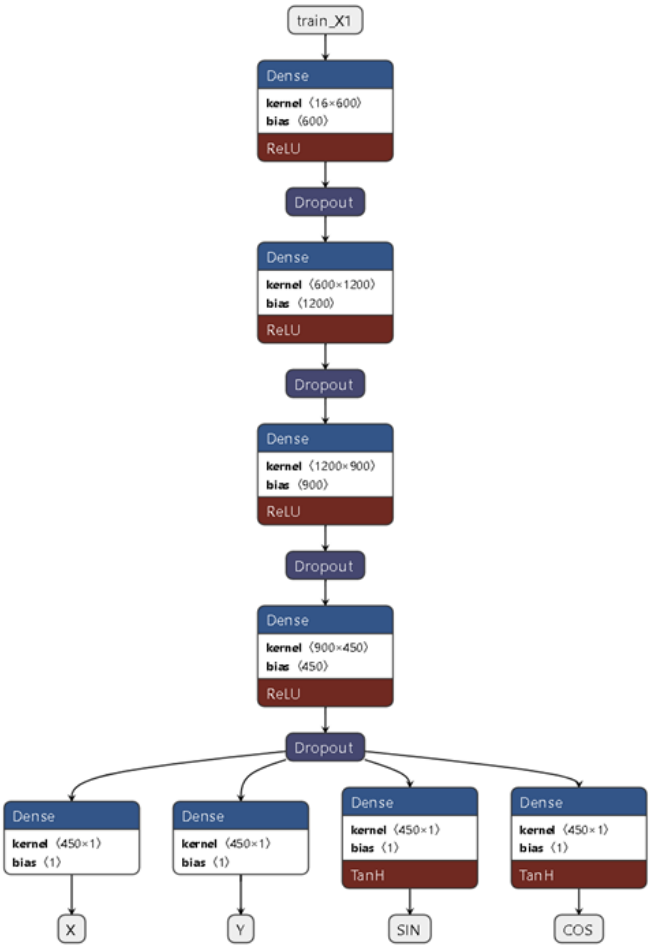


Figure 2: Configuration of the DNN for continuous values of x-y positions and movement directions.

To estimate the user's position, we feed the RSSI from all available BLE beacons for two consecutive time intervals to the trained DNN. The trained DNN will output the estimated position and direction of movement simultaneously from these inputs.

Table 1: Parameters of the DNN model.

Item		Conditions	
Units in each layer		16-600-1200-900-450-4	
Epochs		1000	
Batch size		256	
Dropout		0.5	
Activation	Hidden layers	ReLU	
	Output layers	X, Y	linear
		Dir. (cat.)	softmax
		Dir. (cont.) ($sin\theta$, $cos\theta$)	tanh

In this experiment, the RSSI was measured with no obstacles in the environment. However, since RSSI measurements collected from multiple data sequences are used as the training data of the DNN model, it is expected that estimation with some number of obstacles can be accomplished with a relatively small increase in error. The parameters of the DNN are shown in Table 1. The structure of DNN when the direction of movement is treated as a continuous value is shown in Figure 2.

As stated before, DNNs are used to simultaneously estimate the position and the direction of movement. The units in the input layer are given the RSSI of multiple beacons for two consecutive point-in-time instances in the training data. In the output layer, two units are used for the x- and y-coordinates of the estimated position and additional units are used for the estimation of the movement direction.

We modeled the motion direction as either categorical or continuous. The categorical model outputs one of eight directions (0, 45, 90, 135, 180, 225 270, or 315 °) as the estimate as one-hot output. Note that the categorical estimation values result in an estimation error, if any, of at least 45° [22], however.

On the other hand, the continuous output processes the movement directions as continuous values. The advantage of this is that the numerical difference (error) between the estimated and true values can also be expressed as continuous values. For example, in the experiment in [22], when the direction of movement was treated as a continuous value instead of a categorical value, the results showed that the frequency of fatal errors in the estimation of the direction of movement (*e.g.*, estimates that are the exact opposite of the correct value) was greatly reduced. The estimation of the angle θ, the direction of movement, is calculated according to (1) using $\sin\theta_{estimate}$ and $\cos\theta_{estimate}$ estimated as continuous values, and where $\theta_{estimate}$ is the estimated angle of movement.

$$\theta = \tan^{-1}\frac{\sin\theta_{estimate}}{\cos\theta_{estimate}} \tag{1}$$

We decided to use $\sin\theta$ and $\cos\theta$ to express the movement direction as output from the DNN. This is because these values are continuous, ranging from -1 to $+1$. If they were expressed in raw degrees, we would have to deal with discontinuities between 0 and 360° (if expressed as positive values) or $-180°$ and $+180°$ (if expressed using positive values for the right half, and negative values for the left half). The output values $\sin\theta$ and $\cos\theta$ are converted to movement direction by applying $\tan^{-1}$ to the ratio between $\sin\theta$ and $\cos\theta$ ratios (*i.e.*, $\tan\theta$). We would have to deal with discontinuities at $+90°$ and $-90°$ for this conversion, but this can be done relatively

easily by observing the positive/negative combinations of $\sin\theta$ and $\cos\theta$ values.

For the training data of the DNN, RSSI measurements were conducted when the receiver moved at a constant walking speed in one of eight predetermined directions (described in Section 3). These data were labeled with the correct position and direction. For training, 11,336 sets were used, of which 20% were used as validation data. For the test data, we used 2,434 sets.

The configuration of our proposed indoor position and movement direction estimation system is shown in Figure 3. This system is a client-server type system. First, the user's smartphone collects the RSSIs transmitted from multiple beacons at regular intervals. Next, the smartphone sends the collected RSSIs to the PC server via UDP packets. The server processes the received RSSIs, estimates the position and direction of movement using a DNN model that has been trained in advance, and sends the estimated position and direction back to the smartphone using UDP packets. By repeating the above operation, continuous estimation becomes possible. By providing DNN models that match the environment on the server-side, the computational load on the client-side can be reduced.

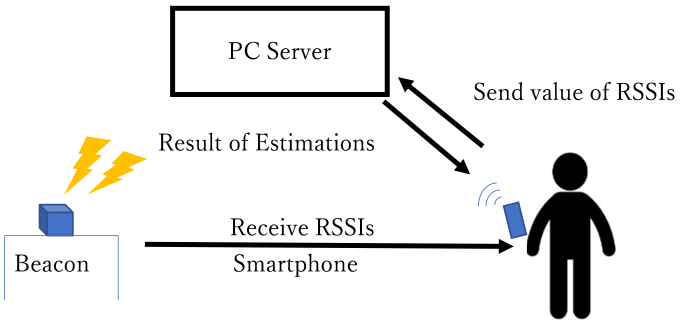


Figure 3: Proposed client-server system configuration.

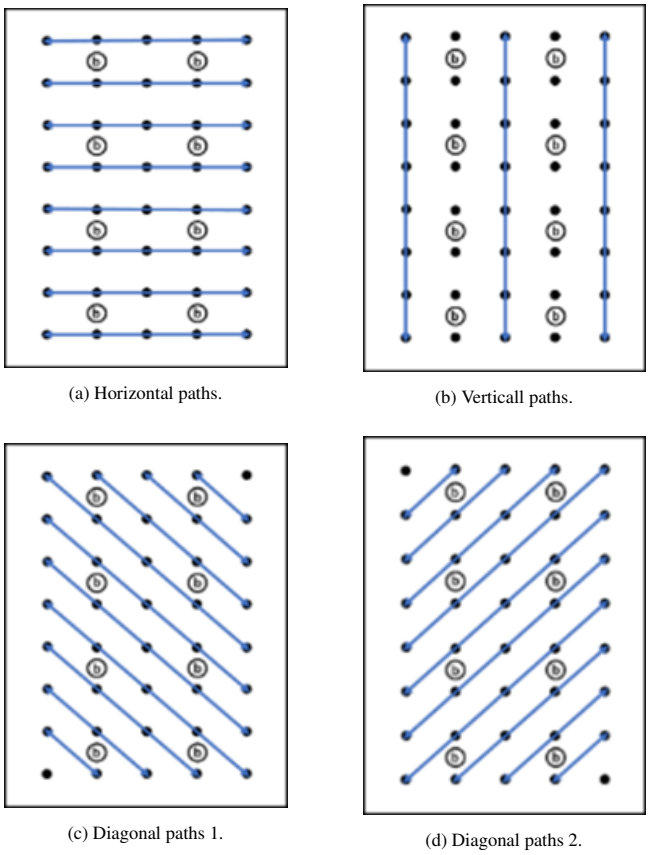
3 Experimental Conditions

We conducted two evaluation experiments explained here. The first experiment (A) was designed to evaluate the accuracy achievable using a fixed beacon position configuration and a fixed number of beacons (eight, which is the maximum number to be considered in this paper). The goal of this experiment is to compare the effect of preprocessing on the RSSIs of the estimation accuracy. The second experiment (B) was designed to evaluate the effect of the number of beacons and their positions on the estimation accuracy. The goal was to find the minimum number of beacons and their positions required for acceptable estimation accuracy.

3.1 *Experiment A: Comparison of estimation accuracy among RSSI processing methods*

The experimental environment is shown in Figure 4. As shown in this figure, eight beacons, shown here as ⓑ, were placed on chairs at regular intervals in a rectangular room (5 m x 8 m). A total of 40 grid points (5 x 8) were set at 1 m intervals. For this environment, we assumed a total of 31 vertical, horizontal, and diagonal linear paths in total. We collected data for each path by moving at a constant speed of 0.5 m/s repeatedly from the edge to the other edge of each path. We used a smartphone with BLE reception capability, SONY Xperia5 (SO-01M), as the receiving device, and Mybeacon-Pro (MB004) from Aplix Ltd. as the BLE beacons. We measured the RSSI data every 0.25 seconds while moving on each path. To acquire the RSSI data at the exact instance, temporal interpolation and extrapolation of the RSSI sequence data are considered in this experiment.

Interpolation is a method to calculate the RSSI on the estimated coordinates using the RSSI measured before and after passing through the estimated grid points, while extrapolation is a method to calculate the RSSI on the estimated grid points using the RSSI acquired before passing through the estimated grid points. When collecting RSSI data in a real environment for experiments, we want to acquire RSSI precisely on the grid point positions we want to estimate (the grid points shown in Figure 4), but this may be difficult due to discrepancies in the RSSI measurement timing. In addition, in rare cases, RSSI may not be collected from some beacons (due to internal processing delays or packet losses), and the collected RSSI sequence data may be collected at unequal time intervals. Therefore, we use interpolation and extrapolation to create RSSI data that are equally spaced in time.



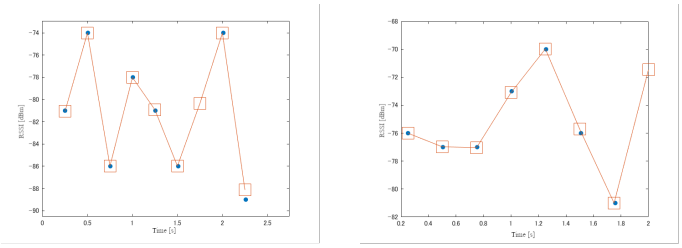
(a) Horizontal paths.

(b) Verticall paths.

(c) Diagonal paths 1.

(d) Diagonal paths 2.

Figure 4: Beacon placement and path configuration for experiment A. Beacons are shown as ⓑ, and the dots indicate the grid points at which the position and the motion direction is evaluated.



(a) An example of interpolation. (b) An example of extrapolation.

Figure 5: Examples of interpolation and extrapolation of RSSI data. The blue dots indicate the measured RSSIs, and the squares indicate the smoothed data used in the later estimation stages using DNNs.

Figure 5 shows an example of interpolation and extrapolation of RSSI data. In this figure, the collected RSSI data is represented by dots and the interpolated and extrapolated data by squares. In this experiment, RSSI data was collected while walking at a constant speed of 0.5 m/s. The subjects are expected to pass over the grid points in Figure 4 every 2 seconds. The RSSI to be acquired at exact timing is acquired from a data sequence that is supplemented at intervals of 0.25 s using interpolation or extrapolation. The interpolation is based on the piecewise cubic Hermitian interpolation polynomial (pchip) [28, 29], and the extrapolation is based on cubic spline interpolation [30, 31].

In the example in Figure 5 (a), the sample at 1.75 s was lost and was interpolated from samples before and after this interval. In Figure 5 (b), the sample at 2 s was not available and was extrapolated from preceding samples. Up to seven preceding samples and one succeeding sample in time are used to interpolate one coordinate in the collected data. Extrapolation uses up to 8 past samples. Since RSSI values in the real environment vary widely, it is not possible to process accurate RSSI data if the number of data used for interpolation and extrapolation is small. Thus, based on preliminary experiments, we decided to use the above-mentioned preceding and succeeding samples for smoothing as a compromise between accuracy and real-time constraints. We will also compare the accuracy without RSSI processing (no interpolation or extrapolation).

We will also examine whether categorization or continuous angular values are appropriate for estimating the direction of movement. The output layer of DNN is different depending on how to handle the movement direction. If the movement direction is output as a categorical value, the output layer will be 10 (x-, y-coordinates, and 8 additional outputs for one of 8 directional categories, *i.e.* $8 \times 45°$, as one-hot output). If the movement direction is output as a continuous value, the output layer of DNN will be 4 (x-, y-coordinates, and $\sin\theta$ and $\cos\theta$ values).

To compare the accuracy of the position estimation, we created a heat map showing the average estimation error for all coordinates and the average estimation error for each coordinate with each RSSI processing method. The average position estimation error, d, is calculated as in (2).

$$d = \sum_{i=1}^{N} \frac{\sqrt{(x_i - x)^2 + (y_i - y)^2}}{N} [m] \tag{2}$$

Here, N is the number of estimations, and x and y are the true coordinates, and x_i and y_i are estimated x and y positions. As stated before, to treat the movement direction as a continuous value in the DNN, the output of the DNN is set to be $\sin\theta$ and $\cos\theta$. Therefore, the error in estimating the direction of movement, θ_{error}, is calculated as shown in (3) and (4).

$$\theta_{estimate} = \tan^{-1} \frac{\sin\theta_{estimate}}{\cos\theta_{estimate}} [°] \tag{3}$$

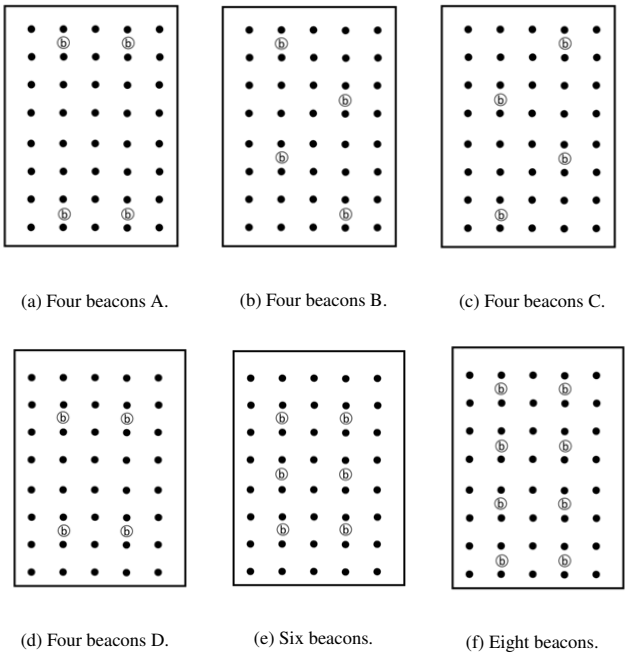
$$\theta_{error} = \frac{1}{N} \sum_{i=1}^{N} |\theta_{true} - \theta_{estimate}| [°] \tag{4}$$

Here, θ_{true} is the true direction of movement, and $\sin\theta_{estimate}$ and $\cos\theta_{estimate}$ are estimated sin and cos values output from the DNN.

3.2 *Experiment B: Comparison of estimation accuracy by placement and number of installed beacons*

The number and placement pattern of the installed beacons were set as shown in Figure 6 in this experiment to investigate the effect of beacon placement and the number of beacons.

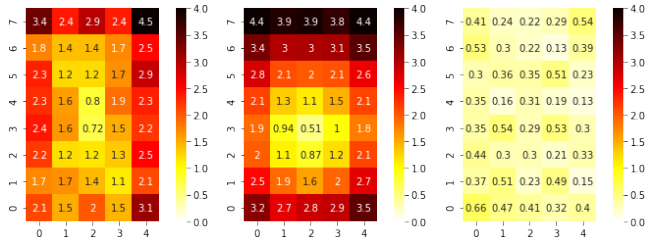
Beacons were placed on the chair at regular intervals as shown in Figure 6, and a total of 40 grid points (5×8) were set at 1 m intervals. As in Experiment A, a total of 31 vertical, horizontal and diagonal linear paths were assumed, and data were collected by moving at a constant speed of 0.5 m/s repeatedly from the edge of each path. The collected RSSI data were interpolated to obtain the exact RSSI measured at the estimated grid points. All movement directions were treated as continuous values. Figures 6 (a) to (d) show the configuration of the beacon placement to compare estimation accuracy for four patterns using four beacons. In Figures 6 (d) to (f), the number of beacons to be placed is varied to compare the estimation accuracy. The calculation of the estimation accuracy is done in the same way as in Experiment A.



(a) Four beacons A. (b) Four beacons B. (c) Four beacons C.

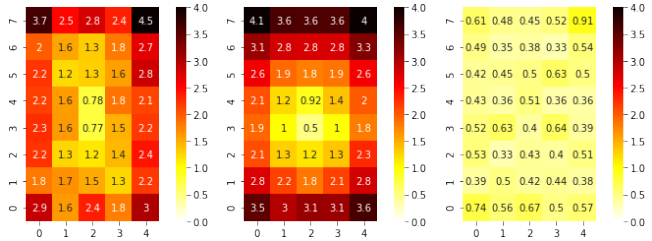
(d) Four beacons D. (e) Six beacons. (f) Eight beacons.

Figure 6: Beacon placement for experiment B. Beacons are shown as ⓑ, and the dots indicate the grid points at which the position and the motion direction is evaluated.



(a) No processing. (b) Extrapolation. (c) Interpolation.

Figure 7: Position estimation error using movement direction with **continuous** values.



(a) No processing. (b) Extrapolation. (c) Interpolation.

Figure 8: Position estimation error using movement direction with **categorical** values.

Table 3: Average direction of movement estimation error with **categorical** direction values.

Direction	Average estimation error [°]		
	No processing	Extrapolation	Interpolation
All	112.60	85.24	**10.61**
L	57.09	38.44	**10.99**
R	45.15	100.83	**12.78**
U	164.79	90.36	**10.52**
D	136.71	80.18	**11.10**
UL	135.48	65.16	**9.34**
LL	119.57	53.65	**12.96**
UR	153.00	98.91	**3.13**
LR	125.68	96.35	**15.28**

Table 4: Direction estimation error distribution by angle error range.

Error range	Occurrences	
	Continuous	Categorical
$157.5 \leq \theta \leq 180.0$	13	9
$112.5 \leq \theta \leq 157.5$	18	25
$67.5 \leq \theta \leq 112.5$	78	102
$22.5 \leq \theta \leq 67.5$	314	269
$0.0 \leq \theta \leq 22.5$	2011	2029

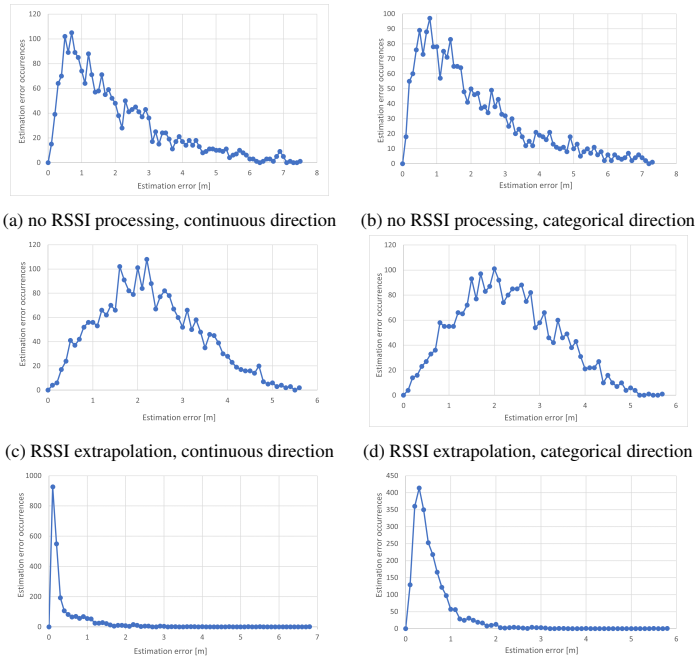
4 Results and Discussions

4.1 Experiment A

The experimental results are shown in Figure 7 through 10 and Tables 2, 3, and 4, respectively.

Table 2: Average direction of movement estimation error with **continuous** direction values. The path direction notations are L:horizontal right to left, R: horizontal left to right, U: vertical bottom to far top, D: vertical top to near bottom, UL: diagonal lower right to upper left, LL: diagonal upper right to lower left, UR: diagonal lower left to upper right, LR: diagonal upper left to lower right, and All: average over all directions.

Direction	Average estimation error [°]		
	No processing	Extrapolation	Interpolation
All	115.96	78.28	**12.96**
L	59.02	73.76	**10.75**
R	57.48	76.02	**14.72**
U	159.49	109.44	**14.38**
D	143.86	73.51	**11.12**
UL	131.55	89.99	**10.27**
LL	130.26	53.65	**15.44**
UR	156.59	90.53	**7.14**
LR	123.78	67.81	**17.68**

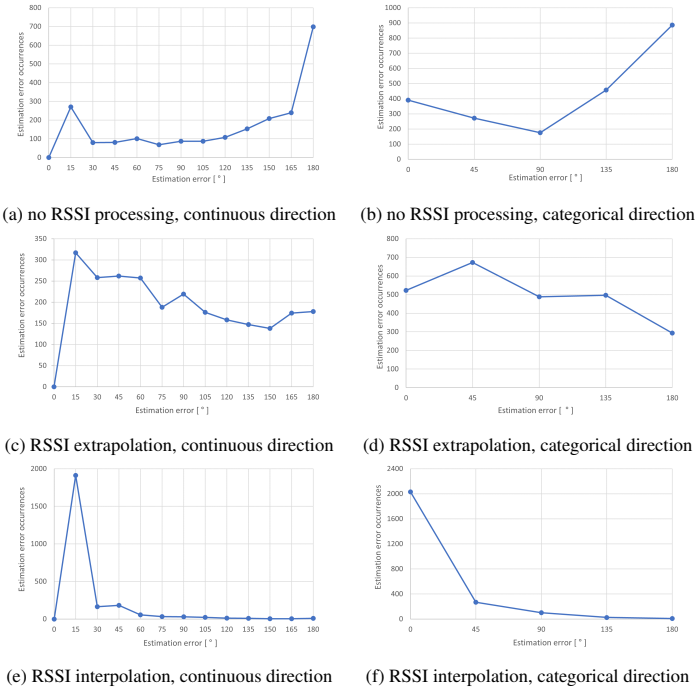


(a) no RSSI processing, continuous direction (b) no RSSI processing, categorical direction

(c) RSSI extrapolation, continuous direction (d) RSSI extrapolation, categorical direction

(e) RSSI interpolation, continuous direction (f) RSSI interpolation, categorical direction

Figure 9: Position estimation error distribution by RSSI processing method and angular expression.



(a) no RSSI processing, continuous direction
(b) no RSSI processing, categorical direction
(c) RSSI extrapolation, continuous direction
(d) RSSI extrapolation, categorical direction
(e) RSSI interpolation, continuous direction
(f) RSSI interpolation, categorical direction

Figure 10: Direction estimation error distribution by RSSI processing method and angular expression.

Figures 7 and 8 show the position estimation results when the direction of movement is treated as a continuous value and when it is treated as a categorical value, respectively. The average error at each grid position is color-coded as shown in the color bars. Figs. (a), (b), and (c) in both figures show the average estimation error at each grid position when RSSI is not processed, extrapolated, and interpolated, respectively. Note that we trained and tested the DNN with the same RSSI sequence handling, *i.e.*, applying no processing on the RSSI data in both training and testing to get (a), applying extrapolation to both the training and test data for (b), and interpolation to both the training and test data for (c), respectively.

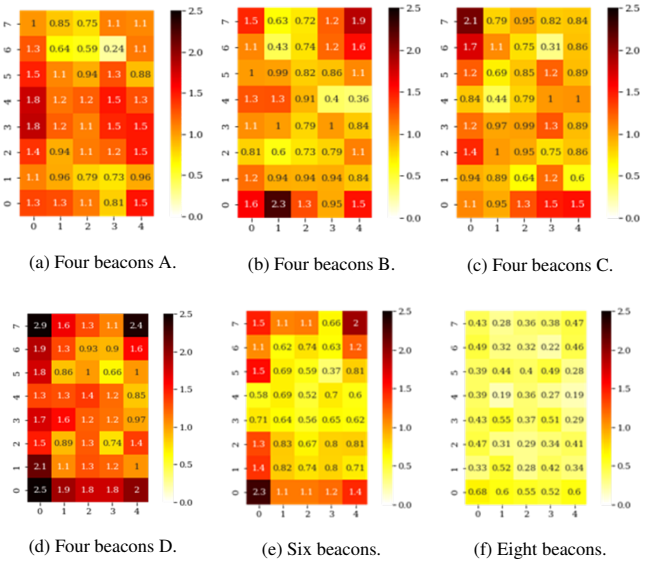
The average position estimation errors for the three methods in Figure 7 were 1.777 m, 2.069 m, and 0.330 m, respectively. The average estimation errors of the three methods in Figure 8 were 1.812 m, 2.061 m, and 0.469 m, respectively. These results show that interpolation can provide much higher position estimation accuracy. On the other hand, the difference in the treatment of the direction of movement (categorical vs. continuous) does not seem to have a significant impact on the accuracy of position estimation.

Tables 2 and 3 show the direction estimation errors when the direction of movement is treated as a continuous value and when it is treated as a categorical value, respectively. Table 4 also compares the error range distribution of both continuous and categorical direction. From Table 2 and 3, we can see that interpolation is still effective in estimating the direction of movement. However, depending on how the direction of movement is treated, the magnitude of the error remains the same for both categorical and continuous values. This is also evident in Table 4, where we can see that both methods generate approximately the same number of instances with errors in each range. It should be added, however, that the advantage of using continuous values is that if the number of categories of directions to be estimated needs to be changed, a new labeling and learning process is required, whereas this is not necessary when using continuous values.

Figures 9 and 10 show the distribution of the position estimation error and the direction estimation error for each combination of RSSI preprocessing and moving direction handling, respectively. These correspond to conditions included in Figures 7 and 8 and also Tables 2 and 3. In these figures, the occurrences of the position estimation errors are plotted for every 0.1 m interval. The error in the direction of movement is plotted in 15° increments for continuous values, and in 45° increments for categorical values. The rather large increment for categorization was because we designed the categorization in these increments, while the continuous value treatment allows arbitrary intervals, for which we arbitrarily used bins in 15° increments for error analysis. From the results, we can see that the interpolation for both position and direction of movement estimation is stable with little error overall. In particular, when the direction of movement is treated as a continuous value and the RSSI values are interpolated, the position estimation error peaks in the range of 0.0 to 0.2 m, as shown in Fig. 9e. For the direction of movement, a peak can be seen in the range of 0 to 15°, as shown in Fig. 10e . This tendency is not seen in the extrapolation or unprocessed cases, indicating the effectiveness of interpolation. The large average position estimation error with extrapolation compared to unprocessed RSSI is probably due to the inaccuracy of RSSI calculation based on past data and the inability to remove RSSI variations by extrapolation alone. However, the extrapolation succeeded in slightly reducing the estimation error of the direction of movement. From these results, we conclude that interpolation is the best method out of the three tested for processing RSSI for both indoor position and direction of movement estimation.

4.2 *Experiment B*

The experimental results of Experiment B are shown in Figures 11 through 14, and Tables 5 and 6.



(a) Four beacons A.
(b) Four beacons B.
(c) Four beacons C.
(d) Four beacons D.
(e) Six beacons.
(f) Eight beacons.

Figure 11: Position estimation error for experiment B.

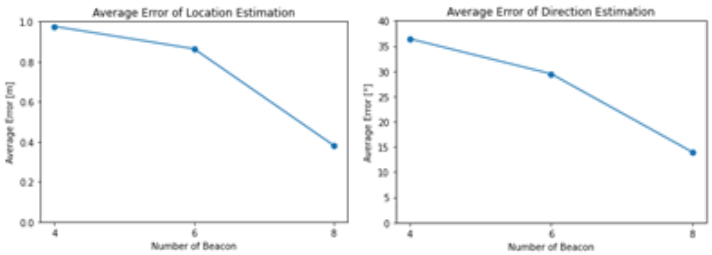
Table 5: Average direction of movement estimation error with various beacon placement pattern.

Direction	Average estimation error [°]			
	(a)	(b)	(c)	(d)
All	38.93	**36.18**	36.56	47.94
L	38.06	35.31	**32.21**	39.65
R	36.82	44.22	43.03	**35.17**
U	46.98	**36.56**	37.11	83.21
D	42.01	**37.12**	47.06	48.26
UL	44.88	39.67	**34.91**	45.81
LL	35.49	37.44	**34.85**	40.98
UR	31.83	**25.40**	29.72	35.20
LR	38.70	**33.24**	34.82	63.49

Table 6: Average direction of movement estimation error vs. the number of beacons.

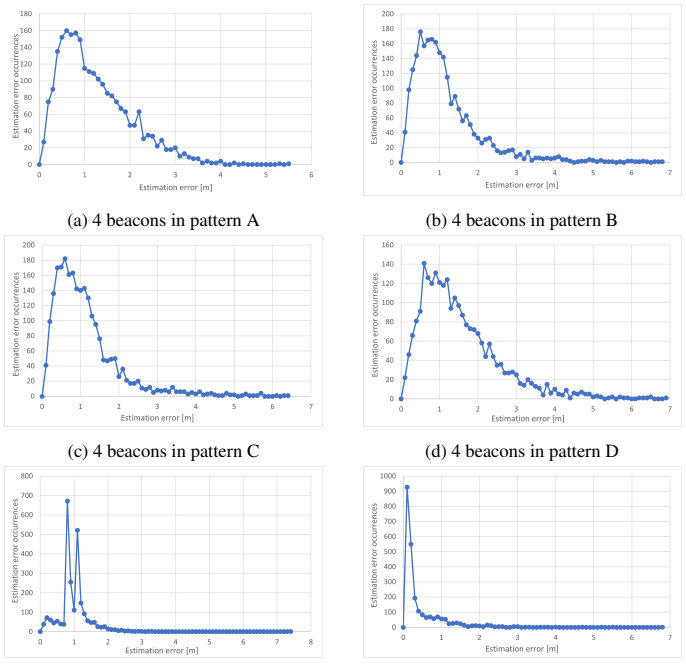
Direction	Average estimation error [°]		
	(d)	(e)	(f)
All	47.94	33.45	**10.61**
L	39.65	26.64	**10.99**
R	35.17	18.63	**12.78**
U	83.21	62.89	**10.52**
D	48.26	25.46	**11.10**
UL	45.81	32.36	**3.13**
LL	40.98	31.60	**15.28**
UR	35.20	29.99	**9.34**
LR	63.49	48.15	**12.96**

Figures 13a through 13d and 14a through 14d show the relationship between the position and direction estimation error and corresponding number of occurrences for the four beacon installation patterns (A, B, C, and D). These correspond to conditions in (a), (b), (c), and (d) in Figure 11 and Table 5. In these Figures, the position estimation error is plotted every 0.1 m, and the error in the direction of movement is plotted in 15° increments. In the case of four beacons used, we can see a large peak in the position estimation error distribution at less than 1 m in all patterns. However, the wide range and the multiple peaks of the estimation error distribution indicate that stable estimation is not possible. As for the moving direction estimation error, a peak can be seen around 15°, again in all patterns. We point out that beacon pattern D contains notable error occurrence distribution up to 180° compared to other patterns, in which the occurrences generally converge to a small number of instances at larger error values.



(a) position estimation error. (b) Direction estimation error.

Figure 12: Position and direction estimation error vs. number of beacons

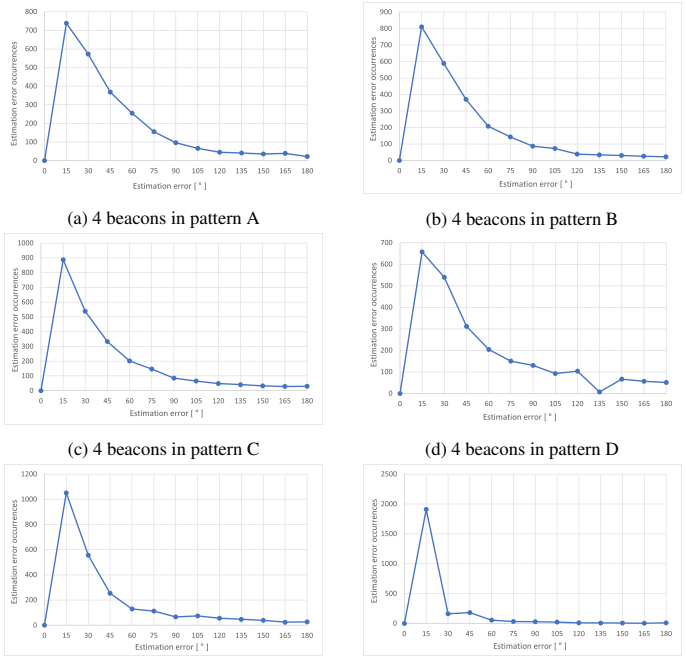


(a) 4 beacons in pattern A (b) 4 beacons in pattern B

(c) 4 beacons in pattern C (d) 4 beacons in pattern D

(e) 6 beacons (f) 8 beacons

Figure 13: Position estimation error distribution by number of beacons and its placement.



(a) 4 beacons in pattern A (b) 4 beacons in pattern B

(c) 4 beacons in pattern C (d) 4 beacons in pattern D

(e) 6 beacons (f) 8 beacons

Figure 14: Direction estimation error distribution by number of beacons and its placement. Directions were treated as continuous values.

Figures 13d through 13f and 14d through 14f show the relationship between the estimation error and the corresponding number of occurrences with 4, 6, and 8 installed beacons. They correspond to conditions in (d), (e), and (f) in Figure 11 and Table 6. In terms of position estimation error, the distribution of estimation error is shown in a rather wide range, often showing a notable number of

Table 7: Comparison of average position estimation errors.

System	Ave. loc. est. error [m]	Room size [m]	No. of APs/beacons
Proposed	0.33	5 x 8	4 to 8
WiFi CSI+Autoencoder [9]	0.9425	4 x 7	1
WiFi fing. + Autoenc. (WiDeep) [8]	1.21	14.5 x 4.5	59
BLE DOD [17]	0.54	16 x 10	2
BLE fing. + 3D-CNN [26]	0.72	3 x 6	8

occurrences up to about 3 m with four beacons. In the case of six installed beacons, there are two large peaks close to 1 m. In the case of eight installed beacons, there is only one large peak close to 0 m. These seem to indicate that increasing the number of beacons contributes to stabilizing the position estimation. As for the error in the estimation of the direction of movement, the peak of the error generated is in the range of 0 to 15° for any number of beacons, but the large estimation error is reduced when the number of beacons is increased. In particular, when there are 8 beacons installed, the percentage of estimation errors of 45° or more for all directions of movement in the test data is 7.3%. We believe this is sufficient accuracy for the intended application of our system.

Table 6 also compares the estimation error of the direction of movement when the number of beacons is 4, 6, and 8. From this Table, we can see that the estimation error becomes smaller as the number of beacons increases, similar to the position estimation results. The estimation accuracy percentage can be calculated using (5), where $\theta_{estimation\_error}$ [°] ($0 \leq \theta_{estimation\_error} \leq 180$) is the angle estimation error.

$$1 - \frac{\theta_{estimation\_error}}{180}[\%] \tag{5}$$

From (5), the estimation accuracy of 4 beacons is 73.4%, that of 6 beacons is 81.4%, and that of 8 beacons is 94.1%. Therefore, the estimation accuracy of 8 beacons is 20.7% higher than that of 4 beacons. From the above results, we can say that increasing the number of beacons contributes significantly to the improvement of stable estimation accuracy, both in terms of position and motion direction. Thus, for a room with a size similar to the one in this study, the average value of the position estimation error can be expected to be well below 1 m, and the average movement direction estimation error can be expected to be around 10° if 8 beacons are used, which can both be considered to be sufficient accuracy for practical indoor navigation.

Although it is quite difficult to compare estimation results with other schemes since they will differ in terms of conditions (room size, number of APs/beacons and their placement, etc.), we have summarized the position estimation error comparison in Table 7. From this table, it seems safe to say that the proposed system achieves comparable or better estimation accuracy compared to the other similar methods. There is no data on movement direction estimation accuracy to compare with the proposed method, however.

5 Conclusion

In this study, we investigated an indoor position and direction estimation system consisting of a smartphone and BLE beacons. The RSSI measurements of multiple BLE beacons were fed into a DNN model that estimated both position and direction of movement simultaneously. Multiple time instances of the RSSI of the BLE beacons were used for the estimation. The post-processing of the RSSI of the BLE beacons and the number of beacons required for accurate estimation were considered. The results showed that temporal interpolation is the most effective post-processing of RSSIs before using them as input data for DNN models. In particular, interpolation contributes to the stabilization of the accuracy of the estimation of the direction of movement. Eight beacons in a space of 5 m × 8 m resulted in an average position estimation error of 0.330 m for all grid positions in this room, and a total error of 10.61° in the estimation of the direction of movement. We believe this is one of the most accurate estimations even among methods that rely on fingerprints. We also showed that the estimation of the direction of movement should be treated as a continuous value to increase the accuracy of the direction of movement estimation. Regarding the placement pattern and the number of beacons required, we confirmed that sparsely and evenly distributed beacons and a sufficient number of beacons contribute to stable estimation. In the measurement tested room, when 8 beacons were installed, the percentage of estimation errors of 45° or more for all directions of movement in the test data was 7.3%, significantly less than with a smaller number of beacons. With this level of accuracy, we believe that the system can be applied to actual indoor navigation systems.

In this experiment, we measured RSSIs with no obstacles in the surroundings. However, when we consider actual usage scenarios, objects and people in the room often become obstacles. Since we used multiple RSSI sequences to train the DNN model, the models may be able to handle some deviation in the RSSI measurements due to object placement owing to the generalization capabilities of the DNN. However, some form of reinforced learning will become necessary to handle an extensive amount of deviations due to the introduction of obstacles. This is currently out of the scope of this paper but is planned to be investigated in the future. We also assume the system to be used on the same floor and thus assume that the z-axis values are fixed. Multiple floor navigation systems will be investigated in the future as well.

Additionally, all user paths modeled in this work were simple straight paths at constant speeds. We would like to further expand our model to handle more complex paths, consisting of different direction and speed combinations.

Conflict of Interest The authors declare no conflict of interest.

Acknowledgment The authors would like to thank the members of the Kondo Laboratory for their useful discussions.

References

[1] H.-G. Shin, Y.-H. Choi, C.-P. Yoon, "Movement Path Data Generation from Wi-Fi Fingerprints for Recurrent Neural Networks," Sensors, **21**(8), 2021, doi:https://doi.org/10.3390/s21082823.

[2] D. Manandhar, K. Okano, M. Ishii, H. Torimoto, S. Kogure, Maeda, Hiroaki, "Development of Ultimate Seamless Positioning System Based on QZSS IMES," in 21st International Technical Meeting of the Satellite Division of The Institute of Navigation (ION GNSS 2008), 1698–1705., Savannah, GA, 2008.

[3] Y.-T. Wang, C.-C. Peng, A. Ravankar, A. Ravankar, "A Single LiDAR-Based Feature Fusion Indoor Localization Algorithm," Sensors, **18**(4), 2018, doi: https://doi.org/10.3390/s18041294.

[4] A. Yassin, Y. Nasser, M. Awad, A. Al-Dubai, R. Liu, C. Yuen, R. Raulefs, E. Aboutanios, "Recent Advances in Indoor Localization: A Survey on Theoretical Approaches and Applications," IEEE Communications Survey & Tutorials, **19**(2), 1327–1346, 2017, doi:https://doi.org/10.1109/COMST.2016.2632427.

[5] A. Correa, M. Barcelo, A. Morell, J. L. Vicario, "A Review of Pedestrian Indoor Positioning Systems for Mass Market Applications," Sensors, **17**(8), 2017, doi:https://doi.org/10.3390/s17081927.

[6] S. H. Marakkalage, B. P. L. Lau, Y. Zhou, L. Ran, C. Yuen, W. Q. Yow, K. H. Chong, "WiFi Fingerprint Clustering for Urban Mobility Analysis," IEEE Access, **9**, 69527–69538, 2021, doi:https://doi.org/10.1109/ACCESS.2021.3077583.

[7] Z. E. Khatab, A. Hajihoseini, S. A. Ghorashi, "A Fingerprint Method for Indoor Localization Using Autoencoder Based Deep Extreme Learning Machine," Sensors, **2**(1), 2018, doi:https://doi.org/1109/LSENS.2017.2787651.

[8] M. Abbas, M. Elhamshary, H. Rizk, "WiDeep: WiFi-based Accurate and Robust Indoor Localization System using Deep Learning," in IEEE International Conference on Pervasive Computing and Communication, Kyoto, Japan, 2019, doi:https://doi.org/10.1109/PERCOM.2019.8767421.

[9] X. Wang, L. Gao, S. Mao, S. Pandey, "CSI-based Fingerprinting for Indoor Localization : A Deep Learning Approach," IEEE Transactions on Vehicular Technology, **66**(1), 763–776, 2017, doi:https://doi.org/10.1109/TVT.2016.2545523.

[10] C. Takahashi, K. Kondo, "Indoor Positioning Method for Augmented Audio Reality Using iBeacons," in 2015 IEEE Global Conference on Consumer Electronics (GCCE 2015), 468–469, Osaka, Japan, 2015, doi:https://doi.org/10.1109/GCCE.2015.7398636.

[11] C. Takahashi, K. Kondo, "Accuracy Evaluations of an Indoor Positioning Method Using iBeacons," in 2016 IEEE Global Confer-ence on Consumer Electronics (GCCE 2016), 425–426, Kyoto, Japan, 2016, doi:https://doi.org/10.1109/GCCE.2016.7800465.

[12] X. Zhao, Z. Xiao, A. Markham, N. Trigoni, Y. Ren, "Does BTLE measure up againstWiFi? A comparison of indoor location performance," in European Wireless 2014, 263–268, Barcelona, Spain, 2014.

[13] Z. Chen, H. Zou, H. Jiang, Q. Zhu, Y. Soh, L. Xie, "Fusion of WiFi, Smartphone Sensors and Landmarks Using the Kalman Filter for Indoor Localization," Sensors, **15**(1), 715–732, 2015, doi:https://doi.org/10.3390/s150100715.

[14] M. Nakano, N. Komuro, K. K., "Indoor Positioning Method based on BLE Location Fingerprint with Statistics Ap-proach," in 2019 IEEE 8th Global Conference on Consumer Electronics (GCCE 2019), 1160–1163, IEEE, Osaka, Japan, 2019, doi:https://doi.org/10.1109/GCCE46687.2019.9015459.

[15] P. Kriz, F. Maly, T. Kozel, "Improving Indoor Localization Using Bluetooth Low Energy Beacons," Mobile Information Systems, **2016**, 1–11, 2016, doi: https://doi.org/10.1155/2016/2083094.

[16] S. Sadowski, P. Spachos, "Optimization of BLE Beacon Density for RSSI-Based Indoor Localization," in 2019 IEEE International Conference on Communications Workshops (ICC Workshops), Shanghai, China, 2019, doi: https://doi.org/10.1109/ICCW.2019.8756989.

[17] K. Kikuchi, R. Tazawa, N. Honma, "DOD-Based Indoor Localization Using BLE Beacons," in 2017 Microsoft Indoor Localization Competition IPSN 2017, Pittsburgh, PA, USA, 2017.

[18] Y. R. Kumar, J.-Y. Pyun, "BLE beacon based Indoor Positioning using Variance-Weighted KNN," in Symposium of the Korean Institute of Communications and Information Sciences, 128–129, Seoul, Korea, 2018.

[19] H. Yanagimoto, K. Hashimoto, T. Matsuo, "Indoor Positioning Estimatio Using BLE Beacons," in The 13th International Joint Symposium on Artificial Intelligence and Natural Language Processing (iSAI-NLP 2018), 124–128, Pattaya, Thailand, 2018.

[20] M. Hadwan, R. U. Khan, K. I. M. Abuzanouneh, "Towards a Smart Campus for Qassim University: An Investigation of Indoor Navigation System," Advances in Science, Technology and Engineering Systems Journal, **5**(6), 831–837, 2020, doi:https://doi.org/10.25046/aj050699.

[21] J. Paek, J. Ko, H. Shin, "A Measurement Study of BLE iBeacon and Geometric Adjustment Scheme for Indoor Location-Based Mobile Applications," Mobile Information Systems, **2016**, 2016, doi:https://doi.org/10.1155/2016/8367638.

[22] K. Echizenya, K. Kondo, "Estimation of indoor position and motion direction for smartphones using DNN to BLE beacon signal strength," in 2020 IEEE International Conference on Consumer Electronics Taiwan, (ICCE-TW), Taoyuan, Taiwan, 2020, doi:https://doi.org/10.1109/ICCE-Taiwan49838.2020.9258282.

[23] K. Echizenya, K. Kondo, "Comparison of RSSI Processing Methods for Improved Estimation Accuracy in BLE Indoor Position and Movement Estimation System Using DNN," in 2021 IEEE Global Conference on Consumer Electronics (GCCE 2021), 171–172, IEEE, Kyoto, Japan, 2021, doi: https://doi.org/10.1109/GCCE53005.2021.9622021.

[24] K. Echizenya, K. Kondo, "The Effect of Density and Placement of BLE Beacons on Indoor Location and Motion Direction Estimation Accuracy," in 2021 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA-ASC), APSIPA, Tokyo, Japan, 2021.

[25] T. Kitagawa, K. Kondo, "Evaluation of a Visually Handicapped Navigation System Using 3D Audio on a Smartphone under Realistic Conditions," in 2018 IEEE Global Conference on Consumer Electronics (GCCE 2018), 467–468, Nara, Japan, 2018, doi:https://doi.org/10.1109/GCCE.2018.8574718.

[26] K. Tasaki, T. Takahashi, S. Ibi, S. Sampei, "3D Convolutional Neural Network-Aided Indoor Positioning Based on Fingerprints of BLE RSSI." in 2020 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA-ASC), Auckland, New Zealand, 2020.

[27] Z. Chen, Q. Zhu, H. Jiang, H. Zou, Y. Soh, L. Xie, R. Jia, C. Spanos, "An iBeacon Assisted Indoor Localization and Tracking System," in The 14th International Conference on Information Processing in Sensor Networks (IPSN '15), Seattle, WA, USA, 2015.

[28] Mathworks, "Piecewise Cubic Hermite Interpolating Polynomial (PCHIP)," https://www.mathworks.com/help/matlab/ref/pchip.html.

[29] F. N. Fritsch, R. E. Carlson, "Monotone Peicewise Cubic Interpolation," SIAM Journal of Numerical Analysis, **17**, 238–246, 1980, doi:https://doi.org/10.1137/0717021.

[30] Mathworks, "Cubic spline data interpolation (spline)," https://www.mathworks.com/help/matlab/ref/spline.html.

[31] C. de Boor, A Practical Guide to Splines, Applied Mathematical Sciences, Springer-Verlag, 1978.

ASTES

Advances in Science, Technology and Engineering Systems Journal Vol. 7, No. 3, 139-151 (2022)

www.astesj.com

Special Issue on Innovation in Computing, Engineering Science and Technology

ASTESJ

ISSN: 2415-6698

Automated Robotic System for Sample Preparation and Measurement of Heavy Metals in Indoor Dust Using Inductively Coupled Plasma Mass Spectrometry (ICP-MS)

Heidi Fleischer[1,*], Sascha Statkevych[2], Janne Widmer[2], Regina Stoll[3], Thomas Roddelkopf[1], Kerstin Thurow[4]

[1]*University of Rostock, Faculty of Computer Science and Electrical Engineering, Institute of Automation, Rostock, 18119, Germany*

[2]*Innerstädtisches Gymnasium (Senior High School), Rostock, 18055, Germany*

[3]*University Medical Center, Institute for Preventive Medicine, Rostock, 18055, Germany*

[4]*Center for Life Science Automation (celisca), Rostock, 18119, Germany*

ARTICLE INFO

Article history:
Received: 05 April, 2022
Accepted: 25 May, 2022
Online: 24 June, 2022

Keywords:
Automated measurement
Compound-oriented measurement
Environmental samples
Heavy metals
ICP-MS
Indoor dust
Robotic system

ABSTRACT

Dust is ubiquitous in our daily environment—outdoor and indoor. In modern times, people often spend the majority of their time at home, in offices, at work or in schools. Suspended particles such as tiny crumbs up to long fibers generate indoor dust deposits. Inhouse sources are the interior releasing abraded fibers from carpets, bedding and clothing as well as the human itself distributing skin cells, lost hairs and food residues. External sources are finest sand, pollen, exhaust particles and microorganisms (e.g., dust mites). An exposure to heavy metals in certain concentrations may affect the human health and may lead to intoxication, allergies or carcinogenic effects. The heavy metals amount in indoor dust depends on the environmental conditions, requiring a sampling with adequate sampling points and numbers. High sampling numbers ensure good coverage of the area to be examined. Therefore, fast and reliable measurement methods for identification and quantification of the elemental composition are needed. To meet these requirements, a robotic system for automated sample preparation and determination of heavy metals in indoor dust using ICP-MS was developed. The values for repeatability, recovery rate, within-laboratory precision, measurement precision and the limits of detection and quantification were determined for both, the manual and the automated process. Furthermore, the individual process steps and times were compared. Besides the processing of certified reference material, inhouse dust samples from different origin were prepared and measured to give a first overview of the inhouse dust composition. The results show, that the robot-assisted system is well-suited for the heavy metal screening in indoor dust.

1. Introduction

This paper is an extension of "Robot-assisted Measurement of Heavy Metals in Indoor Dust Using ICP-MS" presented at the IEEE International Instrumentation and Measurement Technology Conference (I2MTC) 2021 in Glasgow (virtual) [1]. The focus of this study is related to the determination of the elemental composition of dust in our close environment. In the daily life, dust is omnipresent—outdoors and indoors. Especially, the indoor air quality should be focused since their contamination can be higher than outdoors [2, 3]. Today, urban people spend more than 90% of their time indoors—at home, in schools, offices etc. [4]. Furthermore, indoor dust contains suspended particles from internal origins such as smoking, cooking, fuel combustion and decorative material (e.g., paints, carpets, furniture etc.) and external origins as soil, mining, smelting, industrial activities and vehicular emissions [4-6]. And the human itself and pets can be a dust source by distributing food residues, loose hair and skin particles as well as fibrous material from clothes. Both, inorganic and organic contaminants (e.g., biological material, pesticides, PCBs, PAHs, and heavy metals) can be found in indoor dust and may be absorbed and accumulated in the human organism. The heavy metal amount in household dust can be elevated in comparison to outdoor dust such as pavement dust, road dust or garden soil [4, 7]. Heavy metals cannot be degraded. An exposure to individual high concentrations can impact the human health significantly. Allergies, signs of poisoning and cancer diseases may result [2, 4, 8-10]. Several carcinogenic effects to the

*Corresponding Author: Heidi Fleischer, University of Rostock, Institute of Automation, Friedrich-Barnewitz-Strasse 8, 18119 Rostock, Germany, email: Heidi.Fleischer@uni-rostock.de

respiratory organs, the cardiovascular and nervous system as well as to the growth development were reported [2]. Heavy metals can be uptaken into the body in different ways—oral, over the respiratory organs or the skin. Particularly, kids are endangered since they often play on floors [2, 11].

Due to the wide variety of origins, heavy metals cannot completely be avoided in household dust. Therefore, a monitoring of heavy metal concentrations to determine influencing factors and potential health risks is required. A review from the year 2016 gives a comprehensive insight into multiple studies to determine heavy metals in indoor dust [2]. A widespread study from the year 2021 is based on more than 127 articles published between 1985 and 2019 and shows the spatial distribution, sources, and consequences of heavy metal concentrations in indoor dust [4].

The high number of studies in the field of heavy metal determination in indoor dust show an increasing demand for fast and reliable sample preparation and measurement methods able to process multiple samples in a short time frame. Such measurement methods are "compound-oriented measurements" following the "concept of the pre-, intra- and post-sensory selectivity" [12]. In the next section of this publication, this concept is explained regarding the heavy metal determination in household dust samples. Furthermore, an overview of conventional elemental measurement techniques for heavy metal determination as well as of current automation approaches is given. The existing gap and the aim of this study is explained. The system concept and the system design is printed in the third section. The materials and methods are described in the fourth section. The fifth section shows the validation results, a comparison of the automated analytical measurement and the manual process, including the results for the first exemplarily dust samples collected at various inhouse places. A summary and a forecast to future investigations are provided in the last section.

2. Heavy Metal Measurement

2.1. Compound-oriented Measurements

From a metrological perspective, the qualitative and quantitative analysis of heavy metals in indoor dust is assigned to "compound-oriented measurements" [12]. A wide variety of elemental (e.g., heavy metals) and compound species is existing, and no specific single sensors are available for the individual species. Therefore, a generally applicable concept was developed for determination of single elements and chemical compounds in complex mixtures to reach the required selectivity [12]. The "concept of the pre-, intra-, and post-sensory selectivity" is mainly focused to species which are included in a complex matrix causing interfering signals [12]. In our case, the heavy metals to be measured are embedded in dust—particles, crystals, fibers, biologic material, cells etc.

The "pre-sensory selectivity" is reached in the first step of the concept. The heavy metals to be determined must be separated from the matrix dust. This includes sampling, homogenizing by milling and mixing, weighing, a microwave-assisted acid digestion and a final dilution. After these sample preparation steps, the heavy metals were transferred from the solid dust matrix into a liquid measurement solution. This separation step takes place before the sensor system (pre-sensory selectivity). Inductively coupled plasma mass spectrometry (ICP-MS) was used to achieve a high degree of "intra-sensory selectivity". The ICP-MS data must be evaluated, interpreted and visualized in the last step to achieve qualitative (identification of metal species) and quantitative information—the "post-sensory selectivity". Fig. 1 visualizes the individual process steps for heavy metal measurements in dust samples and shows the relationship to the "concept of the pre-, intra- and post-sensory selectivity".

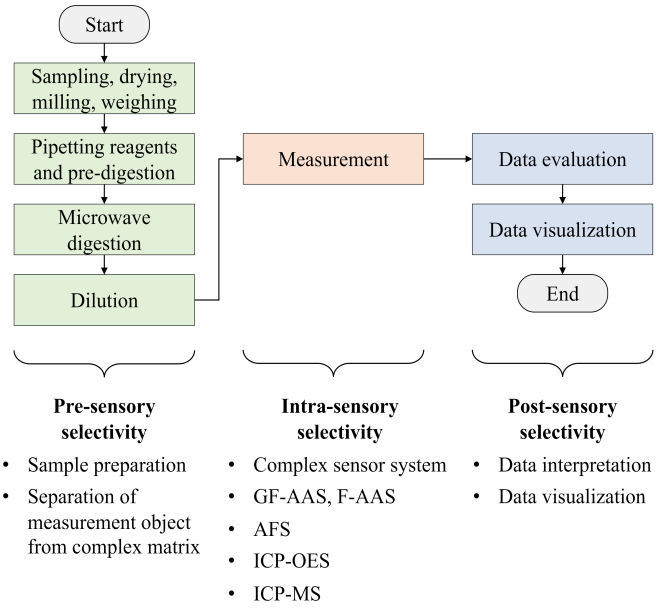


Figure 1: Process steps for heavy metal measurements in dust samples and their relationship to the "concept of the pre-, intra-, and post-sensory selectivity" (abbreviations: GF-AAS—graphite furnace atomic absorption spectroscopy, F-AAS—flame atomic absorption spectroscopy, AFS—atomic fluorescence spectroscopy, ICP-OES—optical emission spectroscopy with inductively coupled plasma, ICP-MS—mass spectrometry with inductively coupled plasma)

2.2. Measurement Methods

The decision for the selection of an analytical measurement technique always depends on the elements (metals) and the aim of investigation [13, 14]. Standardized methods are available from the German Institute for Standardization (Deutsches Institut für Normung e.V., DIN) for the determination of heavy metals in particles of airborne dust, in exhaust gases and atmospheric air [15-17]. The Association of German Engineers (Verein Deutscher Ingenieure, VDI) published standardized methods in German and English for the determination of multiple single elements in particulate matter and in suspended matter in ambient air [18-26]. The solid particles require a special sample preparation before the measurement. Some elements, such as arsenic, antimony, and selenium were separated from the dust matrix into their volatile hydrides for measurement in gaseous form [19]. Mercury can be sorbed into amalgam and then be converted into mercury cold vapor for measurement [25]. For the most elements, such as heavy metals, a digestion is required to transfer them from the solid matrix into a liquid form. An acid digestion is a common procedure performed in different ways such as open, thermal or microwave-assisted digestions using a wide variety of acids or acid mixtures (e.g., hydrofluoric acid, aqua regia or nitric acid with hydrogen peroxide) [13, 23]. The qualitative and quantitative determination of heavy metals is mostly performed using optical spectroscopy

such as graphite-furnace atomic absorption spectroscopy (GF-AAS), flame atomic absorption spectroscopy (F-AAS), atomic fluorescence spectroscopy (AFS), and optical emission spectroscopy with inductively coupled plasma (ICP-OES). Mass spectrometry (ICP-MS) has gained more and more importance in the last decades and have been established for heavy metal measurements [15-26].

2.3. Automated Sample Preparation and Measurement

Today, screening systems show a high degree of automation in the fields of biotechnology, pharmacy and high-throughput applications. In contrast, the automation degree in the field of compound-oriented measurements is still low. Multiple—and often changing—process steps with complex sub-processes, single vessels varying in size and design, harsh processing environments (e.g., high pressures, high temperatures, toxic reagents, high-concentrated acids, organic solvents etc.) and special laboratory devices such as heating plates, coolers, incubators, shakers, centrifuges etc. must be integrated into the automation system. A further challenge is the flexibility—to run multiple applications at one automation system to make it more economically for analytical laboratories.

Only few automation solutions have been established in elemental measurements. Typically, the existing automation systems are often specialized to an individual application. Mercury analyzers are one example for automated sample preparation, sample introduction and measurement using AAS and AFS—sometimes coupled with the cold vapor technique. The advanced mercury analyzer AMA 254 (LECO company, Geleen, Netherlands) was designed for the determination of mercury trace amounts in liquid and solid samples [6]. It uses a technique of direct combustion to decompose the sample in an oxygen-rich environment and to remove interfering elements. The mercury amount is determined using a gold amalgamator trap and a spectrophotometer [27]. The advantage is a very sensitive mercury determination in different samples (solids and liquids) without additional chemical sample preparation [6]. Another automated mercury analyzer is the DMA-80 L (MLS-MWS Laboratory Solutions, Leutkirch im Allgäu, Germany) which uses the cold vapor technique [28]. Both devices are limited to the element mercury and therefore limited in their flexibility.

ICP-OES and ICP-MS are predestinated for measurement of multiple elements in a short time frame, since they allow simultaneous determination of numerous elements. In general, those instruments process samples in liquid and gaseous state. The measurement of solid samples typically requires a previous sample pretreatment, such as a microwave-assisted acid digestion. One example of a commercial automation system is the MiniLab (Rohasys Robotic and Handling Systems & Rouwette Consultancy, Spaubeek, Netherlands) for the automated sample preparation of solid samples using an aqua regia digestion [29]. Mercury and other heavy metals can then be measured using ICP-OES or ICP-MS. Another example for automation in elemental analysis is the implementation of flow-injection systems for automated sample introduction and adding of the internal standard prior to the measurement using ICP-OES [30]. This automation solution significantly increases the instrument's performance and precision and is today integrated into the majority of commercial ICP-MS and ICP-OES instruments.

The automation solutions mentioned before are limited in their flexibility—the systems are designed for a special functionality or for a special application. The aim of the authors of this study is the development of flexible automation systems for "compound-oriented measurements", which are not limited to a special application and which can be applied to elemental and structural measurement tasks. A first version of the robotic system was introduced for the sample preparation of waste wood samples and the determination of mercury using ICP-OES and ICP-MS [31-34]. The automation system was extended for the preparation of animal and human bone samples for the determination of calcium and phosphor using ICP-MS to support the recent osteoporosis research [35]. For a further enhancement of the methodological scope, the system was modified and extended to perform the sample preparation for both elemental and structural measurements. Incrustations of clogged biliary endoprostheses as well as pig bile were automatically prepared for the determination of trace metals using ICP-MS and for the determination of cholesterol using GC/MS to support the development of more compatible biliary endoprostheses [36-39]. The robotic system was recently extended to perform the sample preparation of human tissue samples for subsequent heavy metal measurements to support arthrosis treatment by joint implants [40, 41].

This study is focused on the investigation of the close indoor environment. The robotic system was expanded to include an automated screening setup for heavy metal determination in indoor dust. New labware was designed and integrated. New sample processing procedures for the individual automation stations and process control methods were developed. The method development was challenging since common sample preparation procedures for dust samples include a digestion using hydrofluoric acid—recommended for certain heavy metals [13, 14, 42, 43]. Special trained staff as well as special safety devices for the laboratory equipment and instruments are therefore required. In this study, an automated screening method for the determination of heavy metals in indoor dust samples was developed as previously presented [1, 44]. This includes the development of the manual sample preparation method and ICP-MS measurement method [1, 44]. The sample pretreatment (microwave-assisted digestion) was done with nitric acid and hydrogen peroxide; the use of hydrofluoric acid was avoided. Samples with conspicuous results acquired in the screening can then be confirmed by the manual standard method with hydrofluoric acid. As a result, overall processing times and financial resources can be reduced. Furthermore, this study presents a very flexible robot-assisted sample preparation system, which is adaptable to multiple application areas. This promising automation concept enables small and middle analytical laboratories to install such a robotic system and to use it for different tasks, increasing the overall laboratories' efficiency, economy, and safety.

3. System Concept and System Design

3.1. Robotic Sample Preparation System

The robotic sample preparation system was designed according to the automation concept of the "central system integrator" with a decentralized, open structure [45, 46]. For connecting the

individual automation stations (e.g., storage system, different liquid handling devices etc.) two ORCA laboratory robots (Beckman Coulter, Krefeld, Germany) were integrated. These two articulated robots move on orthogonal linear rails and transfer labware with samples and reagents between the stations. A labware transfer station is mounted at the crossing area of the linear rails, enabling the take-over of labware from one robot to the other one. A barcode reader is included for identification of the labware. Two further stations are used for liquid handling and dosing. The Biomek 2000 (Beckman Coulter, Krefeld, Germany) is a liquid handling station for dosing liquids in volumes up to 1,000 µL. The single-vial liquid handler is used for higher liquid volumes (max. 10 mL). Especially in dilution tasks, this large-volume dispenser reduces pipetting steps and processing time. The automation system was designed in a decentralized structure distributed over three laboratories. The microwave digestion device Mars 5 (CEM, Kamp-Lintfort, Germany) is located in a separate laboratory close to the robotic system. The measurement instrument ICP-MS 7700x (Agilent Technologies, Waldbronn, Germany) is installed in another laboratory. The automation system and the additional instruments are equipped with a labware transfer station. These stations allow the access by both—by laboratory assistants and alternatively by mobile robots. The storage system consists of 196 positions for labware in the standard microplate footprint to provide chemicals, standards and labware required for the process run. Figure 2 shows the robotic sample preparation system as a 3D CAD drawing.

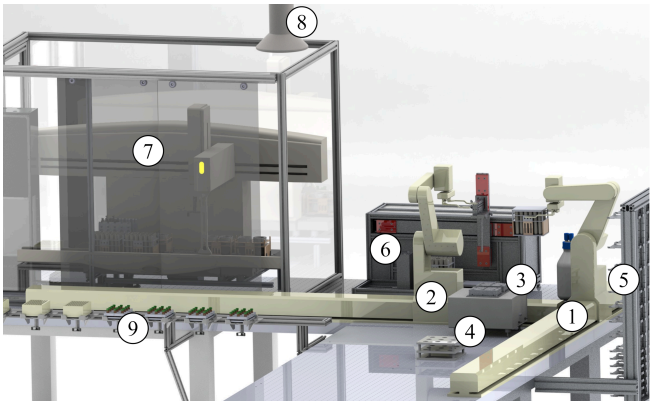
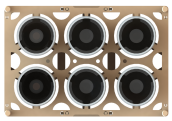
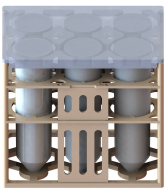
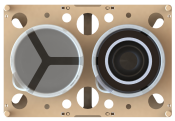
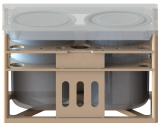
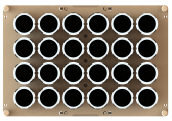
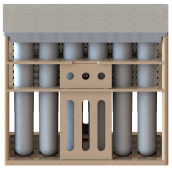


Figure 2: Robotic system for automated sample preparation—laboratory robots moving on linear rail (1) in front of the storage system and (2) in front of liquid handling stations, (3) labware transfer station between two robots, (4) additional rack and thermo shaker on workbench, (5) storage system for chemicals and labware, (6) single-vial liquid handler, (7) liquid handler Biomek 2000 with safety housing and, (8) exhaust system, (9) labware transfer station to laboratory staff or to mobile robots

In general, the labware required in sample preparation tasks for analytical measurements (e.g., microwave vessels, autosampler vials) are not designed in the standard microplate format. Several single vessels, beakers, flasks, and vials with different designs, volumes, lids, and screw caps are required. To ensure a safe labware handling by the ORCA laboratory robots, several racks were designed in the standard microplate format to arrange multiple single vessels. For the sample preparation process of dust samples four racks each with six microwave digestion vessels Xpress (vol. 25 mL, CEM, Kamp-Lintfort, Germany) and one solid lid (cover) for simultaneous covering all six vessels were constructed. These vessels are used to store the powdery samples and for the microwave digestion. Additional covers with six holes were constructed to reduce the evaporation of acid fumes and nitrous gases during the pipetting steps after the microwave digestion [1]. The digestion reagents are provided in another rack with two beakers (PFA, vol. 100 mL) for nitric acid and hydrogen peroxide (the beaker contains a smaller insert with a lower volume of 25 mL). The final—with ultrapure water diluted—measurement solutions are stored in 24 tubes (PP, vol. 14 mL) arranged at a further rack with a solid lid. Table 1 gives an overview of the labware arranged in the special designed racks.

Table 1: Special designed racks in microplate footprint (aluminum body, chemical resistant lid made of PTFE) enabling robotic handling of different numbers of single vessels with various volumes and shapes

Description	CAD drawing	
	unlidded (top view)	closed (front view)
6 microwave vessels Xpress (vol. 25 mL)		
2 beakers for storing nitric acid (left: vol. 100 mL) and hydrogen peroxide (right: insert with 25 mL)		
24 tubes for diluted samples/measurement solutions (vol. 14 mL)		

3.2. Multi-level Software System

A multi-level software system is required to schedule and to control multiple automated workflows supported by robotic systems, on measurement instrumentation as well as in interaction with human operators and mobile transportation robots [45]. The overall process with individual subprocesses—sample preparation of dust specimens including robot-assisted liquid handling and microwave digestion, the ICP-MS measurements, transportation steps as well as data processing and visualization—is managed by the high-level workflow management system [47]. The subprocesses supported by robots were controlled by the process control software "SAMI Workstation Ex 4.0" (Beckman Coulter, Krefeld, Germany) representing the middle-level software. The human operator defines the type of the labware, their source position and their destination. Furthermore, the individual substations (e.g., liquid handling) were selected. Finally, the entire workflow of each subprocess is saved in a SAMI method file, which will be executed in the process run. The lowest level of the software system is formed by the device software of the individual substations. The Biomek 2000 software (Beckman Coulter, Krefeld, Germany) is used to control the liquid handler Biomek 2000. The human operator defines the deck layout as well as the detailed pipetting parameters and saves them in a Biomek method

file. The single vial-liquid handler is controlled by an inhouse-developed software module [45].

3.3. Automation

The entire automated process of dust sample preparation and measurement is divided in multiple subprocesses. In the first subprocess "Pre-digestion", the reagents and samples were transported by the ORCA laboratory robots to the liquid handler Biomek 2000. The reagents are added to the solid samples and the pre-digestion over 20 min is performed using the Biomek 2000. In the second subprocess "Microwave digestion" a human operator closes the microwave vessels, starts the automated digestion run and opens the vessels after cooling down. The third subprocess "Dispensing" is carried out in parallel the robotic system. The ORCA laboratory robots transport empty vials to the single-vial liquid handler and ultrapure water is automatically dispensed. Finally, one ORCA laboratory robot transports the rack with water to the Biomek 2000. In the fourth subprocess "Dilution", the digestion solutions were pipetted into the vials—previously filled with water—and mixed the liquids using the Biomek 2000. After this, the samples are ready for the subprocesses "Measurement" and "Data Evaluation and Visualization". Two top processes were created—for simultaneously processing of 12 and alternatively 24 samples. A more detailed description of the automated subprocesses is given in section 5.2.

4. Materials and Methods

4.1. Standards and Chemicals

The certified reference material (CRM) of indoor dust SRM 2583 (NIST, Gaithersburg (MD), United States) was purchased from Sigma Aldrich/Merck (Darmstadt, Germany). Nitric acid (HNO_3) and 30% stabilized hydrogen peroxide (H_2O_2)—both in suprapure quality (Rotipuran®)—were purchased from Carl-Roth (Karlsruhe, Germany). The ICP multi-element standard IV and ICP single-element standards for As, Hg, Lu and Re were from Merck (Darmstadt, Germany).

4.2. Sample Preparation Using Microwave Digestion

The certified reference material SRM 2583 (fine dust powder) was directly weighted (30 mg) into a microwave digestion vessel Xpress with a volume of 25 mL (CEM, Kamp-Lintfort, Germany). Dust samples—collected by the authors at different inhouse locations—were previously dried and pulverized using an oscillating mill MM2000 (Retsch, Haan, Germany). Three stirring balls (PTFE, diameter 6 mm) from Bola (Grünsfeld, Germany) were added to each sample in a microwave digestion vessel. Nitric acid was used as digestion reagent. To correct evaporation effects during the microwave digestion, Rhenium (Re) was added to the nitric acid as internal standard (ISTD) with a concentration of 1.25 mg/L (after sample processing 80 µg/L). A volume of 2.4 mL of HNO_3 with ISTD and 0.6 mL of H_2O_2 were pipetted to the powdery samples. The vessels were closed after a rest of 20 minutes without lids. The Mars 5 device (CEM, Kamp-Lintfort, Germany) is used for the microwave digestion with the following temperature-time program: room temperature to 180 °C (356 °F) in 20 min, 180 °C to 220 °C (428 °F) in 20 min and a hold time at 220 °C (428 °F) for 20 min. After cooling down and opening the vessels, the samples were diluted with ultrapure water.

4.3. ICP-MS Measurement Parameters

The ICP-MS 7700x was operated with the autosampler ASX-500 (Cetac, Omaha (NE), United States) for automated sample introduction. The data were acquired using the MassHunter Workstation Software for 7700 ICP-MS G7201C Version C.01.04 (Agilent Technologies, Waldbronn, Germany) and the data interpretation and statistical evaluations were done using the in-house developed software module "Analytical Data Evaluation" (ADE)—a web application running at all operating systems [48, 49].

The calibration was generated with five standard solutions (concentrations 1, 5, 10, 50 and 100 µg/L for Li, Be, V, Cr, Co, Ni, As, Sr, Ag, Cd, Tl, and Pb). Mercury was added to the respective standard solution with 0.01, 0.05, 0.10, 0.50 and 1.00 µg/L. The ICP-MS was operated with the following parameters: 1,550 W RF power, 10 mm sample depth, 1.05 L/min nebulizer gas flow, 0.10 rps nebulizer pump speed, 13 °C (55.4 °C) spray chamber temperature, argon as plasma and nebulizer gas, and 4.3 mL/min collision cell helium flow. The data were acquired with three repetitions (0.3 sec integration time). The autosampler probe was flushed after each measurement first with HNO_3/HCl (5%/1%) followed by HNO_3 (8%). After the last measurement, the probe was finally flushed with water.

5. Results and Discussion

5.1. Manual Sample Preparation and Measurement

The manual sample preparation and measurement method was validated before the transfer to the automation system. The validation was focused on several parameters [33, 36, 45]. The repeatability and the recovery rate were determined with 23 standard samples of the certified reference material (CRM) NIST SRM 2583 and one blank—prepared and measured at one day. The within-laboratory precision was determined with ten daily prepared and measured CRM samples—this procedure was performed on five different days. The measurement precision of the ICP-MS was determined by measuring one sample ten times. The limit of detection (LOD) and the limit of quantification (LOQ) were established by measuring ten blanks samples, which previously gone through the whole sample preparation process. The validation results were previously published [44].

5.2. Automated Sample Preparation and Measurement

In the first subprocess "Pre-digestion" the racks providing reagents in beakers and solid samples in microwave digestions vessels were transported from the storage system to the deck of the liquid handling station Biomek 2000 using the two ORCA laboratory robots and the robot transfer station. At the Biomek 2000, the reagents (nitric acid and hydrogen peroxide) were added to the samples. After a rest over 20 min with open vessels, the racks were covered and provided on the sample transfer station. The second subprocess "Microwave digestion" requires the support of a human operator for closing the individual microwave vessels and starting the microwave digestion run. After cooling down, the screw caps of the vessels were manually removed, the vessels arranged in the racks and the racks lidded with a special perforated cover [1].Then the racks were directly provided on the deck of the Biomek 2000. In parallel processing to the microwave digestion, the third subprocess "Dispensing" is carried out using the robotic

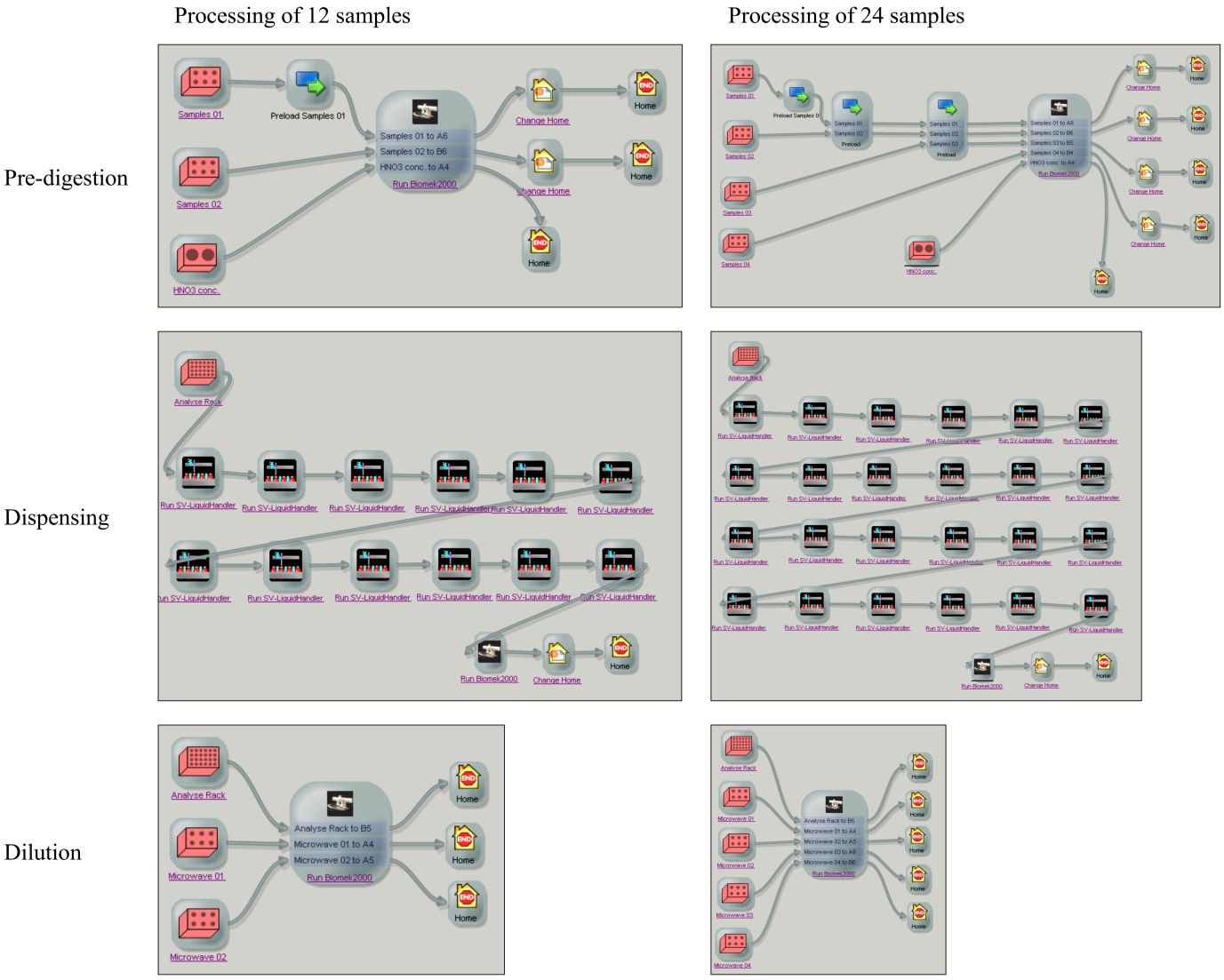


Figure 3: SAMI Ex Editor with subprocesses of the automated preparation of 12 samples arranged on two racks (left) and 24 samples arranged on four racks (right)—Pre-digestion, dispensing and dilution

system. First, the lidded rack with empty vials is transported from the storage system to the deck of the single-vial liquid handler station using the ORCA laboratory robots. The lid is removed by one ORCA laboratory robot, ultrapure water automatically dispensed into the empty vials and the rack is covered again by the robot. Finally, the rack with the water-filled vials is transported by the robot to the deck of the liquid handler Biomek 2000. Figure 3 shows the SAMI methods for the preparation of 12 and 24 samples.

The digestion solutions were diluted with ultrapure water in the fourth subprocess "Dilution" to adjust the acid concentration for the subsequent measurement. The digestion solutions were aspirated through the perforated lids and dispensed into the water-filled vials. The liquids were well mixed by multiple aspiration/dispensing. After this, the rack is lidded and the samples are ready for the subprocesses "Measurement" and "Data Evaluation and Visualization".

The rack with the measurement solutions is transported to the ICP-MS, which is located in a separate laboratory. The transport can be performed either manually or automated using mobile robots [50].

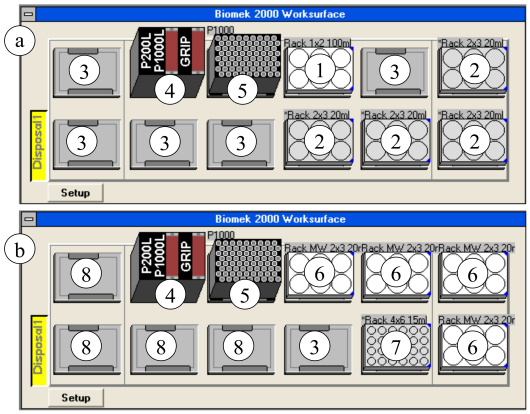


Figure 4: Deck layout of the Biomek 2000 for processing 24 dust samples—(a) pre-digestion, (b) dilution, (1) rack with nitric acid and hydrogen peroxide, (2) lidded racks with samples, (3) free positions for lids, (4) tool rack, (5) tip box, (6) racks with digestion solutions and perforated lids, (7) water filled tubes for measurement solutions, (8) free positions

The sample rack is positioned into the ICP-MS autosampler and the automated sample introduction and measurements are

started. After finishing, the data were automatically evaluated [49]. Figure 4 shows the deck layout of the liquid handler Biomek 2000 for the subprocesses "Pre-digestion" and "Dilution" for the processing of 24 dust samples.

5.3. Validation of the Automated Process

The validation parameters were determined similar to the manual method validation. The repeatability and the recovery rate were determined with 23 CRM samples and one blank arranged in four racks. The repeatability showed coefficients of variation (CV) for the elements chromium, arsenic, cadmium, and mercury which were in good accordance to the certified values. The CV values were slightly higher (+0.26%) for lead. The average recovery rates were determined for arsenic with 93.9%, for cadmium with 82.9% and for lead with 82.04%. The average recovery rates of chromium and mercury were lower, with values of 56.3% (Cr) and 66.4% (Hg). To increase the recovery rates, the use of a stronger acid mixture with hydrofluoric acid in the microwave digestion is recommended [13, 14, 42, 43]. In this automation approach, a hydrofluoric acid digestion was avoided for safety reasons. The automated process is performed on a technical system with mechanical and electronical components sensitive to corrosive acids. The technical equipment and the laboratory staff must be protected, and additional safety equipment would be required.

The within-laboratory precision was determined with 11 CRM samples and one blank prepared at five days. Two racks were used in each method run. The resulting CV values are in good accordance to the certified values for Cr, As, and Cd. The concentration uncertainties of the CRM were given for chromium (±27.5%) arsenic (±22.9%) and cadmium (±50.7%). The CV values for the lead concentration were slightly higher than certified. The method presented is a rapid screening method—therefore, a CV value lower than 15% can be accepted. The measurement precision—determined using ten measurements of one sample—showed results in the expected range (<2%). The limits of detection and quantification were determined with the preparation and measurement of 10 blank samples. The LOD values were determined for the measurement solutions with 180.7 ng/L (Cr), 10.2 ng/L (As), 3.7 ng/L (Cd), 6.5 ng/L (Hg), and 14.5 ng/L (Pb) and for the solid dust material with 225.9 µg/kg (Cr), 12.7 µg/kg (As), 4.6 µg/kg (Cd), 8.1 µg/kg (Hg), and 18.2 µg/kg (Pb). The LOQ values were determined for the measurement solutions with 423.2 ng/L (Cr), 25.8 ng/L (As), 9.2 ng/L (Cd), 12.1 ng/L (Hg), and 27.3 ng/L (Pb) and for the solid dust material with 529.0 µg/kg (Cr), 32.2 µg/kg (As), 11.6 µg/kg (Cd), 15.2 µg/kg (Hg), and 34.2 µg/kg (Pb). The validation results for the certified elements Cr, As, Cd, Hg, and Pb are summarized in Table 2.

In this extended version of the study, the results of additional elements are presented (Li, B, V, Co, Ni, Sr, Ag, and Tl). The average concentrations range from 0.045 mg/kg (Tl) to 65.25 mg/kg (Sr). The repeatability showed coefficients of variation ranging from 3.97% (Li) to 13.28% (Ni). Only Ag has a higher CV value of 36.75%. Certified values for these additional elements are not available for the CRM used in this study. For this reason, the values provided have merely informative character. The precision values of the non-certified elements were similar to the certified elements, and this shows the ability of the automated method to measure more than the certified elements. Additional determination of the recovery rate would be required to confirm the true value.

The measurement precision was determined with 10 repetition measurements of the same sample. The LOD and LOQ values were determined with 10 blank samples for both the measurement solutions and the solid dust material. The entire validation results for the non-certified elements are also included in Table 2.

5.4. Comparison of Manual and Automated Measurements and Literature Values

Comparison of the process steps: A volume of 2.4 mL HNO_3 is given to the powdery samples. Only one step is needed in manual processing. The liquid handler Biomek 2000 has a maximum pipetting volume of 1 mL. For this reason, three steps each with 0.8 mL are required to add the digestion acid. After the microwave digestion, a volume of 2 mL sample solution is manually diluted with 23 mL water. Therefore, water is dispensed in three steps with volumes of 10 mL, 10 mL, and 3 mL. The sample solution is then added in one step and mixed with the water. This dilution step was miniaturized in the automated procedure. The single-vial liquid handler dispenses a volume of 5.75 mL water in one step. Then a volume of 0.5 mL sample solution is added to the water and mixed by the Biomek 2000. This miniaturization has an additional advantage. Due to the smaller total volume of the final measurement solutions—vials with a volume of 30 mL were reduced to vials with 14 mL—more vials can be arranged on one rack.

Only six large volume vials can be placed on one rack, but 24 low-volume vials. This allows a higher sample throughput due to the reduced transportation steps for multiple racks. Table 3 summarizes the numbers of manual and automated process steps.

Comparison of the processing times: The entire processing times of the manual and automated processing were determined with the preparation of 12 samples. The duration of the automated measurement process is 15.5 min longer than the manual procedure executed with trained laboratory staff. Typically, the human operator arranges all required equipment on the workbench. Special transportation steps between a storage system and individual liquid handling stations are not required.

Furthermore, the tool change process of the liquid handler Biomek 2000 (e.g., from gripper tool for lid handling to pipetting tool and return) needs additional time. A human uses it hands for lid handling and operating the pipette. This time delay—between manual and automated operation—can be avoided by a suitable process scheduling with parallel and overlapping runs. Nevertheless, the laboratory assistant is free in the time of the automated run and can perform other tasks in parallel. Furthermore, the automated process can run during the staff's break times. In summary, despite the slightly longer processing time of the automated process, the laboratories efficiency and the throughput can be increased. Table 4 compares the processing times for the individual process steps and for the entire process as well as the operators in the manual and the automated procedure.

Table 2: Validation results of the automated measurement method in comparison to the certified values of the indoor dust CRM NIST SRM 2583

	^{7}Li	^{9}Be	^{51}V	^{52}Cr	^{59}Co	^{60}Ni	^{75}As	^{88}Sr	^{107}Ag	^{111}Cd	^{202}Hg	^{205}Tl	^{208}Pb
Repeatability (n=23)													
Average [mg/kg]	12.48	0.23	13.13	45.02	3.73	44.41	6.57	65.25	1.13	6.05	1.04	0.045	70.48
STDEV [mg/kg]	0.49	0.02	0.54	3.09	0.33	5.90	0.26	5.02	0.41	2.15	0.11	0.004	6.11
CV [%]	3.97	7.24	4.09	6.87	8.76	13.28	3.94	7.70	36.75	35.45	10.47	7.93	8.66
Recovery rate (n=23)													
Average [%]	n.a.	n.a.	n.a.	56.27	n.a.	n.a.	93.90	n.a.	n.a.	82.89	66.44	n.a.	82.04
Min [%]	n.a.	n.a.	n.a.	48.55	n.a.	n.a.	86.97	n.a.	n.a.	56.11	60.58	n.a.	71.73
Max [%]	n.a.	n.a.	n.a.	64.84	n.a.	n.a.	100.33	n.a.	n.a.	171.37	94.01	n.a.	99.73
Within-laboratory precision (n=11, 5 days)													
Max. average [mg/kg]	14.98	0.29	16.15	56.18	4.46	54.10	7.42	81.51	1.29	8.31	1.43	0.061	79.15
Min. average [mg/kg]	13.90	0.25	15.07	50.64	3.87	45.85	6.74	70.69	0.94	6.10	0.92	0.056	68.68
Max. CV [%]	11.29	11.40	11.47	14.18	28.32	30.78	10.28	19.12	40.70	58.92	49.39	10.30	14.69
Min. CV [%]	1.85	5.37	3.16	5.77	4.88	14.65	4.04	1.17	24.50	21.88	5.26	5.20	5.70
Measurement precision (n=1, 10 measurements)													
CV [%]	0.59	5.41	0.94	0.66	0.72	0.77	1.16	0.46	0.74	0.80	1.99	2.58	0.41
Analytical LOD and LOQ (in measurement solution)													
LOD [ng/L]	28.1	3.8	121.5	180.7	6.3	109.4	10.2	178.6	11.7	3.7	6.5	9.8	14.5
LOQ [ng/L]	45.7	9.1	152.5	423.2	13.6	253.9	25.8	418.3	20.8	9.2	12.1	14.2	27.3
Methodical LOD and LOQ (in solid dust samples)													
LOD [µg/kg]	35.1	4.8	151.8	225.9	7.9	136.7	12.7	223.3	14.7	4.6	8.1	12.3	18.2
LOQ [µg/kg]	57.1	11.4	190.6	529.0	16.9	317.4	32.2	522.8	26.0	11.6	15.2	17.8	34.2
Certified values of NIST SRM 2583													
Average [mg/kg]	n.a.	n.a.	n.a.	80	n.a.	n.a.	7	n.a.	n.a.	7.3	1.56	n.a.	85.9
Uncertainty [mg/kg]	n.a.	n.a.	n.a.	±22	n.a.	n.a.	±1.6	n.a.	n.a.	±3.7	±0.19	n.a.	±7.2
Uncertainty [%]	n.a.	n.a.	n.a.	±27.5	n.a.	n.a.	±22.9	n.a.	n.a.	±50.7	±12.80	n.a.	±8.4

Table 3: Manual and automated process steps (1 sample) [1]

Process Step	Manual		Automated	
	Volume [mL]	Steps	Volume [mL]	Steps
Pipetting HNO_3	2.40	1	0.80 0.80 0.80	3
Pipetting H_2O_2	0.60	1	0.60	1
Dispensing H_2O	10.00 10.00 3.00	3	5.75	1
Pipetting digested sample solution	2.00	1	0.50	1

Table 4: Processing times and operators in manual and automated sample preparation and measurement (12 samples) [1]

Process step	Manually		Automated	
	Processing time [min]	Operator	Processing time [min]	Operator
Initial transport (labware, reagents)	2.0	Human laboratory assistant	3.5	2 ORCA laboratory robots
Pipetting HNO_3	1.0	Human laboratory assistant	8.5	Biomek 2000 liquid handler

Process step	Manually		Automated	
	Processing time [min]	Operator	Processing time [min]	Operator
Pipetting H_2O_2	1.0	Human laboratory assistant	2.0	Biomek 2000 liquid handler
Pre-digestion	20.0	Waiting time	20.0	Waiting time
Closing vessels (manually)	2.0	Human laboratory assistant	2.0	Human laboratory assistant
Microwave digestion, cool down	90.0	Microwave device	90.0	Microwave device
Opening vessels (manually)	4.0	Human laboratory assistant	4.0	Human laboratory assistant
Dispensing water (performed in parallel to microwave digestion)	(3.5) not included in calculation	Human laboratory assistant	(6.5) not included in calculation	Single-vial liquid handler
Pipetting of digestion solution, mixing	4.5	Human laboratory assistant	10.0	Biomek 2000 liquid handler
ICP-MS measurement (12 samples)	54.0	ICP-MS equipped with autosampler	54.0	ICP-MS equipped with autosampler
Total processing time	178.5		194.0	

Table 5: Measurement results of dust samples collected in laboratories and offices (samples 1-7, 12), in private households (samples 8-11, 21), in old disused garden houses (samples 13-20), and in high school locations (22-32); (n.a.: not acquired)

No.	^{7}Li [mg/kg]	^{9}Be [mg/kg]	^{51}V [mg/kg]	^{52}Cr [mg/kg]	^{59}Co [mg/kg]	^{60}Ni [mg/kg]	^{75}As [mg/kg]	^{88}Sr [mg/kg]	^{107}Ag [mg/kg]	^{111}Cd [mg/kg]	^{202}Hg [mg/kg]	^{205}Tl [mg/kg]	^{208}Pb [mg/kg]
1	18.29	n.a.	n.a.	34.86	0.99	5.87	n.a.	334.81	0.28	0.21	n.a.	n.a.	9.66
2	26.99	n.a.	n.a.	28.42	0.58	4.43	n.a.	15.73	0.28	0.13	n.a.	n.a.	3.41
3	62.28	n.a.	n.a.	23.79	0.55	14.97	n.a.	14.52	0.35	3.87	n.a.	n.a.	12.39
4	7.33	n.a.	n.a.	97.22	1.11	9.23	n.a.	41.59	0.35	0.19	n.a.	n.a.	13.07
5	12.89	n.a.	n.a.	89.73	1.77	19.60	n.a.	22.86	0.58	0.83	n.a.	n.a.	40.13
6	7.80	n.a.	n.a.	195.61	1.49	6.99	n.a.	17.22	0.39	0.23	n.a.	n.a.	7.07
7	6.48	n.a.	n.a.	63.48	0.81	4.60	n.a.	11.43	0.13	0.17	n.a.	n.a.	3.45
8	7.92	n.a.	n.a.	30.67	1.14	6.87	n.a.	37.33	0.20	0.21	n.a.	n.a.	5.13
9	0.76	n.a.	n.a.	5.47	0.64	5.52	n.a.	13.24	0.18	0.20	n.a.	n.a.	5.79
10	1.07	n.a.	n.a.	5.87	0.47	6.11	n.a.	16.46	0.43	0.15	n.a.	n.a.	6.68
11	2.18	n.a.	n.a.	15.12	1.30	4.03	n.a.	17.73	0.36	0.10	n.a.	n.a.	8.09
12	14.61	n.a.	n.a.	69.44	3.36	45.90	n.a.	91.79	1.59	0.84	n.a.	n.a.	35.28
13	3.20	n.a.	n.a.	39.31	1.86	6.24	n.a.	37.48	0.09	6.11	n.a.	n.a.	2,779.47
14	6.79	n.a.	n.a.	36.77	14.13	14.07	n.a.	180.51	0.23	4.38	n.a.	n.a.	158.43
15	6.21	n.a.	n.a.	100.81	5.33	15.20	n.a.	98.84	0.36	2.47	n.a.	n.a.	114.51
16	3.65	n.a.	n.a.	29.61	5.34	17.89	n.a.	65.91	0.46	5.94	n.a.	n.a.	620.00
17	6.56	n.a.	n.a.	227.95	6.82	129.53	n.a.	127.61	0.08	1.80	n.a.	n.a.	164.50
18	10.90	n.a.	n.a.	29.78	3.83	27.52	n.a.	183.32	0.23	5.86	n.a.	n.a.	68.08
19	4.21	n.a.	n.a.	43.85	2.88	189.28	n.a.	72.15	0.34	15.39	n.a.	n.a.	354.79
20	0.46	n.a.	n.a.	13.93	0.30	1.81	n.a.	55.88	<LOD	0.04	n.a.	n.a.	3.77
21	0.57	n.a.	n.a.	10.59	0.25	2.91	n.a.	11.95	0.55	0.11	n.a.	n.a.	3.20
22	2.88	0.09	5.37	49.84	1.67	48.20	1.15	33.74	2.46	0.75	0.15	0.04	13.93
23	3.12	0.04	2.48	54.06	4.05	13.25	9.89	40.08	2.59	1.12	0.20	0.03	53.30
24	1.96	0.07	2.45	56.36	1.28	14.38	0.97	46.72	5.83	1.79	0.15	0.04	20.70
25	13.68	0.58	10.82	37.93	4.29	10.76	5.18	54.26	1.19	0.70	0.11	0.14	9.92
26	3.38	0.11	3.89	86.86	1.33	8.31	0.94	46.84	0.55	0.26	0.05	0.03	7.74
27	7.37	0.10	4.49	41.27	2.67	78.93	0.97	57.52	4.49	0.54	0.28	0.04	21.93
28	0.60	0.01	0.63	117.17	1.05	8.59	0.20	10.17	2.16	0.75	0.03	0.01	3.43
29	3.18	0.10	3.57	64.98	1.44	16.04	0.84	44.85	0.49	1.09	0.04	0.03	23.94
30	3.56	0.13	6.97	81.12	3.59	16.13	1.92	73.19	0.88	1.17	0.10	0.06	196.30
31	2.62	0.09	2.97	122.83	1.75	8.17	0.68	33.72	1.36	0.62	0.08	0.03	9.87
32	11.97	0.14	7.04	67.86	5.34	263.70	2.30	158.35	34.67	3.83	9.92	0.07	273.90

Comparison of repeatability and recovery: A close correlation between the validation results of the manual and the automated process was reached. The individual validation results were presented and compared in previous publications [1, 44]. These

results show the successful transfer of the manual measurement method to the automation system. In the two measurement processes—manual and automated—relatively high CV values were observed in the determination of the repeatability (35.45%) and the recovery rate (min. 56.11%, 171.37%) for the element cadmium. Reasons can be found in the inhomogeneity of the CRM. The CRM used in this study certifies Cd uncertainties of ±3.7 mg/kg (±50.7%). In the literature, similar results were reported for other dust reference materials such as ERM-CZ120 [14].

5.5. Measurement Results of Collected Dust Samples

The automated measurement system was tested with a wide variety of real samples. Dust samples were collected in rooms of the institute including robotic laboratories, chemical laboratories, analytical laboratories, and offices (samples 1-7, 12), in private households (samples 8-11, 21), in old disused garden houses (samples 13-20), and at several places in a high school including class and storage rooms, cafeteria, auditorium, and the sports hall. The indoor dust was taken with hand gloves and plastic forceps and was stored in plastic containers to avoid contamination during the sampling.

Selected results were previously presented in [1, 44]. Overall, relatively low heavy metal concentrations were acquired, which were within the expected range. Higher concentrations of some heavy metals were found in dust samples collected in old garden houses. Further, elevated heavy metal concentrations were monitored in access-controlled storage cabinets for chemicals. Table 5 summarizes the results—previously published and extended with results of additional heavy metals. In the first 21 samples no recording of the concentrations of Be, V, As, Hg, and Tl is available since these elements were later included into the measurement method to extend the methodological scope. The concentrations of all elements shown in Table 5 were determined for the subsequent sample set (samples 22-32). The measured heavy metal concentrations in the collected indoor dust samples show a wide concentration range. Elevated Pb concentrations were determined in the samples from old garden houses.

Elevated As and Hg concentrations were monitored in dust collected in storage cabinets of chemicals. The concentrations of Fe were higher than the highest calibrations standard in all samples measured. If the aim of investigation is determination of Fe, then an additional dilution step must be included in the sample preparation to adjust the concentration in the measurement solutions to the typical working range of the ICP-MS detector. In this study, the determination of Fe was not the focus and no additional dilution was performed. Figure 5 shows the mass spectra of a calibration standard with an element concentration of 100 µg/L and of a dust sample collected in an old garden house (sample 14). An exemplarily overview of the heavy metal concentrations (Li, Cr, Co, and Sr) in different samples gives Figure 6.

A more detailed exploration of the heavy metal distribution requires the analysis of a higher sample number per location. This will be subject to future investigations. The results of this study show the suitability and performance of the automated sample pretreatment and the analytical measurement setup for its application in the determination of heavy metals in indoor dust—and possibly in outdoor dust.

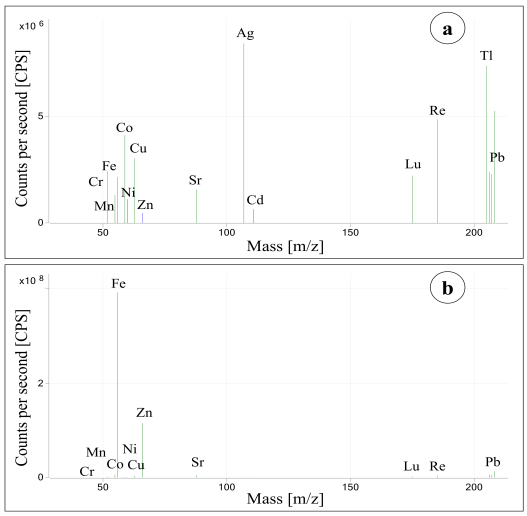


Figure 5: ICP-MS mass spectra—(a) calibration standard with element concentrations of 100 µg/L, (b) collected dust sample 14.

6. Conclusion

In this extended paper, the entire study of the development and application of an automated robot-assisted system for determination of multiple heavy metals in indoor dust was presented. The study started with the development and the validation of the manual sample preparation (microwave-assisted acid digestion) of dust samples and the measurement method (ICP-MS measurements). The manual procedure was validated with the following parameters determined using certified reference material: repeatability, recovery rates, within-laboratory precision, measurement precision as well as the limit of detection (LOD) and the limit of quantification (LOQ). In the next step, the automation system was extended, new racks for the used labware were designed and manufactured and new processes were created in the three-level software system. The manual procedure was not identical transferred to the automation system. The subprocess dilution was miniaturized resulting in a higher sample throughput by using a higher number of smaller vessels. Furthermore, the integration of the single-vial liquid handler allows the dispensing of a large volume ultrapure water in one step, reducing pipetting uncertainties. The automated procedure was identically validated and compared with the manual processing.

The developed measurement method can be understood and used as a screening procedure to prepare and measure a high number of dust samples in a relative short time frame. The processing time of the automated process is approx. 15 min longer than the manual procedure. But while running the automated process, the laboratory staff is free for other tasks operated in parallel. The method validation for both the manual and the automated procedure was performed using certified reference material. The results of the two methods were compared with the certified values for Cr, Cd, As, Hg, and Pb. Close results were reached for As, Cd, and Pb. The recovery rates of Cr and Hg were lower than certified, but they are acceptable for a screening method. Conspicuous samples found in the screening can then be analyzed by the standard method using a stronger acid mixture in the microwave digestion procedure.

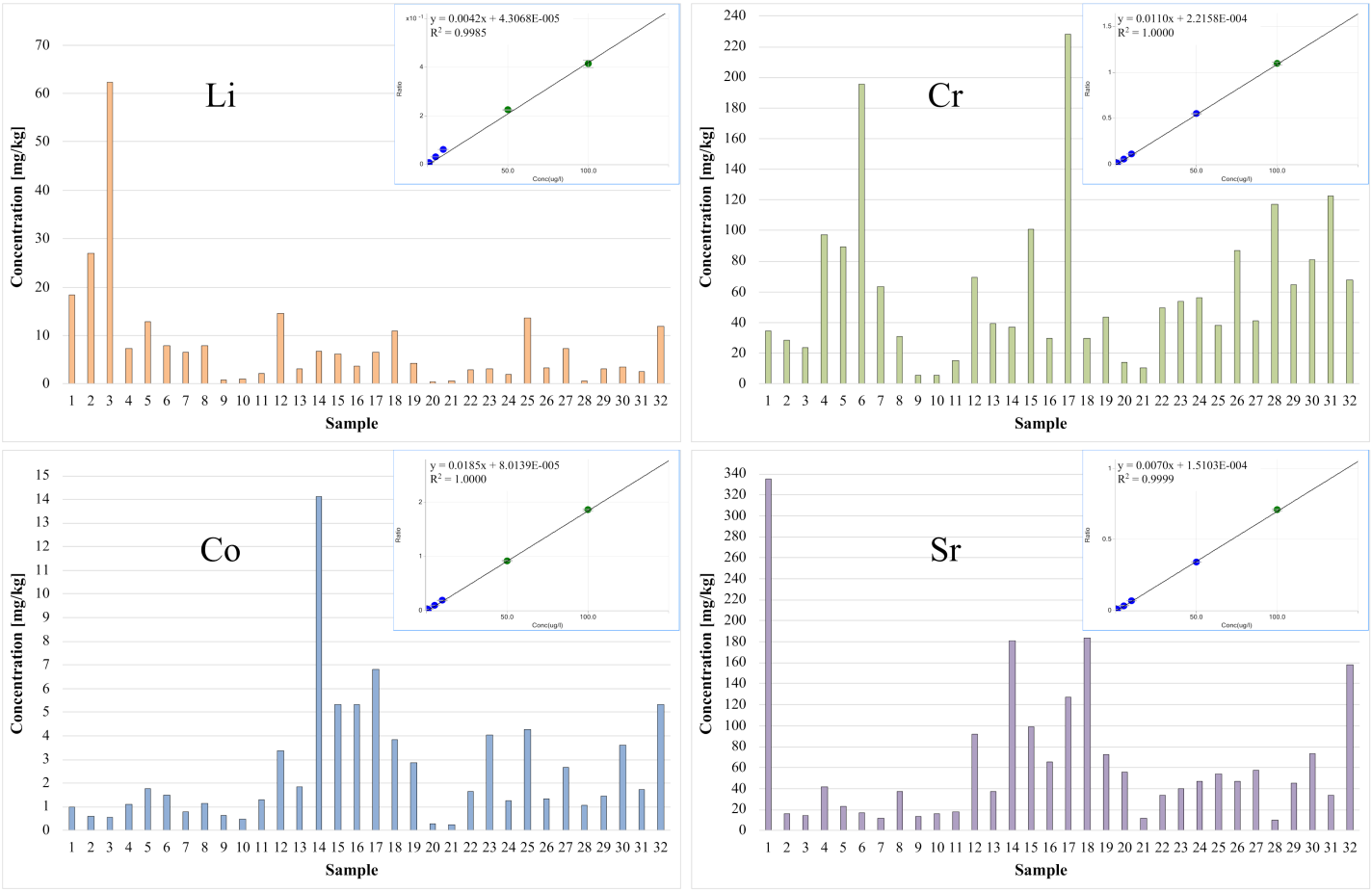


Figure 6: Calibration data and concentrations of the elements Li, Cr, Co, and Sr measured in dust samples which were collected in laboratories and offices (samples 1-7, 12), in private households (samples 8-11, 21), in old disused garden houses (samples 13-20), and in high school locations (samples22-32)

Besides the five elements certified in the CRM, additional elements were measured and validated (except for the determination of the recovery rate). Depending on the element concentration, the recorded CV values show that the automated measurement method is also suitable to screen more than the certified elements.

The automation of the entire process—sample preparation and measurement—has multiple advantages. The human operator is protected from potential hazards caused by harmful reagents such as high concentrated nitric acid and hydrogen peroxide, toxic samples and sample solutions as well as harmful nitrous gases. The laboratory staff is free from highly repetitive tasks like pipetting for multiple times. Other laboratory tasks can be executed in parallel. Overall, this increases the sample throughput as well as the laboratories performance and safety.

The automated measurement system presented is ready for its application in large studies with a high number of samples. In this study, the sample preparation and measurement of collected inhouse dust samples give a first overview of the element concentrations and show the performance of the automation. In further studies, a suitable design of experiment for the desired location is needed together with well-planned sampling methods to acquire large data sets of the area to be investigated. The automation system is here a contribution to the current exploration of the distribution of heavy metals inhouse and for a safer and healthier daily life. Furthermore, the system presented can also be applied in environmental studies screening outdoor dust.

Conflict of Interest

The authors declare no conflict of interest.

Acknowledgment

The authors thank Dr. C. Kupke for organizing the project "Jugend Forscht" (German youth science competition) with the Innerstädtisches Gymnasium (Senior High School) in Rostock, Germany. Special thanks are expressed to H. Engelhardt, Dipl.-Ing. L. Woinar, and Dr. S. Junginger for technical support and suggestions, as well as to S. Horn and B.Sc. A. Reichelt for their excellent laboratory assistance.

References

[1] H. Fleischer, J. Widmer, S. Statkevych, R. Stoll, T. Roddelkopf, K. Thurow, "Robot-assisted Measurement of Heavy Metals in Indoor Dust Using ICP-MS," in 2021 IEEE International Instrumentation and Measurement Technology Conference (I2MTC), 1-6, 2021, doi: 10.1109/I2MTC50364.2021.9459859.

[2] S. Y. Tan, S. M. Praveena, E. Z. Abidin, M. S. Cheema, "A review of heavy metals in indoor dust and its human health-risk implications," Reviews on environmental health, **31**(4), 447-456, 2016, doi: 10.1515/reveh-2016-0026.

[3] N. Bukowiecki, M. Hill, R. Gehrig, C. N. Zwicky, P. Lienemann, F. Hegedüs, G. Falkenberg, E. Weingartner, U. Baltensperger, "Trace Metals in Ambient Air: Hourly Size-Segregated Mass Concentrations Determined by Synchrotron-XRF," Environmental Science & Technology, **39**(15), 5754-5762, 2005, doi: 10.1021/es048089m.

[4] T. Shi, Y. Wang, "Heavy metals in indoor dust: Spatial distribution, influencing factors, and potential health risks," Science of the total Environment, **755**(2021), 2021, doi: 10.1016/j.scitotenv.2020.142367.

[5] S. T. Y. Tong, "Indoor and outdoor household dust contamination in Cincinnati, Ohio, USA," Environmental Geochemistry and Health, **20**(3), 123-133, 1998, doi: 10.1023/A:1006561832381.

[6] R. Millan, T. Schmid, M. J. Sierra, S. Carrasco-Gil, M. Villadóniga, C. Rico, D. M. S. Ledesma, F. J. D. Puente, "Spatial variation of biological and pedological properties in an area affected by a metallurgical mercury plant: Almadenejos (Spain)," Applied Geochemistry, **26**(2), 174-181, 2011, doi: 10.1016/j.apgeochem.2010.11.016.

[7] D. J. A. Davies, J. M. Watt, I. Thornton, "Lead levels in Birmingham dusts and soils," The Science of the Total Environment, **67**(2-3), 177-185, 1987, doi: 10.1016/0048-9697(87)90210-5.

[8] J. Patocka, K. Cerný, "Inorganic lead toxicology," Acta medica (Hradec Králové) / Universitas Carolina, Facultas Medica Hradec Králové, **46**(2), 65-72, 2003, doi: 10.14712/18059694.2019.8.

[9] C. D. Klaassen, J. Liu, B. A. Diwan, "Metallothionein protection of cadmium toxicity," Toxicology and Applied Pharmacology, **238**(3), 215-220, 2009, doi: 10.1016/j.taap.2009.03.026.

[10] M. P. Waalkes, "Cadmium carcinogenesis in review," Journal of Inorganic Biochemistry, **79**(1-4), 241-244, 2000, doi: 10.1016/S0162-0134(00)00009-X.

[11] D. J. A. Davies, I. Thornton, J. M. Watt, E. B. Culbard, P. G. Harvey, H. T. Delves, J. C. Sherlock, G. A. Smart, J. F. A. Thomas, M. J. Quinn, "Lead intake and blood lead in two-year-old U.K. urban children," The Science of the Total Environment, **90**(C), 13-29, 1990, doi: 10.1016/0048-9697(90)90182-T.

[12] H. Fleischer, K. Thurow, "Compound-oriented Measurement Processes: Elements, Molecules, Structures," in 2020 IEEE International Instrumentation and Measurement Technology Conference (I2MTC), 1-6, 2020, doi: 10.1109/I2MTC43012.2020.9128468.

[13] R. Lumpp, M. Klein, E. Bieber, F. Bunzel, U. Eckermann, C. Frels, W. Günther, C. Hagemann, C. Koch, A. Olschewski, C. Temme, "Comparison of different digestion methods for elemental analysis of airborne dust and dust deposition," Gefahrstoffe - Reinhaltung der Luft, 72(1/2), 64-71, 2012 (article in German with an abstract in English).

[14] R. Lumpp, M. Klein, K. Berger, E. Bieber, F. Bunzel, M. Ernst, C. Frels, W. Günther, C. Hagemann, C. Koch, A. Olschewski, "A study on the certified reference material ERM®-CZ120 – Results of round robin tests," Gefahrstoffe - Reinhaltung der Luft, **75**(1/2), 35-40, 2015 (article in German with an abstract in English).

[15] Deutsches Institut für Normung e.V., "Ambient air quality - Standard method for the measurement of Pb, Cd, As and Ni in the PM10 fraction of suspended particulate matter," (DIN EN 14902), 2005.

[16] Deutsches Institut für Normung e.V., "Stationary Source Emissions - Determination of the total emission of As, Cd, Cr, Co, Cu, Mn, Ni, Pb, Sb, Tl and V," (DIN EN 14385), 2004.

[17] Deutsches Institut für Normung e.V., "Ambient air quality - Standard for determination of arsenic, cadmium, lead and nickel in atmospheric deposition," (DIN EN 15841), 2010.

[18] Verein Deutscher Ingenieure, "Chemical Analyis of Particulate Matter - Determination of Ba, Be, Cd, Co, Cr, Cu, Ni, Pb, Sr, V, Zn in Particulate Emissions by Atomic Spectrometric Methods," (VDI 2268 Part 1), 1987.

[19] Verein Deutscher Ingenieure, "Chemical Analysis of Particulate Matter - Determination of Arsenic, Antimony and Selenium in Dust Emissions by Atomic Absorption Spectrometry after Separation of their Volatile Hydrides," (VDI 2268 Part 2), 1990.

[20] Verein Deutscher Ingenieure, "Chemical Analysis of Particulate Matter - Determination of Arsenic, Antimony and Selenium in Dust Emissions by Graphite-Furnace Atomic Absorption Spectrometry," (VDI 2268 Part 4), 1990.

[21] Verein Deutscher Ingenieure, "Determination of suspended matter in ambient air - Measurement of the element concentration after sampling on filters - Determination of Al, As, Ba, Ca, Cd, Co, Cr, Cu, Fe, K, Mg, Mn, Na, Ni, Pb, Sb, Se, Sn, Tl, V, and Zn by GF-AAS, ICP-OES, or ICP-MS," (VDI 2267 Part 1), 2019.

[22] Verein Deutscher Ingenieure, "Determination of suspended matter in ambient air - Measurement of Al, As, Ba, Ca, Cd, Co, Cr, Cu, Fe, K, Mg, Mn, Na, Ni, Pb, Sb, Se, Sn, Tl, V, and Zn as part of the atmospheric deposition after sampling with bulk and wet-only collectors using GF-AAS, ICP-OES, and ICP-MS," (VDI 2267 Part 2), 2019.

[23] Verein Deutscher Ingenieure, "Determination of suspended matter in ambient air - Digestion variants for dust samples for subsequent determination of the mass concentration of Al, As, Ba, Ca, Cd, Co, Cr, Cu, Fe, K, Mg, Mn, Na, Ni, Pb, Sb, Se, Sn, Tl, V and Zn," (VDI 2267 Part 3), 2015.

[24] Verein Deutscher Ingenieure, "Determination of suspended particles in ambient air - Measurement of the mass concentrationof mercury - Sampling by sorption as amalgam and determination by atomic absorption spectrometry (AAS)," (VDI 2267 Part 8), 2000.

[25] Verein Deutscher Ingenieure, "Determination of suspended particulate matter in ambient air - Measurement of the mass concentration of mercury - Sampling by sorption as amalgam and determination by atomic fluorescence spectrometry (AFS)," (VDI 2267 Part 9), 2002.

[26] Verein Deutscher Ingenieure, "Determination of suspended particulate matter in ambient air - Measurement of the mass concentration of As, Ba, Cd, Co, Cr, Cu, Ni, Pb, Sb, V und Zn as part of dust precipitation by atomic absorption spectrometry (AAS)," (VDI 2267 Part 16), 2007.

[27] W. A. Willford, R. J. Hesselberg, H. L. Bergman, "Versatile combustion amalgamation technique for the photometric determination of mercury in fish and environmental samples," Journal of the Association of Official Analytical Chemists, **56**(4), 1008-1014, 1973, doi: 10.1093/jaoac/56.4.1008.

[28] G. Kopp, "Quecksilberanalytik - Für jede Applikation die optimale Methode," LABO - Instrumentelle Analytik, **2012**(6), 20–23, 2012.

[29] Rohasys Robotic and Handling Systems, Rouwette Consultancy. "Probenvorbereitung für Schwermetallanalysen," 08.03.2022; http://issuu.com/rouwetteconsultancy/docs/probenvorbereitung_f__r_schwermetal.

[30] J. W. Milburn, "Automated addition of internal standards for axial-view plasma ICP spectrometry using the Optima 3000 XL," Atomic Spectroscopy, **17**(1), 9–14, 1996.

[31] H. Fleischer, K. Thurow, "Determination of Total Mercury Content and Trace Metal Analysis in Wood Materials Part 1: ICP-OES Using Mercury Cold Vapor," American Laboratory, **45**(8), 6–9, 2013.

[32] H. Fleischer, K. Thurow, "Determination of Total Mercury Content in Wood Materials - Part 2: ICP-MS - A Multi-element Method," American Laboratory, **45**(9), 6–11, 2013.

[33] H. Fleischer, E. Vorberg, K. Thurow, "Determination of Total Mercury Content in Wood Materials - Part 3: Miniaturization Using ICP-MS," American Laboratory, **46**(6), 16–20, 2014.

[34] E. Vorberg, H. Fleischer, S. Junginger, N. Stoll, K. Thurow, "Automated Sample Preparation for Mercury Analysis in Wood Materials," IET Science, Measurement and Technology, **10**(5), 398-404, 2016, doi: 10.1049/iet-smt.2015.0036.

[35] H. Fleischer, E. Vorberg, M. Warkentin, D. Behrend, K. Thurow, "Determination of Calcium and Phosphor in Bones Using Microwave Digestion and ICP-MS: Comparison of Manual and Automated Methods using ICP-MS," 5th IMEKO TC19 Symposium on Environmental Instrumentation and Measurement), 94–99, 2014.

[36] H. Fleischer, K. Ramani, S. Bauer, M. Warkentin, D. Behrend, K. Thurow, "Determination of Calcium and Trace Metals in Clogged Biliary Endoprostheses Using ICP/MS," American Laboratory, **49**(6), 6-9, 2017.

[37] H. Fleischer, K. Ramani, K. Blitti, M. Warkentin, D. Behrend, K. Thurow, "Flexible Automation System for Determination of Elemental Composition of Incrustations in Clogged Biliary Endoprostheses Using ICP-MS," SLAS Technology, **23**(1), 83-96, 2018, doi: 10.1177/2472630317727451.

[38] H. Fleischer, K. Ramani, K. Thurow, "Sample Preparation and Measurement of Cholesterol in Pig Bile Using SPE and GC/MS," American Laboratory, **52**(2), 10-13, 2019.

[39] H. Fleischer, T. Roddelkopf, R. Stoll, K. Thurow, "Automated Analytical Measurement System for Determination of Cholesterol in Pig Bile," in 2019 IEEE International Instrumentation and Measurement Technology Conference (I2MTC), 1510-1515, 2019, doi: 10.1109/I2MTC.2019.8826917.

[40] H. Fleischer, C. Lutter, A. Buttner, W. Mittelmeier, K. Thurow, "Semi-Automated Determination of Heavy Metals in Autopsy Tissue Using Robot-Assisted Sample Preparation and ICP-MS," Molecules, **26**(13, 2021, doi: 10.3390/molecules26133820.

[41] H. Fleischer, T. Roddelkopf, C. Lutter, A. Büttner, W. Mittelmeier, K. Thurow, "Automation in arthrosis research," SLAS Technology, **27**(1), 97-99, 2022, doi: 10.1016/j.slast.2021.10.009.

[42] T. Schwank, K. Pitzke, K. Gusbeth, K. Ashley, D. Breuer, "Comparison of Microwave-Assisted Digestion and Consensus Open-Vessel Digestion Procedures for Evaluation of Metalliferous Airborne Particulate Matter," Annals of Work Exposures and Health, **63**(8), 950-964, 2019, doi: 10.1093/annweh/wxz068.

[43] P. E. Rasmussen, K. S. Subramanian, B. J. Jessiman, "A multi-element profile of house dust in relation to exterior dust and soils in the city of Ottawa, Canada," Science of the total Environment, **267**(1-3), 125-140, 2001, doi:

10.1016/S0048-9697(00)00775-0.
[44] H. Fleischer, S. Statkevych, J. Widmer, K. Thurow, "Allergene und Schwermetalle in Hausstaub - Bestimmung der elementaren Zusammensetzung mittels ICP-MS," GIT Labor-Fachzeitschrift, **65**(5-6), 36-38, 2021.
[45] H. Fleischer, K. Thurow, Automation Solutions for Analytical Measurements - Concepts and Applications, Wiley-VCH, 2017.
[46] H. Fleischer, K. Thurow, "On the Way to Efficient Analytical Measurements: The Future of Robot-Based Measurements," SLAS Technology, **25**(2), 208-211, 2020, doi: 10.1177/2472630319886270.
[47] S. Neubert, X. Gu, B. Göde, T. Roddelkopf, H. Fleischer, N. Stoll, K. Thurow, "Workflow Management System for the Integration of Mobile Robots in Future Labs of Life Sciences," Chemie-Ingenieur-Technik, **91**(3), 294-304, 2019, doi: 10.1002/cite.201800007.
[48] H. Fleischer, M. Adam, K. Thurow, "A Cross-platform Modular Software Solution for Automated Data Evaluation Applied in elemental and Structural Mass Spectrometry," in 2015 IEEE International Conference on Automation Science and Engineering (CASE), 758-763, 2015, doi: 10.1109/CoASE.2015.7294172.
[49] M. Adam, H. Fleischer, K. Thurow, "Generic and Automated Data Evaluation in Analytical Measurement," SLAS Technology, **22**(2), 186-194, 2017, doi: 10.1177/2211068216672613.
[50] H. Liu, N. Stoll, S. Junginger, K. Thurow, "Mobile robot for life science automation," International Journal of Advanced Robotic Systems, **10**, 1-14, 2013, doi: 10.5772/56670.

Advances in Science, Technology and Engineering Systems Journal Vol. 7, No. 3, 152-160 (2022)

www.astesj.com

Special Issue on Multidisciplinary Sciences and Engineering

ASTESJ

ISSN: 2415-6698

Deep Learning Affective Computing to Elicit Sentiment Towards Information Security Policies

Tiny du Toit*, Hennie Kruger, Lynette Drevin, Nicolaas Maree

School of Computer Science and Information Systems, North-West University, Potchefstroom, 2531, South Africa

ARTICLE INFO

Article history:
Received: 31 January, 2022
Accepted: 06 June, 2022
Online: 27 June, 2022

Keywords:
Affective computing
Deep learning
Information security policies
Non-compliance
Sentiment analysis

ABSTRACT

Information security behaviour is an integral part of modern business and has become a central theme in many research studies. One of the essential tools available that can be used to influence information security behaviour is information security policies (ISPs). These types of policies, which is mandatory in most organisations, are formalised rules and regulations which guide the safeguarding of information assets. Despite a significant number of ISP and related studies, a growing number of studies report ISP non-compliance as one of the main factors contributing to undesirable information security behaviour. It is noteworthy that these studies generally do not focus on the opinion of users or employees about the contents of the ISPs that they have to adhere to. The traditional approach to obtain user or employee opinions is to conduct a survey and ask for their opinion. However, surveys present unique challenges in fake answers and response bias, often rendering results unreliable and useless. This paper proposes a deep learning affective computing approach to perform sentiment analysis based on facial expressions. The aim is to address the problem of response bias that may occur during an opinion survey and provide decision-makers with a tool and methodology to evaluate the quality of their ISPs. The proposed affective computing methodology produced positive results in an experimental case study. The deep learning model accurately classified positive, negative, and neutral opinions based on the sentiment conveyed through facial expressions.

1. Introduction

The importance of information security behaviour and the challenges associated with using information security policies (ISPs) as a management tool to ensure that employees and users comply with security requirements is a widely studied discipline. This paper addresses specific concerns and techniques that may assist in evaluating ISPs and is an extension of the work initially presented at the 2020 2nd International Multidisciplinary Information Technology and Engineering Conference (IMITEC) [1]. This paper is also partially based on a master's degree study done in Computer Science [2].

Information security behaviour forms part of the general information security discipline and refers to the protection of information and information technology assets [3]. The human behaviour element of information security has become an integral part of modern enterprises, and considerable amounts of effort are often assigned to ensure that information security awareness, ISPs and other relevant human aspects are sufficiently addressed [4]. Technical solutions for undesirable human information security behaviour play an essential role [5] but are generally inappropriate on their own [6]. Additional measures to address the behaviour problem effectively are necessary. One approach often employed to influence security behaviour is ISPs [7], [8]. The popularity of ISPs as a control measure has inspired many studies with new research that is regularly added to the information security discipline [9]-[11].

Despite a large number of ISP and related studies, there is still a significant number of problems such as the inefficient use or non-compliance to ISPs that are regularly reported in the literature. Behavioural problems are evidenced by phenomena such as the privacy paradox [12] and the knowing-doing gap [13]. Users with a high level of information security awareness are easily persuaded to reveal personal or confidential information. Literature resources also indicate that one of the major contributing factors influencing the effective use of an ISP is the general lack of compliance [14], [15]. The work of [16] also presents a systematic overview of studies related to ISP compliance. Moreover, the lack of ISP compliance has also led to studies investigating the use of

*Corresponding Author: Tiny du Toit, North-West University, South Africa
Tel: +27828472512 E-mail: Tiny.DuToit@nwu.ac.za

https://dx.doi.org/10.25046/aj070317

psychological models to explain information security behaviour [17], [18].

It is clear from the above that many research projects are continuously conducted to evaluate and explain different aspects of ISP compliance. However, despite this large number of studies, little attention is given to the opinion of employees or users about the ISPs that they have to adhere to. For an ISP to be successful, employees should buy into the contents of the ISP and should have a positive attitude towards the contents – if not, non-compliance is likely to remain a reality. Two traditional methods to obtain the opinion of people or workers are to ask them or physically observe their behaviour. However, in addition to logistical difficulties (specifically to monitor employees), both techniques are subjected to biased results. During observation, users may comply with an ISP out of fear or merely because they know it is expected. Direct questioning through interviews or surveys also presents similar problems such as response bias, where answers may be faked [19]. In an attempt to address the bias problem, sentiment analysis, also known as opinion mining [20], is often employed. This technique enables decision-makers to determine whether someone has a positive, negative or neutral opinion or attitude about something through an analysis of personal sentiment information. Text-based sentiment analysis is a popular approach to determine someone's sentiment [21]. However, an ISP may still be subjected to response bias when a user simply writes down what is expected. To address this problem, affective computing may be used to perform sentiment analysis. Affective computing is a computational approach that aims to diagnose and measure emotional expression [22] and then use these measurements to evaluate human behaviour [23]. The technique can determine a user's opinion without asking any questions, thereby removing the risk of social desirability.

In this paper, the aim is to employ affective computing and sentiment analysis to address response bias problems and contribute to evaluating the quality of ISPs. The results would assist management in positively addressing challenges within ISPs and timely assessing and changing the contents of an ISP. The remainder of the paper is structured as follows. In Section 2, a brief overview of ISPs will be given, while background information on sentiment analysis and affective computing will be presented in Section 3. In Section 4, deep learning, which forms the basis of the experimentation, will be addressed. The experimental design of an illustrative case study will be discussed in Section 5, with the results and a reflection presented in Section 6. The paper will be concluded in Section 7 with some final remarks.

2. Information Security Policies

There are several definitions in the literature for an ISP. The authors of [24] provide a basic description by referring to an ISP as a set of rules and regulations that inform users of their responsibilities to safeguard information technology assets. A more formal definition at an organisational level is given by [25] as "a set of formalised procedures, guidelines, roles and responsibilities to which employees are required to adhere to safeguard and use properly the information and technology resources of their organizations". The importance of an ISP is also confirmed in internationally accepted information security standards such as the ISO/IEC 27002 standard which defines the objective of an ISP as "to provide management direction and support for information security in accordance with business requirements and relevant laws and regulations" (Source: www.iso.org/standards.html). These formal information security standards also prescribe ISPs as mandatory for information security management [26], and auditors are regularly advised to review the understanding and compliance of ISPs to ensure that users maintain acceptable levels of information security behaviour [27].

There is a general consensus that an ISP plays a critical role in any organisation. The researchers of [28] argue that effective information security management in organisations is largely dependent on the adherence to ISPs, while [4] state that the long-term success of any organisation in the current global and digitally driven economy is determined by the creation, deployment and enforcement of ISPs. However, there still seems to be an ongoing problem in ISP compliance. Large numbers of studies are found in the literature that try to explain and even predict the non-compliance of ISPs. Examples of such studies include the work of [24], who propose a model to raise the level of ISP compliance amongst end-users; [28], to predict ISP compliance, proposed a theoretical model that links security-related stress, discrete emotions, coping response and ISP compliance; and [29] who performed a study where aspects of the theory of planned behaviour and ISP compliance were investigated. Other examples of studies that employ psychological models to explain non-compliance can be found in [30], [31]. In addition to the existing non-compliance problem, it is also clear from the literature that employees and users are affected by the quality of an ISP. The scholars of [32] argue that the general quality of an ISP will affect employee satisfaction and ultimately plays a significant role in ISP compliance. This poses another question on how to determine employee or user satisfaction with an ISP. As alluded to in the introduction, the answer may be to simply ask employees for their opinion on the ISP. This, however, is not an easy task as different problems such as social desirability may render results invalid.

Social desirability is defined as the tendency to answer questions acceptable rather than truthful [33]. It is a significant problem in situations where opinions are solicited, and numerous studies exist on various aspects of applications and ways to address any adverse effects [34]-[36]. Social desirability is also applicable in information security, such as information security behaviour [19] and information security awareness evaluations [37]. The work by [38] is of particular interest as this research study has proved that response bias exists in current scale measurements used in compliance research. As a result, the findings of several studies in policy compliance may be questionable. To overcome these problems, this paper aims to introduce sentiment analysis and affective computing to exclude possible response bias when evaluating the quality of an ISP. A brief introduction to sentiment analysis and affective computing is presented in the next section.

3. Sentiment Analysis and Affective Computing

Opinions, like emotions, play an important role in human decision-making; thus, emotion recognition and sentiment analysis are critical for determining user or consumer preferences and opinions. Furthermore, sentiment analysis can enhance organizational functions such as sales and marketing by allowing

researchers to better understand consumers' preferences and behaviours [39]. Affective computing is a relatively recent method for computationally identifying and measuring emotions to adapt decisions to support people's emotional states. Therefore, in this paper, affective computing is suggested to analyse the opinions of employees or users towards ISPs. This will allow information security administrators to develop high-quality ISPs with high user satisfaction while excluding problems like social desirability from the opinion survey process.

3.1. Sentiment Analysis

Sentiment analysis is a method for analyzing people's feelings or opinions towards an entity [40]. Text-based sentiment analysis has an extensive body of knowledge, and studies in this field are performed regularly [40], [41]. These studies, however, remain difficult because they require a deep understanding of language, both in terms of semantics and syntax [42]. Therefore, it has become a more common practice to perform sentiment analysis using videos rather than text. The advancement and availability of communication technology (i.e. consumers who tend to record their opinions on products using a webcam and then upload the videos to social media platforms) are two reasons for this trend, according to [39]. Videos also provide multimodal data, such as vocal and visual modalities, contributing to more accurate emotion and sentiment models. The fundamental task of video sentiment analysis is to detect, model, and exploit the sentiment conveyed by facial gestures, as shown in numerous instances in the literature [42], [43]. Extracting emotions for sentiment analysis is a well-known task in affective computing, which will be addressed in more detail in the next section.

Figure 1: Emotions as represented by facial expressions.

3.2. Affective Computing

Affective computing is described by [44] as techniques for detecting, recognising, and predicting human emotions such as anger, fear, disgust, surprise, pleasure, and sadness. It is a branch of artificial intelligence dealing with creating or adapting computational systems to offer decision support depending on an individual's emotional state. Emotions may be identified by observing facial expressions, followed by a feature extraction process, which is then used to classify emotions. Figure 1 (Source: https://www.linkedin.com/pulse/scientific-tactics-boost-non-verbal-communication-body-rokham-fard/) is an example of the six fundamental universally distinctive emotions [45] as represented by facial expressions.

The data used in this paper's experimental case study is similar to the facial expressions presented in Figure 1 and consists of videos of people reading various text passages to prompt a particular sentiment. However, computational requirements dictate that the affective data be converted and represented quantitatively. This quantification process was performed using the Affectiva Software Development Kit (SDK) [46]. The Affectiva system is a reliable affective computing tool trained on more than 7.5 million faces. The Affectiva system processes information in four stages to classify emotional states in videos: detecting faces and 34 facial landmarks, feature extraction from face texture, classification of facial actions, and modelling emotion expression [46]. Figure 2 is an example of Affectiva's 34 identified landmarks, used to calculate 43 numeric metrics to classify emotions. Among the 43 metrics produced are seven emotions (the six identified by [45] in Figure 1 plus the emotion contempt), 21 facial expressions (e.g. brow raise, eye widen, jaw drop, etc.), 13 emojis (e.g. wink, smiley, etc.), and two additional values to represent valence and engagement.

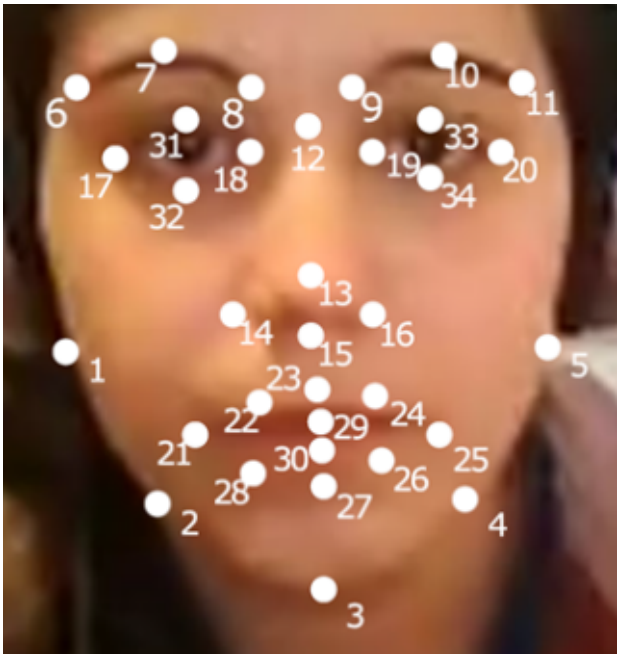


Figure 2: Facial landmarks identified by Affectiva.

Facial expression and emotion recognition research is widespread, and there are numerous relevant research projects in the literature [42], [47] and [48]. For example, two artificial intelligence researchers in Japan reported a practical application where facial expressions and emotion recognition were used to predict future policy changes. The Governor of the Bank of Japan's facial expressions at post-meeting news conferences were analysed in this study. Predicting an impending negative policy shift was possible based on observed signs of emotions such as anger and disgust, which were correlated with negative interest rates. The same researchers performed a follow-up and similar study (Source:

www.japantimes.co.jp), this time analysing the facial expressions of the European Central Bank's Chief. As in the first study, observing signs of sadness in videos recorded at previous press conferences enabled the prediction of negative changes in the bank's monetary policy.

These two examples clearly demonstrate the purpose of this current paper, which is to utilize identified emotions to evaluate if users or employees have a positive, negative, or neutral opinion of an organization's ISP. In this study, the identified emotions are obtained from video recordings of students reading known text passages which elicit specific emotions and their corresponding opinions from the subjects. Subsequently, a deep learning model is built to associate the elicited emotions obtained from facial expressions with the three opinion classes. First, background information on deep learning is presented in the following section. Then, the technique is applied in the illustrative case study of Section 5.

4. Deep learning

Machine learning is a subfield of artificial intelligence that focuses on constructing computer programs that can automatically adapt based on experience [49]. It has a broad field of applications, including, but not limited to, computer vision, speech recognition, natural language processing, and robotics. Until recently, research within machine learning generally employed shallow artificial neural networks, consisting of at most two hidden layers and one input layer [50]. These shallow models proved to be useful in solving basic and well-constrained problems. However, difficulties emerged when they were applied to problems with greater complexity levels, such as processing human voice, language, and real images and sceneries. The processing of raw natural data using shallow artificial neural networks was rather restricted [51]. Extensive domain expertise and careful engineering were necessary to create a machine learning system capable of extracting and transforming raw input data into an internal representation that the classifier could readily utilise to recognise and classify patterns in the input.

In 2006, deep learning originated from research in machine learning and artificial neural networks [50]. It was inspired by the deep architectures of human information processing mechanisms employed to extract complex structures and generate internal representations based on rich sensory inputs. Because deep learning models may convert a representation at one (lower) level into a higher abstracted representation, they can learn complex functions [51]. Starting with the raw input data, this transformation ensures that only the essential characteristics of the classification problem are highlighted while irrelevant aspects are ignored.

According to [52], artificial neural networks are structures of nodes or neurons (densely interconnected processing elements) that can perform many parallel computations. The architecture of a neural network is characterised by the pattern of connections between the neurons, the training or learning algorithm (the method for calculating the weights on the connections) and the activation function [53]. Deep learning is machine learning that uses neural networks with many layers of nonlinear nodes to solve problems. For feature extraction, supervised or unsupervised learning approaches are used at each of the successively higher levels of abstracted layers [42], [50]. In addition, in deep learning models, gradient-based optimisation algorithms such as the backpropagation algorithm modify the network's parameters depending on the output error rate [49]. The latter technique is discussed in more detail next.

4.1. Neural network training

The most fundamental deep learning neural network is a multilayer perceptron (MLP) neural network based on [53], [54]. An MLP comprises an input and an output layer and several hidden layers in between. It takes an input x and maps it to a category y by transferring the input values sequentially from one layer of nodes to the next and is represented as follows:

$$y = f(x, \theta), \tag{1}$$

where θ denotes the parameters, i.e. connection weights and biases, that the MLP uses to learn. It is important to notice that an MLP does not have any connections that transfer higher-level output values to lower-level nodes. Each layer of nodes has parameters that support the MLP in its learning process.

The term *learning* refers to the process of modifying the connection weights inside the MLP to minimise the difference between the desired and produced outputs [54]. The backpropagation algorithm [42], [55] is a frequently used method for training an MLP. The algorithm is given a collection of examples

$$\{\boldsymbol{p}_1, \boldsymbol{t}_1\}, \{\boldsymbol{p}_2, \boldsymbol{t}_2\}, \dots, \{\boldsymbol{p}_Q, \boldsymbol{t}_Q\}, \tag{2}$$

each of which comprises an input vector ($\boldsymbol{p}_Q$) that is mapped to a target output vector ($\boldsymbol{t}_Q$). The MLP adjusts its parameters in response to the calculated mean square error as it processes each of these inputs. This process can be summarised as follows:

1. Propagate the inputs forward through the MLP.
2. Calculate and propagate sensitivities backwards through the MLP.
3. Adjust the MLP's parameters accordingly.

For the first step, the outputs of a layer which is then used as input for the subsequent layer, is expressed as

$$\boldsymbol{a}^{m+1} = \boldsymbol{f}^{m+1}(\boldsymbol{W}^{m+1}\,\boldsymbol{a}^{m} + \boldsymbol{b}^{m+1}), \tag{3}$$

$$\text{for } m = 0, 1, \dots, M-1,$$

where $\boldsymbol{f}$ denotes the activation function, and $\boldsymbol{W}^n$ and $\boldsymbol{b}^n$ denote the weight vector and bias of layer n, respectively. M represents the number of layers in the MLP, and its starting point is denoted by

$$\boldsymbol{a}^0 = \boldsymbol{p}. \tag{4}$$

In (4) $\boldsymbol{p}$ denotes the original input vector, and the MLP's final layer's output represents the MLP's output, i.e.

$$\boldsymbol{a} = \boldsymbol{a}^M. \tag{5}$$

In the second step, the following equations are used to calculate the sensitivities:

$$\boldsymbol{s}^M = -2\dot{\boldsymbol{F}}^M(\boldsymbol{n}^M)(\boldsymbol{t} - \boldsymbol{a}), \tag{6}$$

where $\boldsymbol{n}$ denotes the net input, $\boldsymbol{t}$ represents the target or expected outputs, and

$$\dot{\boldsymbol{F}}^m(\boldsymbol{n}^m) = \begin{bmatrix} \dot{f}^m(n_1^m) & 0 & \cdots & 0 \\ 0 & \dot{f}^m(n_2^m) & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & \dot{f}^m(n_{S^m}^m) \end{bmatrix}, \quad (7)$$

where

$$\boldsymbol{s}^m = \dot{\boldsymbol{F}}^m(\boldsymbol{n}^m)(\boldsymbol{W}^{m+1})^T \boldsymbol{s}^{m+1}. \quad (8)$$

Finally, the MLP's biases and weights may be adjusted. This is accomplished via the use of the mean square error, which is calculated as follows:

$$\boldsymbol{W}^m(k+1) = \boldsymbol{W}^m(k) - \alpha \boldsymbol{s}^m(\boldsymbol{a}^{m-1})^T \text{ and} \quad (9)$$

$$\boldsymbol{b}^m(k+1) = \boldsymbol{b}^m(k) - \alpha \boldsymbol{s}^m, \quad (10)$$

at iteration k, with a learning rate represented by α.

More technical aspects of neural networks and deep learning are excluded due to the paper's scope. The work of [53] and [54] provide further details for interested readers. Constructing the best neural network model manually can be laborious. A neural architecture search methodology can alleviate this problem by finding architectures that perform well for the given data. This methodology is discussed in the following section.

4.2. Neural architecture search

The automation of machine learning model selection, hyperparameter optimization, and model search is called automated machine learning (AutoML) [56]. Neural architecture search (NAS), a subfield of AutoML that automates neural network architecture engineering, has resulted in models that outperform manually designed models [57]. The search space, search strategy, and performance estimation strategy are the three dimensions of a NAS method. Figure 3 depicts a simplified version of such a method.

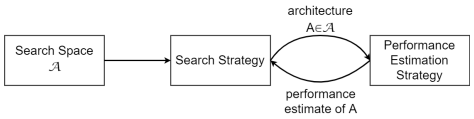


Figure 3: A high-level illustration of neural architecture search [57].

The search space ($\mathcal{A}$) defines all architectures that may be considered. Its size may be reduced by using previous knowledge about comparable task architectures, but this adds an undesired human bias. The maximum number of hidden layers (potentially unbounded), the operation of each layer, and the hyperparameters associated with the process define the search spaces of MLP neural networks and other chain-like neural networks. The choice of the search space determines the complexity of the architecture optimization problem, which is not continuous and has multiple dimensions.

A search strategy is used to explore the search space and identify an architecture $A \in \mathcal{A}$, which is then evaluated by the performance estimation strategy. Premature convergence to a region where suboptimal architectures exist should be avoided to find architectures that perform well. To find a suitable architecture inside the search space, approaches including random search, Bayesian optimisation, evolutionary methods, reinforcement learning, and gradient-based methods may be utilised. A reinforcement learning, evolutionary, and random search approach were compared in research by [58]. They discovered that the latter method outperformed the first two approaches. Furthermore, compared to the other two techniques, the evolutionary method created models with better accuracy throughout the early stages of the process. To develop and choose a suitable architecture in the experiment, a modified version of a regularised evolution approach given by [58] was implemented for the search strategy utilised in this work. This method is summarised in Algorithm 1.

Algorithm 1: Regularised evolution search strategy

```
Result: Highest accuracy model in history
population ← empty queue;
history ← empty list;
while | population | < P do
    model.arch ← RANDOM_ARCHITECTURE();
    model.accuracy ← TRAIN_AND_EVAL(model.arch);
    add model to right of population;
    add model to history;
end
while | history | < C do
    sample ← empty list;
    while | sample | < S do
        candidate ← distinct random element from population;
        add candidate to sample;
    end
    parent ← highest accuracy model in sample;
    child.arch ← MUTATE(parent.arch);
    child.accuracy ← TRAIN_AND_EVAL(child.arch);
    add child to right of population;
    add child to history;
    remove dead from the left of population;
    discard dead;
end
return highest accuracy model in history
```

Throughout the experiment, the method stores a population of previously trained models. At the start of the experiment P models with random architectures, based on the search space outlined above, are introduced to the population. The population is then mutated and added to the history list using C cycles. During each cycle, S candidates are selected at random from the population. After that, the candidate with the best accuracy is selected, mutated, and trained, resulting in a child model. A mutation performs a simple and randomised change in the chosen architecture. To achieve this, randomising one or more of the architecture's hyperparameters is done. The population and history are then updated to include the child model. Finally, the population is adjusted to exclude the oldest model. The performance estimation strategy is kept simple by maximising the model's validation loss. The generated models are configured to finish training when the model's accuracy begins to converge to guarantee that the NAS method makes optimal use of computing resources. In the next section, specific performance metrics used to evaluate the best neural network model found is addressed.

4.3. Performance metrics

According to [59], evaluating the performance of a machine learning model using just one aggregated measurement is insufficient. The researchers of [60], [61], [62], [63] and [64] all utilise or advise using different performance metrics. The following are some of the performance measures:

- Accuracy;
- Precision;
- Recall or sensitivity; and
- *F*-measure, also sometimes referred to as the F_1-measure.

Each sample in a testing process is always labelled with a real and a predicted label [61]. The real label identifies the real class to which the testing sample belongs. The predicted label is the predictor's output. As shown in Table 1, a multiclass confusion matrix can visually represent these label counts.

Table 1: Multiclass confusion matrix [65].

		Predicted		
		Class$_1$ - Class$_{k-1}$	**Class$_k$**	**Class$_{k+1}$ - Class$_n$**
Real	**Class$_{k+1}$ - Class$_n$**	tn_1	fp_1	tn_2
	Class$_k$	fn_1	tp	fn_2
	Class$_1$ - Class$_{k-1}$	tn_3	fp_2	tn_4

All of the above performance measures are based on the values represented by the multiclass confusion matrix. Each of the measures is discussed briefly below, along with a definition. The most common metric is accuracy, which determines how well the model can correctly classify positive and negative samples. To calculate accuracy, the number of correctly classified samples, positive and negative, are divided by the total number of samples.

As a result, it can be formalised as follows:

$$\text{Average accuracy} = \left(\sum_{i=1}^{n} \frac{tp_i + tn_i}{tp_i + tn_i + fp_i + fn_i}\right) / n. \quad (11)$$

The error rate is a measurement of how frequently errors occurred during the prediction phase. It is given as

$$\text{Average error rate} = \left(\sum_{i=1}^{n} \frac{fp_i + fn_i}{tp_i + tn_i + fp_i + fn_i}\right) / n. \quad (12)$$

The precision measure can be used to calculate the proportion of correctly classified true positives versus the total number of predicted positives. As a result, its definition is as follows:

$$\text{Precision}_M = \left(\sum_{i=1}^{n} \frac{tp_i}{tp_i + fp_i}\right) / n. \quad (13)$$

The recall measure calculates the proportion of samples labelled as positive compared to all truly positive samples. Consequently, this metric denotes the model's completeness. It can be defined as follows:

$$\text{Recall}_M = \left(\sum_{i=1}^{n} \frac{tp_i}{tp_i + fn_i}\right) / n. \quad (14)$$

Finally, the *F*-measure, also known as the harmonic mean of precision and recall, is a metric for determining how accurately a model performed on a test. The metric is defined as

$$F_M = \frac{(\beta^2+1) \times Precision_M \times Recall_M}{\beta^2 \times Precision_M + Recall_M}, \text{where } 0 \leq \beta \leq +\infty. \quad (15)$$

The β value is used to balance the importance of precision and recall. *F* becomes the harmonic mean of precision and recall if β is equal to 1 because both measures have the same weight. When β is greater than 1, *F* becomes more recall-oriented. In contrast, *F* becomes more precision-oriented when β is less than 1.

The following section will describe the experimental design to illustrate how deep learning affective computing and sentiment analysis may assist in solving response bias issues in the context of ISPs.

5. Experimental Design

A deep learning neural network approach is proposed to illustrate the concept of affective computing and sentiment analysis. This experimental approach is divided into two components: dataset acquisition and the building and testing of a deep learning neural network architecture.

5.1. Data acquisition

Instead of using publicly accessible videos, it was decided that a small video dataset would be generated as an initial experiment. A group of nine postgraduate Computer Science students agreed to participate in the study and help create facial expression videos. The nine participants were instructed to read three text passages while being recorded. The three text passages were selected to prompt a particular sentiment from the participants, and they were classified as positive, neutral, or negative. A collection of jokes was used to elicit a positive sentiment, and an ordinary neutral news article was used to evoke a neutral sentiment. Finally, a news article about consequences for unlawfully copying online material (which most students frequently do) was used to elicit a negative sentiment. The participants were informed that they would be recorded. Still, the objective of the exercise was not revealed until after the recording to ensure that they were not influenced to respond in a particular manner. The participants were offered the option of withdrawing from the experiment after they learned the purpose of the recordings. Despite this, they all decided to continue to be involved in the research.

The Affectiva SDK [46] was then used to extract 42 features from the 27 videos that were annotated based on the desired sentiment of the text passages. The emotion contempt was omitted from the dataset since it did not correlate with the sentiment. The pre-processing yielded 132 261 data records extracted from the videos. Positive sentiment was represented by 41 934 records,

neutral sentiment by 54 873 records, and negative sentiment by 35 454 records. The complete set of records was randomized and divided into three datasets: training (70%), validation (20%), and test (10%), all of which were utilized to build the deep neural network model.

5.2. Deep learning neural network architecture

To identify and select a suitable deep learning neural network architecture, the Google Colab cloud service was utilized. Then, model search, model selection, and hyperparameter optimisation were performed using the NAS methodology [57] described in Section 4.2. This method yielded a deep learning feed-forward neural network architecture with 42 input nodes (the 42 extracted facial expressions) and three output nodes (positive, neutral, and negative). The final layer used a softmax activation function to determine the sentiment of each input sample. In addition, five hidden layers were constructed, each using the ReLU activation function. Figure 4 shows a graphical representation of the deep learning model that was selected.

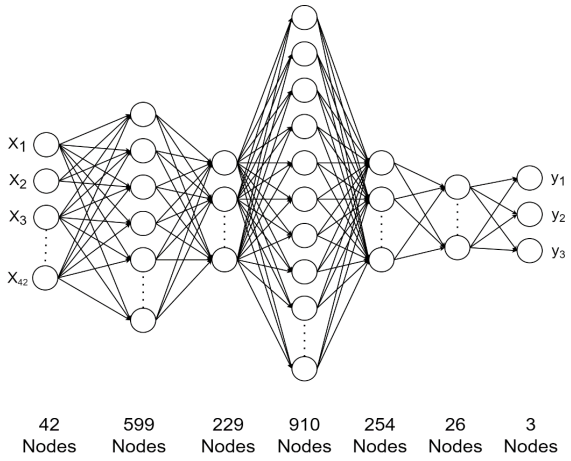


Figure 4: Deep learning neural network architecture.

The model was trained on the Google Colab cloud service for 2 hours and 16 minutes with a batch size of 9376 and 751 epochs. The accuracy was fairly high, as discussed further in the following section.

6. Results and Discussion

The selected model was evaluated on the test dataset to predict the out-of-sample class of each data record after the training and validation process. These predictions had an average accuracy of 96.23 percent, according to the results. The high levels of accuracy achieved with the selected architecture are detailed in a confusion matrix (Table 2) of the test dataset (13226 records). In addition, other calculated metrics, such as precision (average of 94.43 percent), recall (average of 94.19 percent), and *F*-measure (average of 94.31 percent), support the above-average results and the model's ability to perform sentiment classification. The high precision value shows that the model is very effective. In addition, the high recall value indicates that a high fraction of the total number of relevant instances was correctly classified.

The results will be discussed next regarding the selected deep learning model and the implications for ISP compliance.

Table 2: Confusion matrix for the test dataset.

		Predicted		
		Positive	**Neutral**	**Negative**
Real sentiment	**Positive**	3971	65	65
	Neutral	65	5237	213
	Negative	27	306	3277

6.1. Reflection on the deep learning model

The high accuracy result indicates that affective computing and sentiment analysis based on video analysis and an appropriate deep learning neural network architecture is feasible, supporting previous literature studies in this field. However, despite the high accuracy and excellent performance metrics achieved, the results of the particular illustrative experiment and the selected deep learning neural network model reported in this study should be interpreted with caution.

A variety of factors may impact the results, which will be considered in a follow-up study. The exceptionally high accuracy might be attributed to the limited number of participants utilized to create the videos. This means that a dataset with minimal variation was produced, which may aid the learning process in achieving high accuracy results. The minimal variation in the data may be contributed to the fact that all participants had the same study background. It is also uncertain if reading text is the most effective method of prompting a sentiment; maybe viewing a video would provide a more reliable dataset. Further experiments with splitting the dataset into training, validation, and test datasets may reduce overfitting.

Nonetheless, the objective was to show how a dataset including facial expressions might be generated and then used to perform sentiment analysis using a deep learning neural network. The experiment conducted in this paper achieved above-average results, demonstrating the feasibility of the suggested techniques.

6.2. Reflection on information security compliance

As explained previously, non-compliance with ISPs may be attributed partly to employees or users who negatively react to a policy because they disagree with its contents. Employee opinions may be obtained via surveys or text-based sentiment analysis; however, both methods might be biased since opinions can be expressed in a fake manner to meet expectations. When prompting employees for their opinions on the contents of an ISP, affective computing, which is based on emotional expression, offers a different approach that may be utilized to reduce the response bias problem. The dataset generated in this study, together with the selected deep learning neural network model, may be used to address social desirability problems in a similar way as predicting the sentiment of a bank governor based on facial expressions (see Section 3.2). It is no longer necessary to ask individuals their opinions; instead, one may deduce an opinion from their facial expressions. This may be especially significant when it comes to ISP compliance. Management will now understand whether or not employees are satisfied with the context of an ISP in general. It

may also assist in a more specific way by identifying particular areas of concern, leading to new or extra information security training opportunities.

A dataset acquired in the context of ISPs, i.e. employees reading an ISP, would be ideal for training a deep learning neural network model. This is unrealistic, however, since gathering a big enough sample of individuals who read an ISP would be difficult if not impossible. Furthermore, to create a dataset that can be utilized in a supervised learning environment, readers will be asked to indicate whether they found the ISP positive or negative, which puts one back to the response bias problem. The approach used in this paper is similar to that used in practice, i.e., in the example of bank governors, the training set was not constructed using a large number of bank governors but rather a large dataset of everyday videos from which facial expressions could be extracted. This implies that a model trained on regular individuals in videos may detect sentiment based on facial expressions in any other video.

This paper provided an example of the proposed concept. The following steps would be to collect a more extensive and more diverse dataset and test the model on employees that read an ISP.

7. Conclusion

This paper argues that the opinion of users and employees is essential in the creation and maintenance of ISPs. Employees should have a positive attitude toward an ISP and buy into the contents of the ISP to avoid non-compliance. However, obtaining user input on an ISP often poses a social desirability problem. Users are more likely to answer questions in an acceptable rather than truthful way. This study suggested sentiment analysis and affective computing to exclude possible fake responses while evaluating the contents of an ISP to minimize this problem. A deep learning neural network model was constructed to classify sentiment as positive, neutral, or negative in a real-world scenario. The model was trained using a video dataset of individuals reading various text passages to elicit multiple facial expressions. The suggested method proved to be an acceptable choice after achieving high accuracy. The experiment's findings may significantly affect how ISPs are evaluated since it would no longer be required to ask consumers for their opinions, which risks social desirability. Applying the suggested affective computing and sentiment analysis improves the policy evaluation process by making it simpler to gather opinions without the risk of fake answers. Management may identify areas of concern that may be addressed by either changing or correcting the policy's contents or giving extra training to particular (negative) users, or training on specific topics.

The research presented in this paper is an exploratory study, and many opportunities for future investigation have been identified. For example, experiments involving larger populations (participants being recorded), various methods of evoking emotions (i.e. viewing a video instead of reading text), and the use of different neural network architectures are all possibilities.

References

[1] H. Kruger, T. du Toit, L. Drevin, N. Maree, “Acquiring sentiment towards information security policies through affective computing,” in 2nd International Multidisciplinary Information Technology and Engineering Conference (IMITEC), 1-6, 2020, doi:10.1109/IMITEC50163.2020.9334134.

[2] N. Maree, Affective computing and deep learning to perform sentiment analysis, M. Sc. Thesis, North-West University, South Africa, 2020.

[3] R.E. Crossler, A.C. Johnston, P.B. Lowry, Q. Hu, M. Warkentin, R. Baskerville, "Future directions for behavioural information security research," Computers & Security, **32**, 90-101, 2013, doi:10.1016/j.cose.2012.09.010.

[4] W.A. Cram, J. D'Arcy, J.G. Proudfoot, "Seeing the forest and the trees: a meta-analysis of the antecedents to information security policy compliance," MIS Quarterly, **43**(2), 525-554, 2019, doi:10.25300/MISQ/2019/15117.

[5] V.T. Patil, P.R. Patil, V.O. Patil, S.V. Patil, "Performance and information security evolution with firewalls," Journal of Scientific Computing, **8**(4), 1-6, 2019, doi:16.10089.ISC.2019.V8I5.285311.2630.

[6] M. Butavicius, K. Parsons, M. Lillie, A. McCormac, M. Pattinson, D. Calic, "When believing in technology leads to poor cyber security: Development of a trust in technical controls scale," Computers & Security, **98**, 102020, 2020, doi:10.1016/j.cose.2020.102020.

[7] G.D. Moody, M. Siponen, S. Pahnila, "Toward a unified model of information security policy compliance," MIS Quarterly, **42**(1), 285-311, 2018, doi:10.25300/MISQ/2018/13853.

[8] J. C. Sipior, D.R. Lombardi, "The impact of employee organisational commitment on compliance with information security policy," in Proceedings of the 2019 Southern Association for Information Systems Conference (SAIS), 2019.

[9] M. Kang, A. Hovav, "Benchmarking methodology for information security policy (BMISP): Artifact development and evaluation," Information Systems Frontiers, **22**, 221-242, 2020, doi:10.1007/s10796-018-9855-6.

[10] M. Karjalainen, M.T. Siponen, S. Sarker, "Toward a stage theory of the development of employees' information security behaviour," Computers & Security, **93**, 101782, 2020, doi:10.1016/j.cose.2020.101782.

[11] A. Vance, M.T. Siponen, D.W. Straub, "Effects of sanctions, moral beliefs, and neutralization on information security policy violations across cultures," Information & Management, **57**(4), 103212, 2020, doi:10.1016/j.im.2019.103212.

[12] S. Kokolakis, "Privacy attitudes and privacy behavior: a review of current research on the privacy paradox phenomenon," Computers & Security, **64**, 122-134, 2017, doi:10.1016/j.cose.2015.07.002.

[13] J.A. Cox, "Information systems user security: a structured model of the knowing-doing gap," Computers in Human Behavior, **28**(5), 1849-1858, 2012, doi:10.1016/j.chb.2012.05.003.

[14] K.L. Gwebu, J. Wang, M.Y. Hu, "Information security policy noncompliance: An integrative social influence model," Information Systems Journal, **30**(2), 220-269, 2020, doi:10.1111/isj.12257.

[15] J.H. Nord, A. Koohang, K. Floyd, "Impact of habits on information security policy compliance," Issues in Information Systems, **21**(3), 217-226, 2020, doi:/10.48009/3_iis_2020_217-226.

[16] R.A. Alias, "Information security policy compliance: Systematic literature review," Procedia Computer Science, **161**(2019), 1216-1224, 2019, doi:10.1016/j.procs.2019.11.235.

[17] P. Ifinedo, "Understanding information systems security policy compliance: An integration of the theory of planned behavior and the protection motivation theory," Computers & Security, **31**(1), 83-95, 2012, doi:10.1016/j.cose.2011.10.007.

[18] T.B. Lembcke, K. Masuch, S. Trang, S. Hengstler, P. Plics, M. Pamuk, "Fostering information security compliance: Comparing the predictive power of social learning theory and deterrence theory," in Proceedings of the 2019 American Conference on Information Systems (AMCIS), Information Security and privacy (SIGSEC), 2019.

[19] D.P. Snyman, H.A. Kruger, W.D. Kearney, "The lemming effect in information security," in Proceedings of the 2017 International Symposium on Human Aspects of Information Security & Assurance (HAISA), 91-103, 2017.

[20] S. Redhu, S. Srivastava, B. Bansal, G. Gupta, "Sentiment analysis using text mining: a review," International Journal on Data Science and Technology, **4**(2), 49-53, 2018, doi:10.11648/j.ijdst.20180402.12.

[21] G.S. Murthy, S.R. Allu, "Text based sentiment analysis using LSTM," International Journal of Engineering Research & Technology, **9**(5), 299-303, 2020, doi:10.17577/IJERTV9IS050290.

[22] E. Yadegaridehkordi, N.F.B.M. Noor, M.N.B. Ayub, H.B. Affal, N.B. Hussin, "Affective computing in education: a systematic review and future research," Computers & Education, **142**, 2019, doi:10.1016/j.compedu.2019.103649.

[23] S. Richardson, "Affective computing in the modern workplace," Business Information review, **37**(2), 78-85, 2020, doi:10.1177/0266382120930866.

[24] M.J. Alotaibi, S. Furnell, N. Clarke, "A framework for reporting and dealing

with end-user security policy compliance," Information & Computer Security, **27**(1), 2-25, 2019, doi:10.1108/ics-12-2017-0097.
[25] P.B. Lowry, G.D. Moody, "Proposing the control-reactance compliance model (CRCM) to explain opposing motivations to comply with organisational information security policies," Information Systems Journal, **25**(5), 433-463, 2015, doi:10.1111/isj.12043.
[26] H. Paananen, M. Lapke, M. Siponen, "State of the art in information security policy development," Computers & Security, **88**, 2020, doi:10.1016/j.cose.2019.101608.
[27] T. Stafford, G. Deitz, Y. Li, "The role of internal audit and user training in information security policy compliance," Managerial Auditing Journal, **33**(4), 410-424, 2018, doi:10.1108/MAJ-07-2017-1596.
[28] J. D'Arcy, P. The, "Predicting employee information security policy compliance on a daily basis: the interplay of security-related stress, emotions and neutralization," Information & Management, **56**(7), 2019, doi:10.1016/j.im.2019.02.006.
[29] T. Sommestad, H. Karlzen, J. Hallberg, "The theory of planned behaviour and information security policy compliance," Journal of Computer Information Systems, **59**(4), 344-353, 2019, doi:10.1080/08874417.2017.1368421.
[30] M. Rajab, A. Eydgahi, "Evaluating the explanatory power of theoretical frameworks on intention to comply with information security policies in higher education," Computers & Security, **80**, 211-223, 2019, doi:10.1016/j.cose.2018.09.016.
[31] S. Trang, B. Brendel, "A meta-analysis of deterrence theory in information security policy compliance research," Information Systems Frontiers, **21**(6), 1265-1284, 2019, doi:10.1007/s10796-019-09956-4.
[32] A. Alzahrani, C. Johnson, S. Altamimi, "Information security compliance: investigating the role of intrinsic motivation towards policy compliancy in the organisation," in Proceedings of the 2018 International Conference on Information Management (ICIM), 125-132, 2018, doi:10.1109/INFOMAN.2018.8392822.
[33] R.J. Fisher, "Social desirability bias and the validity of indirect questioning," Journal of Consumer Research, **20**(2), 303-315, 1993, doi:10.1086/209351.
[34] N. Bergen, R. Labonte, "Everything is perfect and we have no problems: Detecting and limiting social desirability bias in qualitative research," Qualitative Health Research, **30**(5), 783-792, 2020, doi:10.1177/1049732319889354.
[35] D. Burchett, Y.S. Ben-Porath, "Methodological considerations for developing and evaluating response bias indicators," Psychological Assessment, **31**(12), 1497–1511, 2019, doi:10.1037/pas0000680.
[36] D. Kwak, P. Holtkamp, S.S. Kim, "Measuring and controlling social desirability bias: Applications in information systems research," Journal of the Association for Information Systems, **20**(4), 2019, doi:10.17705/1jais.00537.
[37] A. McCormac, D. Calic, M. Butavicius, K. Parsons, T. Zwaans, M. Pattinson, "A reliable measure of information security awareness and the identification of bias in responses," Australasian Journal of Information Systems, **21**, 1-12, 2017, doi:10.3127/ajis.v21i0.1697.
[38] S. Kurowski, "Response biases in policy compliance research," Information & Computer Security, 2019, doi:10.1108/ICS-02-2019-0025.
[39] S. Poria, N. Majumder, E. Cambria, A. Gelbukh, A. Hussain, "Multimodal sentiment analysis: addressing key issues and setting up the baselines," IEEE Intelligent Systems, **33**(6), 17-25, 2018, doi:10.1109/MIS.2018.2882362.
[40] J.K. Rout, K.-K.R. Choo, A.K. Dash, S. Bakshi, S.K. Jena, K.L. Williams, "A model for sentiment and emotion analysis of unstructured social media text," Electronic Commerce Research, **18**(1), 181-199, 2018, doi:10.1007/s10660-017-9257-8.
[41] D.P. Alamanda, A. Ramdhani, I. Kania, W. Susilawati, E.S. Hadi, "Sentiment analysis using text mining of Indonesia tourism reviews via social media," International Journal of Humanities, Arts and Social Sciences, **5**(2), 72-82, 2019, doi:10.20469/ijhss.5.10004-2.
[42] N. Maree, T. du Toit, L. Drevin, H. Kruger, "Affective computing and deep learning to perform sentiment analysis," in Proceedings of the 2019 Southern Africa Telecommunication Networks and Applications Conference (SATNAC), 94-99, 2019.
[43] S. Poria, E. Cambria, N. Howard, G.-B. Huang, A. Hussain, "Fusing audio, visual and textual clues for sentiment analysis from multimodal content," Neurocomputing, **174**, 50-59, 2016, doi:10.1016/j.neucom.2015.01.095.
[44] B. Kratzwald, S. Ilic, M. Kraus, S. Feuerriegel, H. Prendinger, "Deep learning for affective computing: text-based emotion recognition in decision support," Decision Support Systems, **115**, 24-35, 2018, doi:10.1016/j.dss.2018.09.002.
[45] P. Ekman, Basic emotions. Handbook of cognition and emotion, **98**(45-60), 16, 1999.
[46] D. McDuff, M. Mahmoud, M. Mavadati, J. Amr, J. Turcot, R. Kaliouby, "AFFDEX SDK: a cross-platform real-time multi-face expression recognition toolkit," in Proceedings of the 2016 CHI conference extended abstracts on human factors in computing systems, 3723-3726, 2016, doi:10.1145/2851581.2890247.
[47] S.E. Kahou, X. Bouthillier, P. Lamblin, C. Gulcehre, V. Michalski, K. Konda, S. Jean, P. Froumenty, Y. Dauphin, N. Boulanger-Lewandowski, et al., "Emonets: multimodal deep learning approaches for emotion recognition in video," Journal on Multimodal User Interfaces, **10**(2), 99-111, 2016, doi:10.1007/s12193-015-0195-2.
[48] O.M. Nezami, M. Dras, P. Anderson, L. Hamey, "Face-cap: image captioning using facial expression analysis," Joint European Conference on Machine Learning and Knowledge Discovery in Databases: Springer, 226-240, 2018, doi:10.1007/978-3-030-10925-7_14.
[49] M.I. Jordan, T.M. Mitchell, "Machine learning: Trends, perspectives, and prospects," Science, **349**(6245), 255-260, 2015, doi:10.1126/science.aaa841.
[50] L. Deng, D. Yu, "Deep learning: methods and applications. Foundations and trends in signal processing," **7**(3–4), 197-387, 2014, doi:10.1561/2000000039.
[51] Y. LeCun, Y. Bengio, G. Hinton, "Deep learning," Nature, **521**(7553), 436, 2015, doi:10.1038/nature14539.
[52] I.A. Basheer, M. Hajmeer, "Artificial neural networks: fundamentals, computing, design, and application," Journal of Microbiological Methods, **43**(1), 3-31, 2000, doi:10.1016/S0167-7012(00)00201-3.
[53] I. Goodfellow, Y. Bengio, A. Courville, Deep learning, MIT press, 2016.
[54] H. Ramchoun, M.A.J. Idrissi, Y. Ghanou, M. Ettaouil, "Multilayer Perceptron: Architecture optimization and training," IJIMAI, **4**(1), 26-30, 2016, doi:10.9781/ijimai.2016.415.
[55] M.T. Hagan, H.B. Demuth, M.H. Beale, O. De Jesus, Neural Network Design, Martin Hagan, 2014.
[56] I. Guyon, K. Bennett, G. Cawley, H.J. Escalante, S. Escalera, T.K. Ho, N. Macia, B. Ray, M. Saeed, A. Statnikov, "Design of the 2015 ChaLearn AutoML challenge," in Proceedings of 2015 International Joint Conference on Neural Networks (IJCNN), IEEE, 1-8, 2015, doi:10.1109/IJCNN.2015.7280767.
[57] T. Elsken, J.H. Metzen, F. Hutter, "Neural architecture search: A survey," Journal of Machine Learning Research, **20**(55), 1-21, 2019, doi:10.5555/3322706.3361996.
[58] E. Real, A. Aggarwal, Y. Huang, Q.V. Le, "Regularized evolution for image classifier architecture search," in Proceedings of the 2019 AAAI Conference on Artificial Intelligence, **33**(1), 4780-4789, 2019, doi:10.1609/aaai.v33i01.33014780.
[59] P. Flach, "Performance evaluation in machine learning: The good, the bad, the ugly and the way forward," in Proceedings of 2019 AAAI Conference on Artificial Intelligence, 2019, doi:10.1609/aaai.v33i01.33019808.
[60] A. Tripathy, A. Agrawal, S.K. Rath, "Classification of sentiment reviews using n-gram machine learning approach," Expert Systems with Applications, **57**, 117-126, 2016, doi:10.1016/j.eswa.2016.03.028.
[61] Y. Jiao, P. Du, "Performance measures in evaluating machine learning based bioinformatics predictors for classifications," Quantitative Biology, **4**(4), 320-330, 2016, doi:10.1007/s40484-016-0081-2.
[62] E. Gokgoz, A. Subasi, "Comparison of decision tree algorithms for EMG signal classification using DWT," Biomedical Signal Processing and Control, **18**, 138-144, 2015, doi:10.1016/j.bspc.2014.12.005.
[63] D.M. Powers, "Evaluation: From precision, recall and F-measure to ROC, informedness, markedness and correlation," Journal of Machine Learning Technologies **2**(1), 37-63, 2011, doi:10.48550/arXiv.2010.16061.
[64] M. Sokolova, G. Lapalme, "A systematic analysis of performance measures for classification tasks," Information Processing & Management, **45**(4), 427-437, 2009, doi:10.1016/j.ipm.2009.03.002.
[65] F. Krüger, Activity, context, and plan recognition with computational causal behaviour models, Ph.D Thesis, Universität Rostock, 2016.

ASTES

Advances in Science, Technology and Engineering Systems Journal Vol. 7, No. 3, 161-165 (2022)

www.astesj.com

ASTESJ

ISSN: 2415-6698

NiO Quantum dots Doped Triple Cation Perovskite CsMAFAPbI$_2$Br$_2$ Heterojunction Photodetector with High Responsivity

Yara Abdullah Alwadei[1], Manar Saleh Alshatwi[2 *], Norah Mohammed Alwadai[2], Maymunah Abdullah AlWehaibi [2], Mohammad Faihan Alotaibi [3], Maha Mahmoud Lashin [4], Mohammad Hayal Alotaibi [3]

[1] *Manarat-Alriyadh Schools, P.O. Box 3835, Riyadh 1148, Saudi Arabia*

[2] *Department of Physics, College of Sciences, Princess Nourah bint Abdulrahman University, P.O. Box 84428, Riyadh 11671, Saudi Arabia*

[3] *National Center for Nanotechnology, King Abdulaziz City for Science and Technology, P.O. Box 6086, Riyadh 11442, Saudi Arabia*

[4] *College of Engineering, Princess Nourah bint Abdulrahman University, P.O. Box 84428, Riyadh 11671, Saudi Arabia*

ARTICLE INFO

Article history:
Received: 13 March, 2022
Accepted: 04 June, 2022
Online: 27 June, 2022

Keywords:
Photodetectors
Organic–inorganic perovskites
Quantum Dots
NiO
Responsivity
Heterojunction

ABSTRACT

Optoelectronic devices applications based on Organic–inorganic perovskites are promising and effective low-cost energy materials due to their exceptional physical properties which include high carrier mobility, high optical absorption coefficient, and long carrier diffusion length. In the presented work, a TiO$_2$/NiO+5% Fe quantum dots (QDs)–doped CsMAFAPbI$_2$Br$_2$ perovskite heterojunction broadband photodetector was fabricated on FTO/glass substrate. The photodetector can detect a wide range of wavelengths, from UV to Vis (100–800 nm), has high responsivity (0.99 A/W), and has excellent detectivity (8.9 × 10^{12} Jones). Scanning electron microscopy (SEM), transmission electron microscopy (TEM), X-ray diffraction (XRD), atomic force microscopy (AFM), and electrical (I–V) characterization were used to measure the responsivity and detectivity of the photodetectors. Doping with NiO+5% Fe QDs protected the device from oxygen and moisture and improved the morphology of the perovskite by reducing pit defects. The results showed high performance and the potential of a NiO+5% Fe QDs–doped triple cation perovskite photodetector device.

1. Introduction

In the last decade, hybrid organic–inorganic perovskites (HOIPs) have attracted significant attention because of their chemical variability, structural diversity, and favorable physical properties. HOIPs can be classified into formates, azides, dicyanamides, cyanides, and dicyanometallates. They are a future material for optoelectronic applications [1], optical communication, light-emitting diodes [2], UV detectors [3], gas sensors [4], transistors [5], and solar cells [6]. HOIPs have 400–990 nm coefficient of absorption, long carrier diffusion lengths, and high carrier mobilities [6]. ABX_3 is the main perovskite structure. Within the formula ABX_3, A is an organic or inorganic cation, usually methylammonium $CH_3NH_3^+$, formamidinium $CH_2(NH_2)_2^+$, FA, Cs^+, or Rb^+; B is Pb_2 or Sn_2; and X is a halide anion or mixed halide (to gain stability and resistance to environmental influence). The formula ABX_3 is illustrated in figure 1.

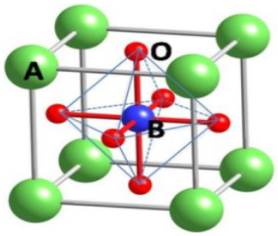


Figure 1: The structure of perovskite.

Varieties of perovskites have been developed via different preparation procedures, varying properties such as the choice of cations or halide elements, and optimizing the morphology. Perovskites with mixed cations and halides have been more thermally and structurally stable than pure perovskites. One of the most efficient mixtures of cations and halides is triple cation perovskite CsFAMA, which is currently a promising material for

*Corresponding Author: Manar Saleh Alshatwi, Manarsalleh@gmail.com

future optoelectronic applications. Triple cation perovskite CsFAMA has high thermal stability, reduced recombination, and increased electron lifetime due to a low percentage of Cs, resulting in highly homogenous grains of pure perovskite [6]. Moreover, partially displacing inorganic cations, such as Cs^+ and Rb^+, enhances the thermal and mechanical stability of MA/FA halide perovskites. The stability and distortion of a crystal perovskite material's structure can be estimated by finely controlling the t value. By combining organic cations, such as FA or MA, with relatively slight amounts of Rb^+ or Cs^+ cations, a more beneficial Goldschmidt tolerance factor (t) value can be achieved, allowing the photoactive perovskite phase to be stabilized over a wider temperature range, leading to stable devices [7–9]. The development of stable and efficient electron transport materials (ETM) and hole transport materials (HTM) is of prime consideration. Heterojunction photodetectors can be produced using metal oxides, such as ZnO, TiO_2, WO_2, or SnO_2, as an electron transport layer (ETL). Triple cation perovskite CsFAMA photodetectors have high-density pits defects, which have the effect of reducing carrier transport, reducing responsivity, and decreasing the crystal quality of the film. Fe doped NiO improves the properties of carrier transport due to better transmission, wide band gap, high stability, and suitable alignment of energy level with perovskite for good hole collection [10–13]. In the presented work, we demonstrate the enhancement of the responsivity and detectivity of triple cation perovskite photodetectors by optimizing the morphology. This is done through fabricated heterojunction photodetectors by doping with iron-doped nickel oxide quantum dots (QDs).

2. Experiment

2.1 Preparation of $CsMAFAPbI_2Br_2$

Using a glove box, a mix of 21.84 mg of CsI, 18.84 mg of $MABr_2$, 62.04 mg of $PbBr_2$, 247.2 mg of FAI, and 722.4 mg of PbI_2 was dissolved in anhydrous dimethylformamide/dimethylsulphoxide at a 4:1 volume ratio to create a perovskite precursor solution.

2.2 Preparation of spiro-OMeTAD

HTM solutions were prepared in a N_2 glove box by mixing 17.8 µL of lithium (Li) with 28.8 µL of 4-Tert-Butylpylpyridine and then adding 85.6 mg of spiro-OMeTAD with 1000 µL chlorobenzene.

2.3. Preparation of Iron-doped NiO Quantum Dots

Chemical precipitation is the normal method for preparing QDs. In this study, iron-doped NiO QDs synthesized by dissolving 13.81 g nickel nitrate, Ni $(NO_3)_2$, and 1.01 g iron nitrate, Fe $(NO_3)_3$, in 50 mL of distilled water at a moderate stirring speed at 50°C. After 15 minutes, 1 g of citric acid was added to the mixture, and it was stirred for another 15 minutes. Then, a drop of ammonium hydrate was added, and precipitate began to form until the combination became a thick gel. The temperature was eventually elevated to obtain a totally dry sample, producing FeO_3-doped NiO QDs. Finally, the samples were heated for 2 hours at 350°C in a furnace.

2.4. Thin films and photodetector fabrication

Ultrasonication in deionized water, acetone, and ethanol was used to clean the fluorine-doped tin oxide (FTO) / glass substrate, followed by UV drying for 15 minutes. Then, a nanoscale blocking layer (TiO_2) was sprayed on the cleaned FTO at 450°C (1:9 volume ratio) using a commercial titanium diisopropoxide bis (acetylacetonate) solution (75% in 2-propanol, Sigma-Aldrich) diluted in anhydrous ethanol. A 150 nm TiO_2 mesoporous layer was spin coated at 4000 rpm for 20 seconds then annealed at 450°C for 30 minutes in dry air (diluted paste, 1:6 wt. ratio of Dyesol 30NRD to ethanol). Manufacturing of the device was done inside a dry air box under carefully controlled settings. In a two-stage process, Spin coating of perovskite solution at 1000 and 6000 rpm was performed. Iron-doped NiO QDs were doped within the perovskite at the same stage. On the spinning substrate, 200 µL of chlorobenzene was dropped in the last 10 seconds of spinning. All samples were annealed at 100°C for 45 minutes. Finally, spiro-OMeTAD was coated as hole transport layer in steps at 4000 rpm for 20 seconds, as shown in figure 2.

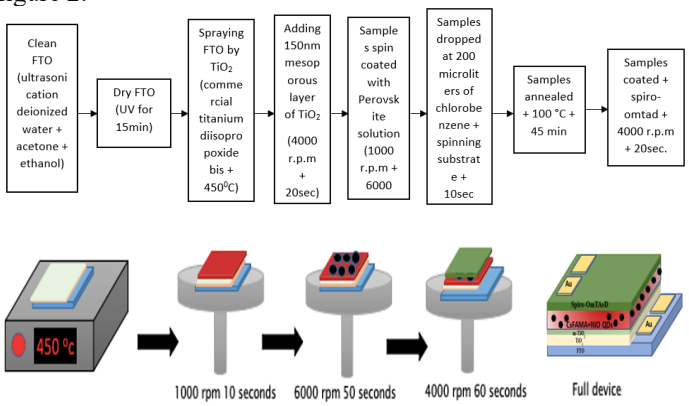


Figure 2. (a) *Steps of photodetector fabrication.*

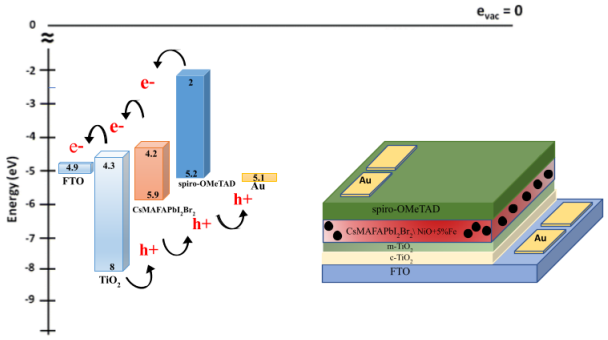
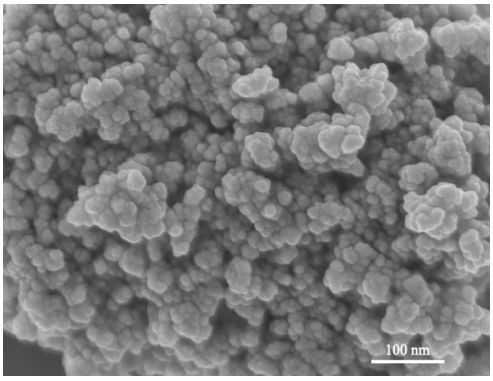


Figure 2. (b) band gap alignment

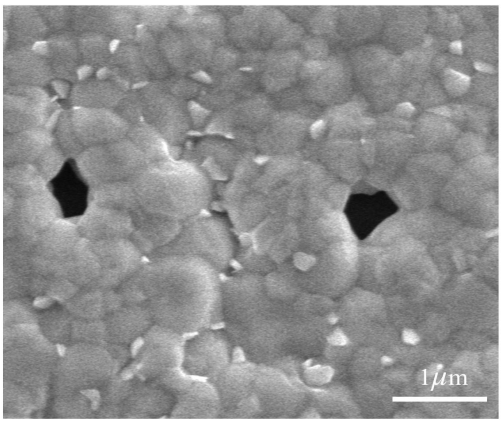
3. Results and Discussion

SEM imaging was used for studying the manufactured NiO+5%Fe QDs–doped triple cation $CsMAFAPbI_2Br_2$ film. Figure 3(a) shows a NiO+5% Fe QDs image where the QDs are nearly spherical with 4.3, 5.6, 9.7, and 11.6 nm sizes. A $CsMAFAPbI_2Br_2$ perovskite film exhibiting many voids and defects between grains is shown in figure 3(b). Doping NiO+5% Fe in the perovskite surface improved morphology and decreased defects. A cross section of the device and image of NiO+5% Fe

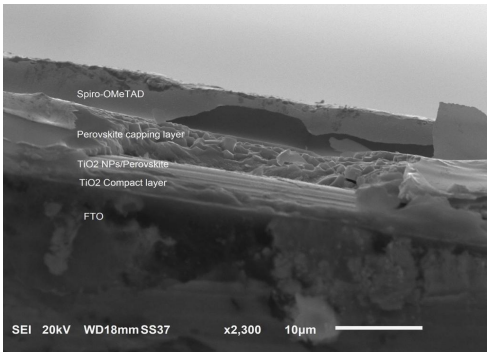
QDs is shown in figure 3(c). Figure 3(d),(e) shows high-resolution TEM (HR-TEM) images of NiO+5% Fe QDs in different scales.



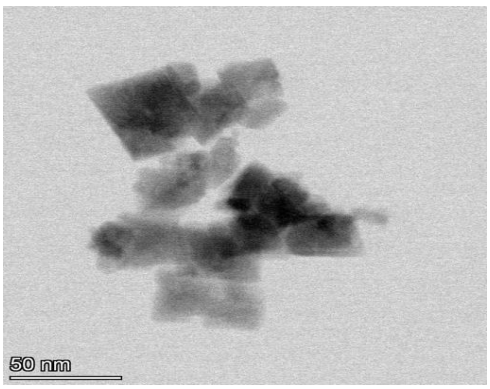
(a)



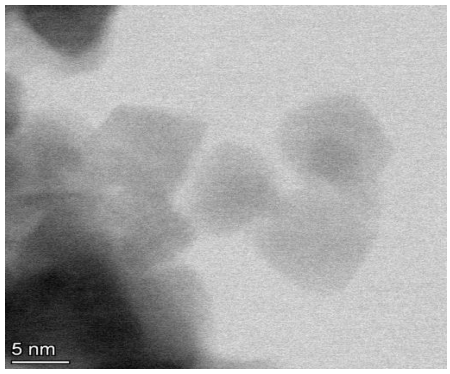
(b)



(c)



(d)



(e)

Figure 3. (a) SEM image of NiO+5% Fe QDs calcined at 350 °C. (b) Illustration of morphology of $CsFAMAPbI_2Br_2$ perovskite. (c) Cross section of full device. (d) HR-TEM images of NiO+5% Fe QDs.

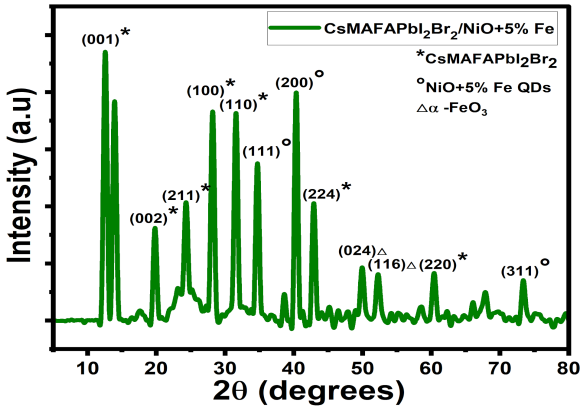
NiO+5% Fe QDs with 5 nm particle size, with XRD that doped NiO+5% Fe QDs by $CsMAFAPbI_2Br_2$ and SEM homogeneously distribution, as shown in figure 4(a). The diffraction peaks were observed at 37.2°, 43.1°, 62.90°, 74.9°, and 79.2° for the NiO+5% Fe QDs that were well-defined for (111), (200), (220), (311), and (222) cubic NiO_3+FeO_3 crystals [14,15].The $CsMAFAPbI_2Br_2$ /NiO+5% Fe film confirmed the formation of α-$Fe2O_3$ phase via the observed peaks (024) and (116) at 50.4° and 52.2°, respectively, which is consistent with previous studies [16,17]. The diffraction peaks of $CsMAFAPbI_2Br_2$ were observed at 12.5°, 19.8°, 24.3°, 28.2°, 31.6°, 34.7°, 40.3°, 42.9°, 49.8°, 60.3°, and 73.2° for (011), (002), (211), (100), (111), (224), (024), (166), (220), and (311), respectively, confirming the diffraction pattern of the cubic perovskite. The NiO_3+FeO_3 QDs size derived from the diffraction peaks using the Scherrer formula was 5.6 nm. The Scherrer formula can be written as follows:

$$D = \frac{k\lambda}{\beta cos\theta} \quad (1)$$

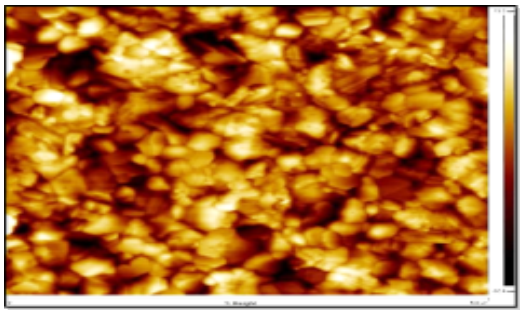
where D is the crystallite size, k = 0.89 is the Scherrer constant depend on the spherical particles and crystals index (hkl), λ= 0.145 is the source wavelength, and β is the full width at half the maximum (FWHM) of the peak in radians.

AFM topography maps of 5 μm $CsMAFAPbI_2Br_2$ perovskite films are shown in figure 4(b). The maps identified a slightly uniform grain distribution with dimensions between 200 and 400 nm and created a unified background film with pit defects, and comparable dimensions, surpass in height and exhibit a morphological contrast on the background layer's surface. The $CsMAFAPbI_2Br_2$ perovskite film's thickness was assessed as 372 nm ± 31, which is similar to the grains' lateral dimensions.

Studies of UV absorption show a blue shift. Figure 5(a) shows the absorption spectra and a decrease in bandgap as a result of an increase in QD size. The UV absorption spectrum's peaks were between 300 and 400 nm, which is consistent with previous research [18,19]. $CsMAFAPbI_2Br_2$ absorption peaks found between 700 and 800 nm in the visible region is consistent with previous studies [20]. Figure 5(b) shows the absorption of $CsMAFAPbI_2Br_2$ doped with NiO QDs in the broadband visible, and the absorption spectrum from UV to visible is shown in figure 5(c).

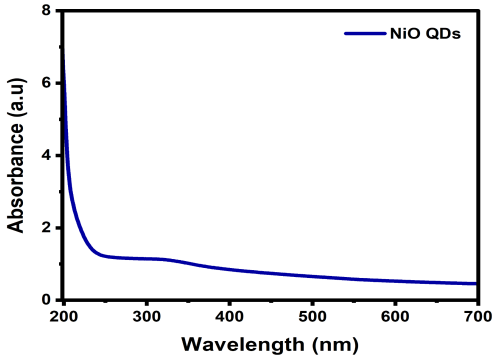


(a)

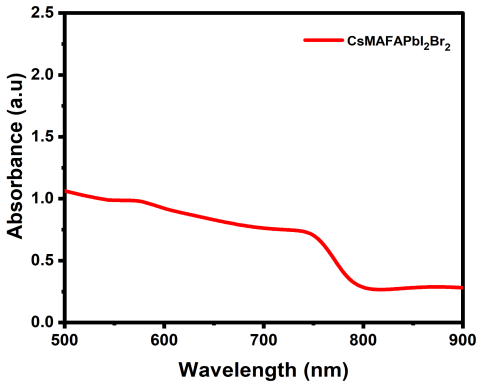


(b)

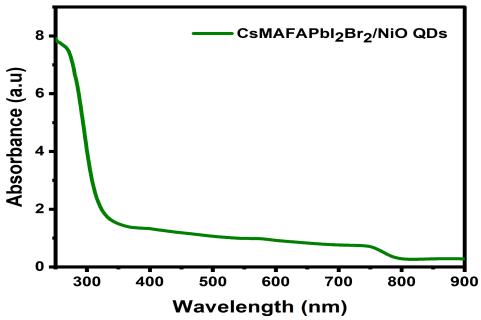
Figure 4.(a)CsMAFAPbI$_2$Br$_2$ /NiO+5% Fe XRD patterns. (b) CsMAFAPbI$_2$Br$_2$ AFM image



(a)



(b)

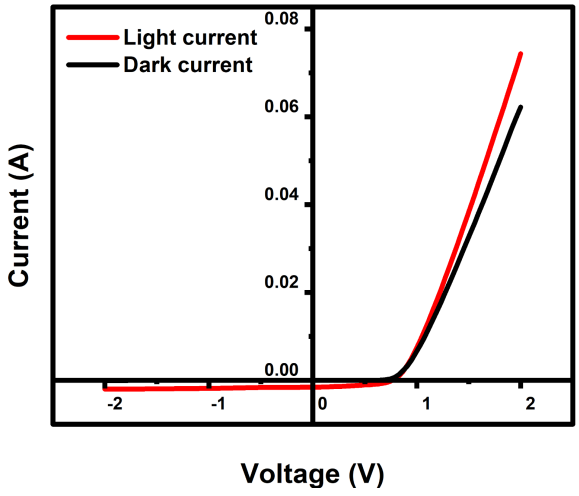


(c)

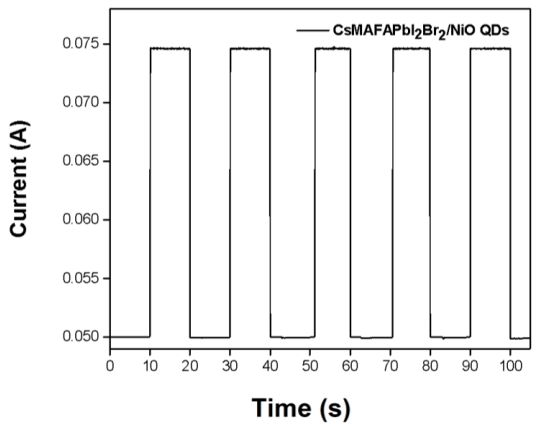
Figure 5. UV–visible absorption spectra of (a) NiO QDs, (b)CsMAFAPbI$_2$Br$_2$, and (c) NiO QDs/CsMAFAPbI$_2$Br$_2$.

Current–voltage (I–V) measurements of NiO+5% Fe QDs–doped CsMAFAPbI$_2$Br$_2$ were measured under dark conditions and illumination with 1 mW/m^2. The performance of the photodetector NiO+5% Fe QDs device is shown in figure 6(a). Equation (2) were used for calculating a photo-responsivity (R) where ΔI denotes the difference between the photo-current and the dark current, P denotes incident power density, and S denotes the effective area illuminated by the white lamp source. Furthermore, we calculated the detectivity (D^*) by equation (3) where q denotes the electron charge ($1.6 \times 10^{-19} C$) and J_d denotes the dark current density. R and D* were calculated depend on that noise in the current noise is mainly dominated by shot noise. Compared with a previous study [21] appeared responsivity 0.99 A/W where the detectivity of 8.9×10^{12} Jones, related to the improved morphology with a decrease in grain size and vacancies, which is reflected in the high performance level of our photodetector. By using on/off ratio as a measurement of photodetector repeatability, we obtained a high capacity that allows the photodetector to be operated repeatedly and with the same efficiency without any damage to the photodetector, as shown in figure 6(b).

$$R = \frac{\Delta I}{PS} \qquad (2) \qquad D^* = \frac{R}{\sqrt{2qJ_d}} \qquad (3)$$



(a)



(b)

Figure 6. (a) I–V characteristics of $CsMAFAPbI_2Br_2$ photodetector heterojunction under illumination of 1 mW/m^2. (b) On/off ratio of $CsMAFAPbI_2Br_2$/NiO+5% Fe QDs photodetector heterojunction.

4. Conclusion

In the presented work, our goals were achieved with enhanced responsivity and detectivity of the photodetector while trying to keep a low dark current. Doping NiO QDs in the perovskite layer significantly reduced perovskite pit defects. The responsivity of the photodetector was enhanced by 47% (0.99 A) with doped NiO+5% Fe QDs and a detectivity of 8.9×10^{12} Jones was achieved, which is the highest value that has been reported. Moreover, the photodetector performance was enhanced by expanding the photodetection range from the UV to the visible spectrum.

Acknowledgement

The authors express their gratitude to Princess Nourah bint Abdulrahman University Researchers Supporting Project (Grant No. PNURSP2022R11), Princess Nourah bint Abdulrahman University, Riyadh, Saudi Arabia. This work was supported by King Abdulaziz City for Science and Technology (KACST).

References

[1] S. Xu, Y. Qin, C. Xu, Y. Wei, R. Yang, Z.L. Wang, "Self-powered nanowire devices," Nature Nanotechnology, **5**(5), 2010, doi:10.1038/nnano.2010.46.

[2] B.O. Jung, Y.H. Kwon, D.J. Seo, D.S. Lee, H.K. Cho, "Ultraviolet light emitting diode based on p-NiO/n-ZnO nanowire heterojunction," Journal of Crystal Growth, **370**, 2013, doi:10.1016/j.jcrysgro.2012.10.037.

[3] S.S. Shinde, K.Y. Rajpure, "Fabrication and performance of N-doped ZnO UV photoconductive detector," Journal of Alloys and Compounds, **522**, 2012, doi:10.1016/j.jallcom.2012.01.118.

[4] R. Bao, C. Wang, L. Dong, R. Yu, K. Zhao, Z.L. Wang, C. Pan, "Flexible and controllable piezo-phototronic pressure mapping sensor matrix by ZnO NW/p-polymer LED array," Advanced Functional Materials, **25**(19), 2015, doi:10.1002/adfm.201500801.

[5] M. Esro, G. Vourlias, C. Somerton, W.I. Milne, G. Adamopoulos, "High-mobility ZnO thin film transistors based on solution-processed hafnium oxide gate dielectrics," Advanced Functional Materials, **25**(1), 2015, doi:10.1002/adfm.201402684.

[6] S. Ašmontas, A. Čerškus, J. Gradauskas, A. Grigucevičienė, K. Leinartas, A. Lučun, K. Petrauskas, A. Selskis, A. Sužiedėlis, E. Širmulis, R. Juškėnas, "Cesium-containing triple cation perovskite solar cells," Coatings, **11**(3), 2021, doi:10.3390/coatings11030279.

[7] R.E. Beal, D.J. Slotcavage, T. Leijtens, A.R. Bowring, R.A. Belisle, W.H. Nguyen, G.F. Burkhard, E.T. Hoke, M.D. McGehee, "Cesium Lead Halide Perovskites with Improved Stability for Tandem Solar Cells," Journal of Physical Chemistry Letters, **7**(5), 2016, doi:10.1021/acs.jpclett.6b00002.

[8] Y. Hu, E.M. Hutter, P. Rieder, I. Grill, J. Hanisch, M.F. Aygüler, A.G. Hufnagel, M. Handloser, T. Bein, A. Hartschuh, K. Tvingstedt, V. Dyakonov, A. Baumann, T.J. Savenije, M.L. Petrus, P. Docampo, "Understanding the Role of Cesium and Rubidium Additives in Perovskite Solar Cells: Trap States, Charge Transport, and Recombination," Advanced Energy Materials, **8**(16), 2018, doi:10.1002/aenm.201703057.

[9] Y. Hu, M.F. Aygüler, M.L. Petrus, T. Bein, P. Docampo, "Impact of Rubidium and Cesium Cations on the Moisture Stability of Multiple-Cation Mixed-Halide Perovskites," ACS Energy Letters, **2**(10), 2017, doi:10.1021/acsenergylett.7b00731.

[10] X. Yin, M. Que, Y. Xing, W. Que, "High efficiency hysteresis-less inverted planar heterojunction perovskite solar cells with a solution-derived NiOx hole contact layer," Journal of Materials Chemistry A, **3**(48), 2015, doi:10.1039/c5ta08193a.

[11] J. Cui, F. Meng, H. Zhang, K. Cao, H. Yuan, Y. Cheng, F. Huang, M. Wang, "CH3NH3PbI3-based planar solar cells with magnetron-sputtered nickel oxide," ACS Applied Materials and Interfaces, **6**(24), 2014, doi:10.1021/am507108u.

[12] I.J. Park, G. Kang, M.A. Park, J.S. Kim, S.W. Seo, D.H. Kim, K. Zhu, T. Park, J.Y. Kim, "Highly Efficient and Uniform 1 cm2 Perovskite Solar Cells with an Electrochemically Deposited NiOx Hole-Extraction Layer," ChemSusChem, **10**(12), 2017, doi:10.1002/cssc.201700612.

[13] X. Yin, J. Liu, J. Ma, C. Zhang, P. Chen, M. Que, Y. Yang, W. Que, C. Niu, J. Shao, "Solvothermal derived crystalline NiOx nanoparticles for high performance perovskite solar cells," Journal of Power Sources, **329**, 2016, doi:10.1016/j.jpowsour.2016.08.102.

[14] P.M. Ponnusamy, S. Agilan, N. Muthukumarasamy, T.S. Senthil, G. Rajesh, M.R. Venkatraman, D. Velauthapillai, "Structural, optical and magnetic properties of undoped NiO and Fe-doped NiO nanoparticles synthesized by wet-chemical process," Materials Characterization, **114**, 2016, doi:10.1016/j.matchar.2016.02.020.

[15] K. Karthik, G.K. Selvan, M. Kanagaraj, S. Arumugam, N.V. Jaya, "Particle size effect on the magnetic properties of NiO nanoparticles prepared by a precipitation method," Journal of Alloys and Compounds, **509**(1), 2011, doi:10.1016/j.jallcom.2010.09.033.

[16] M. Qayoom, K.A. Shah, A.H. Pandit, A. Firdous, G.N. Dar, "Dielectric and electrical studies on iron oxide (α-Fe2O3) nanoparticles synthesized by modified solution combustion reaction for microwave applications," Journal of Electroceramics, **45**(1), 2020, doi:10.1007/s10832-020-00219-2.

[17] S. Suresh, S. Karthikeyan, K. Jayamoorthy, "Effect of bulk and nano-Fe2O3 particles on peanut plant leaves studied by Fourier transform infrared spectral studies Effect of Fe2O3 particles on peanut plant leaves >," Journal of Advanced Research, **7**(5), 739–747, 2016, doi:10.1016/j.jare.2015.10.002.

[18] P.A. Sheena, K.P. Priyanka, N. Aloysius Sabu, S. Ganesh, T. Varghese, "Effect of electron beam irradiation on the structure and optical properties of nickel oxide nanocubes," Bulletin of Materials Science, **38**(4), 2015, doi:10.1007/s12034-015-0953-5.

[19] T. Singh, T. Miyasaka, "Stabilizing the Efficiency Beyond 20% with a Mixed Cation Perovskite Solar Cell Fabricated in Ambient Air under Controlled Humidity," Advanced Energy Materials, **8**(3), 2018, doi:10.1002/aenm.201700677.

[20] S. Tong, C. Gong, C. Zhang, G. Liu, D. Zhang, C. Zhou, J. Sun, S. Xiao, J. He, Y. Gao, J. Yang, "Fully-printed, flexible cesium-doped triple cation perovskite photodetector," Applied Materials Today, **15**, 2019, doi:10.1016/j.apmt.2019.03.001.

[21] G.R. Adams, V.O. Eze, M.A.S. Shohag, R. Simpson, H. Parker, O.I. Okoli, "Fabrication of rapid response self-powered photodetector using solution-processed triple cation lead-halide perovskite," Engineering Research Express, **2**(1), 2020, doi:10.1088/2631-8695/ab7b38.

ASTES

Advances in Science, Technology and Engineering Systems Journal Vol. 7, No. 3, 166-173 (2022)

www.astesj.com

ASTESJ

ISSN: 2415-6698

A Supervised Building Detection Based on Shadow using Segmentation and Texture in High-Resolution Images

Ayoub Benchabana[*, 1], Mohamed-Khireddine Kholladi[2], Ramla Bensaci[3], Belal Khaldi[3]

[1]*Laboratory of Operator Theory and EDP: Foundations and Application, University of El Oued, PB. 789., El Oued 39000, Algeria*

[2]*Department of Computer Science, University of El Oued, El Oued, Algeria;c MISC Laboratory of Constantine 2, University of Constantine 2, Algeria*

[3]*Lab of Artificial Intelligence and Data Science, Kasdi Merbah Ouargla University, PB. 511., Ouargla 30000, Algeria*

ARTICLE INFO

Article history:
Received: 12 March, 2022
Accepted: 19 June, 2022
Online: 28 June, 2022

Keywords:
Building detection
Superpixel
Texture
Colors features
Machine Learning
Shadows
Regional growth

ABSTRACT

Building detection in aerial or satellite imagery is one of the most challenging tasks due to the variety of shapes, sizes, colors, and textures of man-made objects. To this end, in this paper, we propose a novel approach to extracting buildings in high-resolution images based on prior knowledge of the shadow position. Firstly, the image is split into superpixel patches; the colors and texture features are extracted for those patches. Then using the machine learning method (SVM), four classes are made: buildings, roads, trees, and shadows. According to the prior knowledge of shadows position, a seed point initial has been defined along with an adaptive regional growth method to determine the approximate building location. Finally, applying a contouring process included an open morphological operation to extract the final shape of buildings. The performance is tested on aerial images from New Zealand area. The proposed approach demonstrated higher detection rate precision than other related works, exceeding 97% despite the complexity of scenes.

1. Introduction

Detecting and Identifying building locations is vital for varieties of applications such as mapping, military situations (active engagement of forces, counter-terrorism and peacekeeping measures), natural disaster management (flooding, earthquakes, and landslides), environmental preparation, and urban planning [1–6]. It is feasible to distinguish buildings from the images; however, this can be a time-consuming or difficult operation. Therefore, Automatic building extraction from aerial or satellite images is a highly needed and challenging problem due to its complexity. Moreover, with the advanced technology of capturing very high spatial resolution imagery and the increasing need for map revision without the high cost and time-consuming as a consequence of the rapidly growing urbanization. Automatic building detection becomes possible as the ground resolution size of the pixels is much smaller than the average size of objects in those images. During the past decade, many studies have been carried out on building extraction [7–9]. However, it is still difficult to detect buildings in urban areas because of their variety of shapes, sizes, colors, textures, and the similarity between building and non-building objects.

In the field of building detection, most methods have been based on artificial features; like in work [10], due to low-quality RGB geophotos and to reduce the problem of characteristics extracted from those images, they used the Haar feature method to be able to apply machine learning techniques on it. As for work [11], it integrates a set of algorithms inspired by the human visual system with a combination of classical and modern approaches for extracting image descriptors. Then, the feature descriptors are processed with machine learning to identify buildings.

Moreover, Deep learning-based approaches have been proposed recently. Including, In [12] focused on three different ways to use convolutional neural networks for remote sensing imagery. The authors suggested [13]a set of convolutional neural networks for township building identification that can be applied to a pixel-level classification framework. In [14] proposed a general framework for convolutional neural network-based

[*]Corresponding Author: Ayoub Benchabana, ayoub.benchabana@gmail.com

classification. In [15], a building extraction framework is proposed based on a convolutional neural network (CNN), edge detection algorithm, and building structure. The masked R-CNN Fusion Sobe framework was used to extract the building from high-resolution remote sensing images. But the results showed that it works poorly to extract edges and preserve the building instances' integrity.

However, These studies are not without their limitations, especially since CNN cannot successfully learn the features of the hierarchical contextual image due to the lack of data sets, and increasing the number of layers in the deep model leads to more significant training mistakes. CNNs' prediction abilities with less computation come at the cost of reduced output accuracy.

This paper focuses on building detection from 2D aerial (or satellite) imagery. Therefore a novel approach addresses two main issues: (a) the problem with the particular colors for some buildings with a similar color to other objects in urban images, and (b) the difference in color shades on building's rooftops. The paper is organized in the following way: The related work section categorizes and presents works that tackle the issue of building detection. Section 3, details the steps that have been applied to classify the image into four classes. Then in section 4, we describe how to accurate buildings positions. And section 5, shows the results provided by our experiments and a comparison with some other works. Finally, conclusions and perspectives for future work are in section 6.

2. Related Work

As we said before, building detection is one of the most difficult challenges to solve since they have so many different properties. Numerous building detection techniques have been presented throughout the years, having their efficacy measured in various ways. In this section, we review works that attempt to solve those challenges. Most of those researches can be classified according to whether they are supervised or automatic, extract geometric features, or are area-based [16]. Another classification is based on the use of the height data, the simple 2D imagery, or a hybrid [17].

In [8], the authors proposed an automatic approach using level set segmentation constrained by priors known shape models of buildings. In [18], the authors presented a supervised model using the active contour method combined with local texture and edge information by initial seed points on one of the buildings. The authors [16] have developed an automatic tertiary classifier to identify vegetation, buildings, and non-buildings objects using Nadir Aerial Image with one condition that the building has a convex rooftop. They reduce the colors from 255 for each RGB channel to 17. Then, using segmentation and thresholding on the green color channel to identify the vegetation regions and on the difference between the blue and green color channels to identify shadows. Finally, buildings and non-building are detected by measuring the solidity of their regions using the entropy filtering and watershed segmentation. In [19], the authors present an automatic technique using LIDAR data and multispectral imagery. Buildings and trees are separated from other low objects using thresholding for height. After that, they eliminate the trees with the normalized difference vegetation index method from an orthorectified multispectral image. The authors proposed [20] an automatic approach using a digital surface model and multispectral orthophoto. Initially, they created a building mask from the normalized digital surface model that included only areas where the possible locations for buildings. Then, the vegetation was separated from the building mask using a modified vegetation index based on the use of the near-infrared orthophoto and the correction of the vegetation index using the shadow index and the texture analysis. Finally, using Radon transform, they extract the building position. In [21], the authors implement an object-based classification for urban areas using spot height vector data. After the segmentation of the image, they had classifier the obtained result into five class vegetation, shadows, parking lots, roads, and buildings based on the analysis of the combined spectral, textural, morphological, contextual, and class-related features to assign a class membership degree to each segment (object) based on membership functions or the thresholds. In [22], the authors implemented an automatic method based on the similarity between building roofs using a previously defined reference set to generate a grayscale image with an enhanced potential for building location. Then, they assign pixels to possible buildings or nonbuilding locations using the hit-or-miss transform morphology. Finally, after defining the shadow areas, they verified the final location of the buildings. In [23], the authors developed a supervised approach based on shadow position using segmentation and classification of color features. First, they split the image into superpixel patches with the segmentation algorithm. After that, using Linear discriminate analysis (LDA) color features and support vector machines (SVM) with a previous set of chosen patches for three classes: buildings, non-buildings, and shadow. Finally, from the prior knowledge of the shadows' position, they define a seed point location, and with the regional growth method, they determine the positions of the buildings. In [24], the authors proposed an automatic approach based on the rectangle form of the buildings. They enhance the edge contrast using a developed bilateral filter. Then, they apply a line segment detector to extract lines. Finally, a perceptual grouping approach groups previously detected lines into candidate rectangular buildings. In 2016 [25], the authors provided an ontology-based system for slum identification based on the built environment's morphology. In this technique, a segmentation is followed by hierarchical classification utilizing object-oriented image analysis. For each object, spectral values, form, texture, size, and contextual connections are all computed based on the purpose of the classification. In [26] adopt a new object-based filter consisting at first of splitting the image into homogeneous objects using multi-scale segmentation and at the same time extracting their features vector. Considering that each splitter object is in the center of his surrounding adjacent objects and a part of a fully real existing object in the image. Then, there are two possibilities of his location, either in the interior or in the real object. Hence, it has similar features to its surrounding adjacent objects or in the boundary with no similarity between them. Therefore, topology and feature constraints are proposed to select the considered adjacent objects. Finally, the feature of the central object is smoothed by calculating the average of the selected object's feature. In [27], detecting buildings by determining them using a one-class SVM, They proceed with the texture segmentation technique using a conditional threshold value to extract buildings of different colors and shapes. Buildings are identified from the rest of the roads and vegetation regarding the angle of shadows. In [28], the Building Detection with Shadow

Verification (BDSV) approach was introduced, integrating multiple features such as color, shape, and shadow to detect buildings. Because some roofs can be extracted with color features only, such as for buildings with sloped roof tiles, while Non-tile flat roofs depend on the shape features. The shadow properties were also incorporated, (Candidate buildings with close shadows will be considered as actual buildings).

Nonetheless, many challenges remain to be overcome in building extraction. To begin with, some of the buildings have a particular color (green, red…), so the approaches based on color classification can't separate buildings from trees or lawns [16,17,19,20,23]. Rather than assume that buildings have a standard form like rectangles or use a predefined shape database to bring the results closer to a particular format [8,24], we can't cover all those possibilities of their forms. Furthermore, others use texture to solve the color problem. Yet, they only use it to calculate the entropy or the homogeneity of a single-pixel combined with the high data of objects or to detect the whole area of buildings without separating one from another [21,25]. Therefore, those methods can't give us the results we request without the high data or in a complex scene. That being the case, can we benefit from segmentation to a superpixel size unity by applying the texture methods on the entire superpixel, assuming that they are small-sized objects rather than on a single pixel?.

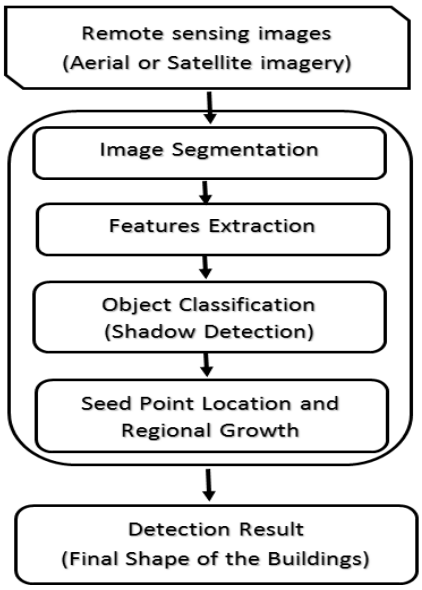


Figure 1. Diagram flow of the proposed algorithm.

3. Classification of image data

Most urban images have six kinds of land covers: trees, grass, shadows, roads, parking lots, and buildings. Trees and grass are usually green, but this depends on the type and the season the image was taken (they may be in red or yellow). Roads and parking lots have, in general, a gray color; unlike the shadows, areas are darker or completely black. And finally, buildings have different color rooftops based on the image's location (the diversity of culture, climatic nature…). In Figure 2, various structures, building patterns, lighting conditions, landscape characteristics, and complex buildings are located within the study area. Visual inspection can quickly discover complicated patterns in buildings, but machine learning cannot.

Consequently, buildings may have a color similar to trees or roads and parking lots. Therefore, we relied on two things to solve this problem: (a) the shadow factor to separate buildings from roads and parking lots, and (b) for trees, the texture features can do the trick. As a result, according to color and texture, four classes pop up: the first one contains buildings, the second is trees and grass, the third are roads and parking lots, and finally shadows. So, to obtain those classes, three steps have been taken into account, We'll walk through them in detail:

Figure 2: Aerial images of buildings in the study area New Zealand from various perspectives

3.1. Superpixel Segmentation Using SLIC

Superpixel segmentation divides an image into a group of connected pixels with similar colors. Instead of working with pixels in a big-sized image, we can reduce it into superpixel patches without losing too much information. The Simple Linear Iterative Clustering (SLIC) algorithm for superpixel segmentation proposed in [29] is a k-means-based local clustering of pixels in the 5-D space [*l, a, b, x, y*], where (*l; a; b*) is the CIELAB color space and (*x; y*) is pixel coordinates. SLIC adapts the k-means clustering approach to efficiently generate superpixels introducing a new distance measure D_s as described in Eq. [1].

$$D_s = d_{lab} + m/S\, d_{xy} \tag{1}$$

$$d_{lab} = \sqrt{(l_k - l_i)^2 + (a_k - a_i)^2 + (b_k - b_i)^2} \tag{2}$$

$$d_{xy} = \sqrt{(x_k - x_i)^2 + (y_k - y_i)^2} \tag{3}$$

where *k* and *i* are respectively the indices of the superpixels center and their surrounding pixels, m is a variable that allows controlling the compactness of superpixels, and *S* is the grid interval between them.

In consideration of the foregoing, this step aims to have homogenous superpixels as much as possible with a sufficient size that allows us to extract texture features disregarding the shape of each one of the superpixels. Therefore, a high value of m makes the spatial distances outweigh the color factors giving more compacted set-sized superpixels that lead to disrespecting the boundaries of the objects in the image. The other way around for a lower value, this will produce small sizes superpixels. So, To get

effective results, we chose a low-value number. Then, merge each superpixel with a smaller size than a given thresholding number of pixels to the nearest similar neighbor. It will turn them into one larger superpixel without losing the adherence to object boundaries.

3.2. Texture Features Extraction Using (ICICM)

Color co-occurrence matrix (CCM) is one of the most efficient yet straightforward texture descriptors. It consists of extracting statistical measurements about the co-existence of different colors from the image. Integrative Color Intensity Co-occurrence Matrix (ICICM) has been introduced in [30] as an extension of CCM to simulate the human perception of textures. They argued that each pixel might be regarded as color or gray-level depending on its intensity level. Thus, two measures, namely W_{col} and W_{gray}, which determine the extents of color and gray, have been extracted from each pixel within the image. After that, these measures have been used to extract four co-occurrence matrices that represent the co-existence of W_{col}/W_{col}, $W_{col} = W_{gray}$, $W_{gray}=W_{col}$, and $W_{gray} = W_{gray}$. Finally, a set of third-order statistical moments have been drawn and used as image descriptors. An improvement of ICICM has been proposed in [31], in which a smooth approach of color/gray-level space quantization has been adopted. Ultimately, ICICM has been used to extract texture features for each final form superpixel.

3.3. Identifying Classes Using SVM

A Support Vector Machine (SVM) is a supervised discriminative classifier formally defined by a separating hyperplane. Given a set of training examples with which class they belong, the SVM training algorithms create a model that can assign the new data to one of those classes. Therefore, the SVM classifier can be trained with the combination vectors between the LAB color features and the texture features of the training superpixel samples to obtain our final class results. The samples are taken from each of the four classes, a simple linear kernel type of SVM was applied, and the results are shown in figure 3(d).

4. Accurate Building Position

From the previous results in figure 3(d), we can see that buildings and trees are entirely separated, unlike some similarities with the roads and the parking lots. As we mentioned before, the prior knowledge of the direction of the shadows lets us distinguish between elevated objects and the ones at ground level. Still, the problem is how to detect the exact shape of the buildings.

Usually, the rooftop of the building has the same color; therefore, a seed point location with the regional growth method may do the trick. However, in many areas around the world and depending on the designs of the buildings, it can cause to show darker sides than the others on the rooftop of the buildings, so we adept our new implementation of the regional growth to resolve this problem.

4.1. Seed Point Location and Regional Growth

At first, an initial superpixels seed points location is defined based on three conditions:

- the superpixels in the shadows class and have a neighbor from the trees class are eliminated;
- the superpixels that have been considered as seed points must be in buildings class that is a neighbor to the one in shadows class after the elimination according to their respective direction (in this case, the up and right sides);
- The superpixels with more neighbors from road class than building class are eliminated.

Next, to make sure that the seed points are located all over the region of the building and resolve the darker side problem, we applied the regional growth method using the [*a, b, ep, h, c, en*] vector with the earlier initial superpixels seed points where a and b are the green–red and blue-yellow color components from the Lab color space; *ep, h, c*, and en are respectively the entropy, homogeneity, correlation, and energy from the texture features. Assuming that V_i is the vector of the initial superpixel seed points S_i and V_s is for the other neighbors' superpixel S_s, so the following logical conditions have been applied, and the results will be our final superpixels seed points

$$V_s - V_i \leq T_s \tag{4}$$

$$S_s \in C_1 \tag{5}$$

In the end, we used another regional growth to get the whole shape of buildings, and this time used the lightness value (L) from the Lab color space for all the pixels P_s of the image starting as an initial with the centers C_i of each superpixel from the final superpixels seed points result take into consideration one condition:

$$L_i - L_s \leq T_g \tag{6}$$

4.2. Accurate The Final Shape of The Buildings

We can see from the last step that neither the boundaries nor the inside of the building are precisely shaped and filled in figure 3(f); there are still some noises and gaps in it. Therefore, some complementary steps are in need. Firstly, an open morphological operation has been applied to remove the noise and fill the gaps. Then, the buildings' boundaries were extracted using a simple contour method. The final results are shown in figure 3.

5. Experimental Results

This section includes three subsections: first, the suitable selection of algorithm parameters has been defined. The second subsection presents the classification results with and without texture. And finally, we analyze the results and compare them with other methods. The entire experiment was applied to a dataset from Land Information New Zealand urban aerial image for Auckland with a size of 6400x9600 for each image and 7.5cm ground resolution. The experiment is implemented in Intel 3.2 GHz CPU with 16G memory

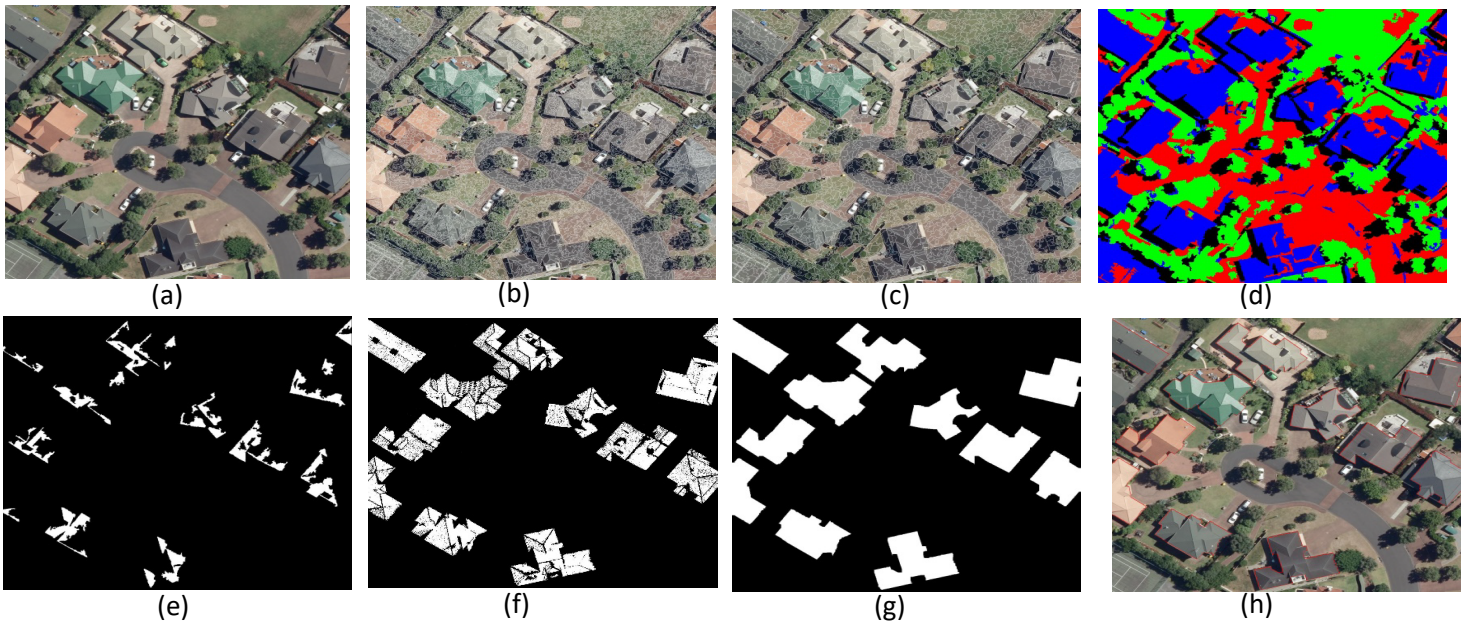


Figure 3. Experimental result of different stages of the algorithm. (a) The original image; (b) Superpixels ; (c) Aggregation Superpixels, (d) classification results, (e) Seed Point Location, (f) after growth (e) Final Shape of The Buildings and (h) The building extraction.

5.1. Parameters Selection

Three values must be determined (n, T_s, T_g) where n is the initialized number of the superpixels in the SLIC algorithm, T_s and T_g are the thresholding of the regional growth in Eqs (4) and (6), respectively. To initial a suitable number n of the superpixels is a two-sided problem. A smaller n leads to fewer calculations and a shorter time in execution. On the other hand, the bigger it is, the better homogeneity we get. That being said, what it depends on is the size and resolution of the image that has been studied. In our test, we use different values of n on a cut from Land Information New Zealand urban aerial image for Auckland with the size of 1286x1249 and 7.5cm ground resolution, and the results reveal that the best value of n is 2500. Therefore, we can define an equation for any image with the exact ground resolution based on that result.

$$n = \frac{\text{size image}}{\text{size test /n test}} \tag{7}$$

In the same way, for the two thresholding, T_s and T_g. We can define their values using a sample test and apply the exact values to the rest of the other images.

5.2. Classification Results

For acceptable outcome classification results, we have to keep an eye on two things. First, as we said before, a proper choice of the initial number of superpixels prevents overlapping groups due to the leak of homogeneity. Second is the selection of training sets for the SVM methods. The more we cover all the possibilities, the better results we get. Under those considerations, figure 4 illustrates the comparison results of the classification with and without adding the texture features. The improvement is much more noticeable when buildings have similar colors to the other classes.

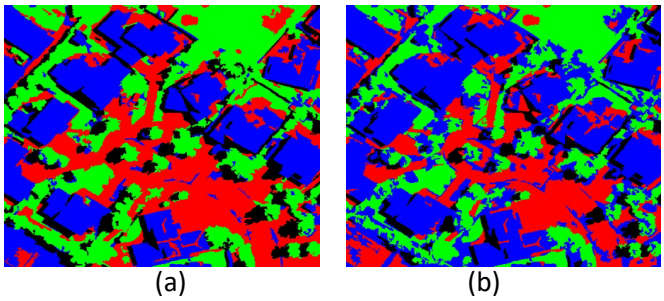


Figure 4. the comparison results of the classification, (a) the classification with the texture features, (b) without adding the texture features, Where the blue represents the buildings, the green the grass and the trees, and the red the road and the sidewalk

5.3. Building Detection Comparison

In this section, the images selected represent diverse building characteristics such as size, the shape of buildings, and different color combinations of their roofs. Our proposed method has been compared to algorithms presented by method [23], a method [26], and method [15] to give a qualitative comparison with our algorithm. As shown in Figure 5, the images on the first row indicate the original input images where we chose three different images, and the second row illustrates the final building extraction results.

Firstly, the following quantities were defined: TP (true positive), the number of correctly detected buildings, FP (false positive), the number of incorrectly detected buildings, and FN (false negative), the number of undetected buildings. To quantify the accuracy of the building extraction results, we used the detection rate (DR or precision) to measure the degree to which detected buildings indeed are actual buildings. Furthermore, the false-negative rate (FNR) measures the degree of missed detections to the total actual buildings. Meanwhile, the completeness of detection (COMP) is the number of correctly detected buildings without decreasing or increasing their boundaries.

Figure 5: Comparison of building detection results obtained by three different algorithms where (row a) Test images, (row b) The results of the proposed method, (row c) The results of the method [15], (row d) The results of the method [23] and (row e) The results of the method [26]. (Red represents building results).

Table 1: Comparison of the building extraction accuracy of the four algorithms (quantitative analysis of Figure 5).

	Our method	**method** [23]	**method** [26]	**method** [15]
TP	3719	3799	3806	3802
FP	110	749	1083	324
FN	83	56	23	27
DR (%)	97.13	83.53	77.85	92.14
FNR (%)	2.18	1.45	0.6	0.7
COMP	3570	2966	2323	3445
COMP (%)	95.99	78.07	61.03	90.61

$$DR = \frac{TP}{TP + FP} \times 100 \quad (8)$$

$$FNR = \frac{FN}{TP + FN} \times 100 \quad (9)$$

Table 1. shows the comparison results of our approach with existing methods; 50 different sizes of images (3883 buildings) have been taken for testing. Our method has a greater missed detection rate than previous strategies (2.18%). However, it has a more efficient detection rate and completeness (97.13%, 95.99%). That is since those algorithms detected elements other than buildings, such as crossroads, lawns, and parking lots, as a building explaining the high detection rate. Furthermore, because of the background interference and other visual objects, 164 buildings were missed by the proposed method from the total number of buildings. Moreover, the small buildings that were tight distributed may have been excised and integrated as a single building. It might also be that small buildings are harder to detect.

5.4. Computation Time

Finally, another important consideration for this method is computing time. The proposed methods were implemented in a MATLAB environment; Table 2 contains further information about computation times. Our suggested approach was applied to 50 test images (3,883 buildings); each line in the table represents the elapsed time for every section. The total time is 2146.18 seconds, and the average time is approximately 42.92 seconds.

Furthermore, segmentation takes only 0.34 seconds on average, which is the quickest of all steps. Conversely, the SVM classification takes substantial time, accounting for 92.1 percent of the overall time. Nevertheless, the average time needed to process an image is 42 seconds, much shorter than the approaches presented in [23] and [26].

Through figure 6, We can see that the segmentation images evaluation took 0.06 seconds longer than the feature extraction time of 0.03 seconds in figure 6a and that the time for zoning is less than feature extraction in figure 6b. This is because the regions in the 50 images that were processed had completely different dimensions (2537*3665, 2961*1761, 3465*5601,...etc.) than the area in the tested image (figure. 6B), which has Dimensions (1286*1249).

6. Conclusion

Almost every building detection method has some limitations due to the restrictions that have been applied. This paper presents a pipeline for building detection in 2D urban images by considering two main problems: the color similarity in urban objects and the difference of color shades in building rooftops. The proposed algorithm combines color and texture features to classify the urban image into four classes to solve the first issue. And the reason this is achievable is due to the advantages of using superpixels instead of single normal pixels.

Table 2: The amount of time each section of the proposed building detection

Section	Total Time (s)	Average time (s)	Percentage (%)
Image segmentation	17.02	0.3404	0.8%
Feature extraction	65.83	1.3166	3.1%
Object classification (shadow detection)	1975.58	39.5116	92.1%
The final shape of the building	87.75	1.755	4.1%
Total	2146.18	42.92.36	100%

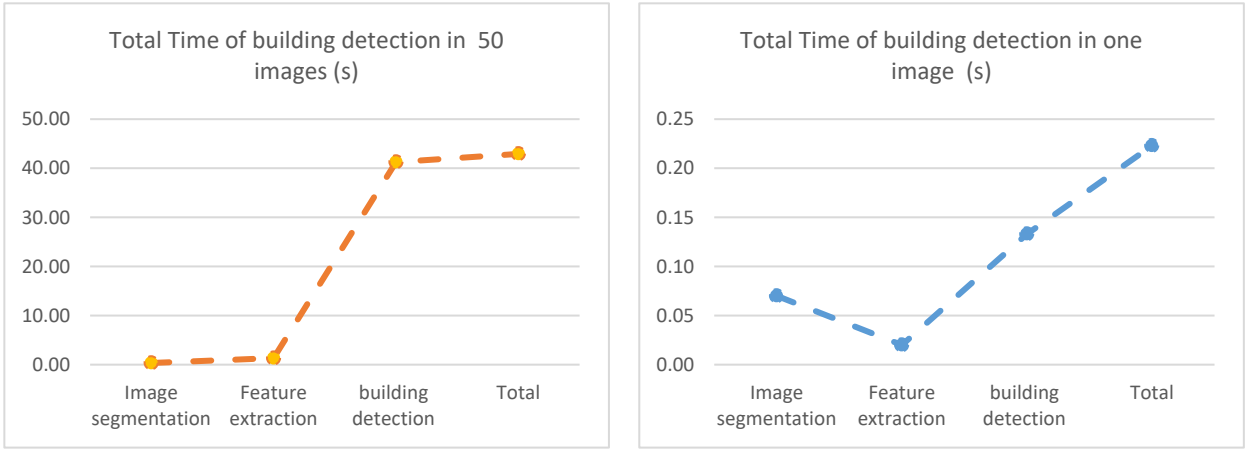


Figure 6. Building detection computational time in sec. Where (a) Total Time of building detection in 50 images and (b) Total Time of building detection in one image.

As for the second one, we use an adaptive regional growth method using only the a and b vectors from the Lab color space with the texture features. After conducting careful analysis, the experimental results revealed a remarkable improvement. Compared to the existing algorithm, the range detection procedure was quick and accurate, and the computing time required to detect the region was also reduced. The automatic detection method used in this work is a reliable methodology that may be used during a catastrophe event. However, some failures are detected in particular cases, like building with different rooftop color parts at once. Therefore, we will target more particular situations to improve detection accuracy in future work.

References

[1] W. Sirko, S. Kashubin, M. Ritter, A. Annkah, Y.S.E. Bouchareb, Y. Dauphin, D. Keysers, M. Neumann, M. Cisse, J. Quinn, "Continental-Scale Building Detection from High Resolution Satellite Imagery," 1–15, 2021.

[2] X. Shen, D. Wang, K. Mao, E. Anagnostou, Y. Hong, "Inundation extent mapping by synthetic aperture radar: A review," Remote Sensing, **11**(7), 1–17, 2019, doi:10.3390/RS11070879.

[3] S.L. Ullo, C. Zarro, K. Wojtowicz, G. Meoli, M. Focareta, "Lidar-based system and optical vhr data for building detection and mapping," Sensors (Switzerland), **20**(5), 1–23, 2020, doi:10.3390/s20051285.

[4] X. Hou, Y. Bai, Y. Li, C. Shang, Q. Shen, "High-resolution triplet network with dynamic multiscale feature for change detection on satellite images," ISPRS Journal of Photogrammetry and Remote Sensing, **177**(May), 103–115, 2021, doi:10.1016/j.isprsjprs.2021.05.001.

[5] U.C. Benz, P. Hofmann, G. Willhauck, I. Lingenfelder, M. Heynen, "Multi-resolution, object-oriented fuzzy analysis of remote sensing data for GIS-ready information," ISPRS Journal of Photogrammetry and Remote Sensing, **58**(3–4), 239–258, 2004, doi:10.1016/j.isprsjprs.2003.10.002.

[6] A. Ghandour, A. Jezzini, "Post-War Building Damage Detection," Proceedings, **2**(7), 359, 2018, doi:10.3390/ecrs-2-05172.

[7] A. Zhang, X. Liu, A. Gros, T. Tiecke, "Building Detection from Satellite Images on a Global Scale," (Nips), 2017.

[8] K. Karantzalos, N. Paragios, "Automatic model-based building detection from single panchromatic high resolution images," International Archives of the Photogrammetry. Remote Sensing & Spatial Information Sciences, **37**, 225–230, 2008.

[9] M. Aamir, Y.F. Pu, Z. Rahman, M. Tahir, H. Naeem, Q. Dai, "A framework for automatic building detection from low-contrast satellite images," Symmetry, **11**(1), 1–19, 2019, doi:10.3390/sym11010003.

[10] J.P. Cohen, W. Ding, C. Kuhlman, A. Chen, L. Di, "Rapid building detection using machine learning," Applied Intelligence, **45**(2), 443–457, 2016, doi:10.1007/s10489-016-0762-6.

[11] A.M. Cretu, P. Payeur, "Building detection in aerial images based on watershed and visual attention feature descriptors," Proceedings - 2013 International Conference on Computer and Robot Vision, CRV 2013, 265–272, 2013, doi:10.1109/CRV.2013.8.

[12] K. Nogueira, O.A.B. Penatti, J.A. dos Santos, "Towards better exploiting convolutional neural networks for remote sensing scene classification," Pattern Recognition, **61**, 539–556, 2017, doi:10.1016/j.patcog.2016.07.001.

[13] Z. Guo, Q. Chen, G. Wu, Y. Xu, R. Shibasaki, X. Shao, "Village building identification based on Ensemble Convolutional Neural Networks," Sensors (Switzerland), **17**(11), 1–22, 2017, doi:10.3390/s17112487.

[14] J. Kang, M. Körner, Y. Wang, H. Taubenböck, X.X. Zhu, "Building instance classification using street view images," ISPRS Journal of Photogrammetry and Remote Sensing, **145**, 44–59, 2018, doi:10.1016/j.isprsjprs.2018.02.006.

[15] L. Zhang, J. Wu, Y. Fan, H. Gao, Y. Shao, "An Efficient Building Extraction Method from High Spatial Resolution Remote Sensing Images Based on Improved Mask R-CNN," Sensors, **20**(5), 1465, 2020, doi:10.3390/s20051465.

[16] N. Shorter, T. Kasparis, "Automatic vegetation identification and building detection from a single nadir aerial image," Remote Sensing, **1**(4), 731–757, 2009, doi:10.3390/rs1040731.

[17] M. Ghanea, P. Moallem, M. Momeni, "Automatic building extraction in dense urban areas through GeoEye multispectral imagery," International Journal of Remote Sensing, **35**(13), 5094–5119, 2014, doi:10.1080/01431161.2014.933278.

[18] M.S. Nosrati, P. Saeedi, "A combined approach for building detection in satellite imageries using active contours," Proceedings of the 2009 International Conference on Image Processing, Computer Vision, and Pattern Recognition, IPCV 2009, **2**, 1012–1017, 2009.

[19] M. Awrangjeb, M. Ravanbakhsh, C.S. Fraser, "Automatic detection of residential buildings using LIDAR data and multispectral imagery," ISPRS Journal of Photogrammetry and Remote Sensing, **65**(5), 457–467, 2010, doi:10.1016/j.isprsjprs.2010.06.001.

[20] D. Grigillo, M. Kosmatin Fras, D. Petrovič, "Automatic extraction and building change detection from digital surface model and multispectral orthophoto," Geodetski Vestnik, **55**(01), 011–027, 2011, doi:10.15292/geodetski-vestnik.2011.01.011-027.

[21] B. Salehi, Y. Zhang, M. Zhong, V. Dey, "Object-based classification of urban areas using VHR imagery and height points ancillary data," Remote Sensing, **4**(8), 2256–2276, 2012, doi:10.3390/rs4082256.

[22] K. Stankov, D.-C. He, "Using the Spectral Similarity Ratio and Morphological Operators for the Detection of Building Locations in Very High Spatial Resolution Images," Journal of Communication and Computer, **10**(March 2013), 309–324, 2013.

[23] D. Chen, S. Shang, C. Wu, "Shadow-based Building Detection and Segmentation in High-resolution Remote Sensing Image," Journal of Multimedia, **9**(1), 181–188, 2014, doi:10.4304/jmm.9.1.181-188.

[24] J. Wang, X. Yang, X. Qin, X. Ye, Q. Qin, "An efficient approach for automatic rectangular building extraction from very high resolution optical satellite imagery," IEEE Geoscience and Remote Sensing Letters, **12**(3), 487–491, 2015, doi:10.1109/LGRS.2014.2347332.

[25] D. Kohli, R. Sliuzas, A. Stein, "Urban slum detection using texture and spatial metrics derived from satellite imagery," Journal of Spatial Science, **61**(2), 405–426, 2016, doi:10.1080/14498596.2016.1138247.

[26] Z. Lv, W. Shi, J.A. Benediktsson, X. Ning, "Novel object-based filter for improving land-cover classification of aerial imagery with very high spatial resolution," Remote Sensing, **8**(12), 2016, doi:10.3390/rs8121023.

[27] P. Manandhar, Z. Aung, P.R. Marpu, "Segmentation based building detection in high resolution satellite images," International Geoscience and Remote Sensing Symposium (IGARSS), **2017-July**, 3783–3786, 2017, doi:10.1109/IGARSS.2017.8127823.

[28] A. Ghandour, A. Jezzini, "Autonomous Building Detection Using Edge Properties and Image Color Invariants," Buildings, **8**(5), 65, 2018, doi:10.3390/buildings8050065.

[29] R. Achanta, A. Shaji, K. Smith, A. Lucchi, P. Fua, S. Susstrunk, "SLIC Superpixels Compared to State-of-the-Art Superpixel Methods," IEEE Trans. on Pat. Anal. and Mach. Intel., **34**(1), 1–8, 2012.

[30] A. Vadivel, S. Sural, A.K. Majumdar, "An Integrated Color and Intensity Co-occurrence Matrix," Pattern Recognition Letters, **28**(8), 974–983, 2007, doi:10.1016/j.patrec.2007.01.004.

[31] B. Khaldi, M.L. Kherfi, "Modified integrative color intensity co-occurrence matrix for texture image representation," Journal of Electronic Imaging, **25**(5), 053007, 2016, doi:10.1117/1.JEI.25.5.053007.

ASTES

Advances in Science, Technology and Engineering Systems Journal
Vol. 7, No. 3, 174-181 (2022)
www.astesj.com
Special Issue on Innovation in Computing, Engineering Science & Technology

ASTES Journal
ISSN: 2415-6698

A Constrained Intelligent Nonlinear Control Method for Redundant Robotic Manipulators

Dinh Manh Hung, Dang Xuan Ba*

Department of Automatic Control, HCMC University of Technology and Education (HCMUTE), Hochiminh City, 71300, Vietnam

ARTICLE INFO

Article history:
Received: 04 April, 2022
Accepted: 23 June, 2022
Online: 26 June, 2022

Keywords:
Redundant Robot
Intelligent Controller
Inverse Kinematics

ABSTRACT

Redundant robotic systems provide great challenges in solving kinematics and control problems, that are yet also open opportunities for exploring new, diverse and intelligent ideas and methods. In this paper, an advanced control method is proposed for position control problems of redundant robots with output constraints. The controller is structured with two control layers. In the high-level control layer, a cost function is first synthesized from the main control objective under constraint conditions. Virtual control signals are then reckoned to optimize the cost function using a soft Momentum-Levenberg-Marquardt approach. To realize the high-level control command, a nonlinear control signal is employed in the low-level control layer throughout a new nonsingular terminal sliding mode control structure. Comparative simulation results verified on a 7-DOF robotic arm model confirmed the effectiveness of the proposed control algorithm.

1 Introduction

Dealing with the limitations of high-order kinematic redundancy robotic models by advanced control methods has always been focuses of research interests for many years. However, solving the inverse kinematics of these flexible models is not inherently simple, making it even more difficult for models with many degrees of freedom [1], [2]. The same facing issues could be observed in robotic control fields.

To solve the inverse-kinematics problems, many interesting methods have been developed. In [2], a genetic algorithm was employed to minimize both end-effector position errors and joint displacements. Promising control results were obtained, but this algorithm took a long time to search the optimal values. In [3], an intelligent controller was developed for a redundant robot using a null-space approach and Bayesian networks. Mao et al. [4] utilized properties of a multi-layered neural network that could form any continuous nonlinear mapping from one domain to another to design an inverse-kinematics neural network to avoid obstacles and solve different solutions. Implementation time of the biomimetic algorithm [3] was impressively fast while the universal ability was exhibited by the neural approach [4]. The inverse-kinematics problem could be also well treated by various optimization methods such as the stretched simulated annealing (SSA) algorithm [5], and Jacobian pseudo inverse method [6], or the damped least-squares (DLS) law [7]. However, physical constraints are open issues of these advanced controllers. To cope with these constraint problems, quadratic programming (QP) solutions were derived using optimization algorithms [8]–[9]. By using generic QP solver, it leads to high computation cost and limit the real-time applicability [10]. Another direction for such the constraint problem is the use of null-space constraint remedies [11]. Priority-task vectors could be adopted to specify the task sequence and saturation functions were used as barriers of the constraint violations [12]. Outstanding control performances were produced but the physical constraints were not strictly consolidated in a smooth manner by using these advanced controllers.

To realize high-level control commands, a vast of controllers could be employed such as linear controllers [13], [14] and nonlinear controllers [15], [16]. Between them, sliding mode control (SMC) methods are favorite by researchers and developers thanks to their simplicity, robustness and acceptable working performances [17], [18]. Conventional SMC approaches however only result in infinite control errors [15], [19]. To further improve the control performances, terminal sliding mode control (TSMC) schemes were proposed and effectively applied for a plenty of robotic applications [19], [20]. Nevertheless, singularity problems were observed in the normal TSMC frameworks [21]. Such the weak issues were treated by a nonsingular terminal sliding mode control (NTSMC) methodologies in which the terminal power was taken into account for the time-derivative of the error signals instead of original ones [22]–[23]. However, the use of the NTSMC algorithms leads to dependence of the control signal directly on the time-derivative ones and it could activate vibration phenomena in noisy control cases

*Corresponding Author: Dang Xuan Ba, HCMC University of Technology and Education (HCMUTE), Hochiminh City, 71300, Vietnam & badx@hcmute.edu.vn

[24]–[25]. To deal with the NTSMC problem in a comprehensive fashion, new design of the NTSMC is required [22], [26], [27].

This paper is an extension of work originally presented in 2021 International Conference on System Science and Engineering (ICSSE) [1]. In this article, we demonstrate the versatility of the previously mentioned intelligent inverse kinematics solution through its successful application to a 7-DOF redundant robot. Structure of the proposed controller includes two control layers with the following contributions:

1. Robustness of the constrained intelligent inverse-kinematics algorithm in the high-level control layer is improved by the Momentum - Levenberg- Marquardt learning technique.

2. A new nonsingular nonlinear terminal sliding mode control method using a flexible power function is proposed for the low-level control layer to enhance the control performance of the closed-loop system.

3. Effectiveness of the proposed control method is verified by comparative simulation results on a 7-DOF robot model.

The outline of the paper is organized as follows. Section 2 discusses problem statements. The modified design of the intelligent inverse-kinematics algorithm is presented in Section 3. Section 4 shows a new NTSMC controller. The intensive simulation results are presented in Section 5. The paper is then concluded in Section 6.

2 Problem Statements

Forward kinematics of a general *n-dof* robotic manipulator could be obtained using homogenous-transformation computation [5], [6], [7]:

$$P_{ee} = f(\theta_1, \theta_2, ..., \theta_n) \tag{1}$$

where $P_{ee} \in R^{N\times 1}$ is the end-effector position, $\theta = [\theta_1, \theta_2, ..., \theta_n]$ is the vector of joint variables, N is the number of task-space variables, and f denotes the forward-kinematics computation.

Behaviors of the joint angles are presented by the following dynamics using the Euler-Lagrange method [28], [27]:

$$M(\theta)\ddot{\theta} + C(\theta, \dot{\theta}) + G(\theta) + F_r(\dot{\theta}) + d = \tau \tag{2}$$

where $\tau \in R^{n\times 1}$ is the vector of joint torques, $M \in R^{n\times n}$ is a symmetric positive-definite mass matrix, $C \in R^{n\times 1}$ is the Coriolis-centripetal vector, $G \in R^{n\times 1}$ is the gravitational vector, $F_r \in R^{n\times 1}$ is the Coulomb friction vector, and d stands for external disturbances.
Remark 1: Note that with a redundant robot, n is larger than N . We define a position error combining from a desired position P_{eed} and the system output P_{ee}. The main control objective here is to figure out proper control signals (τ) to drive the control error to zero complying with specific constraints. To this end, the main controller with a two-layer control structure is used. The complicated dynamics with redundant characteristics and unpredictable external disturbances are however main obstacles in developing the expected controller.

3 High-level Control Layer

The structure of the high-level control layer includes the inverse-kinematics solution stage, the constraint integration stage, and the optimization stage.

3.1 A Basic Inverse-Kinematics Method

The main control objective is formulated from the current and desired positions, as follows:

$$E_1 = 0.5\|P_{ee} - P_{eed}\|^2 \tag{3}$$

where $\|\bullet\|$ is the Euclidean norm of the term ($\bullet$).

By using the Jacobian-transpose-based method to minimize the high-level objective (2), virtual control signal is selected as follows [6], [29]:

$$\dot{\theta}_v = -\eta J^T(P_{ee} - P_{eed}) \tag{4}$$

where $J \in R^{N\times n}$ is the Jacobian matrix of the considering robot, and η is a positive learning constant.
Remark 2: In fact, there are infinite solutions for the problem (3). Hence, it is possible to shape possible solutions inside the expected regions [30].

3.2 Improvements for Joint Constraints

In this subsection, joint constraints are studied as additional control objectives. The former target (3) is modified as:

$$E_2 = 0.5\|P_{ee} - P_{eed}\|^2 + k_1^T \ln\left(\theta - \underline{\theta}\right) + k_2^T \ln\left(\bar{\theta} - \theta\right) \tag{5}$$

where $k_1, k_2 \in R^{n\times 1}$ are vectors of positive constants, and $(\bar{\theta}, \underline{\theta})$ are the upper and lower bounds of joint variables.

With the new objective (5), the virtual control signal (4) is updated as follows:

$$\dot{\theta}_v = -\eta\left(J^T(P_{ee} - P_{eed}) + \text{diag}\,(k_1)\,\frac{1}{\theta - \underline{\theta}} - \text{diag}\,(k_2)\,\frac{1}{\bar{\theta} - \theta}\right) \tag{6}$$

Remark 3: When the function *ln* is added to the expression (5), it interacts with the angular limit by increasing the value as it gets closer. This means that the virtual speed will also be subtracted a significant amount and has a deceleration effect. The parameter will adjust this interaction amount of the binding component *ln*.

However, the uncontrolled increase and decrease of this constraint component will adversely affect the real error update and cause large errors for the whole system when working near the hardware limit. The requirement here is that the additional constraint must be large enough to keep the robot within the allowable area and small enough to not affect the error update. Therefore, a new type of the parameter with great flexibility is needed to adjust the constraint up and down properly. Hence, we develop an alternative solution that makes the parameters automatically change under a nonlinear way:

$$\begin{cases} k_1 = e^{-0.5b_1|\theta - \underline{\theta}|^2} \\ k_2 = e^{-0.5b_2|\theta - \bar{\theta}|^2} \end{cases} \tag{7}$$

where $b_{i|i=2,3}$ are positive constants.

In this virtue, the control signal (6) is updated again as:

$$\begin{aligned}\dot{\theta}_v = &-\eta J^T(P_{ee} - P_{eed}) \\ &-\eta \text{diag}\,(k_1)\left(\frac{1}{\theta-\underline{\theta}} - b_1\left(\theta-\underline{\theta}\right)\ln\left(\theta-\underline{\theta}\right)\right) \\ &-\eta \text{diag}\,(k_2)\left(\frac{1}{\bar{\theta}-\theta} - b_2\left(\theta-\bar{\theta}\right)\ln\left(\bar{\theta}-\theta\right)\right)\end{aligned} \tag{8}$$

Remark 4: At this time, values of the constraint components will be adjusted up by the Gaussian-type gain when the joint variables approach the upper limits or the lower limits. This is caused by the characteristics of the Gaussian function [30], [31] as the variable values approach their centers. Conversely, when the joints move away from the upper and lower limits, depending on the slope of the Gaussian function, the gain will decrease very quickly and push the influence of the constraint down very small. This will satisfy the requirement set forth above.

3.3 Robust Constraint Inverse-Kinematics Approach

From (8), we have reformed the learning of the virtual control signal by using the Momentum-Levenberg–Marquardt method to make it faster while maintaining accuracy:

$$\begin{cases}\dot{\theta}_{v,i} = m\dot{\theta}_{v,i-1} - \frac{\eta L_b}{\alpha+\|L_b\|^2} \\ L_b = J^T(P_{ee} - P_{eed}) + \text{diag}\,(k_1)\left(\frac{1}{\theta-\underline{\theta}} - b_1\left(\theta-\underline{\theta}\right)\ln\left(\theta-\underline{\theta}\right)\right) \\ \quad +\text{diag}\,(k_2)\left(\frac{1}{\bar{\theta}-\theta} - b_2\left(\theta-\bar{\theta}\right)\ln\left(\bar{\theta}-\theta\right)\right)\end{cases} \tag{9}$$

where $(0 < m < 1)$ and α are positive constants, and $\bullet_i$, $\bullet_{i-1}$ denote current and previous states of $(\bullet)$.

Remark 5: The advantage of this learning algorithm is that it could solve the problems of the normal gradient descent learning methods which are slow to converge and easy to get stuck at the local minimum that is difficult or impossible to reach the global minimum [2], [4], [30]. The Momentum approach could help convergence faster by creating momentum behaviors from the previous velocity, while the shortcoming of instability of Momentum methods would be handled by the Levenberg–Marquardt part due to its ability to damp and ensure all the joint velocities not too different.

4 Finite-time Low-level Control Layer

The low-level control layer is a controller that realize the virtual commands (θ_v) generated by the high-level control one. To control the robot joint (θ) following the desired virtual control signal (θ_v) , we define a low-level tracking control error:

$$e = \theta - \theta_v \tag{10}$$

Inspired by terminal sliding mode control theories [17], [18], [19], we then synthesize an indirect control objective as:

$$\begin{cases} s = \dot{e} + \lambda f \\ f = \text{sgn}(e)|e|^{\frac{e^2+\gamma}{e^2+\beta}} \end{cases} \tag{11}$$

where λ and $(\gamma < \beta)$ are positive constants.

To realize the control mission (11) or (12) for the robotic system (10), we propose the following nonlinear terminal control signal:

$$\tau = (C + G + Fr) + M\left(\ddot{\theta}_v - \lambda\dot{f} - K_{dr}s^{\sigma} - K_{ro}\text{sgn}(s)\right) \tag{12}$$

where K_{dr}, K_{ro} and $(0 < \sigma < 1)$ are positive control gains.

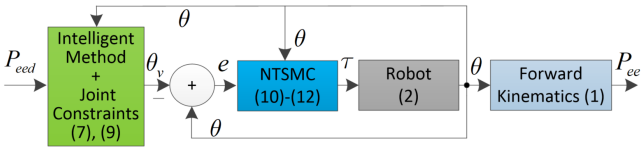


Figure 1: Block diagram of the proposed controller for redundant robots.

Remark 6: The NTSMC method can solve the problem of low settling time and push the error down very small because the larger the error leads to the steeper the selected sliding surface s will be (this is true for both the upper and lower regions of the setpoint); while the steeper the sliding surface provides the faster and more accurate the convergence. This is why the control signal can track the setpoint much better than a Proportional-Integral-Derivative (PID) controller [32], [31]. Block diagram of the proposed controller is presented in **Fig. 1**.

5 Validation Results

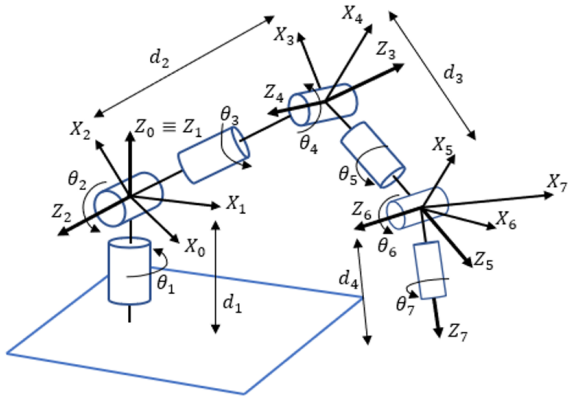


Figure 2: Configuration of a 7DOF robot for investigating.

The proposed controller was verified on a 7DOF redundant robot arm, whose detailed configuration is fully described in **Fig. 2**. The robot had three links (d_1, d_2, d_3) and seven joints $(\theta_{i|i=1..7})$. The desired trajectory of the robot was planned in the Cartesian coordinate system O_{xyz} . The end-effector position $P_{ee} = [x, y, z]^T$ of the robot is expressed as:

$$\begin{cases} x{=}d_1c_1s_2 - d_2(s_4(s_1s_3 - c_1c_2c_3) - c_1c_4s_2) - d_3(c_6(s_4(s_1s_3 - c_1c_2c_3) \\ \quad -c_1c_4s_2){+}s_6(c_5(c_4(s_1s_3 - c_1c_2c_3){+}c_1s_2s_4){+}s_5(c_3s_1{+}c_1c_2s_3))) \\ y{=}d_3(c_6(s_4(c_1s_3{+}c_2c_3s_1){+}c_4s_1s_2){+}s_6(c_5(c_4(c_1s_3{+}c_2c_3s_1) \\ \quad -s_1s_2s_4){+}s_5(c_1c_3 - c_2s_1s_3))) \\ \quad +d_2(s_4(c_1s_3{+}c_2c_3s_1){+}c_4s_1s_2){+}d_1s_1s_2 \\ z{=}d_2(c_2c_4 - c_3s_2s_4) - d_3(s_6(c_5(c_2s_4{+}c_3c_4s_2) - s_2s_3s_5) \\ \quad -c_6(c_2c_4 - c_3s_2s_4)){+}d_1c_2 \end{cases}$$

where $s_{i|i=1..7} := \sin(\theta_i)$ and $c_{i|i=1..7} := \cos(\theta_i)$.

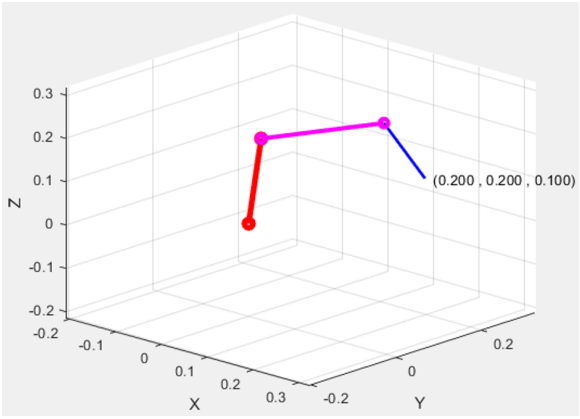


Figure 3: Structure of the 7DOF robot at working position.

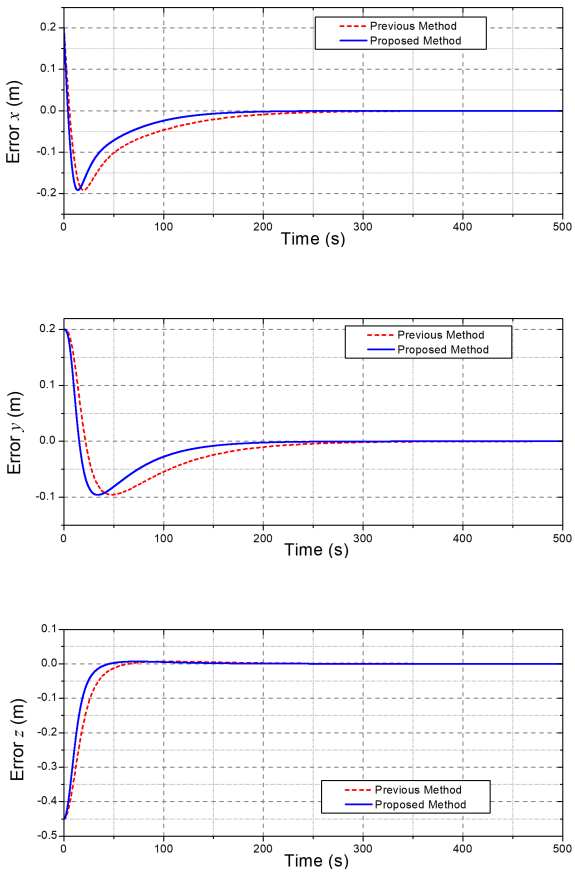


Figure 4: Comparative control errors of the end-effector position obtained by the controllers.

The parameters of the robot model for simulation were selected as follows: $m_1 = 0.5$ kg, $m_2 = 0.5$ kg, $m_3 = 0.5$ kg, $m_4 = 0.3$ kg, $m_5 = 0.3$ kg, $m_6 = 0.2$ kg, $m_7 = 0.2$ kg, $d_1 = d_2 = 0.2$ m, and $d_3 = 0.15$ m . The operating condition was free of interference and had a viscous friction force with a selected friction coefficient $\mu = 20$. The initial values of the joint angles were $\theta_{i|i=1..7} = 0$, and the initial end-effector position of the robot in the Cartesian coordinate was $(0; 0; 0.55)$ (m). The desired position of the end-effector chosen for testing was $(0.2; 0.2; 0.1)$ (m). To clearly evaluate the control performances of the proposed controller, a previous control method [1] was employed to realize the same control mission in the same system under the same testing conditions. Control gains of the previous controller were selected as $b_1 = b_2 = 0.0001, \eta = diag([0.4, 0.4, 0.5, 0.5, 0.6, 0.6, 0.6]), K_P = 200I_7, K_I = 50I_7, K_D = 25I_7$, while those of the proposed controller were manually tuned and obtained as $b_1 = b_2 = 0.0001, \eta = diag([0.4, 0.4, 0.5, 0.5, 0.6, 0.6, 0.6]), \alpha = 1, m = 0.9, \lambda = 10I_7, \gamma_{i|i=1..7} = 0.95, \beta_{i|i=1..7} = 1, \sigma_{i|i=1..7} = 0.9, K_{dr} = 100I_7, K_{ro} = 0.001I_7$.

In the first test, the two controllers were applied to control the robot from the initial position to the desired one with the same low-level control layer but with different high-level control ones. The response posture of the robot working under the proposed controller after the simulation is shown in **Fig. 3**. Control results obtained are shown in **Figs. 4 - 5**.

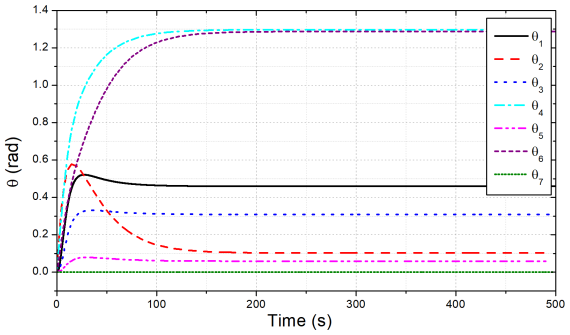


Figure 5: Reference joint angles generated by the proposed controller from the first test.

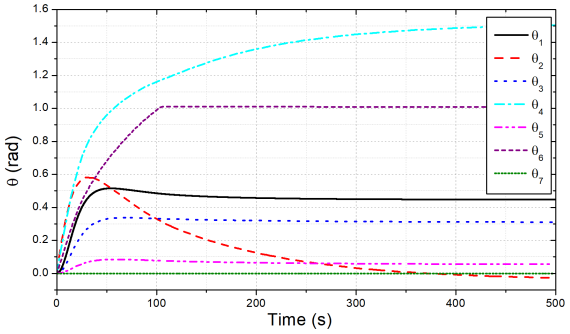


Figure 6: Desired profiles of the joint angles in the joint-constraint test.

The control errors of the end-effector position of the robot accomplished by the two controllers are compared in **Fig. 4**. Generally, the two control systems were stably working with excellent steady-state control errors of about $(6, 7.2, 4.5) \times 10^{-6}$(m) in the x, y, z directions of the end-effector positions, respectively. The figure also shows that the convergence time of the previous controller was about 310 (s) (the red-dot line) while that of the proposed controller was only about 203 (s) (the blue-solid line). The faster results came from the Momentum learning behaviors supported by the Levenberg–Marquardt adaptation scheme (8). Reference joint angles generated by the high-level control layer of the proposed controller are shown in **Fig. 5**. By combining **Figs. 2**, **3** and **5**, it can be easy to understand that to reach the desired end-effector

position, the joints 4 and 6 were hard-working parts with the largest variations (from 0 to 1.29) (rad).

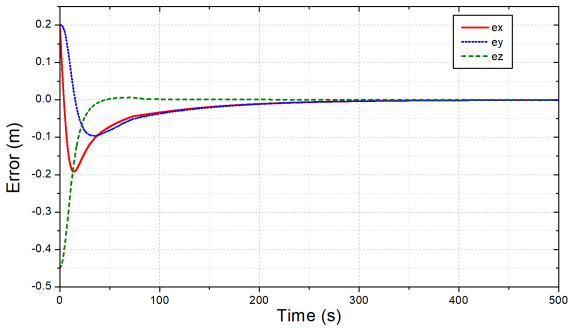


Figure 7: Errors of the end-effector in the joint-constraint test.

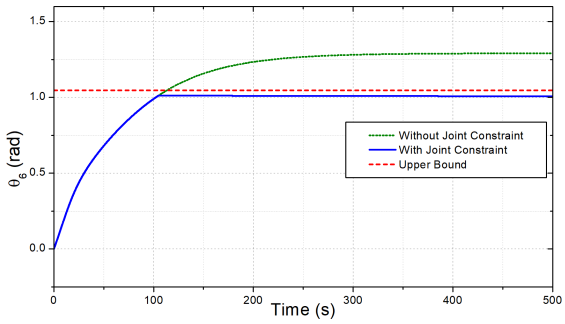


Figure 8: Comparative output at joint 6 in the first and second tests.

To assess the effectiveness of the joint-constraint feature proposed, in the second test, we limited the joint 6 into a range of $(-\pi/3 \leq \theta_6 \leq \pi/3)$. The simulation results achieved by the proposed controller for the old and new testing conditions are presented in **Figs. 6-8**. As seen in **Fig. 7**, the proposed controller still ensured the excellent control quality for the end-effector positions under the constrained working condition. Furthermore, as shown in **Fig. 6**, the profiles of all joint angles would also change to ensure that the data changes of θ_6 was still inside of the given range $(-\pi/3; \pi/3)$. The profiles of the joint variable θ_6 in this test and the last test are compared in **Fig. 8**. It can be seen that, in the case of the constraint defined, the angle θ_6 increased very fast to the upper limit and then stopped increasing and kept a certain distance with the upper limit because the joint velocity was greatly reduced caused by the algorithm. The results would also be similar for the other joints if other joint constraints were applied. Note that, thanks to the redundant robot configuration possessed, even though the joint angle θ_6 was limited, the control burden was shared by the other joint angles, especially by the joint θ_2 and θ_4, that could be observed in this test by comparing the data in **Figs. 5** and **6**.

In the third test, we performed a new simulation to show the importance of the low-level control layer by comparing the response quality between the new controllers with NTSMC and previous PID controllers [1]. The two controllers were used to control the end-effector of the robot to the desired position and their control results are compared in **Figs. 9-12**. For the easier observation, we only took the results of joints 2, 4, 6 because they are folded joints.

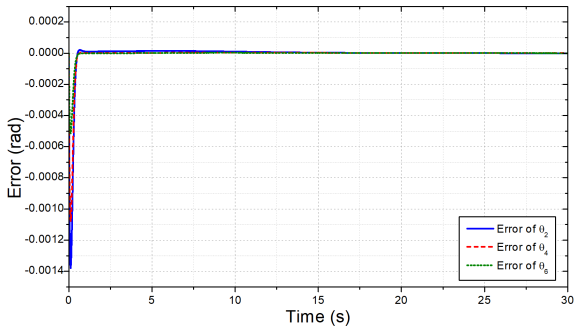


Figure 9: The joint control errors obtained by the new NTSMC controller.

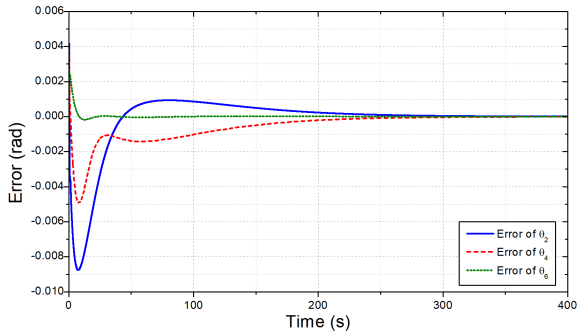


Figure 10: The joint control errors obtained by the previous controller.

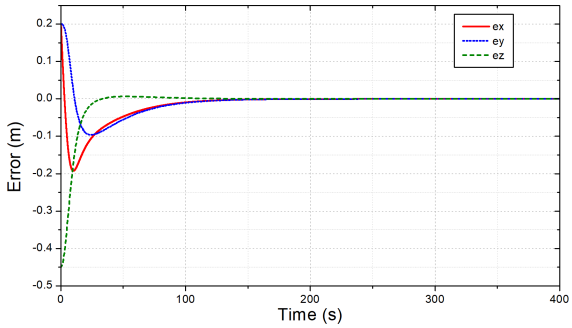


Figure 11: The end-effector control errors obtained the proposed controller.

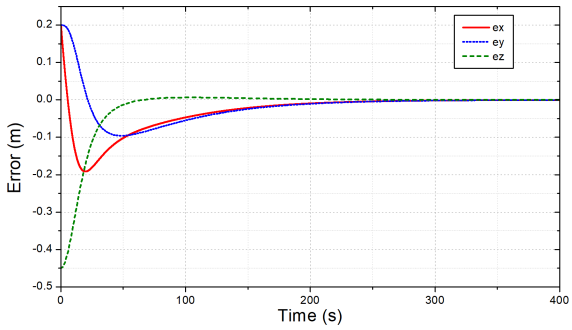


Figure 12: The end-effector control errors obtained the previous controller.

The results in **Figs. 9** and **10** indicate that the settling time and the transient response of the new controller were superior to those of the old one (ten times faster). Achieving a very small error ($\pm 10^{-4}$)

in a very short time has shown the effectiveness of the improved NTSMC method compared to the previous one. The other control results in **Figs. 11** and **12** also imply that with the improvements in the low-level control layer, it is not surprising about the faster response and higher accuracy in the end-effector control space obtained by the proposed controller as comparing to the previous one.

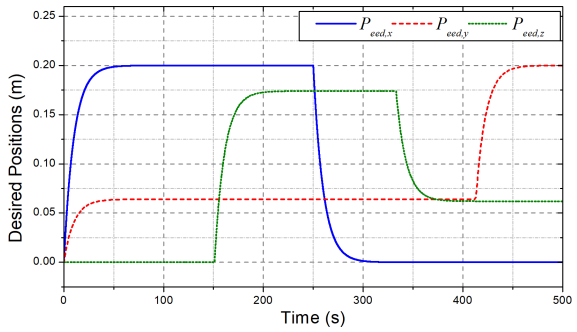


Figure 13: The desired end-effector positions in the fourth test.

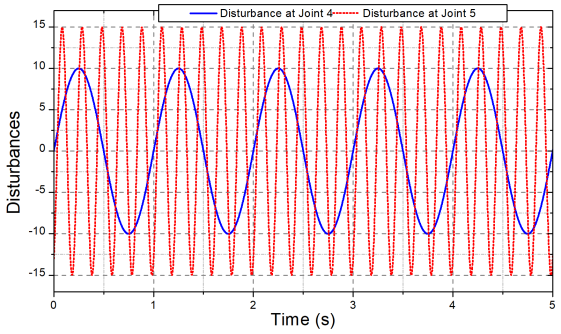
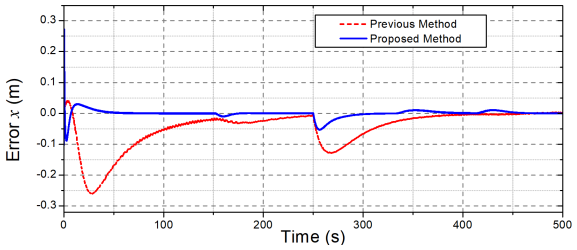
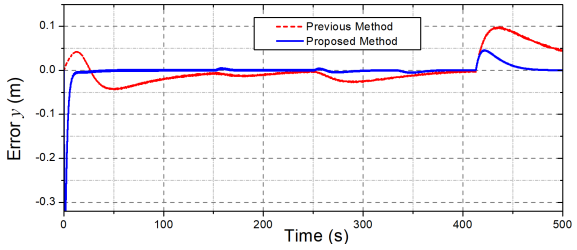


Figure 14: External disturbances in the fourth test.





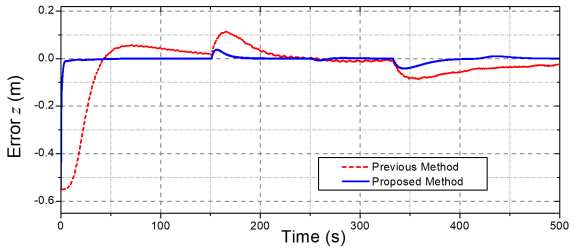
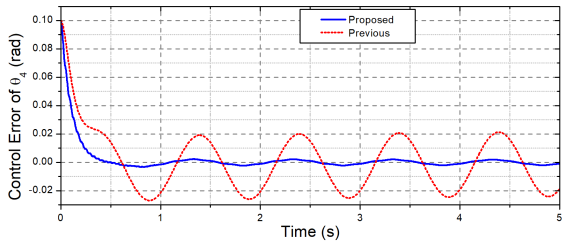


Figure 15: Comparative control errors at the end-effector of the robot.



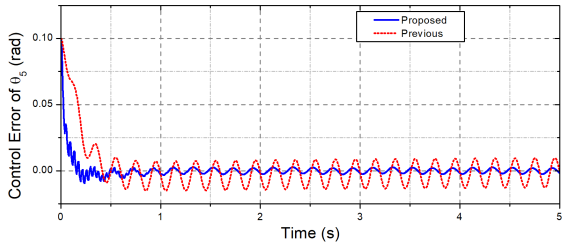


Figure 16: Comparative control errors of the robot joints.

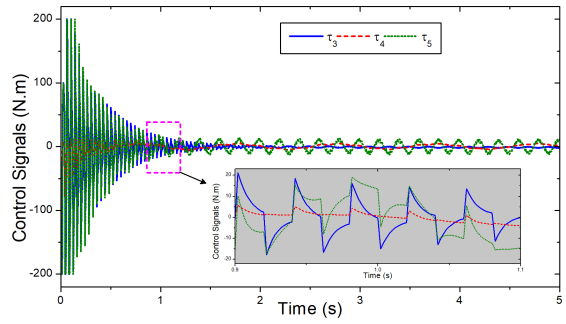


Figure 17: Control signals generated by the proposed controller at the robot joints.

To validate the feasibility of the proposed control approach, in the last simulation, the robot was challenged with new desired end-effector positions of multi-step signals, as depicted in **Fig. 13**, and external disturbances at joints 4 and 5 as presented in **Fig. 14**. The obtained control results of the proposed and previous controllers are illustrated in **Figs. 15 - 17**. Under the new testing conditions, as seen in **Fig. 15**, the designed controller still provided higher control accuracies and faster settling time at the end-effector than the previous one. To this end, the proposed control approach was not only employed the robust intelligent learning control law (7)-(9) but it was also supported by the new NTSMC framework (10)-(12).

Indeed, as shown in **Fig. 16**, the control errors of the new nonlinear low-level controller at joints 4 and 5 in the heavy disturbances were respectively 0.0018 (rad) and 0.002 (rad), while those of the previous one were 0.03 (rad) and 0.015 (rad). **Figure 17** shows the control signals generated by the proposed controller. The feasibility of the proposed control method could be confirmed throughout the data obtained.

6 Conclusions

In this paper, an intelligent two-layer control method for dealing with inverse-kinematics problems of redundant robots has been improved and applied to a 7-DOF robot. In the high-level control layer, the inverse kinematics problem is solved by using the Momentum-Levenberg optimization method. In cases of the joint constraint requirements, an advanced constrained learning feature can be activated to ensure that the robot joints can avoid physical limit collisions. To enhance the control performance of the overall system, a new nonlinear terminal sliding mode control framework is developed in the low-level control layer. The effectiveness of the proposed control algorithm has been consolidated by comparative validation results obtained. In the future, the intelligent method will be integrated more advanced, optimal, flexible working features and verified on a real-time system.

Conflict of Interest The authors declare no conflict of interest.

References

[1] D. M. Hung, D. T. Linh, D. X. Ba, "An Intelligent Control Method for Redundant Robotic Manipulators with Output Constraints," in 2021 International Conference on System Science and Engineering (ICSSE), 116–121, 2021, doi:10.1109/ICSSE52999.2021.9538490.

[2] A. C. Nearchou, "Solving the inverse kinematics problem of redundant robots operating in complex environments via a modified genetic algorithm," Mechanism and Machine Theory, **33**(3), 273–292, 1998, doi:10.1016/S0094-114X(97)00034-7.

[3] P. K. Artemiadis, P. T. Katsiaris, K. J. Kyriakopoulos, "A biomimetic approach to inverse kinematics for a redundant robot arm," Autonomous Robot, **29**, 293–308, 2010, doi:10.1007/s10514-010-9196-x.

[4] Z. Mao, T. C. Hsia, "Obstacle avoidance inverse kinematics solution of redundant robots by neural networks," Robotica, **15**, 3–10, 1997, doi:10.1109/ROBOT.1993.291813.

[5] P. Costa, J. Lima, A. I. Pereira, P. Costa, A. Pinto, "An Optimization Approach for the Inverse Kinematics of a Highly Redundant Robot," in Abraham, A., Wegrzyn-Wolska, K., Hassanien, A., Snasel, V., Alimi, A. (eds) Proceedings of the Second International Afro-European Conference for Industrial Advancement AECIA 2015. Advances in Intelligent Systems and Computing, volume 427, 433–442, 2015, doi:10.1007/978-3-319-29504-6_41.

[6] D. X. Ba, J. B. Bae, "A precise neural-disturbance learning control of constrained robotic manipulators," IEEE Access, **9**, 50381–50390, 2021, doi:10.1109/ACCESS.2021.3069229.

[7] D. S. Binh, K. L. Thanh, D. P. Nhien, D. X. Ba, "An intelligent inverse kinematic solution of universal 6-DOF robots," in Huang, Y. P., Wang, W. J., Quoc, H.A., Giang, L.H., Hung, N.L. (eds) Computational Intelligence Methods for Green Technology and Sustainable Development. GTSD 2020. Advances in Intelligent Systems and Computing, volume 1284, 107–116, 2020, doi:10.1007/978-3-030-62324-1_10.

[8] F. Cheng, T. Chen, Y. Sun, "Resolving manipulator redundancy under inequality constraints," IEEE Transactions on Robotics and Automation, **10**(1), 65–71, 1994, doi:10.1109/70.285587.

[9] A. Escande, N. Mansard, P. B. Wieber, "Fast resolution of hierarchized inverse kinematics with inequality constraints," in 2010 IEEE International Conference on Robotics and Automation, 3733–3738, 2010, doi:10.1109/ROBOT.2010.5509953.

[10] A. Escande, N. Mansard, P. B. Wieber, "Hierarchical quadratic programming: Fast online humanoid-robot motion generation," The International Journal of Robotics Research, **7**(33), 1006–1028, 2014, doi:10.1177/0278364914521306.

[11] F. Flacco, A. D. Luca, O. Khatib, "Prioritized multi-task motion control of redundant robots under hard joint constraints," in 2012 IEEE/RSJ International Conference on Intelligent Robots and Systems, 3970–3977, 2012, doi:10.1109/IROS.2012.6385619.

[12] F. Flacco, A. D. Luca, O. Khatib, "Control of redundant robots under hard joint constraints: saturation in the null space," IEEE Transactions on Robotics, **31**(33), 637–654, 2015, doi:10.1109/TRO.2015.2418582.

[13] V. I. Utkin, "Sliding mode control design principles and applications to electric drives," IEEE Transactions on Industrial Electronics, **40**(1), 23–36, 1993, doi:10.1109/41.184818.

[14] J. Zhang, W. X. Zheng, "Design of adaptive sliding mode controllers for linear systems via output feedback," IEEE Transactions on Industrial Electronics, **61**(7), 3553–3562, 2014, doi:10.1109/TIE.2013.2281161.

[15] D. X. Ba, "A Fast Adaptive Time-delay-estimation Sliding Mode Control for Robot Manipulators," Advances in Science, Technology and Engineering Systems Journal, **5**(6), 904–911, 2020, doi:10.25046/aj0506107.

[16] C. Hu, B. Yao, Q. Wang, "Performance-oriented adaptive robust control of a class of nonlinear systems proceeded by unknown dead zone with comparative experimental results," IEEE/ASME Transactions on Mechatronics, **18**(1), 178–189, 2013, doi:10.1109/TMECH.2011.2162633.

[17] Z. H. Man, A. P. Paplinski, H. R. Wu, "A robust MIMO terminal sliding mode control scheme for rigid robot manipulator," IEEE Transactions on Automatic Control, **39**(12), 2464–2469, 1994, doi:10.1109/9.362847.

[18] L. Fridman, J. Moreno, R. Iriarte, Sliding Modes after the first Decade of the 21st Century, Springer Berlin, Heidelberg, 2012, doi:10.1007/978-3-642-22164-4.

[19] S. T. Venkataraman, S. Gulati, "Control of nonlinear systems using terminal sliding modes," in 1992 American Control Conference, 891–893, 1992, doi:10.23919/ACC.1992.4792209.

[20] Y. Tang, "Terminal sliding mode control for rigid robots," Automatica, **34**(1), 51–56, 1998, doi:10.1109/9.362847.

[21] K. B. Park, J. J. Lee, "Comments on 'A robust MIMO terminal sliding mode control scheme for rigid robotic manipulators," IEEE Transactions on Automatic Control, **41**(5), 761–762, 1996, doi:10.1109/9.362847.

[22] D. X. Ba, H. Yeom, J. B. Bae, "A Direct Robust Nonsingular Terminal Sliding Mode Controller based on an Adaptive Time-delay Estimator for Servomotor Rigid Robots," Mechatronics, **59**, 82–94, 2019, doi:10.1016/j.mechatronics.2019.03.007.

[23] C. K. Lin, "Nonsingular terminal sliding mode control of robot manipulators using fuzzy wavelet networks," IEEE Transactions on Fuzzy Systems, **14**(6), 849–859, 2006, doi:10.1109/TFUZZ.2006.879982.

[24] Z. Ma, G. Sun, "Dual terminal sliding mode control design for rigid robotic manipulator," Journal of the Franklin Institute, **355**(18), 9127–9149, 2018, doi:10.1016/j.jfranklin.2017.01.034.

[25] R. M. Asl, Y. S. Hagh, R. Palm, "Robust control by adaptive nonsingular terminal sliding mode," Engineering Applications of Artificial Intelligence, **59**, 205–217, 2017, doi:10.1016/j.engappai.2017.01.005.

[26] Y. Shtessel, M. Taleb, F. Plestan, "A novel adaptive-gain supertwisting sliding mode controller: Methodology and application," Automatica, **48**(5), 759–769, 2012, doi:10.1016/j.automatica.2012.02.024.

[27] J. Baek, M. Jin, S. Han, "A new adaptive sliding mode control scheme for application to robot manipulators," IEEE Transactions on Industrial Electronics, **63**(6), 3628–3637, 2016, doi:10.1080/00207179308934387.

[28] Y. Feng, X. Yu, Z. Man, "Nonsingular terminal sliding mode control of rigid manipulators," Automatica, **38**(12), 2159–2167, 2002, doi:10.1016/S0005-1098(02)00147-4.

[29] O. Kanoun, F. Lamiraux, P. B. Wieber, "Kinematic control of redundant manipulators: Generalizing the task-priority framework to inequality task," IEEE Transactions on Robotics, **27**(4), 785–792, 2011, doi:10.1109/TRO.2011.2142450.

[30] A. Tringali, S. Cocuzza, "Globally Optimal Inverse Kinematics Method for a Redundant Robot Manipulator with Linear and Nonlinear Constraints," Robotics, **9**(3), 2020, doi:10.3390/robotics9030061.

[31] E. Krastev, "Sliding Mode Control of a Redundant Robot Arm in Motion Planning Subject to Joint Space Constraints," in 2017 International Conference on Control, Artificial Intelligence, Robotics and Optimization (ICCAIRO), 109–114, 2017, doi:10.1109/ICCAIRO.2017.31.

[32] C. M, W. Xu, C. Sun, "On switching manifold design for terminal sliding mode control," Journal of the Franklin Institute, **353**, 1553–1572, 2016, doi:10.1016/j.jfranklin.2016.02.014.