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.

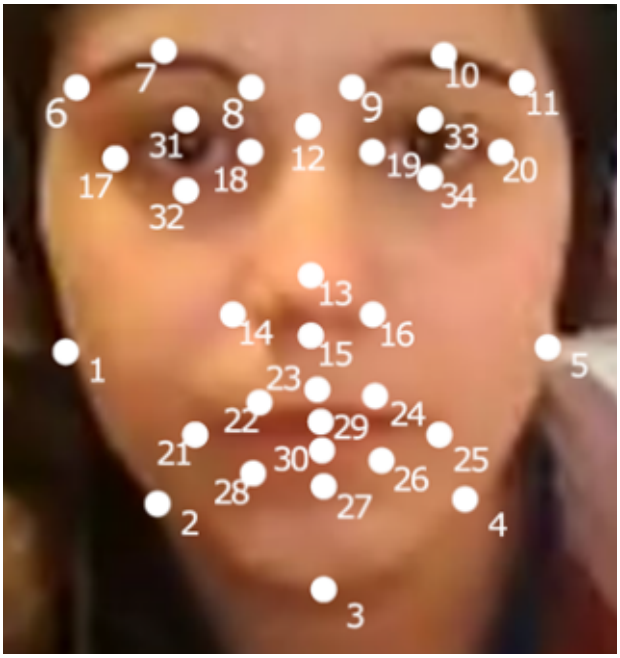

Figure 2: Facial landmarks identified by Affectiva.

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.

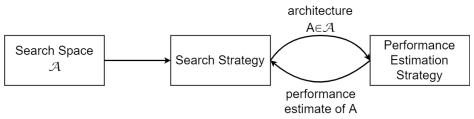

Figure 3: A high-level illustration of neural architecture search [57].

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.

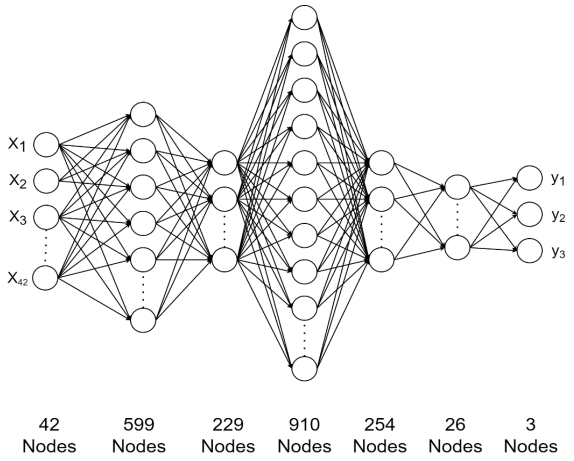

Figure 4: Deep learning neural network architecture.

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.