A Multiplatform Application for Automatic Recognition of Personality Traits in Learning Environments

Víctor Manuel Bátiz Beltrán[*,1], Ramón Zatarain Cabada[1], María Lucía Barrón Estrada[1], Héctor Manuel Cárdenas López[1], Hugo Jair Escalante[2]

[1]*Tecnológico Nacional de México campus Culiacán, Culiacán, Sinaloa, 80220, México*

[2]*Instituto Nacional de Astrofísica, Óptica y Electrónica, Tonantzintla, Puebla, 72840, México*

A R T I C L E   I N F O

A B S T R A C T

The present work shows the development of a data collection platform that allows the researcher to collect new video and voice data sets in Spanish. It also allows the application of a standardized personality test and stores this information to analyze the effectiveness of the automatic personality recognizers concerning the results of a standardized personality test of the same participant. Thus, it has elements to improve the evaluated models. These optimized models can then be integrated into intelligent learning environments to personalize and adapt the content presented to students based on their dominant personality traits. To evaluate the developed platform, an intervention was conducted to apply the standardized personality test and record videos of the participants. The data collected were also used to evaluate three machine learning models for automatic personality recognition.

1.   Introduction

This paper is an extension of a discussion paper originally presented in [1]. One of the most widely used and accepted models for determining personality based on written tests are the trait-based models and specifically the Big-Five model. This model is usually represented by the acronym OCEAN where each letter refers to a term that represents each of the five personality traits: Openness to experience, Conscientiousness, Extraversion, Agreeableness and Neuroticism [2].

One of the most relevant efforts regarding the definition of the questions (items, as they are known in the field of psychology) to be used for the Big-Five model, is the one conducted by the International Personality Item Pool (IPIP), which we can consider as a scientific laboratory for the development of advanced measures of personality traits and other individual differences that are in the public domain thanks to its Web site (http://ipip.ori.org/). This site maintains an inventory of thousands of items and hundreds of scales for the measurement of personality traits and is generally based on the studies conducted by Goldberg [3–5].

In recent years, research has been conducted with the aim of implementing automatic personality recognizers through machine learning. These studies have focused mainly on using the Big-Five model to detect apparent personality based on text, voice, or facial features. The main challenge facing these investigations is the difficulty of having a representative dataset, the need to label the images, and the fact that these efforts are typically not in the public domain and therefore it is difficult to reproduce their results [2].

The main contribution of this work is the development of an integrated environment that allows assessing the personality traits of an individual by using a standardized test based on the Big-Five model and allows capturing video interactions in Spanish. Deep learning based automatic recognizers uses these videos and seek to determine the same personality aspects. The above, to be able to evaluate the effectiveness of such automatic recognizers with respect to the standardized test and thus have relevant information to improve the model used by the recognizers that can be integrated into different intelligent learning systems to add new features such as personalized instruction and feedback to students.

This paper is structured in the following order: Section 2 presents related works in the areas of standardized testing and automatic personality recognition; Section 3 presents an analysis of the proposed data collection platform; Section 4 describes the experiments, results, and discussion; and finally, Section 5 presents conclusions and future work.

*Corresponding Author: Víctor Manuel Bátiz Beltrán, Email:
victor.bb@culiacan.tecnm.mx

2. Related Works

In this section we describe some research works related to the area of standardized tests and automatic personality recognition. These works, although separate efforts, are related to elements of the present research and were considered as the foundation for the development and integration of this project.

2.1. Standardized Personality Tests

Studies have shown that one of the best approaches to personality detection is the Big Five model. Its strength lies in the general acceptance that personality traits, although they may exhibit some changes, remain relatively stable throughout a person's life [6].

In recent years, several studies have been carried out that present adaptations to different languages of the items provided by the IPIP, to evaluate their applicability in different cultures, finding positive results. As an example of the above, we found the study conducted in [7] for the adaptation and contextualization of 100 items of the IPIP repository in the Argentine environment, obtaining satisfactory results in their reliability studies, and on the other hand the adaptation made by the authors in [6] for the application of a reduced version of the IPIP questionnaire in French-speaking participants. This version consisted of 20 items in total, where each of the personality traits was evaluated with 4 items, obtaining as a result the confirmation of the cross-cultural relevance of the personality indicators, of the model of the Big-Five in participants with diverse idiomatic and cultural backgrounds.

2.2. Automatic Recognition of Personality

In the field of automatic recognition, several approaches to apparent personality recognition have been proposed in recent years. Some studies have worked on automatic recognition based on textual information, such as information generated by users on social networks like Facebook, Twitter, and YouTube. Such is the case of the study presented in [8] where diverse approaches such as multivariate regression and univariate approaches such as decision trees and support vector machines are analyzed for automatic recognition.

Other studies have worked on the recognition of apparent personality based on the voice of participants. In [9] the authors propose a system based on a convolutional neural network that evaluates a voice signal and returns values for the five personality traits of the Big-Five model. They conclude that the correlation between different dimensions of a voice signal can help infer personality traits.

Likewise, research has been carried out for the detection of apparent personality based on images extracted from videos of the participants, using various models of neural networks [2]. In the research work conducted in [10], the authors present an apparent personality recognition model based on convolutional neural networks using images extracted from short video clips. They conclude that facial information plays a key role in predicting personality traits.

Renewed interest in the world of artificial intelligence and machine learning, as well as the existence of competencies such as those conducted by ChaLearn Looking at People, have helped the development of various neural network models for apparent personality detection based on first impression [2,11].

3. Data Collection Platform

As we are working with sensitive data from individuals, it is important to emphasize that participants are notified that the data collected from the standardized test and the videos are used internally for the experiments by the team of researchers. Therefore, no information that reveals or compromises their identity is published without prior consent. For this purpose, the platform always requests their registration to have their contact information.

3.1. Architecture

Data collection presents a challenge, as we must establish a system for storing and consulting the information. Nowadays, thanks to the advancement of technology, we can develop environments that make use of the Internet and thus be able to reach more people regardless of their location or the device they use to connect to the Internet. Therefore, we chose to develop a cloud platform that would work on any device and that would allow us to store the information in a repository located on the Internet to facilitate the study of the data.

We chose to use a layered architectural model on a client/server architecture. Three layers are defined: presentation, application logic and data. In Figure 1 we can appreciate the logical view of the platform.

The presentation layer shows users a graphical interface that offers them the option of registering with the system or logging in. The presentation layer uses the application logic layer to execute the operations supported by the system. The application logic layer, in turn, connects to the data layer which contains the database that stores the information regarding user identification, IPIP test and automatic recognizer results and in this layer, we have the file storage whose function is to store video files (including audio) related to the users.

The developed application is a cross-platform system hosted on the Internet cloud, using the free hosting offered by Google Firebase as part of its services.

For the development of the user interface, we decided to use React (https://es.reactjs.org/) and for the logic part of the application and storage of information and files, we opted for the free services of Google Firebase. In addition to availability, another advantage of using these services is the support in privacy and data security offered by Firebase, since it is certified in the main security and privacy standards (https://firebase.google.com/support/privacy).

In Figure 2 we can appreciate a partial view of the application of the standardized personality test in the data collection platform. Answering the test is the first step the user must take after logging into the system. The resulting scores are used as the actual values of the participant's personality traits.
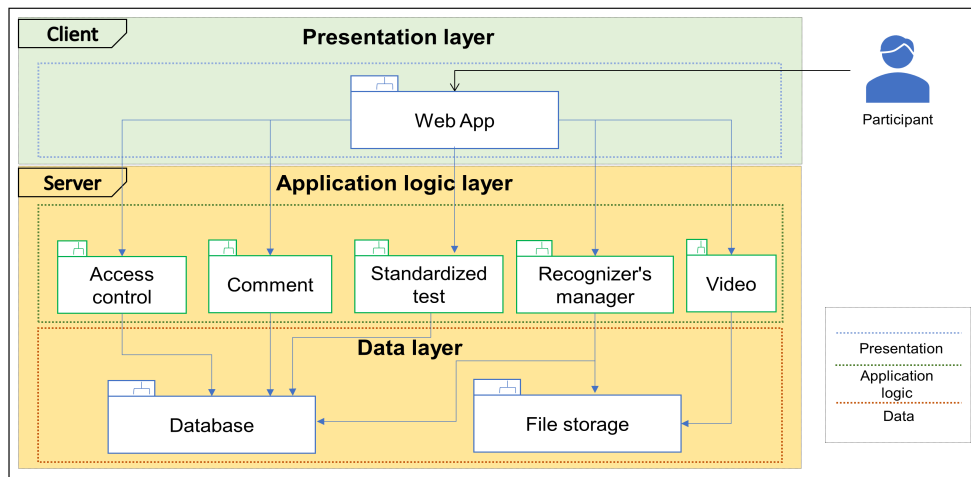
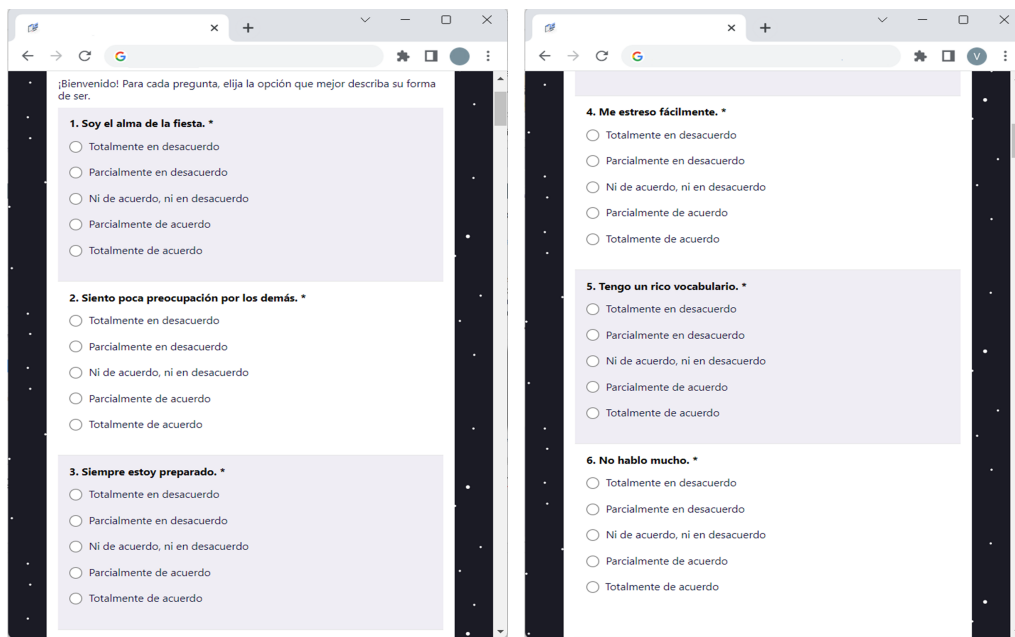Figure 1: Logical view of the platform.

Figure 2: Partial view of standardized personality test.

In Figure 3 we can see the interface where participants record a video talking about a topic of their choice.

3.2. Automatic Recognition of Personality

As an example of the use of our platform, we have decided to evaluate three automatic recognition models based on deep learning, using convolutional neural networks (CNN), and Long Short-Term Memory (LSTM) neural networks. These automatic recognizers were trained using ChaLearn's personality dataset which contains 10,000 videos with an approximate duration of 15 seconds each one [12]. The following is a description of the architectures of these automatic recognizers.

3.2.1. Discrete convolutional residual neural network (TNMCUL1)

The architecture used by this evaluated automatic recognizer (see Figure 4) is a discrete convolutional residual neural network (ResNet). We have our input layer of size 500x500x3.
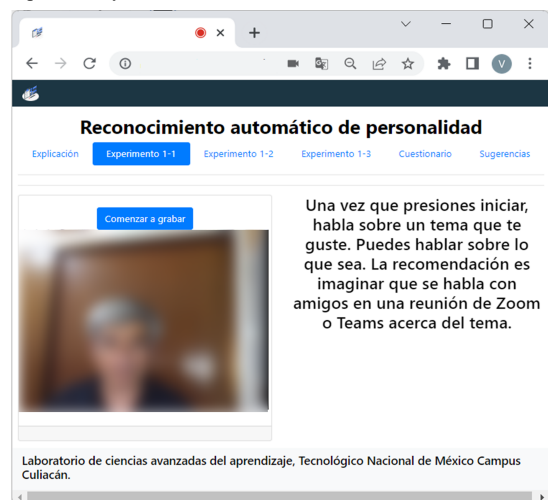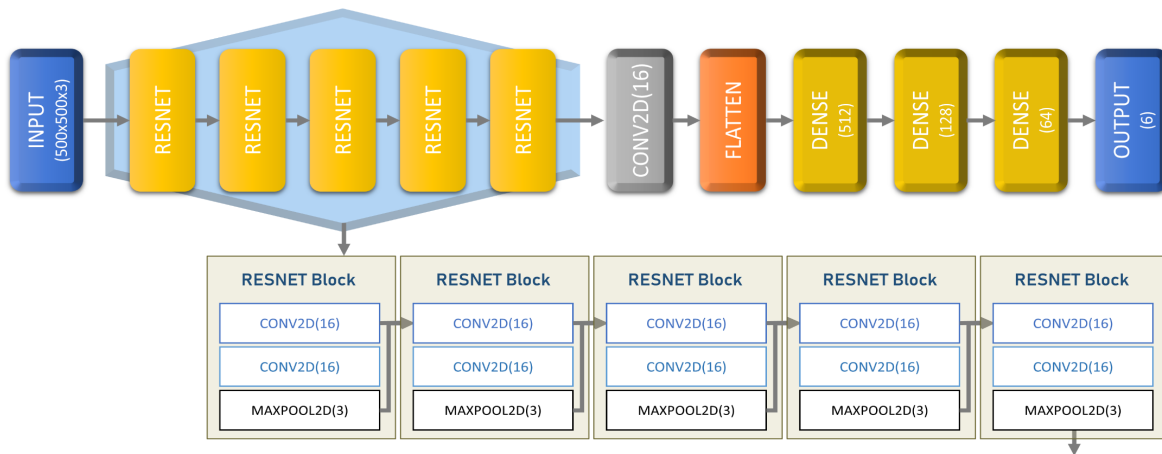

Figure 3: Video Recording.

Figure 4: Discrete convolutional neural network topology.

Then we created 5 ResNet modules, where each module contains a two-dimensional convolutional layer (Conv2D) connected to another Conv2D layer, with 16 filters in 3 dimensions in both, and finally connected to a two-dimensional maximum grouping layer (Maxpool2D) with a stride of 3. A concatenation layer was used to add the characteristics of each ResNet block. Then a single Conv2D layer was used with 16 filters in 3 dimensions. We flattened the features vector and connected 4 densely connected layers, each with 512, 128, 64 and 6 neurons respectively. All layers used ReLU activation except the last one that used sigmoid activation for regression. Loss was measured using the mean absolute error (MAE).

3.2.2. Continuous convolutional residual neural network (TNMCUL2)

The second considered architecture is a time-distributed convolutional neural networks. We used sets of 30 images taken from each video creating vectors of dimensions 30x500x500x3.

Then, we proceeded to use 5 ResNet modules to process the data from the image vectors and use the attribute vectors created by convolution as input for a 3 layered neural network for feature classification with a 6-neuron output layer for regression.

The full architecture is detailed in this way: first, we used an input layer of size 30x500x500x3. Then, we created a time distributed layer to wrap 5 ResNet modules with the same configuration, as explained on TNMCUL1. Finally, we used a global average pooling 3D layer, a flattening layer for the feature vector, and 4 densely connected layers, with 512, 256, 128 and 6 neurons, respectively. The hidden layers used ReLU activation, and the output layer used sigmoid activation for regression.

This architecture was designed to explore the use of time distributed layers in our study to add another dimension for the previous architecture. We aimed to use the full video image data on a full array, different to our previous architecture that only used individual frames. Figure 5 shows the topology of this neural network.
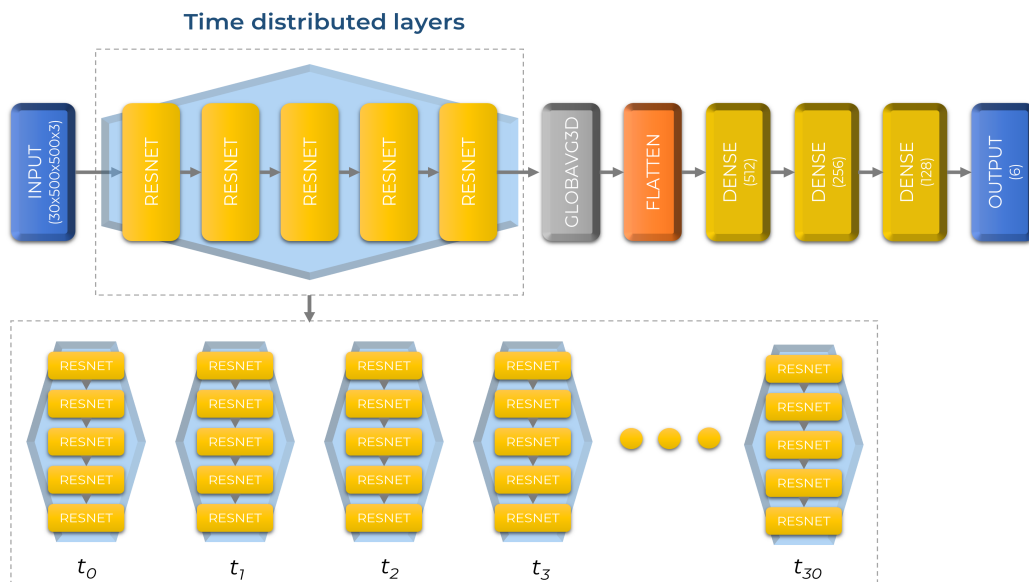

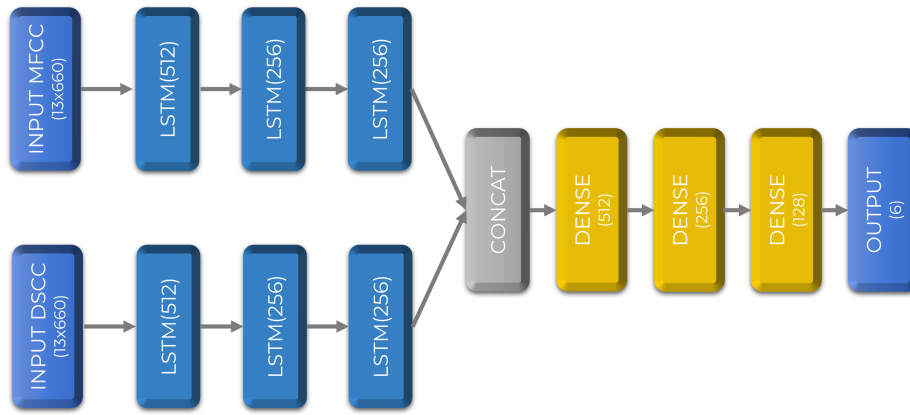Figure 5: Continuous convolutional residual neural network topology.

Figure 6: Speech MFCC + DSCC LSTM neural network topology.

3.2.3. Speech MFCC + DSCC LSTM neural network (TNMCUL3)

The third architecture used Long Short-Term Memory (LSTM) neural networks for audio processing. We used an audio file MFCC and DSCC vectors per video, using dual input layers connected to LSTM layers. Then, we used our two feature vectors for feature concatenation and used a 4 layered neural network for feature classification with a 6-neuron output layer for regression.

The full architecture is structured as follows: we used two twin feature extraction LSTM. First, we created input layers for MFCC and DSCC both used as input layers with a dimension of 13x660, connected to 3 LSTM layers of 512, 256 and 256 neurons, respectively. We then used a concatenation layer for the feature fusion and added 4 densely connected layers with 512, 256, 128 and 6 neurons, respectively. The hidden layers used ReLU and the output layer sigmoid activation for the final regression.

This architecture was designed to explore the use of LSTM layers in our study to explore a different modality than the ones used before. We aimed to use only audio data from the videos. Figure 6 shows the topology of this neural network.

Table 1 shows the accuracy results of the three models used (called TNMCUL1, TNMCUL2, and TNMCUL3) and its comparison against state-of-the-art approaches, included in the publications of the best participants in the apparent personality recognition contests based on First Impressions of ChaLearn [11,13]. TNMCUL2 and TNMCUL1 obtained an accuracy of 0.942215 and 0.936158 respectively, slightly surpassing the other models. TNMCUL3 obtained an accuracy of 0.864853, below the rest of the models.

Table 1: Comparison between our models and other state-of-the-art approaches (prepared with our own data and results in [13]).

Name	Technique	Accuracy
TNMCUL2	**CNN Continuos**	**0.942215**
TNMCUL1	**CNN Discrete**	**0.936158**
NJU-LAMDA	Deep Multi-Modal Regression	0.912968
evolgen	Multi-modal LSTM Neural Network with Randomized Training	0.912063
DCC	Multi-modal Deep ResNet 2D kernels	0.910933
Ucas	AlexNET, VGG, ResNet with HOG3D, LBP-TOP	0.909824
BU-NKU	Deep feature extraction with regularized regression and feature level fusion	0.909387
Pandora	Multi-modal deep feature extraction single frame and late fusion	0.906275
Pilab	Speech features 1000 forest random trees regression	0.893602
Kaizoku	Multi-modal parallel CNN	0.882571
TNMCUL3	**LSTM**	**0.864853**

3.3. Standardized Personality Test

For the standardized test we used a 50-item IPIP representation of the markers mentioned by Goldberg for the factorial structure of the Big-Five model [3]. Each of the five personality traits is evaluated by means of 10 items, which in turn are rated by the participant on a 5-element Likert scale (strongly disagree; partially disagree; neither in agreement, nor in disagreement; partially agree and fully agree) based on participant level of agreement or disagreement with respect to each statement displayed. Each option has a value of 1 to 5 points, so 50 is the maximum score per trait. In the end, we convert the score obtained to a value between 0 and 1. This information is stored in the cloud repository and registered to which user it belongs. These values are used to compare them against the results of automatic recognizers.

3.4. Workflow for Data Collection

Figure 7 shows the workflow used for data collection: as a first step, the participant must register on the platform and log in. Once inside the platform, the participant must answer the standardized 50-item IPIP test. Next, the participant must record videos with an approximate duration of one minute each.

The platform stores the recorded videos in our cloud repository. Subsequently, the collected videos are processed and evaluated using the automatic recognizers and the generated information is stored in the cloud repository, linking the corresponding data to each user.
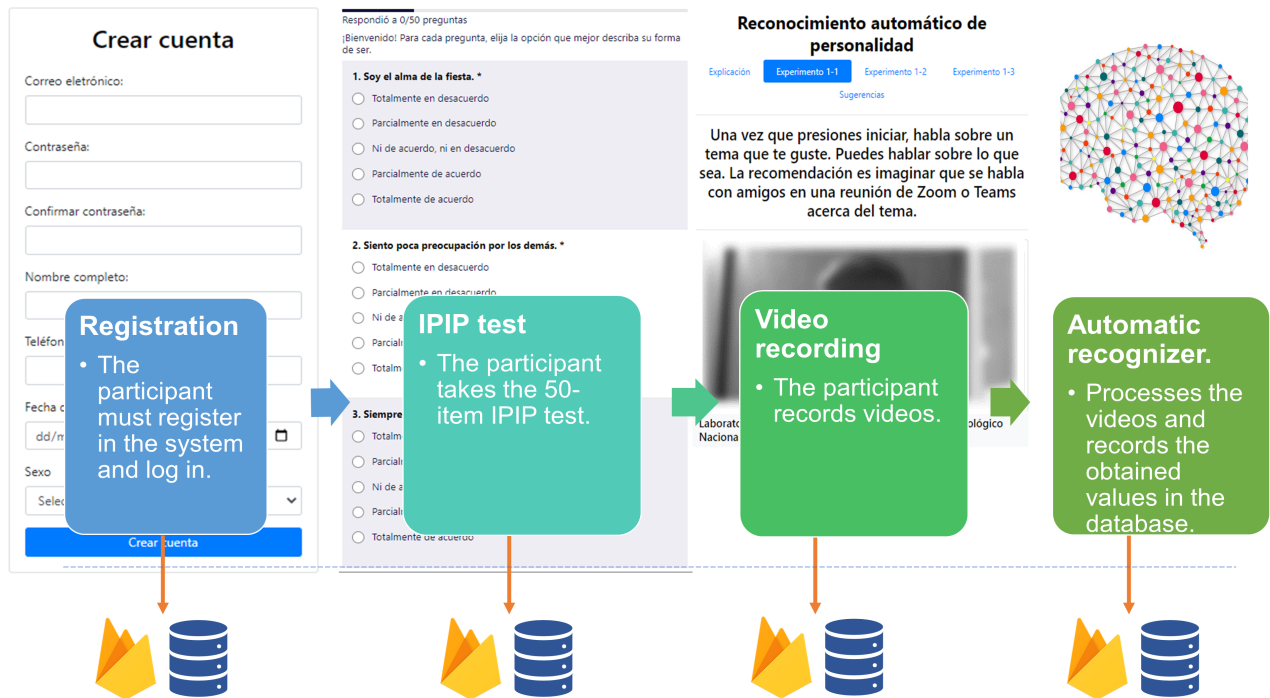
Figure 7: Workflow for data collection.

4. Results and Discussion

In this section we present the initial experiment, the tests, and the results obtained.

4.1. Data Collection Details

Each participant was invited to answer the IPIP test and then record, for one minute, a video where they were asked to speak freely about any topic. Each video is used to extract the images and audios that feed the automatic recognizers and the results are stored directly in the database.

Table 2: Descriptive Statistics of the IPIP Tests.

Trait	Participants	Mean	Standard deviation
Openness	32	0.7344	0.1450
Conscientiousness	32	0.6844	0.1568
Extraversion	32	0.5969	0.1750
Agreeableness	32	0.7906	0.1376
Neuroticism	32	0.5656	0.2598

4.2. Intervention Results

Thirty-two individuals participated in the intervention with the IPIP test, of whom 15 were male and 17 were female. All participants ranged from 23 to 44 years of age. Table 2 shows the descriptive statistics of the data collected for each of the personality traits. It can be observed that the personality traits with the highest mean value were agreeableness with a mean of 0.7906 and openness with a mean of 0.7344. Both traits also presented the least variation with standard deviations of 0.1376 and 0.1450, respectively. The factor with the lowest mean value was neuroticism.

For the evaluation of the selected automatic recognizers of apparent personality, 84 videos were collected. The videos were the product of the intervention of 21 participants (11 of the original participants did not record a video), of which 13 are male and 8 are female. The age range of the participants is between 23 and 40 years old. Table 3 shows the mean absolute error (MAE) values obtained by comparing each value of the personality traits predicted by the automatic apparent personality recognizers against the corresponding value for the participant based on the IPIP test.

Analyzing the results, it was possible to detect that the mean absolute error (MAE) was lower in the extraversion factor and higher in agreeableness. However, in all personality traits the value is too high, so it is not possible to consider that the automatic recognition models evaluated have made an adequate prediction. An interesting aspect is that TNMCUL3 scores better in 4 of the 5 personality traits. TNMCUL2 scores better in Extraversion.

Table 3: Mean Absolute Error (MAE) of each Personality Trait.

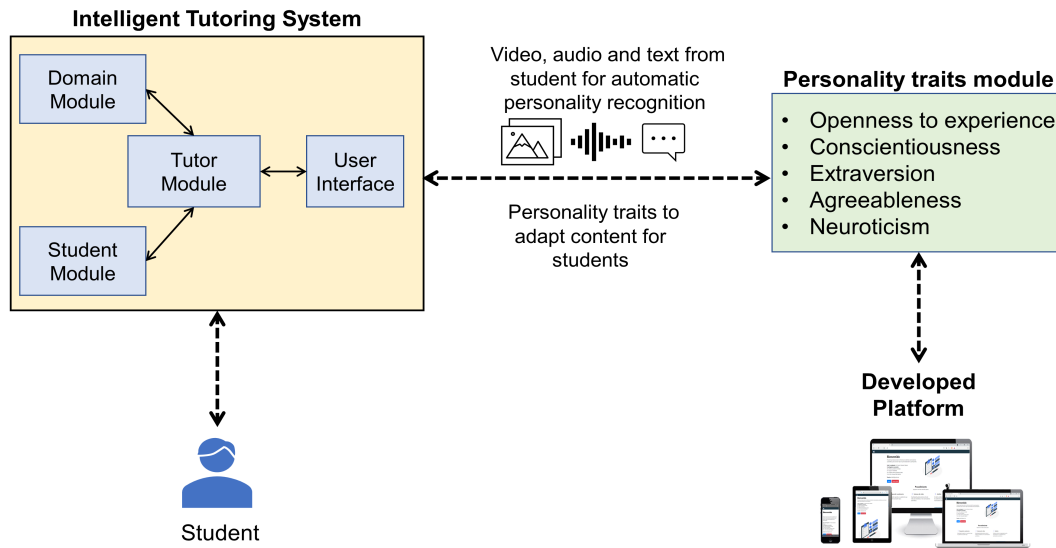| Model	Videos	Technique	Openness	Conscientiousness	Extraversion	Agreeableness	Neuroticism
TNMCUL1	84	CNN Discrete	0.2683	0.2684	0.1941	0.3698	0.2806
TNMCUL2	84	CNN Continuous	0.2483	0.2200	0.1786	0.2200	0.2427
TNMCUL3	84	LSTM	0.2262	0.2150	0.2087	0.2150	0.2426

Figure 8: Platform integration with Learning and Tutoring Systems.

4.3. Personality Recognition for Intelligent Learning Environments

Our proposal is to use the information on personality traits and videos collected with the help of the developed platform to evaluate and optimize automatic personality recognition models that can be integrated into intelligent learning environments. The use of an automatic personality recognition model in an intelligent learning environment or tutoring system would allow exploring the idea of presenting adaptive content in real time to the learner based on their dominant personality traits with the goal of achieving the greatest possible impact on learners during their cognitive process.

In Figure 8 we show the proposal to combine an intelligent tutoring system with a personality traits module that makes use of the bank of automatic personality recognizers optimized with our platform.

The learning or tutoring systems communicate with the personality traits module and send it video, image, audio, or text information of the learner which will be used as input to the automatic recognizers. The personality traits module returns as output the presence or absence of the student's Big-Five personality traits. This information can be used by the intelligent tutoring system to make decisions about the content presented to the student.

5. Conclusions and Future Work

The developed platform allows quite a simple and applicable data collection through any device with Internet access from any location and supports the immediate availability of the collected data for analysis.

We have added as a secondary contribution, the evaluation of three automatic recognition models to review the functionality of the platform. In this first exercise, we have found that the evaluated recognizers present a gap in the results with respect to the IPIP test.

The construction of a dataset of Spanish language videos and personality test results is also considered a relevant contribution that can serve as a starting point for future studies.

Additionally, we presented a proposal to use our platform for improving automatic recognizers that could be integrated into tools such as intelligent learning environments or tutoring systems to personalize instruction and feedback based on the detected personality of the participant.

As future work, it is proposed to continue the improvement of the assessed recognizers using the collected dataset and the results of the IPIP tests and hyperparameter optimization techniques. It is suggested to contemplate the evaluation of automatic recognizers of apparent personality based on text to corroborate if the results are like those of the standardized test and, failing that, to work on the retraining of these models, taking advantage of the dataset that is being formed with videos in Spanish.

Another approach that can be addressed is the use of classification algorithms to determine the presence or absence of each personality trait.

Conflict of Interest

The authors declare no conflict of interest.

Acknowledgment

References

[1] V.M. Bátiz Beltrán, R. Zatarain Cabada, M.L. Barrón Estrada, H.M. Cárdenas López, H.J. Escalante, "A multiplatform application for automatic recognition of personality traits for Learning Environments," in 2022 International Conference on Advanced Learning Technologies (ICALT), 49–50, 2022, doi:10.1109/ICALT55010.2022.00022.

[2] J.C.S. Jacques Junior, Y. Gucluturk, M. Perez, U. Guclu, C. Andujar, X. Baro, H.J. Escalante, I. Guyon, M.A.J. van Gerven, R. van Lier, S. Escalera, "First Impressions: A Survey on Vision-Based Apparent Personality Trait Analysis," IEEE Transactions on Affective Computing, **13**(1), 75–95, 2022, doi:10.1109/TAFFC.2019.2930058.

[3] L.R. Goldberg, "The development of markers for the Big-Five factor structure.," Psychological Assessment, **4**(1), 26–42, 1992, doi:10.1037/1040-3590.4.1.26.

[4] L.R. Goldberg, "A broad-bandwidth, public domain, personality inventory measuring the lower-level facets of several five-factor models," Personality Psychology in Europe, **7**(1), 7–28, 1999.

[5] L.R. Goldberg, J.A. Johnson, H.W. Eber, R. Hogan, M.C. Ashton, C.R. Cloninger, H.G. Gough, "The international personality item pool and the future of public-domain personality measures," Journal of Research in Personality, **40**(1), 84–96, 2006, doi:10.1016/j.jrp.2005.08.007.

[6] O. Laverdière, D. Gamache, A.J.S. Morin, L. Diguer, "French adaptation of the Mini-IPIP: A short measure of the Big Five," European Review of Applied Psychology, **70**(3), 100512, 2020, doi:10.1016/J.ERAP.2019.100512.

[7] M. Gross, M. Cupani, "Adaptation of the 100 IPIP items measuring the big five factors," Revista Mexicana de Psicología, **33**, 17–29, 2016.

[8] G. Farnadi, G. Sitaraman, S. Sushmita, F. Celli, M. Kosinski, D. Stillwell, S. Davalos, M.-F. Moens, M. de Cock, "Computational personality recognition in social media," User Modeling and User-Adapted Interaction, **26**(2–3), 109–142, 2016, doi:10.1007/s11257-016-9171-0.

[9] J. Yu, K. Markov, A. Karpov, Speaking Style Based Apparent Personality Recognition, 540–548, 2019, doi:10.1007/978-3-030-26061-3_55.

[10] C. Ventura, D. Masip, A. Lapedriza, "Interpreting cnn models for apparent personality trait regression," in Proceedings of the IEEE conference on computer vision and pattern recognition workshops, 55–63, 2017.

[11] V. Ponce-López, B. Chen, M. Oliu, C. Corneanu, A. Clapés, I. Guyon, X. Baró, H.J. Escalante, S. Escalera, ChaLearn LAP 2016: First Round Challenge on First Impressions - Dataset and Results, Springer International Publishing, Cham: 400–418, 2016, doi:10.1007/978-3-319-49409-8_32.

[12] H.J. Escalante, H. Kaya, A.A. Salah, S. Escalera, Y. Gucluturk, U. Guclu, X. Baro, I. Guyon, J.C.S.J. Junior, M. Madadi, S. Ayache, E. Viegas, F. Gurpnar, A.S. Wicaksana, C.C.S. Liem, M.A.J. van Gerven, R. van Lier, "Modeling, Recognizing, and Explaining Apparent Personality From Videos," IEEE Transactions on Affective Computing, **13**(2), 894–911, 2022, doi:10.1109/TAFFC.2020.2973984.

[13] A. Subramaniam, V. Patel, A. Mishra, P. Balasubramanian, A. Mittal, Bi-modal First Impressions Recognition Using Temporally Ordered Deep Audio and Stochastic Visual Features, Springer International Publishing, Cham: 337–348, 2016, doi:10.1007/978-3-319-49409-8_27.