International Journal of Combinatorial Optimization Problems and Informatics, 16(4), Sep-Dec 2025, 319-334. ISSN: 2007-1558.
https://doi.org/10.61467/2007.1558.2025.v16i4.1018

www.editada.org

SensorApp: A Move Speaks Louder Than Words: Personality recognition using a dataset obtained from sensors and mobile device usage habits

Néstor Leyva López[1], *Ramón Zatarain Cabada*[1], *Hugo Jair Escalante Balderas*[2], *María Lucía Barrón Estrada*[1].
[1] TecNM Instituto Tecnológico de Culiacán, México.
[2] Instituto Nacional de Astrofísica, Óptica y Electrónica, México.
nestor.ll@culiacan.tecnm.mx, ramon.zc@culiacan.tecnm.mx, hugojair@inaoep.mx,
lucia.be@culiacan.tecnm.mx

Abstract. Personality encompasses a range of traits that define individual emotions, behaviors, and patterns of interaction. Given that these behavioral patterns likely influence how people engage with mobile devices, which are deeply integrated into daily life, this study explores the feasibility of assessing personality attributes through data collected from mobile sensors, specifically accelerometers. We present the methodology for data acquisition and preprocessing, along with the implementation of artificial neural networks and a multilayer perceptron for automatic personality prediction. Preliminary results indicate promising potential for inferring personality traits based on mobile usage data, highlighting accelerometer-derived features as valuable predictors in the absence of traditional surveys.

Keywords: Mobile, Sensors, Personality Recognition, ANN, MLP, Dataset.

Article Info
Received February 25, 2025
Accepted July 6, 2025

1 Introduction

Personality refers to the long-lasting traits and behavior that make up a person's unique adaptation. This includes major personality traits, interests, impulses, beliefs, self-concepts, abilities, emotions, and behavioral patterns (Gary, 2015). Studying personality can help in different areas. For example, in the workplace it can identify candidates with personality profiles that fit the requirements of specific positions. In education it can be used to improve learning by allowing teachers to adapt teaching methods and materials depending on the different personalities of students. In marketing and sales, it assists in the creation of products and services that fit target consumers. In the area of criminology, it helps professionals to assess risks of recurrence of criminal activity of different individuals (Picca, et al., 2023).

Therefore, personality recognition becomes an extremely important task for the development of different areas. The study of personality recognition consists of inferring the personalities of different users by accounting for users' self-reported data. Since the usual approaches tend to use precisely self-reported methods, these may present information biases (Rendon-Macías, et al., 2020), since users tend to respond on their self-concept (Concha, et al., 2012), making the values not reliable at any given time.

Therefore, the focus of this paper is to explore how the information obtained from usage habits of mobile devices like tablets and smartphones, can be used to recognize the personality of users, considering both passively obtained data from mobile device sensors and self-reported data from mobile device usage. For this, a methodology for data collection through mobile devices was designed (Zatarain, et al., 2024). These data consist of measurements provided by mobile accelerometer and gyroscope sensors, as well as usage data reported by the user. Additionally, the user has the possibility to record an audio to achieve a better reference framework and correlation between her/his personality and how this person uses her/his cell phone. The model used for personality characterization is the Big-Five model also known as OCEAN personality model (Wiggins, 1996), which measures five different personality traits in values between 0 and 1. Currently, these five personality traits can be calculated using the International Personality Item Pool (IPIP) test answering self-reported questionnaires.

This paper presents an approach to profile personality prediction based on the user's personality reported in the OCEAN model, by considering the information of mobile sensors instead of IPIP test. For this, we specifically propose and evaluate different neural network architectures that leverage users' mobile usage patterns to estimate personality traits. To demonstrate the feasibility of this approach, we focus on predicting International Personality Item Pool (IPIP) test scores. In this study, we focus on sensor-based input, specifically using only accelerometer data for model training and evaluation.

This article is organized as follows: section 2 presents a summary of related works on the subject. Section 3 describes both the personality model, and the instrument employed to measure personality attributes. Section 4 briefly elaborates the complete structure of the system, as well as how the data is stored. On section 5 we present the steps and considerations to collect the data to generate the dataset. Section 6 explains how the data was prepared to be used in artificial intelligence algorithms. Then, section 7 discusses the machine learning and deep learning models selected and the reasons for their choice, as well as explaining their structures in a more detailed manner. Section 8 presents the different clusters and predictions obtained from the application of the machine learning networks with the proposed pre-processing in previous sections, and finally we present Conclusions and Future work in section 9.

2 Related Work

For this work, we investigated several previous studies related to the scope of this paper, such as the ones on automatic personality recognition or APR, encompassing both classification and prediction approaches. The field of APR has evolved significantly over time. Many studies have explored various machine learning and deep learning techniques for feature extraction, as well as the use of different types of data. One such study is presented by Talaat et al. (2023), who had proposed a system that improves both personality and emotion recognition using text data gathered from social networks. The system comprises four modules: data acquisition, preprocessing, personality and emotion recognition. Feature selection and hyperparameter tuning are optimized using the Gray Wolf Optimization (GWO) algorithm using random forest and decision trees. The results showed that the model achieved approximately 99% accuracy in personality detection using text from users in twitter, outperforming other contemporary methods. On the other hand, the research done by Fu et al. (2024) introduced a method that incorporates data interpolation with heterogeneous conversational graph neural networks (HC-GNNs). Their approach focuses on capturing the interdependencies between speakers within dialogues, and although the authors did not report specific numerical results, they indicate that greater speaker diversity significantly improves personality recognition, leading to performance improvements over reference models, such as Multi-Layer Perceptron and Relational Graph Convolutional Networks. In a different modality, Arfaoui et al. (2025) explored the use of Natural Language Processing (NLP) and Deep Learning (DL) techniques to detect and extract personality traits from text. They employed deep learning architectures such as Convolutional Neural Networks (CNNs), Long Short-Term Memory (LSTM) networks, and BERT transformers in conjunction with Big-Five personality data. After applying standard preprocessing and data augmentation techniques to a Big-Five Essays dataset, the study evaluated multiple deep learning pipelines, revealing that combinations like BERT-CNN and BERT-LSTM achieved average accuracies of 50.6% and 48.5%, respectively across the five personality dimensions. Although moderate performance on accuracy, these results demonstrate the feasibility of integrating deep learning models for personality trait extraction. In a related study, Guenzel et al. (2025) focused on the extraction of personality traits from facial images sourced from the LinkedIn social network. Unlike text-based methods, this study used CNNs to extract visual features, enabling the prediction of Big-Five personality traits in a way that is comparable to traditional surveys. This approach has the added advantage of being less susceptible to manipulation, making it particularly applicable in corporate recruitment settings. However, this study does not report any specific numerical results or evaluation metrics, limiting the assessment of its predictive performance.

Studies have also been conducted on the collection of data from mobile devices with the aim of using this data for psychological and behavioral analysis. One of the earliest works in this field was conducted by Ware et al. (2020), who investigated whether smartphone data could be used to predict symptoms of depression. Location data and phone usage time were collected from 182 undergraduate students; they analyzed this information using unspecified machine learning models. The study successfully predicted cognitive and behavioral symptoms, achieving an F1-score of 0.86. This demonstrated the feasibility of continuous mental health monitoring using passive data collection. Following this, Moshe et al. (2021) extended the scope of data sources by incorporating wearable devices to predict symptoms of both anxiety and depression. The study involved 60 participants and collected data on GPS location, physical activity, and sleep analyzed through multilevel modeling statistical models. Results indicated that variations in location and sleep patterns significantly contributed to the prediction of mental health symptoms, reinforcing the potential of mobile sensing for early detection of psychological conditions. On the other hand, Wampfler et al. (2022) introduced a novel approach to affective state classification using data gathered from smartphones in naturalistic settings. They extracted color maps from tactile events and motion sensors and fed this data into a CNN. The study showed that affective states could be accurately classified using sensor data, and notably emphasized the method's lower invasiveness, making it more suitable for real-world applications in privacy-sensitive environments. However, the reported classification performance was moderate, with Area Under the Curve (AUC) values below 0.85 and classification accuracy under 70%, suggesting room for improvement in model precision and generalizability. Most recently, Xenakis et al. (2023) proposed a system that passively and anonymously collects data from smartphone sensors to identify personality traits using the HEXACO personality model which is based on the OCEAN model For this study, users installed an app on Android or iOS devices and over a two-week period, the app gather data from the mobile sensors (accelerometer, gyroscope, apps usage, etc.). The study found correlations between smartphone usage patterns and the HEXACO personality dimensions, although the authors did not report numerical results, they suggest that such sensor data could be used effectively for personality detection.

Considering the subject of recognizing personality, emotions, and human activity using data obtained from cell phones, we found that recent research has increasingly focused on using mobile sensor data to predict personality traits, emotional states, and human activities. Bera et al. (2023) proposed a real-time user activity recognition model based on Recurrent Convolutional Neural Networks (RCNN), leveraging accelerometer and gyroscope data from smartphones. Their approach achieved over 90% accuracy in classifying activities such as walking and running, demonstrating the effectiveness of deep learning methods in interpreting motion-based data. Expanding the focus to personality prediction, Sze et al. (2024) explored the use of smartphone sensor data to infer traits based on the OCEAN personality model. Activity data was collected from 144 participants, and Random Forest and XGBoost algorithms were used to perform binary and multimodal classification. Results showed promising performance, with binary F1 scores of 0.638 and 0.622 for Random Forest and XGBoost respectively, indicating that personality traits can be reliably inferred without requiring traditional surveys. Similarly, Webb et al. (2024) focused on predicting negative affective states in individuals with mental illness by passively collecting data from wearable devices and smartphones. Their study found that GPS data were particularly predictive, and a combination of logistic regression and SVM outperformed other models reaching an AUC of 0.79. This work emphasized the importance of tailored modeling approaches to improve predictive accuracy for mental health applications. Finally, Yang et al. (2024) proposed a multimodal deep learning framework using a Transformer Encoder and DistilBERT model. Their system integrates mobile sensor data, such as sleep patterns and physical activity, with self-reported emotional states collected via smartwatches. The DistilBERT Transformer demonstrated over 80% accuracy in forecasting both positive and negative affective states one week in advance, highlighting the value of personalization in affective prediction.

Compared to previous studies, what sets our work apart is the use of both data collected from mobile device sensors, and self-reported data on mobile phone use. In addition, we have the personality attribute values from the Big-Five model for each user that was collected by the IPIP test. This data is used to train models that find a correlation between mobile phone use and user personality, allowing us to predict users' personality attributes using only data on their usage habits. The aim is to remove the subjectivity of traditional methods like questionaries and predict personality in a fast, in-real-time manner.

3 Big-Five personality model

The Big-Five personality model, also known as the Five Factor Model (FFM), is one of the most widely accepted frameworks in recent psychology for describing and assessing human personality (Goldberg, 1993). It proposes that personality can be understood through five general and relatively stable personality attributes. These are *Openness*, *Conscientiousness*, *Extraversion*, *Agreeableness*, and *Neuroticism*, which are often abbreviated as OCEAN (from the initial letter of each factor). Each of these traits is quantified as a continuous real-valued score within a defined range, in this particular case, any value ranging between 0 and 1, which reflects the degree to which that trait is present in an individual. As a result, predicting personality traits in this context constitutes a regression task, where models aim to estimate continuous trait values rather than classify discrete categories. The strength of this model lies in its scientific basis and multicultural validity, as it has been systematically reproduced in diverse population groups and has been associated with various social aspects, such as academic success and mental health (Mezquita, 2019). For this reason, it remains a fundamental tool in both research and practice in areas such as psychological assessment and automatic personality recognition (APR).

3.1 IPIP Test

One of the most influential initiatives in the field of personality assessment is the development of the International Personality Item Pool, better known as IPIP. This is a widely available set of items for measuring a broad range of personality traits, commonly using the OCEAN model (Goldberg, 2006). This is achieved by measuring personality attributes using self-reported questionnaires consisting of questions that people answer based on how they would describe themselves on a Likert scale. The IPIP-50 test (Goldberg, 1992) uses 50 items, where each 10 items belong to one specific personality trait. These are rated by the participant on a 5-element Likert scale. Each option has a value of 1 to 5 points, so 50 is the maximum score per trait. At the end, the value obtained is transformed to a value between 0 and 1 for each personality trait (*Openness*, *Neuroticism*, etc.). The IPIP is a robust and valid method for quantitatively representing personality. This is shown through values ranging from 0 (personality trait not present) to 1 (total presence of a personality trait). In the field of APR, assessments conducted using the IPIP framework serve as a standardized basis for training and validating machine learning algorithms.

4 Architecture

In this section we describe the software architecture developed for the SensorApp system. The system uses a layered client-server model for retrieving sensor information in both mobile and web modes. The SensorApp architecture (Fig. 1) contains in the presentation layer both the web application and the mobile application. On the server side it contains an application logic layer and a data layer. This logical layer can be separated into web application and mobile application.

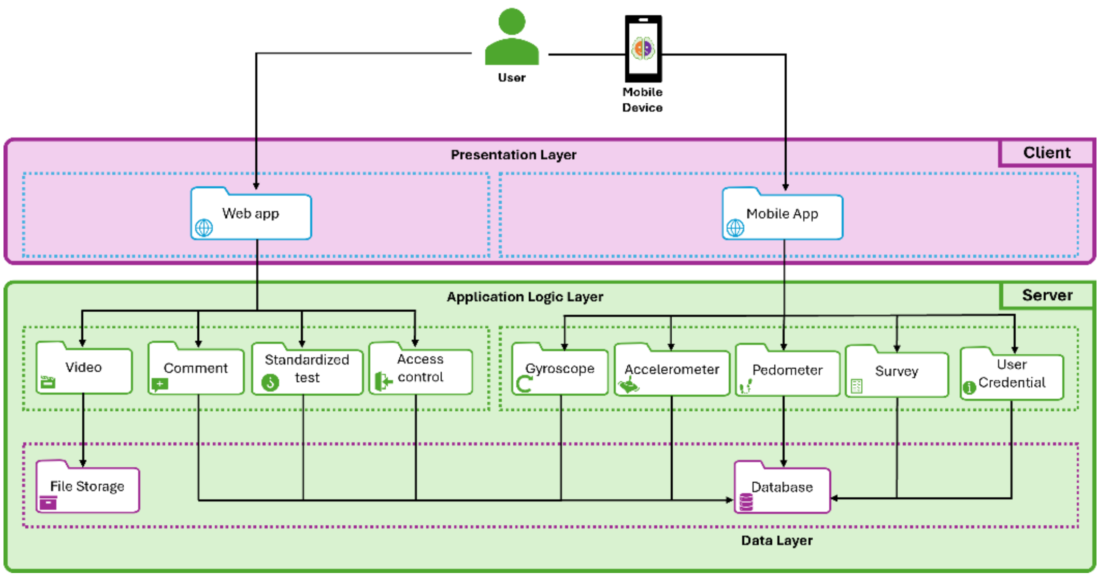

Fig. 1. SensorApp Architecture.

In the logic layer of the mobile application, we can find separate components for data collection from sensors (gyroscope, accelerometer and microphone). As well as a questionnaire that users must answer, the questions in this one are related to the usage habits of the users. While in the logical layer of the web application, we can find the following components: i) Access control: Keeps history of user login, as well as user authentication; ii) Comment: Captures text entered by the user; iii) Standardized test: It uses the IPIP-50 questionnaire to characterize the user's personality with the OCEAN model frame of reference; iv) Video: Record video of users answering some additional auxiliary questions for personality attribute extraction.

After the logical application layer, there is the data layer shared by both parts of the logical application layer, consisting of a non-relational database mounted on firestore, which is a service offered by Firebase. It stores data in different collections (i.e. Users, accel, gyro, survey, etc.).

5 Data Collection methodology

To obtain the required data to train the neural networks, a set of processes, techniques and tools were used to obtain the information in an organized manner. The methodology for data collection (Fig. 2) consists of the user presenting the IPIP-50 test in SensorApp Web, obtaining and extracting the OCEAN personality data. During the application frame of this test, both video and audio are also recorded, in order to obtain additional information to better typify the personality of the individual. All this data is sent to a database (DB) in the cloud via Google Firebase services. Once registered on the web platform, the test subject uses SensorApp Mobile, in which they take the daily survey for a period of 2 weeks. This consists of 6 questions related to their mobile device usage habits. During the survey, data corresponding to different sensors of the device (gyroscope, accelerometer and microphone) are recorded.

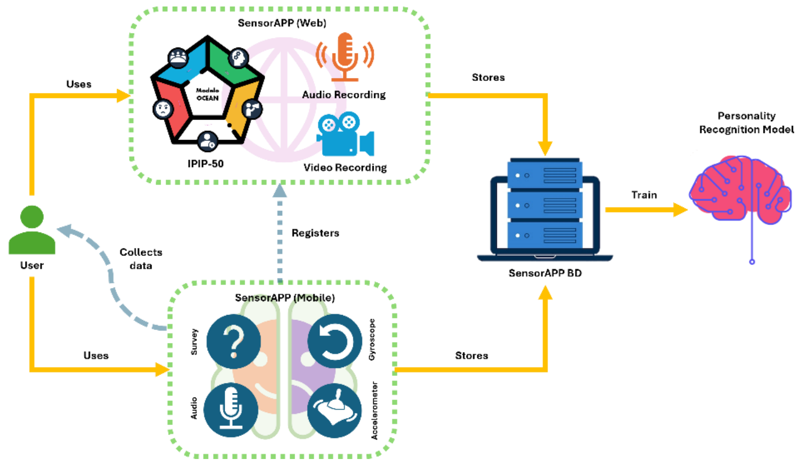

Fig. 2. SensorApp data collection methodology.

5.1 Web Platform

SensorApp Web (SensorApp web, n.d.) consists of an online platform where the user has to register to authenticate his data (Fig. 3). The user takes the standardized IPIP-50 test to recognize his or her personality attributes on the OCEAN scale. This allows obtaining reference data of the user's personality to compare with the data obtained from the mobile application and create a correlation between them.

Fig. 3. SensorApp web main interface.

5.2 Mobile App

SensorApp Mobile (SensorApp Mobile, n.d.) consists of a mobile application for android devices that collects information from the accelerometer and gyroscope of the device on which it is installed (Fig. 4). In addition to this, the application includes a questionnaire about the user's mobile device usage habits. Furthermore, it has the possibility of recording an audio at the end of the questionnaire where the user is invited to talk about a significant experience during the day. It should be noted that the sensor data is only collected during the time the user is answering the questions, once the user finishes answering, the application stops collecting data from the sensors. The purpose of this is to obtain significant samples from the sensors, as it ensures that the data collection happens when the user is actively using its device, allowing us to gather significant moving patterns from their hands while using the phone. This approach also helps prevent users from feeling that their privacy has been compromised, as data is only gathered when they choose to use SensorApp.

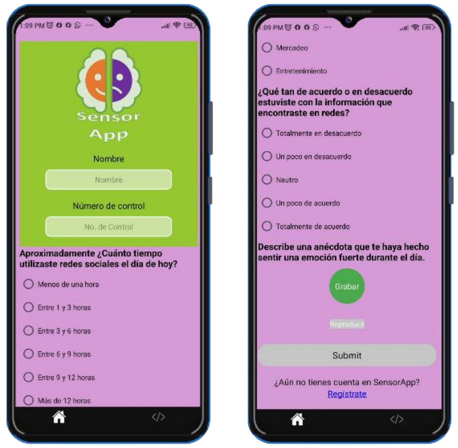

Fig. 4. SensorApp Mobile app interface to gather student's information.

5.3 Questionnaire Design

A questionnaire is defined as a collection of questions used to obtain quantitative and/or qualitative data from users about their attitudes, experiences and/or opinions. The form to be completed by participants contains questions based on previous research (Gil de Zuñiga, et al., 2017; Stachl, et al., 2017; Stade, et al., 2019; Valanarasu, 2021; De Choudhury, et al., 2012) focused on identifying key aspects of personality. This ensures that the questionnaire meets the standard of obtaining necessary and meaningful data for automatic personality recognition. This instrument addresses self-reported aspects related to cell phone usage habits, including social network management. These questions are answered as a form, with multiple-choice questions. Consisting of the following questions: 1. Approximately how much time have you used social networks today?; 2. What type of applications have you used the most during the last 24 hours?; 3.Which social network have you used the most during the last 24 hours?; 4. What emotions did you feel while using social networks?; 5. What content did you mainly consume while browsing social networks?; 6. How much did you agree or disagree with the information you found on social networks?

5.4 Sample group description

To begin the data collection process, three groups of undergraduate students from different semesters and different careers, specifically computer systems engineering and mechatronics engineering were selected from the Tecnológico Nacional de México: Campus Culiacán. Prior to the beginning of the data collection, the project was presented to the participants of each group. The research goal and their role as participants were explained in detail, and it was also made clear that the data collected would be anonymized. The participation of the individuals was completely voluntary. The two digital tools mentioned in this section were provided for data collection. The volunteers were told that before starting the tests on the mobile app they should take the standardized IPIP-50 test, which is performed on the SensorApp web platform. After this, they were instructed to use the SensorApp mobile application, where they should answer the questionnaire that's consists of 6 questions and the possibility of recording an audio about a significant event in their day. This mobile application must be used for a period of 15 days to obtain data.

5.5 Data Formatting

The data collected and its format depends greatly on the nature of the database used, as Firebase provides us with the structure of a non-relational database and the data is presented in JSON format. These can be configured in such a way that the data can be accessed as if they were collections, making data access easier and more traceable.

6 Data Processing

Data processing is commonly defined as a series of operations performed on a set of data to extract, transform and analyze them. For this work we will detail below what was done to be able to use the accelerometer data for the prediction of OCEAN values.

6.1 Extraction of accelerometer data and IPIP values

To treat the different values, only users who have both types, accelerometer and IPIP data are considered. To collect the data in the best way, the first thing is to collect the coordinates of the *X*, *Y* and *Z* axes (previously separated in individual sessions as shown in Fig. 5) and save them in a list of accelerometer data. Simultaneously, a list of Big-Five IPIP values is generated for each user.

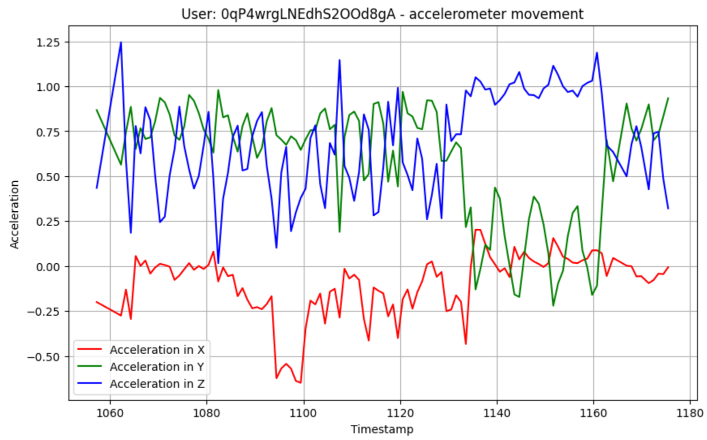

Fig. 5. Example of an individual session of accelerometer motion.

6.2 Conversion to DataFrame

The list of accelerometer and personality trait data is then converted to DataFrame type structure, which consists of a table-like format with rows and columns, where each column represents a specific feature (e.g., accelerometer, gyroscope, neuroticism trait), and each row corresponds to an individual data entry. Both the accelerometer coordinates data (X, Y, and Z) and the IPIP personality traits are assigned to separate columns for each one of the sessions.

6.3 Feature normalization

Subsequently the accelerometer data is normalized. This fits the features so that the data arrays have a similar scale (Fig. 6). In this case, a normalization process was applied that removed the mean and standardized the dispersion, this helps filter out outlier data that may introduce atypical information when training the model. As a result, a new dataset was generated with the accelerometer coordinates properly normalized. This results in a new dataset where the accelerometer coordinates are normalized.

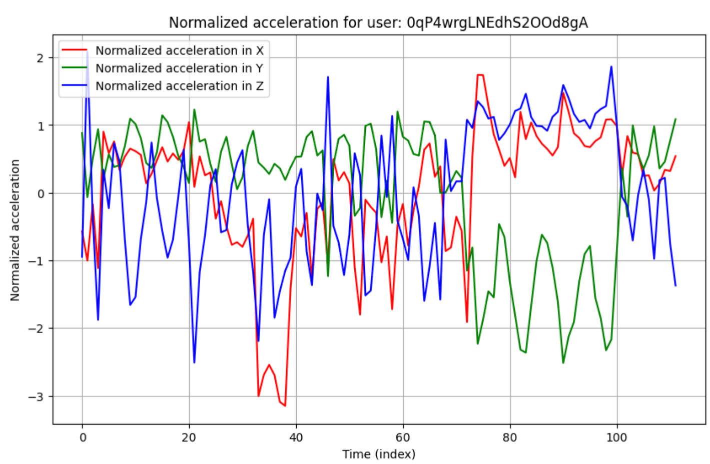

Fig. 6. Example of a standardized acceleration session.

6.4 Data splitting

Then the data is divided into training and test set. The normalized dataset and personality traits are divided into 4 sets. The size of the test set is set with 20% (32 samples) of the data. The remaining 80% (128 samples) is used for training. It's worth mentioning that Cross-validation was also considered as an alternative evaluation strategy. However, since the difference in average accuracy and F1-score between cross-validation and the 80/20 split was minimal and cross-validation is computationally expensive, the 80/20 split was chosen, as shown Table 1.

Table 1. Comparison between 80/20 split and Cross-Validation results

Evaluation method	Accuracy	F1-Score
80/20 split	0.5422	0.9108
Cross validation	0.5269	0.9031

In this way the data is ready to be used in IPIP prediction tests from values extracted from the accelerometer of a mobile device (fig. 7).

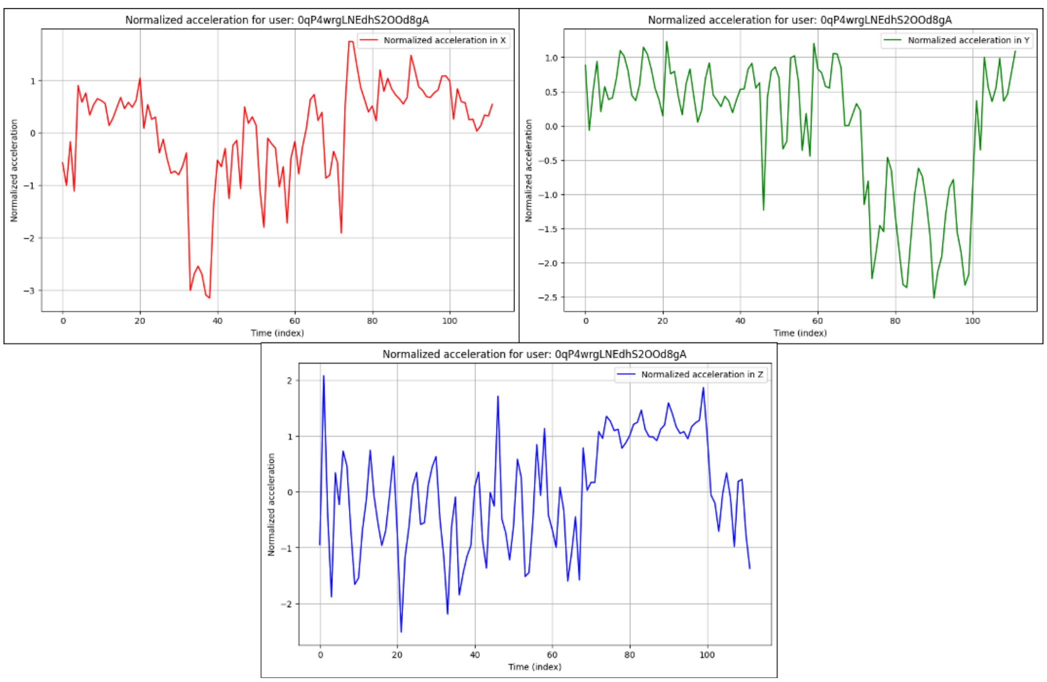

Fig. 7. Example of a standardized acceleration session.

7 Machine Learning Models

To perform the prediction of IPIP values using data gathered from the sensors of the mobile device, two different networks were chosen and trained to have comparison values between them.

7.1 Dataset Description

The used dataset is structured in a dataframe format, containing columns with the following data: a unique user identifier, the values corresponding to each personality trait included in the OCEAN model, which were obtained from the IPIP-50 test. It also contains the responses associated with each of the questions in the SensorApp Mobile questionnaire regarding usage habits and content consumption on mobile devices. The responses were numerically coded using one-hot encoding. Finally, it stores the movement readings on the three axes of the accelerometer for each usage session.

7.2 Fully Connected Feedforward Neural Network

One of the networks defined for the tests was a Feedforward Artificial Neural Network (ANN) (Incio-Flores, et al., 2023), which has a one-way flow of information from input to output. This network, illustrated in Figure 8, has an input layer with an input shape of 3 for the accelerometer axes data. As for the hidden layers, it includes an initial one with 64 neurons, and a second one with 32 neurons, both with ReLU activation function. For the output layer, 5 neurons were defined with a Sigmoid activation function, suggesting that the outputs are in the numerical range between 0 and 1, as it is the scale used for personality traits in OCEAN using the IPIP test as instrument. The loss function used is the mean square error. Unlike multi-label classification problem, this task is treated as a regression problem, in which each personality trait is predicted as a continuous value between 0 and 1. Therefore, the appropriate loss function is Mean Squared Error (MSE), which penalizes the squared difference between the predicted and actual values. Categorical cross-entropy is typically used in classification tasks and is not currently applicable to this study, as our focus is on predicting continuous personality trait scores rather than assigning them to discrete categories.

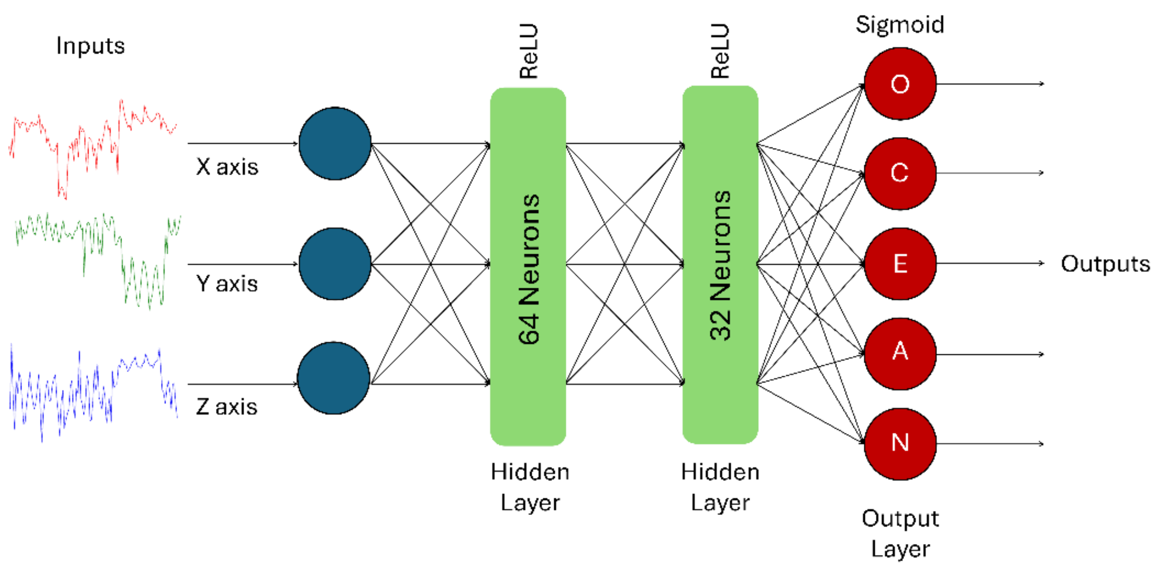

Fig. 8. Fully connected feedforward neural network for Prediction.

7.3 Multi-Layer Perceptron

As presented in Figure 9, the other selected neural network was a Multi-Layer Perceptron (MLP) (Desai & Shah, 2021), with a slightly different approach than the one proposed previously. The peculiarity with this approach is that in this case, independent neural networks were trained to predict each personality trait individually. Allowing the specialization of the prediction in each of the traits, thus seeking to optimize the performance of the model for each OCEAN trait, ranging values from 0 to 1.

This MLP network has an input layer that uses the Glorot Uniform Initialization (also known as Xavier Uniform), and the algorithm used for training is Adam Optimizer. This layer receives a three-dimensional array as input (accelerometer axes). As for the hidden layers, this one was proposed with two of them. The first one with 64 neurons, and the second one with 32 neurons, both with a ReLU activation function. As for the output layer, since these are individual networks for each of the personality traits, there is only one with a sigmoid activation function. The evaluation metric selected is the mean absolute error (MAE), measured individually for each independent network. Additionally, for complementary analysis, predicted values were discretized using a 0.5 threshold to compute classification metrics such as Accuracy and F1-Score.

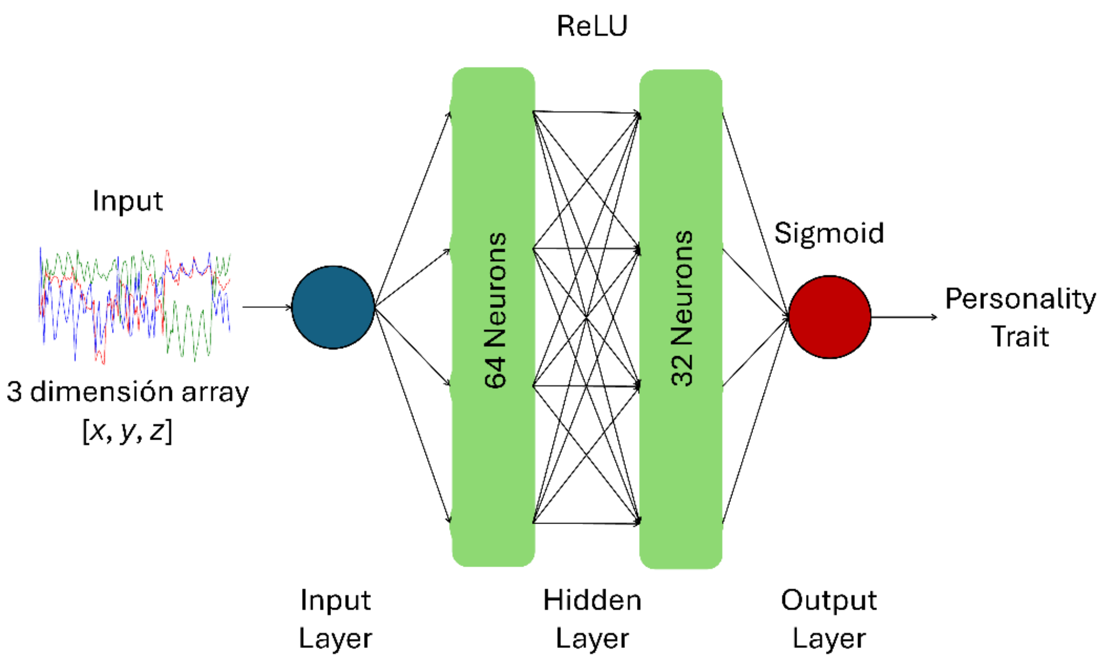

Fig. 9. Schematic of Multi-Layer Perceptron network for prediction.

8 Results

This section presents the relevant results achieved after carrying out the tests of what was previously proposed. The data collected, the tests performed on them, and their interpretations are detailed.

8.1 Collected Data

The data gathered after the users performed the tests in both SensorApp Mobile and Web are presented in Table 2. Of the 126 users who registered data on both platforms, only 16 users carried out the complete scheme, referring to the complete scheme as registering their IPIP-50 on the web platform, as well as entering data in the questionnaire of the mobile application. Those 16 users generated 160 records using the application in 10 days average per user. On the other hand, 110 users did only part of the scheme. 52 individuals did the IPIP-50 in SensorApp Web, without submitting the mobile application test. This was mainly due to 2 causes: some users did not feel comfortable performing a test in which usage data is recorded from their device, while other users do not have an Android device. A total of 58 individuals only used the mobile application to record their usage habits, generating 198 records.

Table 2. Data collected from the SensorApp application

Users	Subset	IPIP-50	Accelerometer	Gyroscope	Questionnaire	Total of Records
16	Complete Scheme (Web-Mobile)	✓	✓	✓	✓	160
52	SensorApp Web	✓				0
58	SensorApp Mobile		✓	✓	✓	198

After getting this data, a clustering analysis was performed (Fig. 10) to check the integrity of the data obtained from the accelerometer movements, where it can be seen the movements of all users that make up the dataset. The valid data (presented in blue) remain consistent in a range between coordinate values -3 and 3.5 on the *Y*-axis, and a range between coordinate values -2 and 3 on the *X*-axis. Except for some outliers (shown in red), as these outliers are usually due to times when the sensor reading can be affected by external variables (i.e. device drops, abrupt user movements, etc.).

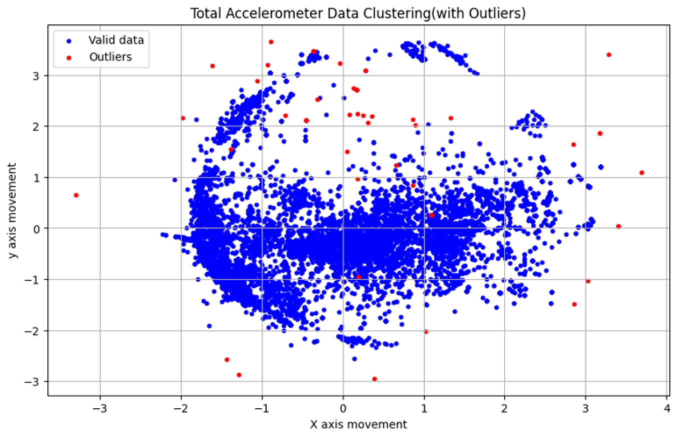

Fig. 10. Clustering of user's mobile movement data.

After clustering the data, it was observed that the users who had a more consistent interaction with the web application are those who present responsibility as the predominant personality trait, meaning it has the highest value. As shown in Fig. 11, the distribution of users' movements (where each color represents a different user) shows a certain tendency or similar movement patterns among those users, varying in range values on *Y*-axis between 1.25 and -1. While the values on the *X*-axis range in values between -1 and 1.5. Each of the colors in the graph represents the movements of a user with responsibility as the dominant trait attribute.

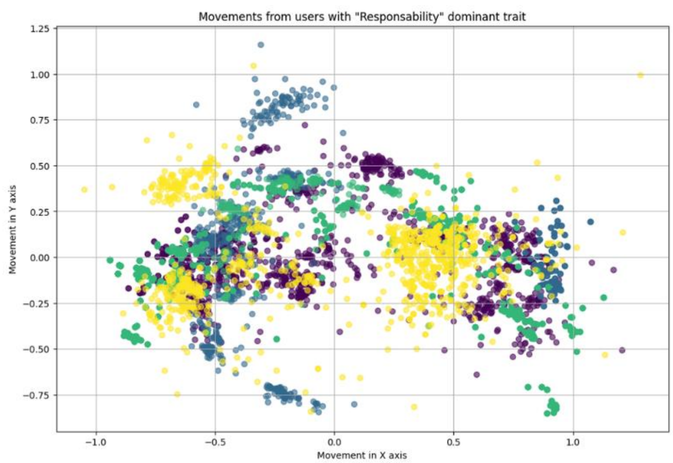

Fig. 11. Movement distribution cluster plot of users with the dominant personality trait of *Conscientiousness*.

8.2 Predictions with neural network models

This section presents the results obtained from the predictions made with the trained neural networks (ANN & MLP). It should be emphasized that, to carry out this task, the performance of each neural network on the personality traits in the OCEAN scale was evaluated from the data collected from the accelerometer of the mobile devices.

8.2.1 Prediction with Fully Connected Feedforward Neural Network

After training the fully connected feedforward neural network, we evaluated the model's performance on the test set. As shown in Table 3, the accuracy was approximately 0.545, and the F1-score was high, around 0.905. We recall the reader that these metrics are complimentary and were obtained after thresholding the predictions of the network. The loss value was notably low at 0.015, indicating a minimal discrepancy between the predicted and actual values. The mean absolute error (MAE) was 0.088, reflecting a low average prediction error. These results suggest that despite moderate accuracy likely influenced by class imbalance, the model consistently and reliably makes predictions.

Table 3. ANN model performance metrics with the test set

Metric	Value
Accuracy	0.5451
F1-Score	0.9059
Loss	0.0152
Mean Absolute Error	0.0885

Likewise, Fig. 12 shows the five graphs corresponding to each of the predicted personality attributes. The effectiveness of the proposed ANN can be observed, since using the data obtained from the accelerometer and the IPIP data, there is a high similarity between both sets of data. This is remarkable because the actual scores maintain proximity with the predictions.

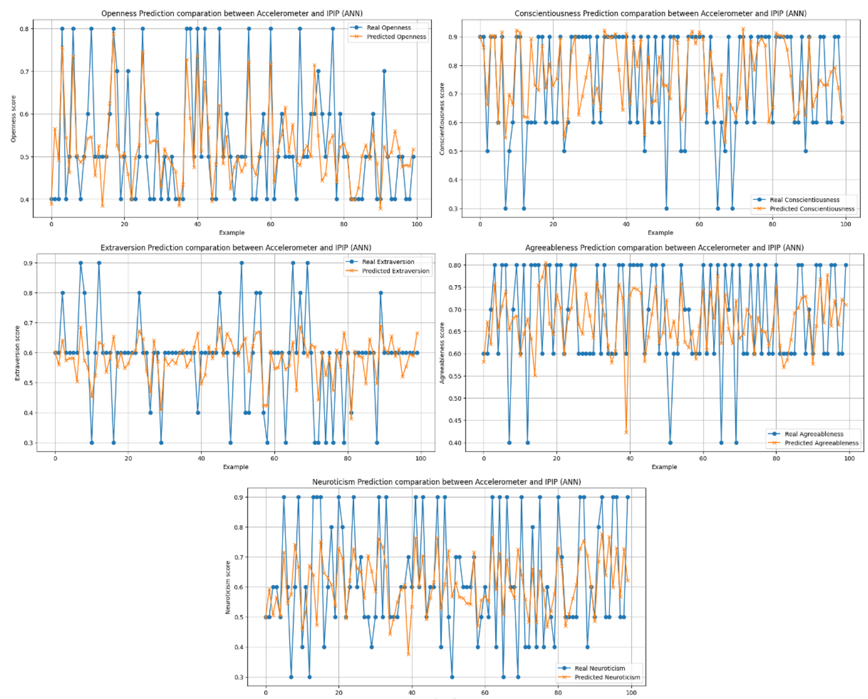

Fig. 12. Personality attribute prediction plots using ANN.

8.2.2 Prediction with Multi-Layer Perceptron

After training and evaluating the model across all five personality traits, the results are summarized in Table 4, with an average accuracy of 0.635 indicating moderate correct classification after binarization and a high F1-score. The F1-score demonstrates that, even in the presence of data imbalance, the predictions remain consistent. This may have impacted on the overall accuracy, but not on the model's ability to correctly identify dominant patterns. The average mean squared error (MSE) was 0.0151 and the mean absolute error (MAE) was 0.0894. Both suggest low prediction errors in the regression output. Despite the moderate accuracy, these results show that the model maintains a high level of predictive precision and reliability.

Table 4. MLP model average performance metrics with the test set

Metric	Value
Average Accuracy	0.6350
Average F1-Score	0.8983
Average MSE	0.0151
Average MAE	0.0894

The performance metrics for each of the personality traits were also evaluated separately, obtaining the results shown in Table 5. We set a threshold of 0.5 to discretized the scores for the continuous traits; values greater than 0.5 indicate the "presence" of a trait, and values equal or below 0.5 indicate "absence" of the trait. Looking at the results, it can be inferred that the predictions provided by the network remain consistent across the five personality traits. In addition, the low MAE values suggest a good level of accuracy for the predictions of the traits individually.

Table 5. Performance metrics for each personality trait

Personality Trait	Accuracy	F1-Score
Openness	0.7125	0.7195
Conscientiousness	0.9625	0.9691
Extraversion	0.8563	0.8844
Agreeableness	0.9625	0.9771
Neuroticism	0.8562	0.9170

In addition, when looking at Fig. 13, where the comparative graphs of the actual and predicted values of the five personality attributes of the OCEAN model are presented, it is easy to notice the similarities between them. In each of these, a blue line is drawn for the actual values, and one in orange for the predicted values. Reviewing the set of images, they demonstrate that the MLP network can approximately and consistently predict the various personality attributes assessed from the accelerometer data alone.

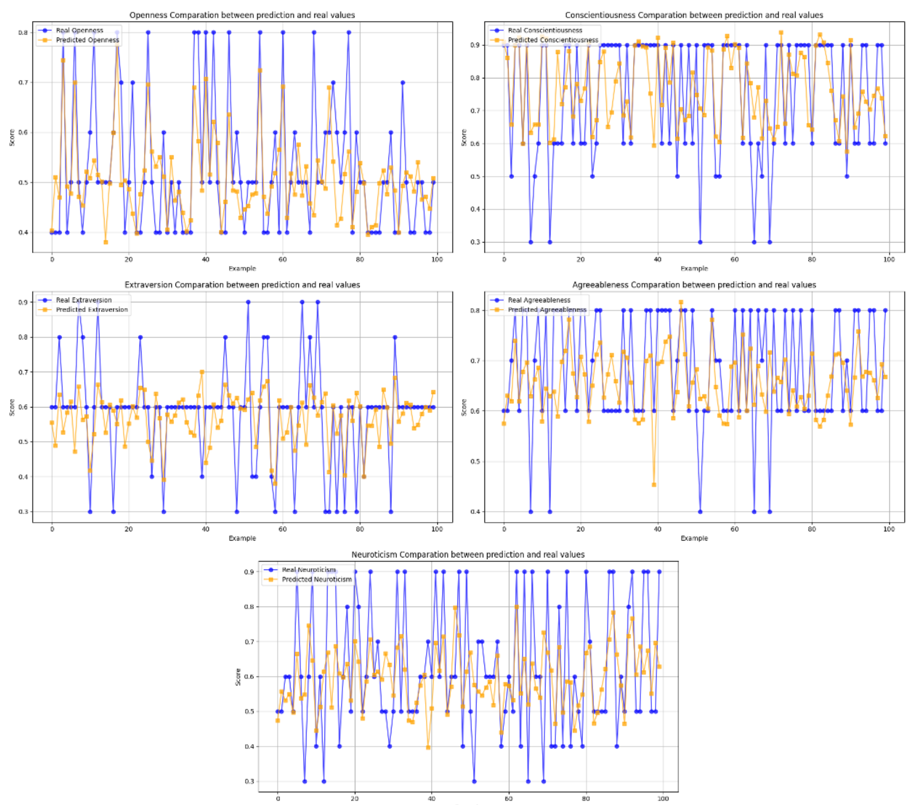

Fig. 13. Personality attribute prediction graphs using MLP.

Furthermore, as shown in Figure 14, we generated confusion matrices for each trait of the OCEAN model.

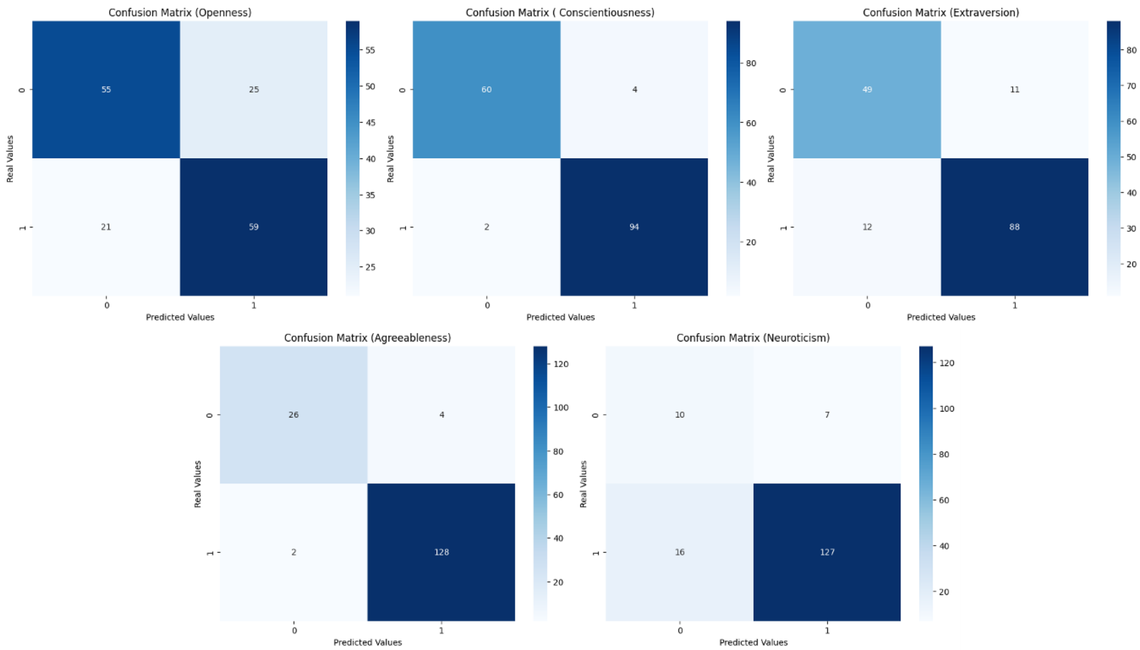

Fig. 14. Confusion matrices for each personality trait.

These matrices provide a more detailed view of the model's classification performance. In particular, the model performed well to classify two traits such as *conscientiousness* and *agreeableness*, with low error rates and clear separation between classes. Two other traits *extraversion* and *neuroticism* also performed well, but with a slightly higher incidence of false positives and false negatives, suggesting some overlap in class boundaries. On the other hand, the model showed difficulty to accurately classify the trait *openness*, especially when identifying its absence, which may indicate a less clear pattern in the input data for this trait. These results reflect that the model is effective in capturing patterns associated with certain traits, but that its performance may vary depending on how clearly these traits manifest themselves in the discretized data.

9 Conclusions and Future Work

This section presents the conclusion of the development of the work. It also sets out future deployment activities to use the outcomes presented here.

9.1 Conclusions

The results of this work confirm the feasibility of assessing personality attributes from data collected from mobile devices, in this case the accelerometer. By observing the data obtained from the different sensors and by analyzing the motion patterns captured from them, it has been demonstrated that this type of dataset is effective in predicting personality attributes with accuracy and consistency. By using different neural networks to obtain predictive data, they highlight the feasibility and potential of using neural networks for the analysis of human behavior. This opens up new avenues for applications in areas such as psychology, mental health and behavioral studies.

However, certain limitations must be considered. One important factor contributing to data imbalance is the tendency for students with high conscientiousness and extraversion traits to participate in these experiments, while more introverted individuals often lack motivation to engage. This selective participation may influence the diversity of data collected and the generalizability of findings. Additionally, the number of participants was limited due to various constraints: some users were uncomfortable with a test involving the recording of usage data from their device, while others did not have an Android device compatible with the experiment. Another key point to highlight is the importance of integrating user self-reported data together with automatically obtained data by means of sensors. In this way it is possible to perform a variety of approaches to personality recognition using data extracted from mobile devices having such a well-documented reference as the IPIP-50.

9.2 Future Work

As future work, we plan to continue expanding the dataset to create a more comprehensive and robust foundation for increasingly accurate predictions in personality recognition. In the next stages of research, we will integrate data from gyroscope readings, mobile device usage questionnaires, and sensor logs, alongside audio and video recordings, allowing for deeper behavioral analysis and richer contextual understanding. This expanded dataset will enable a broader set of experiments utilizing both traditional and advanced machine learning models, as well as deep learning techniques, to improve reliability and precision in personality assessment. By incorporating a wider variety of reference data, we expect predictions to align more closely with real-world personality traits. In addition, we plan to explore the use of categorical cross-entropy as a possible loss function in future experiments, especially in scenarios where personality traits are discretized into categories, in order to evaluate its impact on model performance. This approach will increase experimental diversity, enabling comparisons across different neural network architectures and optimizing model performance for greater accuracy in personality recognition.

References

Arfaoui, N., Mkhini, M., Sidibe, A. S., Baron, B., & Wallelign, S. (2025). Unveiling human essence: Deep learning in personality traits detection. *International Journal of Computer Information Systems and Industrial Management Applications*, *17*.

Bera, S., Pati, A., & Parhi, M. (2023). Human activity detection using smartphone sensor data in real-time based on deep learning approaches. In *AIP Conference Proceedings* (*Vol. 2819*). https://doi.org/10.1063/5.0137525

Concha, D., Bilbao, M. Á., Gallardo, I., Páez, D., & Fresno, A. (2012). Cognitive biases and their relationship to subjective well-being. Salud & Sociedad, 3(2).

De Choudhury, M., Counts, S., & Gamon, M. (2012). Not all moods are created equal! Exploring human emotional states in social media. In *Proceedings of the International AAAI Conference on Web and Social Media* (*Vol. 6*, No. 1, p. 66).

Desai, M., & Shah, M. (2021). An anatomization on breast cancer detection and diagnosis employing multi-layer perceptron neural network (MLP) and convolutional neural network (CNN). *Clinical eHealth*, *4*, 1–11. https://doi.org/10.1016/j.ceh.2020.11.002

Fu, Y., Song, H., Zhao, T., & Kawahara, T. (2024). Enhancing personality recognition in dialogue by data augmentation and heterogeneous conversational graph networks. *arXiv*. http://arxiv.org/abs/2401.05871

Gil de Zúñiga, H., Diehl, T., Huber, B., & Liu, J. (2017). Personality traits and social media use in 20 countries: How personality relates to frequency of social media use, social media news use, and social media use for social interaction. *Cyberpsychology, Behavior, and Social Networking*, *20*(9), 540–552.

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

Goldberg, L. R. (1993). The structure of phenotypic personality traits. *American Psychologist*, *48*(1), 26–34. https://doi.org/10.1037/0003-066X.48.1.26

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

Guenzel, M., Kogan, S., Niessner, M., & Shue, K. (2025). *AI personality extraction from faces: Labor market implications*. SSRN. https://doi.org/10.2139/ssrn.5089827

Incio-Flores, F. A., Capuñay-Sanchez, D. L., & Estela-Urbina, R. O. (2023). Artificial neural network model to predict academic results in Mathematics II. *Revista Electrónica Educare*, *27*(1). https://doi.org/10.15359/ree.27-1.14516

Mezquita, L., Bravo, A. J., Morizot, J., Pilatti, A., Pearson, M. R., Ibáñez, M. I., & Ortet, G. (2019). Cross-cultural examination of the Big Five Personality Trait Short Questionnaire: Measurement invariance testing and associations with mental health. *PLoS ONE*, *14*(12). https://doi.org/10.1371/journal.pone.0226223

Moshe, I., Terhorst, Y., Opoku Asare, K., Sander, L. B., Ferreira, D., Baumeister, H., Mohr, D. C., & Pulkki-Råback, L. (2021). Predicting symptoms of depression and anxiety using smartphone and wearable data. *Frontiers in Psychiatry*, *12*. https://doi.org/10.3389/fpsyt.2021.625247

Picca, D., & Pitteloud, J. (2023). Personality recognition in digital humanities: A review of computational approaches in the humanities. *Digital Scholarship in the Humanities*, *38*(4), 1646–1658. https://doi.org/10.1093/llc/fqad047

Rendón-Macías, M. E., Valenzuela, M., & Villasís-Keever, M. Á. (2020). Biases in diagnostic test studies: Impact on estimating sensitivity and specificity. *Revista Alergia México*, *67*(2), 165–173. https://doi.org/10.29262/RAM.V67I2.771

SensorApp Mobile (n.d). Retrieved June 1, 2025, from https://catalabs.mx/apps/sensorapp/

SensorApp Web. (n.d.). Retrieved June 1, 2025, from https://sensorapp-itc.web.app/

Stachl, C., Hilbert, S., Au, J. Q., Buschek, D., De Luca, A., Bischl, B., Hussmann, H., & Bühner, M. (2017). Personality traits predict smartphone usage. European Journal of Personality, 31(6), 701–722. https://doi.org/10.1002/per.2113

Stade, M., Scherr, S. A., Mennig, P., Elberzhager, F., & Seyff, N. (2019). Don't worry, be happy: Exploring users' emotions during app usage for requirements engineering. In *Proceedings of the 2019 IEEE 27th International Requirements Engineering Conference (RE)* (pp. 375–380). IEEE. https://doi.org/10.1109/RE.2019.00053

Sze, W. Y. S., Herrero, M. P., & Garriga, R. (2024). *Personality trait inference via mobile phone sensors: A machine learning approach*. arXiv. https://doi.org/10.48550/arXiv.2401.10305

Talaat, F. M., El-Gendy, E. M., Saafan, M. M., & Gamel, S. A. (2023). Utilizing social media and machine learning for personality and emotion recognition using PERS. *Neural Computing and Applications*, *35*(33), 23927–23941. https://doi.org/10.1007/s00521-023-08962-7

Valanarasu, R. (2021). Comparative analysis for personality prediction by digital footprints in social media. *Journal of Information Technology and Digital World*, *3*(2), 77–91. https://doi.org/10.36548/jitdw.2021.2.002

VandenBos, G. R. (Ed.). (2015). *APA dictionary of psychology* (2nd ed.). American Psychological Association. https://doi.org/10.1037/14646-000

Wampfler, R., Klingler, S., Solenthaler, B., Schinazi, V. R., Gross, M., & Holz, C. (2022). *Affective state prediction from smartphone touch and sensor data in the wild*. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems* (pp. 1–14). Association for Computing Machinery. https://doi.org/10.1145/3491102.3501835

Ware, S., Yue, C., Morillo, R., Lu, J., Shang, C., Bi, J., Kamath, J., Russell, A., Bamis, A., & Wang, B. (2020). Predicting depressive symptoms using smartphone data. *Smart Health*, *15*, 100093. https://doi.org/10.1016/j.smhl.2019.100093

Webb, C. A., Ren, B., Rahimi-Eichi, H., Gillis, B. W., Chung, Y., & Baker, J. T. (2025). *Personalized prediction of negative affect in individuals with serious mental illness followed using long-term multimodal mobile phenotyping*. *Translational Psychiatry*, *15*, 174. https://doi.org/10.1038/s41398-025-03394-4

Wiggins, J. S. (Ed.). (1996). *The five-factor model of personality: Theoretical perspectives*. Guilford Press.

Xenakis, D., Samikwa, E., Ajayi, J., Di Maio, A., Braun, T., & Schlegel, K. (2023). *Towards personality detection and prediction using smartphone sensor data*. In *Proceedings of the 21st Mediterranean Communication and Computer Networking Conference (MedComNet 2023)* (pp. 121–130). IEEE. https://doi.org/10.1109/MedComNet58619.2023.10168869

Yang, Z., Wang, Y., Yamashita, K. S., Khatibi, E., Azimi, I., Dutt, N., Borelli, J. L., & Rahmani, A. M. (2024). *Integrating wearable sensor data and self-reported diaries for personalized affect forecasting*. *Smart Health*, *32*, 100464. https://doi.org/10.1016/j.smhl.2024.100464

Zatarain Cabada, R., Leyva López, N., Barrón Estrada, M. L., Bátiz Beltrán, V. M., & Escalante Balderas, H. J. (2024). SensorAPP: Una Metodología Para Crear Un Conjunto De Datos Para Reconocimiento De Personalidad SensorAPP: A Methodology to Create a Dataset for Personality Recognition Using Mobile Devices. Manuscript accepted for publication in *Encuentro Nacional de Computación 2024*.