❒   1534

Application of artificial neural networks for personality traits prediction based on handwriting

Ahmed Remaida[1], Benyoussef Abdellaoui[1], Mohamed Amine Lafraxo[2], Zineb Sabri[3], Hamza Nouib[1], Younes El Bouzekri El Idrissi[1], Aniss Moumen[1]

[1]Laboratory of Engineering Sciences, National School of Applied Sciences, Ibn Tofaïl University, Kenitra, Morocco
[2]Laboratory of Biology and Health, Faculty of Sciences, Ibn Tofaïl University, Kenitra, Morocco
[3]Laboratory of Computer Science, Faculty of Sciences, Ibn Tofaïl University, Kenitra, Morocco

Article Info

ABSTRACT

The automatization of personality traits prediction still brings considerable research these days, especially when the detection could be achieved without administrating personality assessment instruments. With the evolution of computational intelligence, a variety of deep machine learning techniques were developed and proposed for that purpose. Nevertheless, proposing robust and rapid systems to solve this problem remains a challenging task. The process of feature extraction is the main key. This paper presents an effective method for extracting five graphological features from handwriting ensuring the prediction of personality traits based on the big five personality traits model. We started by collecting both handwriting samples and big five questionnaires, then the feature extraction process, after that the data preparation and finished with the application of several popular deep machine learning models to achieve the prediction. Experimental results indicate the remarkable performance of the multi-layer perceptron (MLP) compared to other classifiers, the model was 100% precise and classification accuracy attained 100% for trained data and 72.73% for new tested data. With only 100 participants, we strongly believe that our proposed method is simple and promising, and better results will be attained with a larger dataset.

Corresponding Author:

Ahmed Remaida
Laboratory of Engineering Sciences, National School of Applied Sciences, Ibn Tofaïl University
Kenitra, 14000, Morocco
Email: ahmed.remaida@uit.ac.ma

1. INTRODUCTION

Considered as an interdisciplinary research field that incorporates psychology, cognitive science and computer science [1], personality computing consists of applying artificial intelligence technology in personality assessment for automatizing personality prediction from different sources such as body language, handwriting, face [2], voice, natural language, and social networks. Among those sources, handwriting gained considerable attention in the literature, and until today it still gain major interest among researchers [3]. Personality assessment on the basis of handwriting is also known as graphology, and it has been applied as an effective assessment tool in diverse fields, such as psychology, employment [4], [5], education, medicine [6], criminal detection and marriage guidance. This is due to the fact that individuals tend to falsify the scores of measurement tools when they know that their personality is being analysed.

With the increasing evolution of computational intelligence and the emergence of deep machine learning algorithms, automated personality assessment through handwriting gained more interests in the last decade [7], [8]. Many researchers are proposing interesting novel architectures under this scope of research. Table 1 gives a quick overview of some current personality prediction based on handwriting feature extraction:

Table 1. Recent related works

No	Works	Results discussion
1	[9]	This research focused on persian handwriting and the minnesota multiphasic personality inventory psychometric test results in order to create a comprehensive dataset for handwriting personality predicition. Authors trained and tested two classifiers for personality prediction: multi-layer perceptron (MLP) and hidden Markov model. The resulted system had 76% accuracy in training and 61% accuracy in testing.
2	[10]	Introduced a novel three-layer system for predicting the big five personality traits of an individual based on his or her handwriting. Authors followed two methodologies to validate their system: inter-subject and intrasubject, higher accuracies were obtained in the second one with 84.4% and above for openness, extraversion, and neuroticism and with 77% for conscientiousness and agreeableness.
3	[11]	This study presents the largest hadnwriting dataset under this scope of research with 1800 handwriting samples. However, the proposed system is only ment to detect suitable personalities based on graphological features for employement purpose. Authors implemented different machine learning algorithms such as support vector machine (SVM), Naïve bayes, and random forest (RF), the highest prediction accuracy was obtained with the SVM with 97%.
4	[12]	Authors of this work proposed a novel architecture for predicting the big five personality traits from handwriting. Their dataset consisted of neo five-factor inventory-3 test results, handwriting samples in addition to written letter "t" samples. The obtained accuracies were between 65% and 85%, where extraversion scored the highest value and conscientiousness was the lowest value.
5	[13]	This work proposed a convolutional neural networks (CNN) based architecture that predicts personality from both English and Hindi handwriting without selecting any significant handwriting features. Despite the low accuracy rates (average of 43%), the authors intend to improve their architecture's accuracy in future works.

Despite the growing interests toword handwriting personality prediction, extracting graphical features like margins, spaces, baseline, pressure of the pen, alphabet morphology from a given handwriting sample in order to link them to the psychological state of the writer still remain a hard task to achieve [14]. The choice and implementation of the image processing and segmentation techniques is crucial to a successful personality traits prediction. Therefore, we conducted this study to proposes an effective solution to this problem and to make a contribution to this scope of research. By exploring the mapping between five handwriting features (top margin, baseline, line spacing, word spacing, and letter size) and the Big Five personality traits (openness, conscientiousness, extraversion, agreeableness and neuroticism), we developed simple yet effective methods for feature extraction and evaluated deep machine learning classifiers in order to achieve the big five personality traits prediction based on handwriting.

The rest of the paper is organized as following: section 2 highlights the different methods and techniques we used in this work. Followed by section 3, where we discuss the results and performances of the proposed method and compare it to the existing works. Finally, we resume with a conclusion and a future research planning in section 4.

2.   METHOD

In order to achieve personality traits recognition based on handwriting, we had to acquire both: personality traits and handwriting features from participants. For personality traits we referred to a brief big five questionnaire, as for handwriting we chose the London letter model. After collecting the data from 100 participants, we applied several methods for handwriting preprocessing, feature extraction, and classification. For better understanding of this process, Figure 1 shows an illustration of opur proposed architecture. First, we apply image processing on the handwriting sample to enhance the feature extraction. Then, the extracted features and the five factor model personality traits are both stored in the dataset. Finally, after completing the database we implemented classification using five classifiers one to predict each of the personality traits.
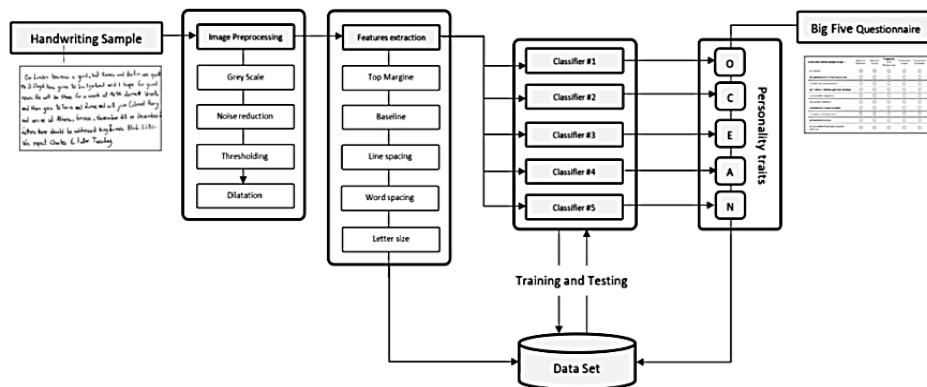

Figure 1. The proposed architecture for personality handwriting prediction

2.1. Personality test

Many psychological studies describe the personality structure as a set of dimensions or traits that could be measured using self-report questionnaires or rating scales, which most often appear in the form of a statement or question with a 'yes-no', 'true-false' or an extended scale like the 'Likert-type' scale. A variety of efficient and validated well-known questionnaires in the literature are used to accomplish this task, such as: the sixteen personality factor questionnaire (16PF) [15], eysenck personality questionnaire (EPQ-R) [16], myers briggs type indicator (MBTI) [17] and big five personality traits inventory (BFI) [18]. Table 2 shows keywords descriptors for each one of the big five personality traits. To acquire participants' personality traits, we chose the BFI questionnaire, precisely, a brief version of it with only 10 items [19]. This short and brief inventory questionnaire even if it is inferior to the standard big five inventory (44 items), it has reached adequate levels in the evaluation process as described by its authors.

Table 2. Big five traits and keywords descriptors

BF traits	Keywords descriptors
Openness	Curious – imaginative – artistic – broad interests – excitable – unconventional
Conscientiousness	Efficient – organized – not careless – thorough – not lazy – not impulsive
Extraversion	Sociable – forceful – energetic – adventurous – enthusiastic – outgoing
Agreeableness	Forgiving –undemanding – warm – not stubborn – not show-off – sympathetic
Neuroticism	Tense – irritable – not contented – shy – moody – not self-confident

2.2. Handwriting samples

For the handwriting samples acquisition, we asked participants with their consent, to rewrite the London letter model in the dedicated box for writing. This model choice was due to the fact that it contains letters of the alphabet in both uppercase and lowercase form, all numerals in different combinations, and many other characteristics that make this choice suitable for handwriting analysis. After scaning all the handwriting samples with a haigh resultion scanner, we obtained images in png format that we stored as our dataset. The next step consisted of applying the differents pre-processing thechniques in order to extract the features that will be used for the personality traits prediction.

2.2.1. Handwriting pre-processing

Pre-processing methods are an elementary step for making image data optimized for segmentation and feature extraction. Denoising, morphological transforming, and binarizing are all methods used to enhance the image quality and make it suitable for the next steps. For this purpose, we have used the OpenCV library [20] in Python and started by reading the image sample in RGB space, then converting it to gray scale. The next steps in preprocessing were implemented using the same library. In the following, we highlight each one of the used methods in this phase.

a) Noise reduction

The obtained handwriting samples contained unwanted noise, and applying filters. Like bilateral filter can resolve that problem very well. It combines domain and range filters and it consists of choosing a location in the image and applying the bilateral filter function [21] to the values of the input image in a small neighbourhood of the same location.

b) Binarization

Binarization is the transforming process of any entity data feature into a binary number vector in order to enhance classifier algorithm efficiency. Binarizing a grey scale handwriting sample would allow us to transform it from the 0-255 spectrum to a 0-1 spectrum. To do so, we applied an inverted thresholding [22] to the filtered handwriting samples where in case the pixel with a greater value than the thresh it is converted to 0 and to 1 in the other case.

c) Dilatation

Dilatation is considered as an important step in image processing, like erosion, it is a morphological operation that processes the shape of an image and both are to be considered as the most basic operations in this process. For dilatation, we need two inputs the original image and the structuring element (kernel) that determines the shape of a pixel neighbourhood over which the maximum is taken. We used dilatation to enlarge the lines in the handwriting sample, the bright area of the line is dilated around the black regions of the background leading to linking words together. Which will allow us to contour the dilated lines and calculate its baseline angle, rotate it and then extract it in the line segmentation step. Dilatation is also used in word segmentation in each line, enlarging the letters of a word will be a key factor to word segmentation and space calculation.

2.2.2. Feature extraction

In order to extract the five handwriting features, we applied several existing techniques: beginning with contouring and rotating, which allowed us to calculate the baseline angle and straighten the lines. Then, horizontal projecting to calculate the top margin, line spacing and segmenting lines. After that, we used vertical projection to calculate both letter size and word spacing. A detailed explanation of the process is presented for each feature.

a) Baseline

For the base line angle, we contoured the dilated lines with the OpenCV contours algorithm [23] which seeks curve joining of all continuous points that have the same color. Next, we draw bounding boxes with those boundary points. After getting all the rotation angles of the bounding rectangles, we calculated the average as the baseline angle of the handwriting sample. To rotate the lines, we applied the affine transformation with linear transformation (matrix multiplication) and translation (vector addition). As result, we had straightened handwriting samples.

b) Top margin

For top margin calculation, line spacing, and segmentation, we used the horizontal projection algorithm. By counting the sum of white pixels in each line, we trace zero values that represent the start and end space indices. Top margin is calculated based on the first two indices.

c) Line spacing

As for line spacing, we traced the max values of the line pixels sum as indexes representing the baseline. Spaces between each two baseline indexes are calculated and the average of all spaces represents the line spacing of the handwriting sample. As mentioned top margin, in the top margin calculation, horizontal projection was used to determine the space indices between lines. After dividing the handwriting samples on the basis of those indices, we got the segmented lines.

d) Letter size

In general, letter size is determined as categories based on middle zone calculation: average, large, or small compared to the average letter size. Although, we chose to calculate this feature in another way, we traced pixels that represent the start and end of the letter shape. The differences between those indices are used to calculate the average letter size of the line and then the average of all lines will represent the letter size feature of the handwriting sample.

e) Word spacing

To calculate only the space between words and not letters, we had to use dilatation on segmented lines and then apply vertical projection on those lines. Once again, like in line space, we traced the zero values of the sum of vertical line pixels, and with start and end indices we determined the spaces between words. After that, we calculated the space average for each line, then the average for the handwriting sample.

2.3. Classification algorithms

To attain the aim of this work, we implemented and evaluated five well-known supervised machine learning algorithms to create models. Capable of predicting each trait of the big five personality traits. Here we give a non-exhaustive description of each one of the implemented algorithms.

2.3.1. Gradient boosting

Considered one of the most powerful techniques used to build predictive models, gradient boosting comes with the idea of testing if a weak learner can be optimized to become better. The first successful realization of boosting was with the adaptive boosting (AdaBoost) [24]. Another successful realization of the boosting technique is a scalable tree boosting system called extreme gradient boosting (XGBoost) [25]. Due to its computational efficiency and high effectiveness, we chose this model for the prediction of personality traits.

2.3.2. Random forest

Random forest is an ensemble learning model estimator that fits a collection of decision tree classifiers (1) to various subsamples of the dataset using average calculation in order to improve the accuracy of prediction and control the over-fitting phenomena [26]. Being more accurate than decision tree and considered as a strong classification algorithm, we chose to implement this algorithm in this work.

$$\{h(x, \Theta k), k = 1, \dots\} \tag{1}$$

2.3.3. Support vector machine

Since the time they were developed in the 1990s, support vector machines were and continues to be one of the most popular machine learning algorithms [27]. In practice, SVM algorithms' implementation uses kernels, where prediction for a new input is done using the dot product between the input (x) and all support vectors (xi) in the training data (2) where B0 and ai are estimated from the training data by the learning

algorithm. SVM is characterized by its high-performing algorithm with little tuning, which makes it suitable for this work.

$$f(x) = B0 + sum(ai * (x, xi))$$ (2)

2.3.4. Logistic regression

Like linear regression, logistic regression (LR) uses the same equation representation in (3) where input features (x) are linearly combined using weights or coefficient values in order to predict an output class (y). The only difference from linear regression is that the outputted value is a binary value (0 or 1) rather than a numeric value. Logistic regression models are very adequate in binary classification, the reason why we chose to implement it in this study.

$$y = e^{\wedge}(b0 + b1 * x) / (1 + e^{\wedge}(b0 + b1 * x))$$ (3)

2.3.5. Artificial neural network (multi-layer perceptron)

Unlike other classification algorithms such as SVM classifier, MLP classifier is an underlying DNN that performs the task of classification or regression by learning from a non-linear function approximator. It is different from logistic regression having one or multiple non-linear layers (hidden layers) between the input and the output layer. In general, the MLP architecture has a defined a number of dimensions for inputted data and outputted classes, and this process is called Feed Forward propagation. Whereas, the process that updates the weights of each neuron, and represents the core of the learning algorithms, is called the Back-propagation process [28]. Figure 2 shows an exemple of a one hidden layer perceptron with X_n features input and one scalar output.

a)   Input layer: consists of a set of neurons (4) representing the input features:

$$\{xi | x1, x2, \ldots, xm\}$$ (4)

b)   Hidden layer: each neuron in the hidden layer transforms the values from the previous layer with a weighted linear summation (5) followed by a non-linear activation function (6).

$$w1x1 + w2x2 + \ldots + wmxm$$ (5)

c)   Output layer: receives the values from the last hidden layer and transforms them into output values using an activation function (6).

$$G(\cdot): R \rightarrow R$$ (6)

d)   Backpropagation algorithm: it aims to minimize the cost function by adjusting the network's weights using the gradients of the cost function (7).
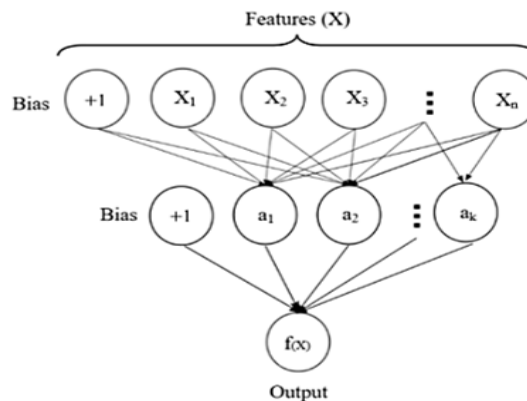
$$Wn = Wo - (\alpha * dL/dw)$$ (7)

Figure 2. Illustration of one hidden layer perceptron

2.3.6. Evaluation metrics

The evaluation of the different classifiers was done using four main evaluation scores: accuracy, recall, precision, and f1-score. We generated the confused matrix table for each model in order to calculate those scores:

a) Accuracy score: it is the fraction of predictions where the model has the right prediction in (8).

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \qquad (8)$$

b) Precision: it is the division of correct positive results by the predicted positive results in (9).

$$Precision = \frac{TP}{TP+FP} \qquad (9)$$

c) Recall: it is the division of correct positive results by all relevant samples in (10).

$$Recall = TP\ (TP + FN) \qquad (10)$$

d) F1 score: it represents harmonic mean between precision and recall. It represents the robustness of the model in (11).

$$F1 = \frac{2}{\frac{1}{Precision} * \frac{1}{Recall}} \qquad (11)$$

3.   RESULTS AND DISCUSSION

As described in the proposed architecture, the final constructed dataset contains both handwriting features and big five personality traits results of 100 participants in the age between 20 and 25 years with 48 males and 52 females. Both results will be used to train and test machine learning models for predicting the big five traits (openness, conscientiousness, extraversion, agreeableness, neuroticism) of a person on the basis of his/her handwriting features (baseline, top margin, line spacing, word spacing, letter size). A clear visualization of the dataset variables distribution is shown in Figure 3. Which has 2 parts: Figure 3(a) handwriting features variables and Figure 3(b) personality traits variables. In extraversion, agreableness, baseline, line spacing, word spacing and letter size data distribution we see the middle clustering of most data points as the rest of data taper off symmetrically toward the extremes, this shows the normal probability distribution in the five personality traits extraversion and agreablness, also in the handwriting features baseline and line spacing. The rest of personality traits and handwriting features in openness, conscentiousness, neuroticism and top margin are slightly right or left skewed showing that data are not normally distributed. We used the resulted dataset as it is and no under or over sampling are applied before the classification step.
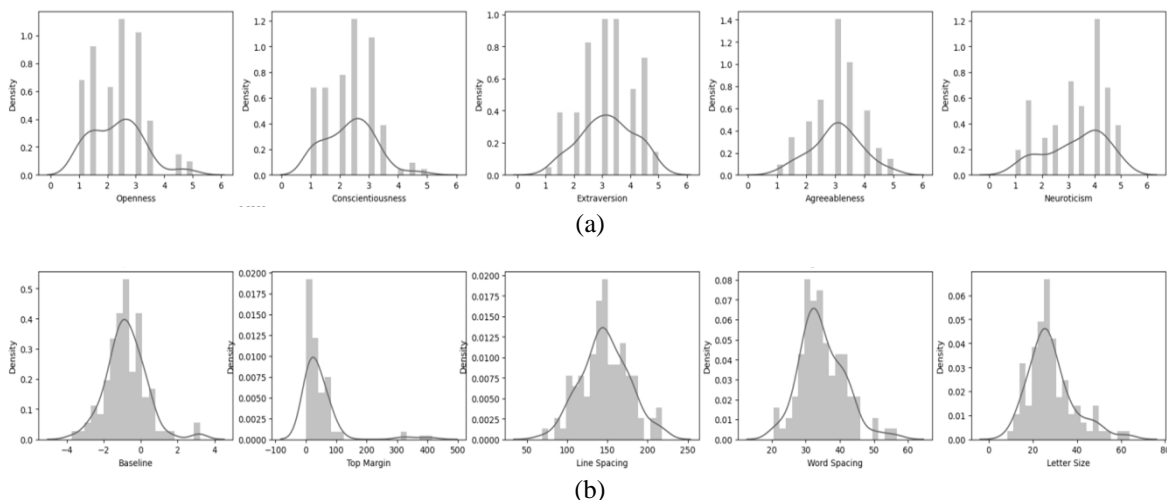

(a)

(b)

Figure 3. Dataset variables distribution visualization; (a) handwriting features variables and
(b) personality traits variables

The five handwriting features in Table 3 are used as input data, while the five personality traits in Table 4 are used as output data, one at a time. To enhance the classification task by different classifiers, we had to standardize the inputted handwriting features and change the values to fit in the form of a unified scale. In order to achieve that, we chose the robust scaler technique [29] which is comparable to the min-max scaler. This scale expels the middle and scales information as indicated by the interquartile range (IQR). We also altered each value of the personality traits to a yes (1) or no (0) value based on median to ensure the binary classification by the different classifiers.

<div style="display:flex">

Table 3. Handwriting features description

Handwriting features	Mean	Median	Range
Baseline	-0.78	-0.8	(-3.74) - 3.22
Top Margin	45.91	25	0 - 415
Line Spacing	147.05	145.67	68.75 - 218.12
Word Spacing	28.23	25.94	8.44 - 63.82
Letter Size	34.74	33.67	20.47 - 56.79

Table 4. Personality traits description

Personality traits	Mean	Median	Range
Openness	2.33	2.5	1 - 5
Conscientiousness	2.34	2.5	1 - 5
Extraversion	3.18	3	1 - 5
Agreeableness	3.04	3	1 - 5
Neuroticism	3.30	3.5	1 - 5

</div>

After the standardization phase, we choose to randomly split the dataset into two sets: 80% of data as training data and the remaining 20% as testing data. The test set has for the purpose to explore the classifier prediction accuracy on unseen handwriting features data. For implementing, training and testing all selected classifiers we carried out experiments on a Lenovo ThinkPad Laptop powered by an 8-core Intel i5 10 Gen CPU of 2.30 GHz, 16 GB of RAM, 8 GB of Intel GPU and 500 GB of Storage access. We choosed Python 3.7 programming language which offers the well-known machine learning library scikit-learn [30]. After fine-tuning each one of the selected classifiers in order to achieve the best performance, we observed different scenarios of trial and error for each one of the classifiers and we have concluded that:

a) LR with bigger values of the inverse of regularization strength (C hyperparameter) showed more efficiency than smaller values where the optimal value was around 100. Higher values are a good indicator that training data reflects the real-world data.

b) RF with 100 decision trees (n_estimators) showed better results for classification than higher or lower values of estimators. A three split allocation for each decision tree (max_depth) was optimal for avoiding model overfitting or underfitting.

c) XGBoost models with smaller learning rate (eta=0.01) and smaller number split allocation (max_depth=3) showed better result, making those models less complex and more robust.

d) SVM with higher regularization strength or ability to reduce weights (C parameter) leads to better performance with less training errors. Also, choosing Gaussian radial basis function (RBF) as kernel type gaves haigher training and testing accuracies in the contrary of the sigmoid kernel type wich only had effect on the testing accuracy. This is due to the fact that SVM model is dealing with non-linear data with no prior knowledge about it. In general SVM classifier is capable of handling unseen data and achieving higher accuracy rate in testing phase than the training phase.

e) MLP with stochastic gradient descent (SGD) optimization algorithms performed better than MLP with ADAM optimizer. Better classification performances were observed when using constant learning rates with momentum and nesterov momentum [31] as shown in Figure 4. As for the number of hidden layers and neurones in each layer, the best choice for achieving high training and testing accuracies for all personality traits was a one hidden layer MLP with only 15 neurones, exept for agreeableness we kept the default one hidden layer MLP with 100 neurones. Neither adding more hidden layers nor augmenting the number of neurons did enhance the performances MLP classifiers.

We see clearly in both Tables 5 and 6 that the MLP classifier outperformed the other classifiers, this observation could be refered to the dynamic non-linearity, high-speed computational capability and also the robust learning algorithms of artificial neural networks models [32]. Although, in the testing phase the MLP model didn't achieve higher accuracy, this could be referred to the insufficient data in our dataset.

In general, we could say that the best choice among the list of proposed classifiers is the MLP, even considering the fact that the presented dataset in this work had unbalanced classes especially for openness, conscientiousness and neuroticism. In Figure 5, the MLP classifier has shown more precision and robusteness (F1 score) on an unseen data. Thus, we mark the high sensitivity (recall) of the MLP compared to other classifiers, and we strongly believe that enlarging the dataset with new handwriting samples in the future will lead and solve this problem. Table 7 gives a detailed comparison of our method with the existing works on big five personality traits prediction based on handwriting features.
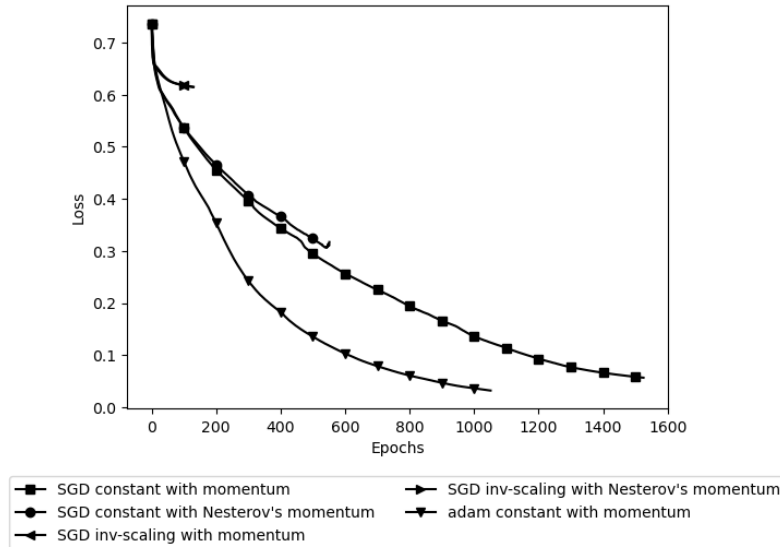
Figure 4. Loss comparison of MLPs classifiers for predicting openness

Table 5. Training accuracy of different classifiers

Big five	LR	RF	XGB	SVM	MLP
O	58.70	84.78	84.78	96.74	95.70
C	61.96	98.91	86.96	**100**	**100**
E	67.39	**100**	79.35	97.83	98.91
A	68.48	80.43	**100**	98.91	**100**
N	59.78	86.96	80.43	92.39	**97.83**

Table 6. Testing accuracy of different classifiers

Big five	LR	RF	XGB	SVM	MLP
O	54.55	**72.73**	**72.73**	**72.73**	**81.80**
C	**63.64**	45.45	54.55	45.45	45.45
E	**72.73**	**72.73**	63.64	**72.73**	**81.82**
A	54.55	54.55	**63.64**	54.55	**63.64**
N	54.55	**72.73**	**72.73**	54.55	**72.73**
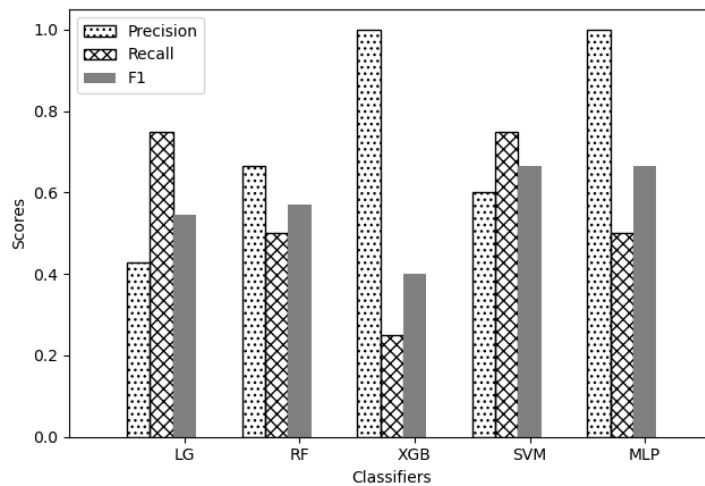
Figure 5. Classifiers metrics evaluation on new handwriting features for the openness personality traits prediction

Table 7. Comparison of the proposed work with the existing works in the literature

Work	Handwriting samples	Handwriting features	Classification techniques	Big five traits prediction accuracy (%)				
				O	C	E	A	N
Gavrilescu and Vizireanu [10]	128	Baseline angle - Word slant - Pen pressure - Connecting strokes - Lower case "t" - Lower case "f" - Line spacing	Feed forward neural network (FFNN)	84	77	84	77	84
Elngar *et al.* [12]	125	Baseline - Pen pressure - Word spacing - lLne spacing - Lower case "t"	Feed forward neural network (FFNN)	85	65	85	75	70
This paper	100	Top margin - Baseline angle - Line spacing - Word spacing - Letter size	Multi-layer perceptron (MLP)	82	64	82	64	73
Chaubey and Arjaria [13]	110	No Feature Selection	Convolutional neural networks	43	46	41	40	46

4. CONCLUSION

Automatic personality prediction from handwriting remains a challenging task, and developing accurate prediction methods would provide a great solution to many different problems such as employment selection, criminal profiling, and early-stage disease detection. Here, in this paper, we have proposed a simple yet effective method for predicting the big five personality traits on the basis of handwriting samples. We asked 100 participants to provide questionnaire responses and handwriting samples in order to create our dataset. We used a variety of techniques in the preprocessing and feature extraction phases before the implementation of personality classification. Then, we implemented and compared several popular machine learning classification algorithms. Experimental results have marked the outperformance of the MLP for both training and testing data. We achieved satisfactory accuracy rates in comparison with the existing works, however, more improvements in the future are needed. As a future work, we plan to implement a user interface of the handwriting feature extraction process based on the proposed method. A more complex feature distribution could be explored, and enlarging the dataset is for sure a crucial target to attain.

REFERENCES

[1] W. Wang, H. Ning, F. Shi, S. Dhelim, W. Zhang, and L. Chen, "A survey of hybrid human-artificial intelligence for social computing," *IEEE Transactions on Human-Machine Systems*, vol. 52, no. 3, pp. 468–480, Jun. 2022, doi: 10.1109/THMS.2021.3131683.
[2] B. Abdellaoui, A. Moumen, Y. El Idrissi, and R. El Bouzekri, "Face detection to recognize students' emotion and their engagement: A systematic review," in *2020 IEEE 2nd International Conference on Electronics, Control, Optimization and Computer Science, ICECOCS 2020*, Dec. 2020, pp. 1–6, doi: 10.1109/ICECOCS50124.2020.9314600.
[3] E. Gómez-Jiménez, "An introduction to graphology: definition, theoretical background and levels of analysis," *Miscelanea*, vol. 51, pp. 71–85, 2015.
[4] A. Moumen, E. H. Bouchama, and Y. E. B. El Idirissi, "Data mining techniques for employability: systematic literature review," *2020 IEEE 2nd International Conference on Electronics, Control, Optimization and Computer Science, ICECOCS 2020*, 2020, doi: 10.1109/ICECOCS50124.2020.9314555.
[5] A. Remaida, A. Moumen, Y. E. B. El Idrissi, B. Abdellaoui, and Y. Harraki, "The use of personality tests as a pre-employment tool: a comparative study," *SHS Web of Conferences*, vol. 119, p. 05007, Aug. 2021, doi: 10.1051/shsconf/202111905007.
[6] S. Ramli and S. Nordin, "Personality prediction based on iris position classification using support vector machines," *Indonesian Journal of Electrical Engineering and Computer Science (IJEECS)*, vol. 9, no. 3, pp. 667–672, Mar. 2018, doi: 10.11591/ijeecs.v9.i3.pp667-672.
[7] K. Chaudhari and A. Thakkar, "Survey on handwriting-based personality trait identification," *Expert Systems with Applications*, vol. 124, pp. 282–308, Jun. 2019, doi: 10.1016/j.eswa.2019.01.028.
[8] A. Remaida, A. Moumen, Y. E. B. El Idrissi, and Z. Sabri, "Handwriting recognition with artificial neural networks a decade literature review," in *Proceedings of the 3rd International Conference on Networking, Information Systems and Security*, Mar. 2020, pp. 1–5, doi: 10.1145/3386723.3387884.
[9] B. Fallah and H. Khotanlou, "Identify human personality parameters based on handwriting using neural network," in *2016 Artificial Intelligence and Robotics (IRANOPEN)*, Apr. 2016, pp. 120–126, doi: 10.1109/RIOS.2016.7529501.
[10] M. Gavrilescu and N. Vizireanu, "Predicting the big five personality traits from handwriting," *EURASIP Journal on Image and Video Processing*, vol. 2018, no. 1, p. 57, Dec. 2018, doi: 10.1186/s13640-018-0297-3.
[11] P. Joshi, P. Ghaskadbi, and S. Tendulkar, "A machine learning approach to employability evaluation using handwriting analysis," in *Advanced Informatics for Computing Research: Second International Conference, ICAICR 2018*, 2019, pp. 253–263, doi: 10.1007/978-981-13-3140-4_23.
[12] A. A. Elngar, N. Jain, D. Sharma, H. Negi, A. Trehan, and A. Srivastava, "A deep learning based analysis of the big five personality traits from handwriting samples using image processing," *Journal of Information Technology Management*, vol. 12, pp. 3–35, 2021, doi: 10.22059/JITM.2020.78884.
[13] G. Chaubey and S. K. Arjaria, "Personality prediction through handwriting analysis using convolutional neural networks," in *Proceedings of International Conference on Computational Intelligence: ICCI 2020*, 2022, pp. 59–70, doi: 10.1007/978-981-16-3802-2_5.
[14] A. H. Garoot, M. Safar, and C. Y. Suen, "A comprehensive survey on handwriting and computerized graphology," in *2017 14th IAPR International Conference on Document Analysis and Recognition (ICDAR)*, Nov. 2017, pp. 621–626, doi: 10.1109/ICDAR.2017.107.
[15] H. E. P. Cattell and A. D. Mead, "The sixteen personality factor questionnaire (16PF)," in *The SAGE Handbook of Personality Theory and Assessment: Volume 2 — Personality Measurement and Testing*, 1 Oliver's Yard, 55 City Road, London EC1Y 1SP United Kingdom: SAGE Publications Ltd, 2008, pp. 135–159.
[16] H. J. Eysenck and S. B. G. Eysenck, *Eysenck personality inventory*. London: University of London Press, 1971.
[17] I. B. Myers, *MBTI manual : a guide to the development and use of the myers-briggs type indicator*. Consulting Psychologists Press, 1998.
[18] G. V. Caprara, C. Barbaranelli, L. Borgogni, and M. Perugini, "The 'big five questionnaire': A new questionnaire to assess the five factor model," *Personality and Individual Differences*, vol. 15, no. 3, pp. 281–288, Sep. 1993, doi: 10.1016/0191-8869(93)90218-R.
[19] S. D. Gosling, P. J. Rentfrow, and W. B. Swann, "A very brief measure of the big-five personality domains," *Journal of Research in Personality*, vol. 37, no. 6, pp. 504–528, Dec. 2003, doi: 10.1016/S0092-6566(03)00046-1.
[20] G. Bradski, "The OpenCV library," *Dr. Dobb's Journal of Software Tools*, 2020.
[21] C. Tomasi and R. Manduchi, "Bilateral filtering for gray and color images," in *Sixth International Conference on Computer Vision (IEEE Cat. No.98CH36271)*, 1998, pp. 839–846, doi: 10.1109/ICCV.1998.710815.
[22] N. Chaki, S. H. Shaikh, and K. Saeed, "A comprehensive Survey on image binarization techniques," *Exploring Image Binarization Techniques*, pp. 5–15, 2014, doi: 10.1007/978-81-322-1907-1_2.
[23] S. Suzuki, "Topological structural analysis of digitized binary images by border following," *Computer vision, graphics, and image processing*, vol. 30, no. 1, pp. 32–46, 1985.
[24] Y. Freund and R. E. Schapire, "A decision-theoretic generalization of on-line learning and an application to boosting," *Journal of Computer and System Sciences*, vol. 55, no. 1, pp. 119–139, Aug. 1997, doi: 10.1006/jcss.1997.1504.

[25] T. Chen and C. Guestrin, "XGBoost: a scalable tree boosting system," in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, Aug. 2016, pp. 785–794, doi: 10.1145/2939672.2939785.

[26] L. Breiman, "Random forests," *Machine Learning*, vol. 45, no. 1, pp. 5–32, 2001, doi: 10.1023/A:1010933404324.

[27] C. J. C. Burges, "A tutorial on support vector machines for pattern recognition," *Data Mining and Knowledge Discovery*, vol. 2, no. 2, pp. 121–167, 1998, doi: 10.1023/A:1009715923555.

[28] A. K. Jain, Jianchang Mao, and K. M. Mohiuddin, "Artificial neural networks: a tutorial," *Computer*, vol. 29, no. 3, pp. 31–44, Mar. 1996, doi: 10.1109/2.485891.

[29] V. N. G. Raju, K. P. Lakshmi, V. M. Jain, A. Kalidindi, and V. Padma, "Study the influence of normalization/transformation process on the accuracy of supervised classification," in *2020 Third International Conference on Smart Systems and Inventive Technology (ICSSIT)*, Aug. 2020, pp. 729–735, doi: 10.1109/ICSSIT48917.2020.9214160.

[30] F. Pedregosa and E. Al., "Scikit-learn: machine learning in python," *The Journal of machine Learning research*, vol. 12, pp. 2825–2830, 2011.

[31] I. Sutskever, J. Martens, G. Dahl, and G. Hinton, "On the importance of initialization and momentum in deep learning," in *30th International Conference on Machine Learning, ICML 2013*, 2013, pp. 2176–2184.

[32] A. Prieto *et al.*, "Neural networks: An overview of early research, current frameworks and new challenges," *Neurocomputing*, vol. 214, pp. 242–268, Nov. 2016, doi: 10.1016/j.neucom.2016.06.014.

BIOGRAPHIES OF AUTHORS

Ahmed Remaida ⓘ 🗗 ᴄ ⓒ is a Ph.D. student at the laboratory of engineering science, department of informatics, logistics and mathematics at the National School of Applied Sciences at Ibn Tofail University in Kenitra. He has a master's degree in educational technology from the "école normale supérieure, ENS" of Martil, Morocco. His research interests include deep machine learning and personality assessment. He can be contacted at email: ahmed.remaida@uit.ac.ma.

Benyoussef Abdellaoui ⓘ 🗗 ᴄ ⓒ is a Ph.D. student at the laboratory of engineering science, department of informatics, logistics and mathematics at the national school of applied sciences at Ibn Tofail University in Kenitra. He has a master's degree in internet of things IoT from the National School of Computer Science and Systems Analysis (ENSIAs) one of the ten establishments of the University Mohammed V, Morocco. His research interests include face detection, emotional state, and engagement with deep learning. He can be contacted at email: benyoussef.abdellaoui@uit.ac.ma.

Mohamed Amine Lafraxo ⓘ 🗗 ᴄ ⓒ is a Ph.D. research scholar in health sciences and medical informatics, at the laboratory of biology and health at the faculty of sciences, Ibn Tofail University. He has a Master's degree in educational technology from the "École Normale Supérieure, ENS" of Martil, Morocco in 2016, and two bachelor's degrees in computer engineering and medical informatics. His research interests include e-learning, software engineering, big data, biometry, epidemiology, public health, stress, and burnout syndrome. He can be contacted at email: lafraxo.ma@gmail.com.

Zineb Sabri ⓘ 🗗 ᴄ ⓒ is a Ph.D. student at the laboratory of research in computer science at faculty of sciences, Ibn Tofail University in Kenitra. She has obtained her master's degree in information systems engineering from University Cadi Ayad, Morocco. Her research disciplines are education, games, machine learning, and employability. She can be contacted at email: zineb.sabri@uit.ac.ma.

Hamza Nouib 🆔 🔗 SC ⟳ Hamza Nouib is a Ph.D. student at the laboratory of engineering science, department of informatics, logistics and mathematics at the national school of applied sciences at Ibn Tofail University in Kenitra. His research interests include machine learning and employability. He can be contacted at email: hamza.nouib@uit.ac.ma.

Younes El Bouzekri El Idrissi 🆔 🔗 SC ⟳ is a research scholar, he is a member of the laboratory of engineering science, department of informatics, logistics and mathematics. at the national school of applied sciences at Ibn Tofail University in Kenitra. He can be contacted at email: y.elbouzekri@gmail.com.

Aniss Moumen 🆔 🔗 SC ⟳ is a research scholar, he is a member of the laboratory of engineering science, department of informatics, logistics and mathematics. At the National School of Applied Sciences at Ibn Tofail University in Kenitra. He can be contacted at email: aniss.oumoumen@uit.ac.ma.