Journal of Applied Informatics and Computing (JAIC)
Vol.10, No.2, April 2026, pp. 1220~1228
e-ISSN: 2548-6861

Big Five Prediction from Handwriting Images Using ResNet-50

Januar Firnando [1]*, Derry Alamsyah [2]*
* Informatics Study Program, Multi Data Palembang University
januar.firnando@mhs.mdp.ac.id [1], derry@mdp.ac.id [2]

Article Info

Article history:

Received 2026-02-04
Revised 2026-02-23
Accepted 2026-04-08

Keyword:

Big Five,
CNN,
Handwriting,
ResNet-50.

ABSTRACT

Personality plays an important role in determining job performance, as mismatches between personality traits and job roles can lead to disengagement and reduced work performance. One of the most widely accepted models for understanding personality is the Five-Factor Model (FFM) or Big Five personality traits, which provides a comprehensive yet parsimonious representation of human personality. Although the Big Five is commonly measured using self-report questionnaires, this method is prone to response bias. Therefore, alternative approaches based on behavioral cues, such as graphology, have gained research attention. This study proposes a CNN-based approach using the ResNet architecture to predict Big Five personality traits from offline handwriting images. The objectives of this study are to evaluate the effectiveness of the ResNet-based model, analyze the impact of Otsu Thresholding preprocessing under normal and uneven lighting conditions, and determine optimal hyperparameter configurations. Experimental results show that the proposed model can learn personality-related patterns from handwriting images, with the best performance achieved using original images without preprocessing, resulting in an MAE of 7.94, RMSE of 10.71, and three-class classification accuracy of 66%. The findings indicate that Otsu Thresholding does not consistently improve performance and may remove important handwriting details. Overall, the results demonstrate that CNN-based ResNet models are effective for offline handwriting-based Big Five personality prediction.

This is an open access article under the CC–BY-SA license.

I. INTRODUCTION

Personality refers to a set of fundamental characteristics that influence how individuals perceive situations, regulate emotions, and behave. Every individual exhibits a unique combination of traits that defines their behavioral patterns. In organizational settings, a candidate's personality plays a significant role in determining job performance [1]. Employees whose personality traits are mismatched with their job roles are more likely to experience disengagement and reduced performance. This mismatch can disrupt teamwork and increase the costs of recruitment and training, which in turn reduces the overall effectiveness of an organization. [2].

The Five-Factor Model (FFM), also known as the Big Five and introduced by McCrae and Costa, is one of the most prominent frameworks used in the study of personality. This model describes personality as a multidimensional construct that reflects individual's emotional responses, cognitive processes, and behavioral tendencies. The FFM is widely used to explain personality traits and their relationship with behavior because it is straightforward to apply while still providing consistent and reliable results across different cultures, age groups, and time periods. [3].

The Big Five Personality framework identifies five fundamental dimensions of personality, namely Openness, Conscientiousness, Extraversion, Agreeableness, and Neuroticism. [4]. Conscientiousness reflects an individual's tendency to be disciplined, well-organized, persistent, and focused on achieving goals, whereas neuroticism refers to a disposition toward emotional vulnerability, including feelings such as anxiety, sadness, irritability, and emotional tension. High levels of neuroticism are commonly associated with emotional instability and difficulty in regulating negative emotions. Extraversion, on the other hand, describes a personality orientation characterized by sociability, enthusiasm, and active engagement in social interactions. Openness to experience represents a tendency toward imaginative thinking, intellectual curiosity, creativity, and openness to new ideas and perspectives. Lastly, agreeableness is associated with prosocial qualities, including kindness, empathy, trustfulness, generosity, and a cooperative attitude toward others [5].

Although Big Five personality traits are typically measured through self-report questionnaires, this approach is intrinsically prone to various forms of response bias. Such biases include acquiescence, in which participants systematically agree with items irrespective of their meaning, as well as social desirability bias, where responses are shaped by socially approved expectations rather than reflecting individuals' genuine personality traits [6]. As a result, alternative methods that rely on behavioral or expressive cues have gained increasing attention. One such approach is graphology.

Graphology is the psychological study of handwriting aimed at identifying an individual's personality traits, psychosocial characteristics, temperament, and behavioral patterns. Handwriting is considered a form of expressive behavior that reflects underlying cognitive, emotional, and neurological processes, allowing personality-related information to be inferred from written patterns [7]. As a complex neurophysiological activity, handwriting involves coordinated motor, cognitive, and emotional mechanisms, extending beyond its primary communicative function.

Within graphology, features such as size, pressure, slant, and spacing are traditionally interpreted as possible indicators of personality traits [8]. However, despite its application in various practical settings, the scientific validity of graphological assessments remains controversial and has not been conclusively supported by empirical evidence [9]. Therefore, rather than adopting symbolic or rule-based interpretations, this study employs a data-driven deep learning approach to investigate whether statistically learnable visual patterns in handwriting images are associated with Big Five personality traits.

Handwriting data acquisition methods can be classified based on the manner in which the data are collected, namely online and offline approaches. In the online method, handwriting is captured using specialized devices such as tablets or digital pens, allowing dynamic information to be recorded during the writing process [10]. In contrast, the offline method relies on handwriting samples written on paper, which are subsequently digitized using scanners or digital cameras. As a result, offline handwriting data suffer from information loss because they are limited to static two-dimensional images and require more complex image processing techniques [11].

Moreover, offline handwriting images are susceptible to variations in image quality, lighting conditions, and background noise, which can negatively impact the performance of computational models [12]. To address these challenges, preprocessing techniques such as binarization are commonly applied. Otsu Thresholding is one of the most widely used automatic thresholding methods for image binarization, which separates foreground from background by determining an optimal threshold value [13]. Otsu thresholding is employed in this work to reduce illumination related variations in handwriting images, enabling an analysis of how such preprocessing affects the personality prediction results.

To reduce subjectivity and improve consistency, recent studies have increasingly explored computational approaches for handwriting-based personality analysis. Convolutional Neural Networks (CNNs) are extensively utilized in visual data processing as they are capable of learning hierarchical features directly from input images without manual feature design. Through convolutional operations and shared filters, CNNs can effectively capture spatial patterns while remaining robust to variations in feature location within an image. These properties make CNNs particularly suitable for tasks involving visual pattern recognition and image-focused machine learning applications [14] [15].

Several studies have applied Convolutional Neural Networks (CNNs) for handwriting-based personality classification. One study investigated offline handwriting images collected from university students, consisting of handwriting samples in two different languages. The experiments were conducted separately on each language and on a combined dataset, yielding classification accuracies of approximately 42–43% across the three scenarios [16]. Another study utilized CNN-based architectures on a large-scale public handwriting dataset comprising over 15,518 handwritten images of the word "AND," categorized into 15 distinct classes. Several CNN models were developed to capture different feature representations and were organized into three model groups. The evaluation results demonstrated that the best-performing model achieved a classification accuracy of 80.88% [17]. The results indicate that CNN-based methods have strong potential in the field of handwriting analysis.

Despite these promising results, existing CNN-based approaches for handwriting-based personality prediction still face several limitations. The study by [16] achieved relatively low classification accuracy, indicating challenges in effectively learning personality-related features from handwriting. While the study by [17] demonstrated higher accuracy, it was limited to a single-word dataset which may not adequately capture the diversity and complexity of natural handwriting patterns. Furthermore, previous studies have not systematically investigated the effects of image preprocessing techniques or varying data quality conditions on model performance. These gaps motivate the need for a more comprehensive evaluation of CNN-based personality prediction systems.

To address the limitations of previous approaches and enhance feature extraction capability, this study adopts a Residual Network (ResNet) architecture. ResNet introduces residual learning through skip connections, which enable the model to learn complex images representation across multiple abstraction levels while mitigating gradient vanishing problems commonly encountered in deep neural networks [18]. This characteristic makes ResNet particularly suitable for capturing subtle graphological features from handwriting images that may correlate with personality traits.

Based on the limitations of questionnaire-based personality assessments and the gaps identified in existing computational approaches, this study proposes a comprehensive evaluation of CNN-based personality prediction from offline handwriting images. The proposed method leverages a Residual Network (ResNet) architecture combined with systematic hyperparameter tuning to optimize prediction performance.

The primary objectives of this research are threefold: (1) to evaluate the effectiveness of a ResNet-based CNN model in predicting the five dimensions of Big Five personality traits from offline handwriting images, (2) to investigate the impact of Otsu Thresholding preprocessing on model performance under normal and uneven lighting conditions, and (3) to identify optimal hyperparameter configurations that maximize prediction accuracy. By addressing these objectives, this study aims to provide a more robust, scalable, and objective alternative for personality assessment in organizational and recruitment contexts, while contributing empirical insights into the role of image preprocessing in handwriting-based personality prediction systems.

II. METODE

This study applies a convolutional neural network (CNN)–based method to analyze the relationship between offline handwriting images and the Big Five personality traits. CNNs are selected because they can automatically learn hierarchical visual representations from image data, which makes them well suited for handwriting analysis. The proposed methodology consists of several main stages, including offline handwriting data acquisition, image preprocessing, CNN-based model construction using a Residual Network (ResNet) architecture, and performance evaluation. An overview of the complete workflow is illustrated in Figure 1.

A. Dataset

The dataset used in this study consists of 100 primary samples collected from undergraduate students, as shown in Figure 2. Participants were recruited voluntarily and completed the data collection process in two stages.

In the first stage, personality assessment was conducted using a publicly available online Big Five personality test based on the Five-Factor Model (FFM). Each personality dimension was scored on a percentage scale ranging from 0 to 100, with higher values indicating stronger expression of the corresponding trait. Participants completed the assessment independently and submitted their results, providing continuous scores for each of the five personality dimensions.

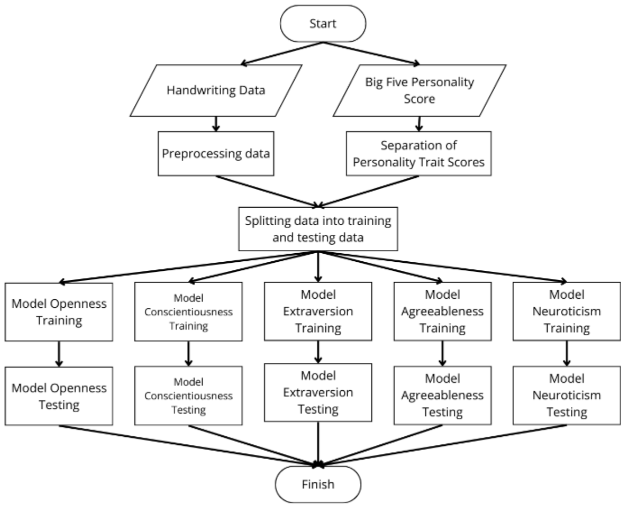


Figure 1. Research Workflow

In the second stage, handwriting samples were collected from the same participants. Each participant was instructed to write a predefined text on plain white A4 paper using a standard ballpoint pen. The handwritten documents were digitized using a digital camera and stored in PNG format for subsequent processing.

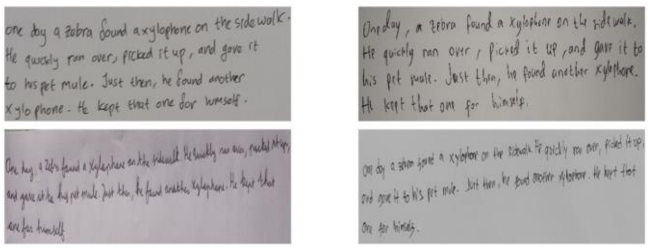


Figure 2. Example of Handwriting Dataset

TABLE 1
EXAMPLE OF BIG FIVE SCORE DATASET

Filename	O	C	E	A	N
img_001	60	69	40	54	46
img_002	81	60	73	62	44
img_003	85	23	62	77	65
img_004	69	60	44	58	46
img_005	56	42	62	31	78

Note: O = Openness, C = Conscientiousness, E = Extraversion, A = Agreeableness, N = Neuroticism

Personality labels were represented using the Big Five Personality Traits, namely Openness, Conscientiousness, Extraversion, Agreeableness, and Neuroticism. Each trait score ranges from 0 to 100, indicating the degree to which an individual exhibits each personality dimension. The handwriting images and personality scores were paired on a per-participant basis and organized into a structured dataset.

An example of the Big Five personality score dataset is shown in Figure 3.

All participants were informed about the purpose of the study and provided informed consent prior to data collection. The collected data were anonymized to ensure that no personally identifiable information was associated with the handwriting samples or personality scores. The study was conducted in accordance with ethical research principles for psychological data handling, ensuring confidentiality and voluntary participation throughout the process.

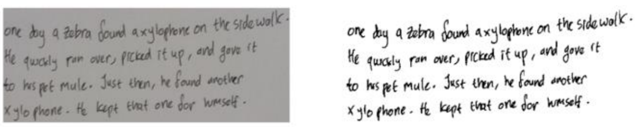
The dataset was divided into training and testing sets using a 90:10 split, resulting in 90 samples for training and 10 samples for testing. The test set was strictly held out during training and used exclusively for performance evaluation.

B. Preprocessing

Before being processed by the model, the collected handwriting images were subjected to several preprocessing steps to improve image quality and reduce irrelevant information. First, image cropping was applied to remove unnecessary background areas around the handwriting. This step helps minimize noise that could interfere with the learning process and potentially reduce prediction accuracy.

After cropping, all images were resized to a fixed resolution of 560 × 224 pixels. This size was chosen to maintain the original rectangular aspect ratio of the handwriting images, which is approximately 5:2. Resizing the images to a square format, such as 224 × 224 pixels commonly used in ResNet-based models, would distort the handwriting structure. Therefore, a rectangular input size was considered more suitable for preserving stroke proportions.

Next, the images were converted to grayscale to simplify the input data and focus on the structural patterns of handwriting. The grayscale images were then transformed into binary images using Otsu thresholding, an automatic thresholding technique that converts a grayscale image into a binary image based on calculated pixel intensity values. The reason for using the Otsu method is its effectiveness in producing an appropriate threshold to distinguish objects from their background [19]. The results of the Otsu thresholding process can be seen in Figure 3.



(a) (b)

Figure 3. Effect of Otsu Thresholding preprocessing: (a) Original Image Handwriting, (b) Handwriting Image after Otsu Thresholding

C. Model

This study employs a convolutional neural network (CNN) model utilizing the ResNet architecture for feature extraction and classification. In general, a CNN architecture is composed of several layers, including convolutional layers, subsampling (pooling) layers, and fully connected layers. The overall structure of the CNN architecture is presented in Figure 4.

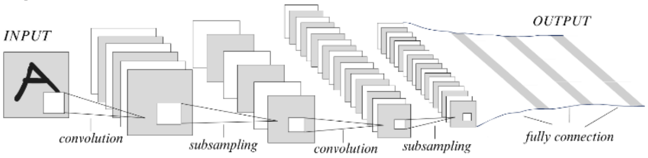


Figure 4. CNN Structure

The convolutional layers extract features from the input image while preserving spatial relationships between pixels, enabling the model to learn various visual characteristics such as edges, curves, and other discriminative features through the use of filters (kernels) that generate different feature maps. Subsampling layers are used to downsample the feature maps, which helps reduce their spatial size and lowers the number of parameters involved in the training process. Subsequently, the fully connected layers carry out the classification task by using the learned features to identify the most suitable class and generate probability scores for each output category [20].

One of the CNN architectures developed to enhance learning in deeper networks is the Residual Network (ResNet). ResNet adopts a residual learning framework by introducing shortcut connections between layers, allowing direct information flow and facilitating optimization. This design addresses common challenges in deep neural network training, such as information loss and performance degradation, enabling ResNet to maintain high accuracy even as network depth increases [21]. ResNet is available in multiple variants, such as ResNet-18, ResNet-34, ResNet-50, ResNet-101, and ResNet-152 [22], each distinguished by its network depth. This study employs ResNet-50, a deep convolutional neural network composed of 50 layers [23]. The detailed architecture of ResNet-50 is presented in Figure 5.

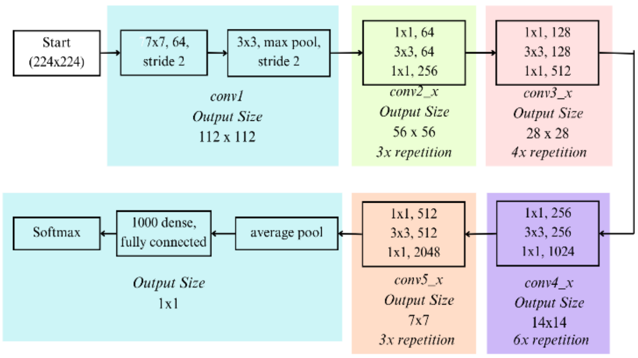


Figure 5. Architecture of ResNet-50

Transfer learning was employed to enhance learning efficiency and mitigate the limitations of a relatively small dataset. The ResNet-50 architecture was initialized with pretrained weights from ImageNet, a large-scale dataset containing approximately 1.2 million images. By leveraging previously learned generic visual representations, the model benefits from robust feature extraction capabilities that have been trained on diverse image patterns. This approach reduces the need for extensive domain-specific data and allows the network to adapt more effectively to the handwriting-based personality prediction task [24].

D. Training Strategy

To evaluate the performance of the proposed approach in predicting personality traits from offline handwriting images, a series of experiments was conducted using different hyperparameter configurations. The experiments involved varying the number of neurons in the first dense layer (128, 256, and 512), the number of neurons in the second dense layer (128, 256, and 512), the batch size (2, 4, and 6), and the number of training epochs (10, 20, and 30). These combinations resulted in a total of 27 experimental configurations for each personality trait model.

Considering that the Big Five Personality model consists of five distinct personality dimensions, five independent CNN-based models were trained, where each model was designed to predict one specific personality trait. This independent modelling strategy allows each network to focus on learning trait-specific handwriting patterns without being influenced by other dimensions. All models were trained using identical experimental settings to ensure consistency and fair comparison across personality traits.

To analyze model robustness and the impact of preprocessing, model performance was evaluated under four different data quality scenarios: (1) original handwriting images without preprocessing, (2) handwriting images processed using Otsu thresholding, (3) handwriting images with uneven illumination without preprocessing, and (4) handwriting images with uneven illumination processed using Otsu thresholding. These experimental scenarios were designed to investigate whether Otsu thresholding improves model performance under varying input conditions.

E. Evaluation

The evaluation stage is conducted to measure how well the proposed model predicts Big Five personality traits based on handwriting images. The model's performance is assessed using a combination of regression metrics and categorical classification measures to obtain a thorough evaluation of its prediction ability. Regression accuracy is quantified using Mean Absolute Error (MAE) and Root Mean Square Error (RMSE), both of which are widely applied indicators for assessing regression-based models.

Mean Absolute Error (MAE) measures the average magnitude of the differences between the predicted values and the actual ground truth values.. This metric provides an intuitive and fair evaluation of model performance, as all prediction errors are treated equally regardless of their magnitude [25]. MAE is particularly suitable when an overall measure of prediction accuracy is required without emphasizing extreme errors. The MAE is calculated using Equation (1).

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|f_i - y_i| \quad (1)$$

In addition to MAE, Root Mean Square Error (RMSE) is employed to quantify the difference between predicted outputs and their corresponding actual values. Because RMSE squares the error terms before averaging, it places greater emphasis on larger discrepancies, making it effective for detecting models that generate substantial prediction errors. The RMSE formula is shown in Equation (2).

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(f_i - y_i)^2} \quad (2)$$

Besides regression evaluation, a category-based classification evaluation was conducted to analyze the model's ability to distinguish personality trait levels. Both the predicted scores and the ground-truth Big Five scores, ranging from 0 to 100, were converted into three personality categories: low, moderate, and high.

The categorization follows the standard interpretation of the Big Five personality scale, where scores greater than or equal to 60 indicate a high tendency toward a particular personality trait, scores between 40 and 60 represent a moderate tendency, and scores less than or equal to 40 indicate a low tendency. These threshold values were determined based on expert consultation to ensure alignment with practical interpretation of personality assessment results. After conversion, classification accuracy was computed by comparing the predicted categories with the corresponding ground-truth categories. This transformation allows the regression outputs to be interpreted categorically and provides additional insight into the model's discriminative capability.

Model performance in the classification task was evaluated using accuracy. Accuracy measures how correctly a model classifies objects. Accuracy is computed by comparing the number of correct predictions to the total predictions produced by the model. This measure reflects the proportion of predictions that are correctly classified out of all predicted outcomes [26]. The formula used to calculate model accuracy is presented in Equation (3).

$$accuracy = \frac{sum\ of\ right\ prediction}{sum\ of\ all\ data} \quad (3)$$

III. RESULT AND DISCUSSION

In the Results and Discussion section, the model's performance in predicting the five dimensions of the Big Five personality traits based on handwritten images is evaluated. Five independent ResNet-based CNN models were trained and tested across 27 hyperparameter configurations under four data quality scenarios: raw data, Otsu-preprocessed data, noisy data (uneven lighting), and noisy data with Otsu preprocessing, as described in the Methodology section. The model's performance was assessed using the Mean Absolute Error (MAE) as the primary metric, along with Root Mean Square Error (RMSE) and three-class classification accuracy as supporting metrics. This analysis aims to evaluate the overall performance of the model and to examine the influence of applying Otsu Thresholding on prediction accuracy across different data conditions.

A. Model Perfomance on Raw Data

Based on the evaluation results on raw handwritten data without preprocessing that are presented in Table 2, differences in the model's performance in predicting each of the Big Five personality traits are observed. Using Mean Absolute Error (MAE) as the primary metric, the Agreeableness trait exhibits the best performance with an MAE value of 5.76 and an RMSE of 7.69, indicating that the model can learn the characteristics of this trait with relatively high accuracy. Conversely, the Extraversion trait shows the lowest performance with an MAE of 12.84 and an RMSE of 17.11, along with a three-class classification accuracy of 51%. These results suggest that Extraversion is the most challenging trait to predict from handwritten images compared to the other traits.

Although the Openness trait does not have the lowest MAE, it yields the highest classification accuracy of 81.11%. This indicates that, despite relatively larger numerical prediction errors, the model still manages to distinguish the Openness personality categories quite well. The Conscientiousness trait demonstrates moderate performance with an MAE of 9.25 and an RMSE of 12.82, achieving a classification accuracy of 58%. Meanwhile, Neuroticism shows an MAE of 7.16 and an RMSE of 9.01, with a classification accuracy of 53%.

TABLE 2
PERFORMANCE METRICS ON RAW HANDWRITING DATA

Personality Trait	MAE	RMSE	Accuracy
Openness	9.17	11.51	81%
Conscientiousness	9.25	12.82	58%
Extraversion	12.84	17.11	51%
Agreeableness	5.76	7.69	72%
Neuroticism	7.16	9.01	53%

Overall, the testing results on raw data demonstrate that the model's performance varies across personality traits, with Agreeableness being the most consistently predicted trait and Extraversion being the most challenging for the model

B. Model Performance on Data with Otsu Thresholding Preprocessing

As shown in Table 3, the evaluation results show that applying Otsu Thresholding preprocessing does not consistently improve the model's performance in predicting personality traits from handwritten images. Several traits actually experienced performance declines compared to the raw data without preprocessing.

TABLE 3
PERFORMANCE METRICS ON DATA WITH OTSU THRESHOLDING PREPROCESSING

Personality Trait	MAE	RMSE	Accuracy
Openness	8.44	10.77	72%
Conscientiousness	11.28	15.56	52%
Extraversion	11.98	16.46	50%
Agreeableness	6.85	8.32	62%
Neuroticism	8.46	10.13	41%

Specifically, the Conscientiousness trait exhibits a noticeable decrease in performance, with MAE increasing by 2.03, RMSE rising by 2.74, and classification accuracy dropping by 6%. Similarly, Neuroticism shows degraded performance with MAE increasing by 1.30, RMSE by 1.12, and accuracy decreasing significantly by 13%. The Agreeableness trait, despite maintaining relatively good performance, experienced an increase in MAE by 1.09 and RMSE by 0.63. However, its classification accuracy showed a decline of 10%. The Extraversion trait shows marginal improvement with MAE decreasing by 0.86 and RMSE by 0.65, although classification accuracy remained stable with no significant change. Interestingly, the Openness trait demonstrated improvement with MAE decreasing by 0.73 and RMSE by 0.74 after preprocessing, though its classification accuracy dropped by 8%.

Overall, the average model performance on data processed with Otsu Thresholding indicates increased errors, with average MAE of 9.40 compared to 8.84 on raw data, and decreased accuracy of 57% compared to 63% on raw data. These findings suggest that binarization of images using Otsu Thresholding potentially removes important information from handwriting patterns, negatively affecting the model's ability to comprehensively learn personality characteristics.

C. Model Perfomance on Noisy Data

The performance evaluation on noisy data, as presented in Table 4, reveals that compared to clean handwriting images, uneven lighting conditions generally degrade the model's performance across most traits, indicating the sensitivity of offline handwriting analysis to illumination variations.

TABLE 4
PERFORMANCE METRICS ON NOISY DATA

Personality Trait	MAE	RMSE	Accuracy
Openness	9.66	11.72	78%
Conscientiousness	10.06	13.31	46%
Extraversion	12.78	16.56	47%
Agreeableness	5.87	8.01	67%
Neuroticism	8.64	10.81	43%

The model's performance on data with uneven lighting conditions tends to decline, particularly for the traits Conscientiousness and Neuroticism. Conscientiousness experienced an increase in MAE of 0.81, RMSE of 0.49, and a decrease in category accuracy of 12%. Meanwhile, Neuroticism showed an increase in MAE of 1.47, RMSE of 1.8, and a decrease in category accuracy of 11%. Despite this, the Extraversion trait maintained relatively stable performance, with a slight increase in MAE of 0.06, RMSE of 0.55, and a decrease in category accuracy of 4%. Overall, the model's performance on the data with uneven lighting conditions exhibited an increase in MAE of 0.56, RMSE of 0.46, and a 7% decline in accuracy compared to the original data without such conditions.

When Otsu Thresholding was applied to handwriting images under uneven lighting, mixed results were observed across personality traits, as shown in Table 5. as shown in Table X. The preprocessing effect differed substantially from its impact on clean data, revealing complex interactions between lighting conditions and binarization.

TABLE 5
PERFORMANCE METRICS ON NOISY DATA WITH OTSU THRESHOLDING PREPROCESSING

Personality Trait	MAE	RMSE	Accuracy
Openness	8.96	11.19	72%
Conscientiousness	11.04	15.67	46%
Extraversion	12.8	17.25	47%
Agreeableness	4.55	5.86	83%
Neuroticism	11.64	14.85	42%

The Openness trait experienced a decrease in MAE of 0.70, a decrease in RMSE of 0.58, and an increase in accuracy of 6%. Meanwhile, the Agreeableness trait showed a decrease in MAE of 1.31, RMSE of 2.35, and an increase in accuracy of 1%. However, several other traits exhibited a decline in model performance, as indicated by increases in MAE and RMSE values and a decrease in accuracy following preprocessing, particularly for Conscientiousness and Neuroticism traits. Conscientiousness experienced an MAE increase of 0.98 and RMSE increase of 2.36, with its accuracy remaining unchanged. Neuroticism showed an MAE increase of 3, RMSE increase of 4.04, and a 1% decrease in accuracy.

Overall, the model's performance deteriorated after applying Otsu Thresholding preprocessing to data with uneven lighting, evidenced by increases in MAE of 0.41 and RMSE of 0.83. However, the classification accuracy improved by 2%.

D. Analysis of Result

A comparative analysis of average error metrics, namely MAE and RMSE, across all Big Five personality traits under different experimental scenarios is presented in Figures X and X. The results indicate that both Otsu Thresholding preprocessing and the presence of uneven illumination generally lead to increased prediction errors compared to the original, unprocessed data. The highest error values are observed in the scenario combining uneven illumination with Otsu Thresholding, yielding an average MAE of 9.57 and RMSE of 12.96. Conversely, the lowest error values are consistently achieved using the original data without preprocessing, with an average MAE of 8.84 and RMSE of 11.63.

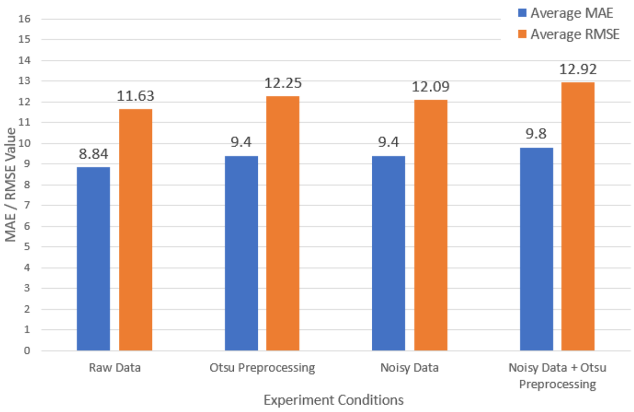


Figure 6. Comparison of Average Error Metrics Across Different Experimental Conditions

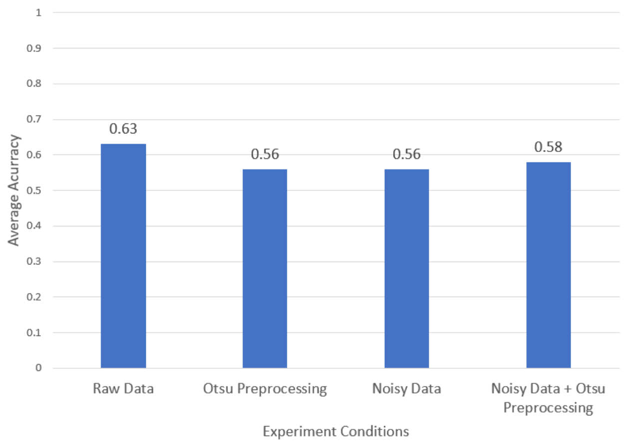


Figure 7. Comparison of Average Accuracy Classification Across Different Experimental Conditions

The comparison of average classification accuracy across the evaluated scenarios shows that the highest accuracy, reaching 63%, was obtained with the original data without preprocessing. Conversely, scenarios involving Otsu Thresholding and uneven illumination resulted in a noticeable decrease in classification accuracy, with the lowest accuracy of 56% observed in the noisy data condition. The combination of noise and Otsu preprocessing yielded an accuracy of 58%, showing slight improvement over noisy data alone but still below the raw data baseline. These findings suggest that alterations in pixel intensity distribution caused by preprocessing or illumination disturbances negatively impact the stability of categorical personality predictions.

Overall, the results demonstrate that Otsu Thresholding does not offer a performance advantage for handwriting based personality prediction using convolutional neural networks. While some minor improvements were observed for particular traits under specific conditions, the general trend indicates increased prediction errors and decreased classification accuracy.

Among all the hyperparameter configurations tested with the original handwriting data, the best-performing model was achieved with 512 neurons in both dense layers, a batch size of 4, and 20 training epochs. This configuration obtained the lowest average prediction error across the five Big Five traits, with an MAE of 7.94, an RMSE of 10.71, and a classification accuracy of 66%. While this result indicates a moderate level of discriminative performance, it suggests that the model is capable of capturing certain statistically learnable visual patterns from handwriting images.

Building upon these findings, the proposed ResNet50-based approach achieves 66% classification accuracy on the five Big Five personality dimensions, representing a substantial improvement over the study by [16], which reported accuracies of 42–43% using CNN on bilingual handwriting datasets. While our performance is lower than the 80.88% reported by [17], several important differences in the experimental setup should be taken into consideration. In particular, [17] employed a single-word dataset ("AND") with 15 classes focused on writer identification rather than personality trait prediction, which represents a fundamentally different task with potentially less complex feature requirements. In contrast, our study addresses the more challenging problem of predicting psychological constructs across five continuous personality dimensions from natural handwriting samples.

Nevertheless, the predictive capability of the model is constrained by several limitations. The relatively limited dataset size, possible subjectivity in self-reported personality assessments, and the transformation of continuous regression outputs into three categorical classes may increase the risk of misclassification, especially around decision thresholds. Furthermore, potential cultural and stylistic influences in handwriting must be taken into account. As all participants were recruited from a single private university, the dataset may not sufficiently capture the diversity of writing styles shaped by varying cultural backgrounds, educational experiences, or regional conventions. Variations in handwriting structure, stroke patterns, and stylistic tendencies across different populations could therefore affect the model's generalizability and introduce unintended bias in predicting personality traits.

These limitations are further compounded by the validation approach employed: model performance was evaluated using a 90:10 train-test split with a fixed random seed, where 90 samples were used for training and only 10 samples were reserved for testing. The small test set size provides limited statistical power for reliable performance estimation and may result in high variance in reported metrics. Therefore, the reported accuracy should be interpreted cautiously as preliminary findings rather than definitive performance indicators. Future research should address these limitations by collecting substantially larger datasets that enable both adequate model training and more robust validation strategies, such as k-fold cross-validation or independent held-out test sets, to obtain more stable and generalizable performance estimates.

Another important limitation concerns model interpretability, the current study does not incorporate feature visualization or explainability analysis to identify which handwriting characteristics contribute most to personality predictions. As the model operates as a deep learning black box, the specific visual features learned by the network and their associations with personality traits remain unclear. This limitation is particularly evident when interpreting the preprocessing results. While Otsu Thresholding was found to degrade model performance—especially for certain traits such as Conscientiousness and Neuroticism—the underlying reasons for this degradation have not been deeply investigated.

The performance decline suggests that the binarization process may eliminate important visual information present in the grayscale images, such as variations in stroke intensity, subtle texture patterns, or other fine-grained characteristics that the model relies upon for prediction. However, systematically analyzing which specific features are lost during preprocessing and how they relate to different personality dimensions would require additional investigation beyond the scope of this study. Future research should explore methods to better understand both what the model learns from handwriting images and how different preprocessing operations affect these learned patterns, which could inform the development of more effective approaches for handwriting-based personality assessment.

Beyond technical limitations, the potential use of automated personality prediction systems in recruitment contexts raises important ethical considerations. Given the moderate predictive performance observed in this study, the proposed model should not be used as a sole decision-making tool in employment selection processes. Misclassification may result in unfair evaluation of candidates, particularly if personality predictions are interpreted as definitive indicators. Additionally, handwriting characteristics may be influenced by cultural, educational, or contextual factors unrelated to intrinsic personality traits. Therefore, any practical deployment of such systems should involve human oversight, transparency, and careful consideration of fairness and bias.

IV. CONCLUSION

This study assessed the performance of a CNN-based ResNet architecture in predicting Big Five personality traits using offline handwriting images across various data conditions. Experimental results show that the proposed model is capable of learning personality-related patterns from handwriting images, although prediction performance varies across personality traits. Using the original handwriting images without preprocessing, the proposed model achieved its best overall performance, with an MAE of 7.94, RMSE of 10.71, and three-class classification accuracy of 66% across the five personality traits. Among the evaluated traits, Agreeableness consistently exhibited the lowest prediction error, while Extraversion remained the most challenging trait to predict.

The impact of Otsu Thresholding as a preprocessing technique was also examined. The results indicate that Otsu Thresholding does not consistently improve model performance. Although minor improvements were observed for certain traits under specific conditions, the general trend shows increased prediction error and reduced classification accuracy compared to using the original handwriting images. This suggests that binarization may remove fine-grained stroke and texture information that is important for handwriting-based personality prediction. Furthermore, experiments conducted under uneven illumination conditions reveal that offline handwriting-based personality prediction is sensitive to lighting variations. The application of Otsu Thresholding under such conditions did not provide a reliable performance improvement.

Overall, the findings confirm that a CNN-based ResNet model can be effectively applied to offline handwriting images for Big Five personality prediction, with the most reliable performance obtained from original images without additional preprocessing.

REFERENCES

[1] H. Perera and L. Costa, “MBTI Personality Classification through Analyzing CVs/ Personal Statements for e-Recruitment,” *International Journal of Machine Learning*, vol. 14, no. 1, 2024, doi: 10.18178/ijml.2024.14.1.1153.

[2] H. W. Khan, M. U. Sattar, S. Noor, and M. I. Alyousef, “A Personality-Informed Candidate Recommendation Framework for Recruitment Using MBTI Typology,” *Information*, vol. 16, no. 10, p. 863, Oct. 2025, doi: 10.3390/info16100863.

[3] G. Angelini, “Big five model personality traits and job burnout: a systematic literature review,” *BMC Psychol.*, vol. 11, no. 1, p. 49, Feb. 2023, doi: 10.1186/s40359-023-01056-y.

[4] N. A. Utami, W. Maharani, and I. Atastina, “Personality Classification of Facebook Users According to Big Five Personality Using SVM (Support Vector Machine) Method,” *Procedia Comput. Sci.*, vol. 179, pp. 177–184, 2021, doi: 10.1016/j.procs.2020.12.023.

[5] S. Mustafa, Y. Qiao, X. Yan, A. Anwar, T. Hao, and S. Rana, “Digital Students’ Satisfaction With and Intention to Use Online Teaching Modes, Role of Big Five Personality Traits,” *Front. Psychol.*, vol. 13, Jul. 2022, doi: 10.3389/fpsyg.2022.956281.

[6] A. Salecha, M. E. Ireland, S. Subrahmanya, J. Sedoc, L. H. Ungar, and J. C. Eichstaedt, “Large language models display human-like social desirability biases in Big Five personality surveys,” *PNAS Nexus*, vol. 3, no. 12, Nov. 2024, doi: 10.1093/pnasnexus/pgae533.

[7] J. Tiwari and R. Sadiwala, “Personality Prediction from Handwriting using Fine-tuned Transfer Learning Models,” *SAMRIDDHI : A Journal of Physical Sciences, Engineering and Technology*, vol. 15, no. 01, pp. 38–44, Jan. 2023, doi: 10.18090/samriddhi.v15i01.06.

[8] G. Marano *et al.*, “The Neuroscience Behind Writing: Handwriting vs. Typing—Who Wins the Battle?,” ., vol. 15, no. 3, p. 345, Feb. 2025, doi: 10.3390/life15030345.

[9] C. Dazzi and L. Pedrabissi, “Graphology and Personality: An Empirical Study on Validity of Handwriting Analysis,” *Psychol. Rep.*, vol. 105, no. 3_suppl, pp. 1255–1268, Dec. 2009, doi: 10.2466/PR0.105.F.1255-1268.

[10] F. Alaei and A. Alaei, “Handwriting analysis: Applications in person identification and forensic,” *Breakthroughs in Digital Biometrics and Forensics*, pp. 147–165, 2022, doi: 10.1007/978-3-031-10706-1_7.

[11] J. F. Vargas, M. A. Ferrer, C. M. Travieso, and J. B. Alonso, “Off-line signature verification based on grey level information using texture features,” *Pattern Recognit.*, vol. 44, no. 2, pp. 375–385, Feb. 2011, doi: 10.1016/j.patcog.2010.07.028.

[12] S. Singh, N. K. Garg, and M. Kumar, “Feature extraction and classification techniques for handwritten Devanagari text recognition: a survey,” *Multimed. Tools Appl.*, vol. 82, no. 1, pp. 747–775, Jan. 2023, doi: 10.1007/s11042-022-13318-9.

[13] A. Riadi and R. Sulaehani, “Analisis implementasi preprocessing dengan Otsu-Gaussian pada pengenalan wajah.,” *ILKOM Jurnal Ilmiah*, vol. 11, no. 3, pp. 200–205, Dec. 2019, doi: 10.33096/ilkom.v11i3.457.200-205.

[14] D. Gagiu and D. Sendrescu, “Detection of Personality Traits Using Handwriting and Deep Learning,” *Applied Sciences*, vol. 15, no. 4, p. 2154, Feb. 2025, doi: 10.3390/app15042154.

[15] A. A. Elngar *et al.*, “Image classification based on CNN: A survey,” *Journal of Cybersecurity and Information Management (JCIM)*, vol. 6, no. 1, pp. 18–50, 2021.

[16] G. Chaubey and S. K. Arjaria, “Personality prediction through handwriting analysis using convolutional neural networks,” in *In Proceedings of International Conference on Computational Intelligence: ICCI 2020*, Singapore: Springer Singapore, 2022, pp. 59–70. doi: 10.1007/978-981-16-3802-2_5.

[17] D. Alamsyah, Samsuryadi, W. Widhiarso, and S. Hasan, “Handwriting analysis for personality trait features identification using CNN,” in *2022 International Conference on Data Science and Its Applications (ICoDSA)*, IEEE, Jul. 2022, pp. 232–238. doi: 10.1109/ICoDSA55874.2022.9862910.

[18] M. Shafiq and Z. Gu, “Deep Residual Learning for Image Recognition: A Survey,” *Applied Sciences*, vol. 12, no. 18, p. 8972, Sep. 2022, doi: 10.3390/app12188972.

[19] D. Indra, T. Hasanuddin, R. Satra, and N. R. Wibowo, “Eggs Detection Using Otsu Thresholding Method,” in *2018 2nd East Indonesia Conference on Computer and Information Technology (EIConCIT)*, IEEE, Nov. 2018, pp. 10–13. doi: 10.1109/EIConCIT.2018.8878517.

[20] X. Zhang, X. Zhang, and W. Wang, “Convolutional Neural Network,” in *Intelligent Information Processing with Matlab*, Singapore: Springer Nature Singapore, 2023, pp. 39–71. doi: 10.1007/978-981-99-6449-9_2.

[21] Rikendry and A. Maharil, “Perbandingan arsitektur VGG16 dan ResNet50 untuk rekognisi tulisan tangan aksara lampung,” *Jurnal Informatika dan Rekayasa Perangkat Lunak (JATIKA)*, vol. 3, no. 2, pp. 236–243, 2022, [Online]. Available: http://jim.teknokrat.ac.id/index.php/informatika

[22] Al Fathir Rizal Januar, J. Indra, D. S. Kusumaningrum, and S. Faisal, “Application of Convolutional Neural Network (CNN) Algorithm with ResNet-101 Architecture for Monkey Pox Detection in Human,” *Journal of Applied Informatics and Computing*, vol. 9, no. 3, pp. 1006–1012, Jun. 2025, doi: 10.30871/jaic.v9i3.9621.

[23] Q. A. Al-Haija and A. Adebanjo, “Breast Cancer Diagnosis in Histopathological Images Using ResNet-50 Convolutional Neural Network,” in *2020 IEEE International IOT, Electronics and Mechatronics Conference (IEMTRONICS)*, IEEE, Sep. 2020, pp. 1–7. doi: 10.1109/IEMTRONICS51293.2020.9216455.

[24] A. Shabbir *et al.*, “Satellite and scene image classification based on transfer learning and fine tuning of ResNet50,” *Math. Probl. Eng.*, vol. 2021, pp. 1–18, Jul. 2021, doi: 10.1155/2021/5843816.

[25] A. A. Suryanto, “Penerapan metode Mean Absolute Error (MAE) dalam algoritma regresi linear untuk prediksi produksi padi,” *SAINTEKBU*, vol. 11, no. 1, pp. 78–83, Feb. 2019, doi: 10.32764/saintekbu.v11i1.298.

[26] F. Rustam *et al.*, “Wireless capsule endoscopy bleeding images classification using CNN based model,” *IEEE Access*, vol. 9, pp. 33675–33688, 2021, doi: 10.1109/ACCESS.2021.3061592.