Journal of Artificial Intelligence and Engineering Applications

Website: https://ioinformatic.org/

June 2026. Vol. 5. No. 3; e-ISSN: 2808-4519

Classification of Handwriting Margin Patterns Using Ensemble Bagging Decision Tree

Rista Ifanka[1]*, Soffiana Agustin[2]

[1,2]*Department of Informatics, Faculty of Engineering, Universitas Muhammadiyah Gresik, Indonesia*
ristaifff@gmail.com

Abstract

Analysis of handwriting margins plays an important role in graphology, as margin patterns are often associated with individual behavioral tendencies and personality traits. Therefore, detecting and classifying margin characteristics is essential to support automated handwriting analysis using computational approaches. This study uses a computer vision-based approach to classify left-margin handwriting patterns into widening and narrowing categories. The classification is performed by analyzing margin characteristics extracted from scanned handwriting images. The processing pipeline consists of image preprocessing, hybrid feature extraction, and classification using an Ensemble Bagging Decision Tree model. The preprocessing stage enhances image quality through grayscale conversion, contrast adjustment, adaptive thresholding, and noise removal, followed by Region of Interest extraction to focus on the handwriting area. The feature extraction stage applies a hybrid strategy that combines line-based margin analysis and global spatial features to capture both local variations and overall structure. Model performance is evaluated using stratified 5-fold cross-validation to ensure reliable and unbiased results. The experimental findings show that the method achieves an average accuracy of 84.91%, with relatively balanced precision, recall, and F1-score across both classes. These results indicate that margin-based features are effective for representing handwriting patterns in classification tasks. However, variations in writing style and noise from the scanning process may influence performance. Overall, this study demonstrates that the applied approach provides reliable classification results and has potential for further improvement through feature expansion and more advanced learning models.

Keywords*: Bagging, Feature Extraction; Handwriting Analysis; Image Processing; Margin Classification*

1. Introduction

Personality represents a psychological construct that encompasses an individual's overall thinking patterns, behavioral tendencies, attitudes, and affective responses when confronting various life circumstances. A thorough understanding of one's personality dimensions yields broad practical benefits, ranging from appropriate workforce placement and self-potential development to the cultivation of more productive interpersonal relationships [1]. One academically established approach to non-verbal personality assessment is graphology, a discipline that examines the characteristic patterns of handwriting as a reflection of the writer's psychological condition. The scientific basis of graphology rests on the fact that the act of writing engages the neuromuscular system in ways that are largely governed by subconscious processes, causing the graphic traces left on the writing medium to implicitly encode the individual's emotional state and personality at the time of writing [2]. The practical relevance of graphology has been demonstrated across multiple domains, including talent identification, human resource selection, clinical psychological assessment, and forensic document examination [3].

Among the various graphological elements that have been studied, page margins stand out as one of the most informative indicators of a writer's psychological orientation. A margin is defined as the blank space between the edge of the paper and the beginning of the written strokes, and its formation is considered to reflect the manner in which a writer subconsciously manages the spatial dimensions of their life [1]. Within the graphological framework, the left margin is specifically associated with the writer's relationship to the past, while the right margin is linked to future orientation [1]. The dynamic shift in left margin position from the first to the last line of writing carries significant psychological meaning. A progressively widening left margin pattern signals growing enthusiasm, a tendency to distance oneself from past burdens, extroversion, and optimism; whereas a consistently narrowing pattern reflects an excessive preoccupation with past experiences, extreme caution, egocentrism, and anxiety about the future [1]. The richness of psychological information embedded in the spatial dynamics of the left margin makes it a highly relevant object of analysis for the development of computationally based graphology systems.

Conventional graphology practice conducted manually by trained experts suffers from several fundamental limitations. Manual measurement is time-consuming, not scalable for large datasets, and highly susceptible to inter-rater inconsistency, as two graphologists analyzing the same handwriting sample may arrive at differing interpretations [1]. These limitations have motivated the research community to develop automated graphology systems grounded in digital image processing and machine learning as a more objective and efficient alternative [4]. Such systems generally follow a workflow encompassing image acquisition, preprocessing, feature extraction, and personality inference through classification algorithms [5]. Within this pipeline, the quality of the preprocessing and feature extraction stages serves as the primary determinant of system success, particularly in producing valid and discriminative numerical representations of handwriting patterns [6].

A number of prior studies have explored the automation of graphological analysis with varying approaches and outcomes. Fadhilla et al. [7] developed a handwriting-based personality classification system incorporating left margin, right margin, letter size, and slant using neural network methods, where Learning Vector Quantization achieved 90% accuracy. Rosyidah and Rochmawati [8] built a personality analysis system based on six handwriting features using Support Vector Machine. Wijaya et al. [9] constructed an Android-based personality detection application relying on handwriting margin features using SVM with an average accuracy of 82.738%. Deore et al. [4] developed a Convolutional Neural Network-based system for real-time personality trait recognition from handwriting samples, attaining 82% accuracy. Esposito et al. [10] examined dynamic handwriting and drawing features from 62 participants and identified significant correlations between graphic characteristics and Big Five personality dimensions. Gumilang and Agustin [1] specifically classified left margin patterns of handwriting into four personality classes using Random Forest with geometry-based bounding box feature extraction, achieving 95% accuracy on the IAM Handwriting dataset. Despite these contributions, the majority of existing studies either combine multiple graphological characteristics simultaneously or employ globally oriented margin representations that fail to capture the per-line positional dynamics of the margin, which constitutes the core of graphological interpretation.

A review of the existing literature reveals several research gaps that remain unaddressed. Feature extraction approaches in prior works generally capture margin representation only at a global level, without recording the line-by-line positional changes in the left margin that are central to the graphological interpretation of widening and narrowing patterns. Furthermore, no adaptive feature extraction mechanism has been established for situations in which line segmentation partially fails due to real-world image quality variations, causing problematic images to yield invalid feature representations. The class distribution imbalance problem, which is commonly encountered in real-world handwriting datasets, has also not been adequately handled; whereas empirical evidence consistently demonstrates that combining oversampling techniques with ensemble models achieves significantly better classification performance under imbalanced conditions than single-model approaches [11][12]. These gaps have not been comprehensively addressed by existing research.

Drawing from the identified research gaps, this study constructs an automated classification system for left margin patterns of handwriting into two personality classes, namely widening and narrowing, by implementing a hybrid feature extraction approach that adaptively integrates a line-based segmentation method with a global feature strategy as a fallback mechanism when line segmentation fails to meet minimum quality requirements. An Ensemble Bagging model built upon decision trees is combined with random oversampling to address class distribution imbalance, and evaluated using a 5-Fold Stratified Cross Validation scheme. This study aims to produce a graphology-based personality inference system that is adaptive to image quality variations while capturing per-line margin positional dynamics as a more granular and representative feature basis, thereby achieving more reliable classification performance than previously established approaches.

2. Literature Review

This study involves several concepts and theories that serve as the foundation in the system design process. These theories include graphology, handwriting margin analysis, digital image preprocessing which consists of grayscale conversion, adaptive thresholding, complement operation, and morphological processing, line segmentation based on horizontal projection, statistical margin feature extraction, Ensemble Bagging, as well as evaluation metrics based on the confusion matrix.

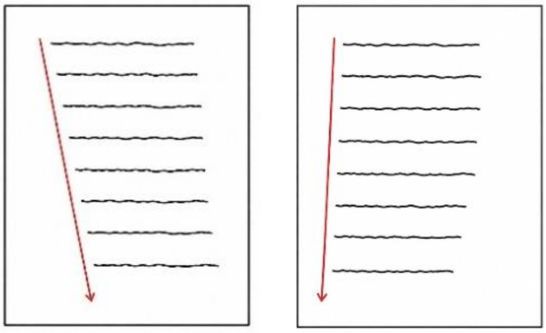
2.1. Graphology

Graphology is a scientific discipline that examines handwriting characteristics as a means of understanding the personality traits of the individual who produces it. Personality is understood as a dynamic psychophysical system that shapes patterns of behavior and ways of thinking that characterize an individual [1]. The foundation of graphology lies in the fact that writing activity is not entirely under full conscious control, but also involves cognitive processes beyond awareness, so that the graphical traces produced indirectly contain psychological information about the writer [13]. In practice, graphology has been applied in various domains such as individual potential assessment, employee selection, psychological evaluation, and forensic document examination [1].

2.2. Margin

Margin in handwriting refers to the distance between the edge of the paper and the starting position of written strokes, which naturally forms when a person writes without explicit boundary guidelines [1]. From a graphological perspective, the margin space is not merely an aesthetic choice, but rather reflects how individuals subconsciously position themselves within their social and emotional contexts [1]. The left margin, in particular, is considered a representation of the writer's relationship with past experiences, whereas the right margin is associated with the writer's orientation toward the future [1].

In graphology, the patterns of the left margin in handwriting can vary, including normal margin, wide margin, narrow margin, no margin, widening margin, and narrowing margin, each of which carries its own psychological interpretation [1]. The dynamic changes in the position of the left margin throughout the page serve as a sensitive indicator of the writer's psychological condition.



(a) (b)

Fig. 1: Patterns of Left Margin in Handwriting: (a) Widening Margin, (b) Narrowing Margin

Based on Figure 1, Figure 1(a) shows the widening pattern, where the starting position of each line gradually shifts to the right. In contrast, Figure 1(b) shows the narrowing pattern, where the starting position shifts progressively to the left.. The widening pattern is interpreted as a tendency of the writer to detach from past influences, which is associated with openness, enthusiasm, initiative, and optimism [1]. In contrast, the narrowing pattern indicates that the writer remains attached to past experiences, reflected in excessive cautiousness, egocentric tendencies, and concerns about the future [1].

Table 1: Graphological Interpretation of Left Margin Patterns in Handwriting

Margin Pattern	Indication
Widening	The writer tends to be open to new experiences, enthusiastic in life, socially oriented, proactive, and optimistic in viewing the future.
Narrowing	The writer tends to be fixated on past experiences, withdraws easily from social environments, exhibits excessive caution, shows egocentric tendencies, and is pessimistic toward change.

2.3. Pre Processing

Image preprocessing is a series of processing operations applied to digital images before the subsequent analysis stage, with the aim of improving image quality so that relevant information can be extracted more accurately and reliably [14].

2.3.1. Greyscale

Grayscale is an image representation in which each pixel has a single intensity value that represents brightness, in contrast to color images that consist of three color channels [14]. The conversion from a color image to grayscale is generally performed using perceptual weighting of the red, green, and blue components based on human visual sensitivity, as expressed in (1).

$$G(x,y) = 0.2989R(x,y) + 0.5870G(x,y) + 0.1140B(x,y) \tag{1}$$

where R(x,y), G(x,y), and B(x,y) denote the intensity values of the red, green, and blue channels at pixel coordinate (x,y).

2.3.2. Contrast Enhancement

Contrast enhancement is an operation aimed at expanding the range of pixel intensity distribution in an image, thereby making the visual distinction between the object and the background more pronounced [14]. This technique is particularly useful for images with a narrow intensity range or uneven pixel distribution, as it improves the sharpness of object details prior to the binarization process [14].

2.3.3. Adaptive Thresholding

Binarization is the process of converting a grayscale image into a binary image consisting of only two pixel values, namely 0 as the background and 1 as the object [14]. Adaptive thresholding is a binarization method that determines the threshold value locally for each region of the image based on the intensity characteristics of its surrounding pixels, making it more robust to non-uniform illumination compared to global thresholding methods [15]. The adaptive threshold value is computed as expressed in (2):

$$T(x,y) = \mu(x,y) - C \tag{2}$$

2.3.4. Complement Operation

The complement operation is a transformation applied to binary images that inverts all pixel values, where pixels with a value of 0 are converted to 1 and vice versa [1]. This operation is formulated as expressed in (3):

$$B'(x,y) = 1 - B(x,y) \tag{3}$$

where B(x,y) represents the original pixel value at coordinate (x,y), and B'(x,y) denotes the complemented pixel value. This operation is commonly applied when the binarization result produces an inverted condition between the object and the background, so that the desired object can be represented as the foreground in subsequent analysis [1].

2.3.5. Morphological Operations

Morphological operations are image processing techniques that manipulate the geometric structure of objects based on their shape, using two main components, namely the input image A and the structuring element B [16]. The two fundamental morphological operations are dilation and erosion, which are defined as follows in (4) and (5) [17]:

$$A \oplus B = \{z \mid (\hat{B})\_z \cap A \neq \emptyset\} \tag{4}$$

$$A \ominus B = \{z \mid (B)\_z \subseteq A\} \tag{5}$$

From these fundamental operations, closing and opening are derived. The closing operation, which is defined in (6), is a combination of dilation followed by erosion and is used to fill small gaps and holes in objects [17]:

$$A \cdot B = (A \oplus B) \ominus B \tag{6}$$

The opening operation, which is defined in (7), is a combination of erosion followed by dilation and is used to remove small noise that does not belong to the main object [17]:

$$A \circ B = (A \ominus B) \oplus B \tag{7}$$

2.3.6. Region of Interest

Region of Interest (ROI) is a concept in image processing that refers to a specific area of an image selected as the focus of analysis, while other regions that are considered irrelevant are excluded [14].

2.4. Feature Extraction

Feature extraction is the process of obtaining quantitative information from digital images to represent object characteristics in the form of numerical vectors that can be processed by machine learning algorithms. The quality of the extracted features plays a crucial role in determining the model's ability to distinguish between different classes.

2.4.1. Line Segmentation

Line segmentation is a technique used to separate each line of text as an independent unit of analysis in document images [18]. The horizontal projection-based method computes the number of active pixels in each row of the image to generate a distribution profile of text presence, as expressed in (8) [19]:

$$P(r) = \sum_{c=1}^{W} I(r, c) \tag{8}$$

2.4.2. Statistical Margin Features

Statistical features are extracted from a set of left margin position values m_i obtained from each line of handwriting i, where each margin value is normalized by the width of the Region of Interest (ROI) W [10]. The normalized margin value is defined as in (9):

$$m_i = \frac{x_{left,i}}{W} \tag{9}$$

where $x_{left,i}$ represents the position of the leftmost pixel in the i-th line, and W denotes the width of the ROI.

Five statistical features are derived from the margin values, as defined in (10)–(14) [10]:

Mean Margin
$$\mu_m = \frac{1}{n}\sum_{i=1}^{n} m_i \tag{10}$$

Slope Margin
$$slope = \frac{n\sum i \cdot m_i - \sum i \sum m_i}{n\sum i^2 - \left(\sum i\right)^2} \tag{11}$$

Range Margin
$$range_m = \max(m_i - \min(m_i)) \tag{12}$$

Standard Deviation
$$\sigma_m = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(m_i - \mu_m)^2} \tag{13}$$

Frst-last Difference
$$FLD = m_n - m_1 \tag{14}$$

2.4.3. Global Features as an Alternative

In cases where line segmentation fails to produce an adequate representation, a global feature-based approach becomes a relevant alternative. Global features are numerical representations extracted from the overall pixel distribution of the object in the image without relying on the structure of text lines [10]. These features capture general spatial characteristics of the handwriting.

Several commonly used global features include the leftmost pixel position as an indicator of the horizontal starting point of the object, the horizontal centroid representing the spatial distribution of the object, the width of the bounding box describing the horizontal spread of the text, and the pixel density within a specific region reflecting the ink distribution in the analyzed area [10]. These features are defined as follows in (15)–(18):

$$\text{LeftMostNorm} = \frac{\min(c)}{W} \tag{15}$$

$$\text{CenterXNorm} = \frac{\bar{c}}{W} \tag{16}$$

$$\text{BBoxWidthNorm} = \frac{\max(c) - \min(c) + 1}{W} \tag{17}$$

$$\text{InkDensityLeft} = \frac{\sum_{x,y \in Z_L} I(x,y)}{|Z_L|} \tag{18}$$

where c denotes the set of column coordinates of handwriting pixels, W represents the width of the Region of Interest (ROI), $| Z_L |$ is the left region of the image, and $| Z_L |$ indicates the total number of pixels within that region.

2.5. Ensemble Bagging

Ensemble Bagging (Bootstrap Aggregating) is a machine learning method that belongs to the ensemble learning category, which combines the predictions of multiple base models to produce a final decision that is more stable and accurate compared to a single model. This method operates by constructing multiple decision trees in parallel, where each tree is trained using a subset of data randomly sampled with replacement from the original dataset through the bootstrap sampling technique. The final prediction is obtained through a majority voting mechanism from all decision trees.

2.6. Confusion Matrix and Evaluation Metrics

The confusion matrix is a performance evaluation tool for classification models that represents the number of correct and incorrect predictions for each class in a matrix form. Based on the confusion matrix, several commonly used evaluation metrics can be derived, including precision, recall, F1-score, and accuracy, which provide a comprehensive assessment of model performance.

3. Research Method

This study aims to classify left-margin handwriting patterns into two categories, namely widening and narrowing. The proposed system consists of several main stages, including dataset preparation, image preprocessing, feature extraction, classification, and performance evaluation.

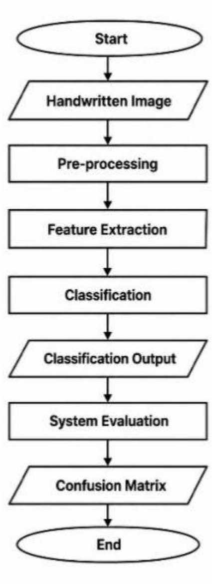


Fig. 2: Research Workflow

3.1. Dataset Preparation

The dataset was collected using white A4 paper and a 2B pencil. Participants were not allowed to use erasers to preserve the natural characteristics of handwriting. The handwritten documents were digitized using an Epson scanner with a resolution of 600 DPI. The participants were drawn from multiple study programs to ensure variability in handwriting styles.

The dataset consists of a total of 192 handwritten images, which are equally distributed into two classes: 96 images labeled as widening and 96 images labeled as narrowing.

Each image was manually labeled into one of two classes, namely widening and narrowing. The labels were stored in an Excel file containing file names and their corresponding class annotations. annotations.

3.2. Pre processing

Image preprocessing was performed to enhance the quality of the input images before further analysis. The process included grayscale conversion, contrast enhancement, adaptive thresholding, and morphological operations. In addition, a Region of Interest (ROI) was extracted to focus the analysis on the handwriting area while excluding irrelevant background regions.

3.3. Feature Extraction

Feature extraction was conducted to transform the handwriting images into numerical representations suitable for machine learning. A hybrid approach was employed, combining line-based margin analysis and global spatial features.

From the line-based analysis, several statistical margin features were derived, including MeanMargin, SlopeMargin, RangeMargin, StdMargin, and FirstLastDiff. In addition, global features such as LeftMostNorm, CenterXNorm, BBoxWidthNorm, and InkDensityLeft were also extracted to represent overall spatial characteristics of the handwriting.

3.4. Classification and Validation

The classification process was carried out using the Ensemble Bagging method with decision tree learners. To evaluate the robustness of the model, stratified 5-fold cross-validation was applied. In each iteration, approximately 80% of the data were used for training and 20% for testing, with class proportions preserved across all folds.

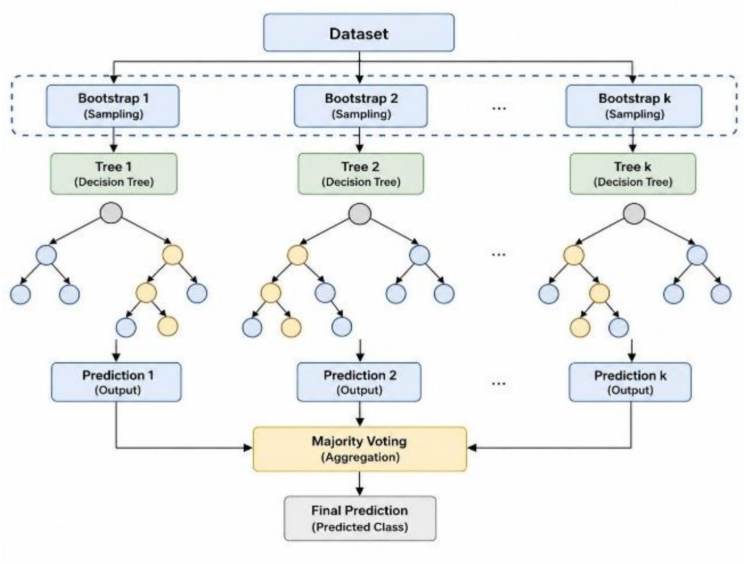


Fig. 3: Ensemble Bagging Model

3.5. System Evaluation

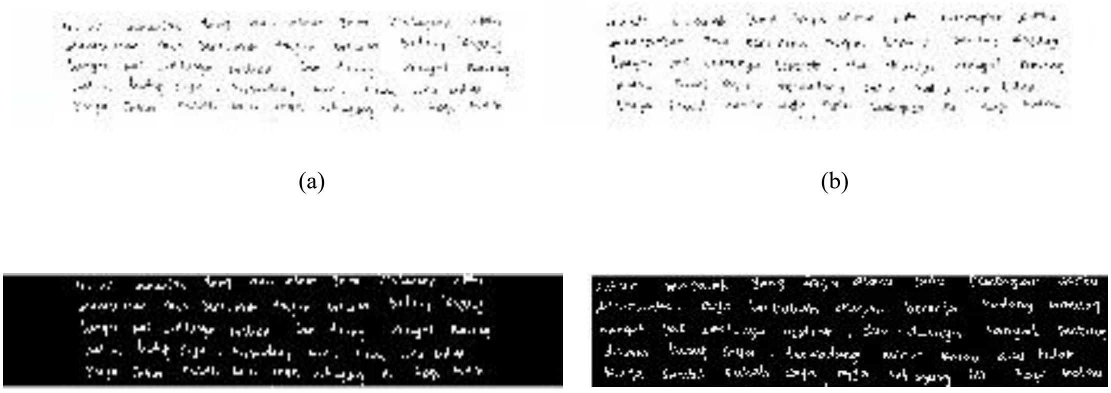
The performance of the classification model was assessed using several evaluation metrics, including accuracy, confusion matrix, precision, recall, and F1-score. These metrics provide a comprehensive understanding of the model's classification capability across different classes.

4. Result and Discussion

This section presents the experimental results and the discussion of each stage. The analysis begins with the preprocessing stage, followed by feature extraction and classification performance evaluation. Each stage is analyzed to show how the method processes handwriting images and contributes to the overall classification results.

4.1. Pre Processing

The preprocessing stage is performed to enhance image quality and ensure that relevant handwriting information can be accurately extracted. The process begins with the original scanned image, which may contain variations in illumination and background noise.



(a) (b)

(c) (d)

Fig. 4: Preprocessing stages: (a) original image, (b) grayscale, (c) binary image, (d) Region of Interest (ROI)

The first transformation applied is grayscale conversion, which reduces the image from a multi-channel color representation into a single intensity channel. This step simplifies the image structure and reduces computational complexity while preserving the essential features of the handwriting. Following this, contrast enhancement is performed to improve the distinction between handwriting strokes and the background. This step is particularly important for scanned images where stroke visibility may vary due to uneven lighting conditions.

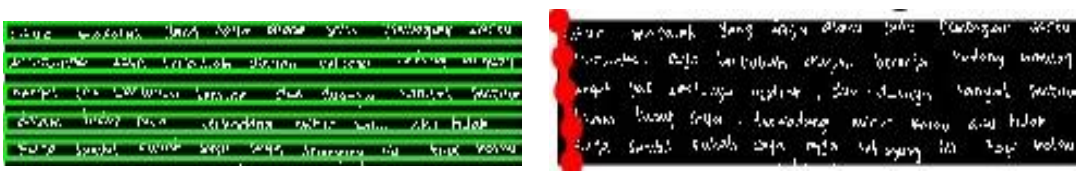
Subsequently, adaptive thresholding is applied to convert the grayscale image into a binary representation. This method determines the threshold value locally for different regions of the image, allowing it to handle variations in intensity more effectively than global thresholding techniques. As a result, handwriting strokes are clearly separated from the background. The binary image is then complemented so that the handwriting becomes the foreground object, which facilitates further processing.

To refine the binary image, noise removal is performed to eliminate small irrelevant components that do not belong to the handwriting. This is followed by morphological operations, including closing and opening, which are applied to improve the structural consistency of the handwriting. Closing helps to connect broken strokes and fill small gaps, while opening removes small isolated noise elements that remain after thresholding.

Finally, the Region of Interest (ROI) is extracted by identifying the active rows and columns that contain handwriting pixels. This step effectively crops the image to focus only on the relevant writing area while excluding unnecessary background regions. By isolating the ROI, the system ensures that subsequent feature extraction is performed on meaningful data, thereby improving the overall reliability of the classification process.

4.2. Feature Extraction

The feature extraction stage is conducted after the preprocessing process to transform the handwriting image into numerical features that can be used by the classification model. In this study, the extracted features are focused on the left margin pattern of handwriting because the margin movement can indicate whether the writing pattern tends to widen or narrow.



(a) (b)

Fig. 5: Feature extraction process: (a) line segmentation result, (c) left margin point detection

The feature extraction process begins from the Region of Interest (ROI) obtained in the preprocessing stage. The ROI is used as the input for detecting handwriting lines. Each handwriting line is identified using horizontal projection, which calculates the distribution of active pixels in each row of the image. Rows containing handwriting pixels are then grouped as text-line regions. After the text lines are detected, the system determines the leftmost pixel position from each line. This position represents the starting point of handwriting in that line and is used as the left margin value. Since each image may have a different width, the margin value is normalized based on the ROI width so that the extracted features become more consistent across different samples.

From the detected margin values, several statistical margin features are extracted, namely MeanMargin, SlopeMargin, RangeMargin, StdMargin, and FirstLastDiff. These features are used to describe the general position, direction of change, variation, and difference between the initial and final margin positions. The SlopeMargin feature is particularly important because it represents the direction of margin movement from the upper line to the lower line. In addition to line-based statistical margin features, global spatial features are also extracted to support the representation of handwriting structure. These features include LeftMostNorm, CenterXNorm, BBoxWidthNorm, and InkDensityLeft. The use of global features is intended to maintain feature representation when the handwriting line segmentation is not sufficiently clear.

The feature extraction stage produces nine numerical features for each handwriting image. These features are then arranged into a feature vector and combined with the corresponding manual label. The resulting feature vector becomes the input data for the classification stage using the Ensemble Bagging Decision Tree method

4.3. Classification

The classification stage was conducted after all handwriting images had been converted into numerical feature vectors. The extracted features were used as input data, while the manually assigned labels, namely widening and narrowing, were used as the target classes. The classification process employed the Ensemble Bagging Decision Tree method because this method combines multiple decision tree models to obtain a more stable prediction result. To evaluate the model, stratified 5-fold cross-validation was applied. In this validation scheme, the dataset was divided into five folds while preserving the class proportion in each fold. In each iteration, four folds were used as training data and one fold was used as testing data. This process was repeated five times so that each fold had the opportunity to become testing data once.

Before model training was performed in each fold, oversampling was applied to the training data to reduce the effect of class imbalance. This step was used to prevent the model from being biased toward the majority class. The oversampling process was only applied to the training data, while the testing data remained unchanged to maintain the validity of the evaluation.

Table 2: Accuracy Result

Fold	Accuracy
Fold 1	78.95 %
Fold 2	79.49 %
Fold 3	87.18 %
Fold 4	92.11 %
Fold 5	86.84 %
Average	84.91 %

The accuracy values presented in Table 2 reflect the classification performance obtained in each fold of the cross-validation process. The average accuracy represents the overall effectiveness of the model across all validation iterations. The relatively consistent performance observed among the folds indicates that the model demonstrates satisfactory generalization capability and is not overly sensitive to specific data partitions.

4.4. System Evaluation

To provide a more comprehensive evaluation of the classification performance, a confusion matrix was constructed based on the aggregated predictions from all cross-validation folds.

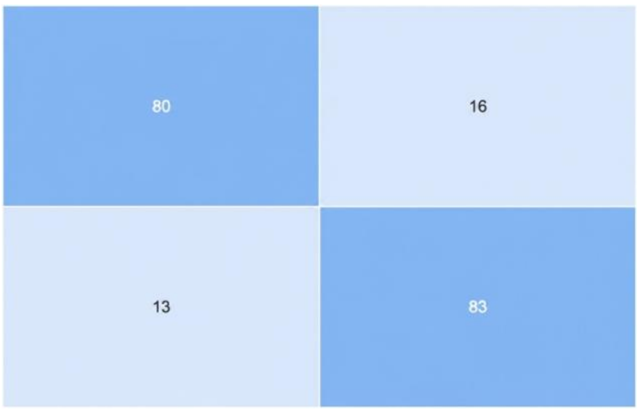


Fig. 6: Confusion Matrix

The confusion matrix illustrates the number of correctly and incorrectly classified samples for each class. From this matrix, several evaluation metrics were derived, including precision, recall, and F1-score. The performance metrics for each class are presented in Table 3.

Table 3: Performance Metrics

Class	Precision	Recall	F1-Score
Widening	0.8333	0.8602	0.8466
Narrowing	0.8646	0.8384	0.8513

The results indicate that both classes are recognized with relatively balanced performance. The widening class achieves a higher recall value, suggesting that the model is more effective in identifying true instances of this class. Conversely, the narrowing class exhibits higher precision, indicating a lower rate of false positive predictions.

Furthermore, the F1-score values for both classes are comparable, demonstrating that the model maintains a balanced trade-off between precision and recall. This balance is essential in ensuring that the classification model performs consistently across different margin patterns.

5. Conclusion

This study uses a method to classify left-margin handwriting patterns into widening and narrowing categories by combining image preprocessing, hybrid feature extraction, and Ensemble Bagging Decision Tree classification. The preprocessing stage improves the quality of handwritten images, while the hybrid feature extraction approach captures both local margin variations and global spatial characteristics.

Based on the experimental results, this study achieves an average classification accuracy of 84.91% using stratified 5-fold cross-validation. The evaluation metrics, including precision, recall, and F1-score, indicate that both classes are classified with relatively balanced performance. These results show that margin-based features, particularly those derived from line-based analysis, are effective in representing handwriting patterns for classification tasks.

Despite the satisfactory performance, several limitations remain. Variations in handwriting style, inconsistencies in margin formation, and noise introduced during the scanning process may affect feature quality and classification accuracy. In addition, relying solely on left-margin features may not fully represent the complexity of handwriting characteristics.

For future work, it is recommended to incorporate additional features such as right margin, word spacing, or stroke-based characteristics to improve classification performance. Exploring more advanced classification approaches, including deep learning methods, may also enhance the ability to capture more complex patterns. Furthermore, increasing the size and diversity of the dataset is expected to improve the generalization capability of the model.r article.

Acknowledgement

The authors would like to express their gratitude to all parties who contributed to the completion of this study. Special thanks are extended to the participants who provided handwriting samples and to the academic environment that supported the research process.

References

[1] A. Gumilang and S. Agustin, "Deteksi Kepribadian Melalui Margin Pada Tulisan Tangan Menggunakan Random Forest," *Jurnal Inovtek Polbeng - Seri Informatika*, vol. 9, no. 1, pp. 311–325, 2024.

[2] S. P. Deore, "Human Behavior Identification Based on Graphology Using Artificial Neural Network," *Acadlore Trans. Mach. Learn.*, vol. 1, no. 2, pp. 101–108, 2022, doi: 10.56578/ataiml010204.

[3] A. G. Purwanto, R. Y. Wijaya, Timotius, I. J. Agustian, and I. B. Trisno, "Analisis dan Desain Intelligent Agent Menentukan Kepribadian Berdasarkan Tulisan Tangan," *J. Nas. Komputasi Teknol. Inf.*, vol. 5, no. 2, pp. 243–247, Apr. 2022.

[4] S. Deore, P. Kalokhe, S. Patil, S. Nagarkar, and V. Oswal, "Identifying the Personality Traits Using Handwriting Recognition in a Real-Time Environment," *Ingénierie des Systèmes d'Inf.*, vol. 30, no. 1, pp. 203–211, Jan. 2025, doi: 10.18280/isi.300117.

[5] Samsuryadi, R. Kurniawan, J. Supardi, Sukemi, and F. S. Mohamad, "A Framework for Determining the Big Five Personality Traits Using Machine Learning Classification through Graphology," *J. Electr. Comput. Eng.*, vol. 2023, pp. 1–15, Jan. 2023, doi: 10.1155/2023/1249004.

[6] D. Gagiu and D. Sendrescu, "Detection of Personality Traits Using Handwriting and Deep Learning," *Appl. Sci.*, vol. 15, no. 4, p. 2154, Feb. 2025, doi: 10.3390/app15042154.

[7] M. Fadhilla, M. R. A. Saf, and D. S. S. Sahid, "Pengenalan Kepribadian Seseorang Berdasarkan Pola Tulisan Tangan Menggunakan Jaringan Saraf Tiruan," *J. Nas. Tek. Elektro dan Teknol. Inf. (JNTETI)*, vol. 6, no. 3, pp. 365–373, Aug. 2017.

[8] U. Rosyidah and N. Rochmawati, "Analisis Kepribadian Melalui Tulisan Tangan Menggunakan Metode Support Vector Machine," *J. Informatics Comput. Sci. (JINACS)*, vol. 1, no. 2, pp. 91–96, 2019.

[9] W. Wijaya, H. Tolle, and F. Utaminingrum, "Personality Analysis through Handwriting Detection Using Android Based Mobile Device," *J. Inf. Technol. Comput. Sci. (JITeCS)*, vol. 2, no. 2, pp. 114–128, 2017.

[10] A. Esposito, T. Amorese, M. Buonanno, M. Cuciniello, A. M. Esposito, M. Faundez-Zanuy, L. Likforman-Sulem, M. T. Riviello, C. Spagnuolo, A. Troncone, and G. Cordasco, "Handwriting and Drawing Features for Detecting Personality Traits: An Analysis on Big Five Sub-dimensions," Acta Polytech. Hung., vol. 19, no. 11, pp. 65–84, 2022.

[11] A. A. Khan, O. Chaudhari, and R. Chandra, "A review of ensemble learning and data augmentation models for class imbalanced problems: Combination, implementation and evaluation," Expert Syst. Appl., vol. 244, p. 122778, Jan. 2024, doi: 10.1016/j.eswa.2023.122778.

[12] A. Ayodele, "A comparative study of ensemble learning techniques for imbalanced classification problems," World J. Adv. Res. Rev., vol. 19, no. 1, pp. 1633–1643, 2023, doi: 10.30574/wjarr.2023.19.1.1202.

[13] M. Khemchandani, N. S. Ghali, D. D. Haldankar, and R. K. Sonkar, "Personality and Behavior Identification Based on Handwriting Analysis," J.

Commun. Eng. Syst., vol. 12, no. 1, pp. 1–6, 2022.
[14] D. T. Anggraeni, "Perbaikan Citra Dokumen Hasil Pindai Menggunakan Metode Simple, Adaptive-Gaussian, dan Otsu Binarization Thresholding," EXPERT J. Manaj. Sist. Inf. Teknol., vol. 11, no. 2, pp. 71–77, Dec. 2021, doi: 10.36448/expert.v11i2.2170.
[15] G. A. Sirait, N. Yudistira, and A. Ridok, "Segmentasi Citra Makanan pada Tray Box menggunakan Metode Otsu Thresholding dengan Ruang Warna," J. Pengemb. Teknol. Inf. Ilmu Komput., vol. 6, no. 2, pp. 649–657, Feb. 2022.
[16] A. Susanto, "Penerapan Operasi Morfologi Matematika Citra Digital untuk Ekstraksi Area Plat Nomor Kendaraan Bermotor," *J. Pseudocode*, vol. 6, no. 1, pp. 49–57, Feb. 2019.
[17] G. A. Trianto, F. J. Sinaga, M. F. Marzuki, and Q. A. Qorni, "Operasi Opening dan Closing pada Pengolahan Citra Digital Menggunakan Matlab," in *Proc. MDP Student Conf. (MSC) 2022*, Palembang, 2022, pp. 104–110, ISBN: 978-602-51717-7-2.
[18] R. Ptak, B. Żygadło, and O. Unold, "Projection-based text line segmentation with a variable threshold," *Int. J. Appl. Math. Comput. Sci.*, vol. 27, no. 1, pp. 195–206, 2017, doi: 10.1515/amcs-2017-0014.
[19] S. Mawaddah and N. Suciati, "Pengenalan karakter tulisan tangan menggunakan ekstraksi fitur bentuk berbasis chain code," *Jurnal Teknologi Informasi dan Ilmu Komputer (JTIIK)*, vol. 7, no. 4, pp. 683–692, Aug. 2020, doi: 10.25126/jtiik.202072022.