

Automated handwriting analysis based on pattern recognition: A survey

Samsuryadi¹, Rudi Kurniawan², Fatma Susilawati Mohamad³

¹Faculty of Computer Science, Universitas Sriwijaya, Indonesia

²Faculty of Computer Universitas Bina Insan, Indonesia & Faculty of Engineering, Universitas Sriwijaya, Indonesia

³Faculty of Informatics and Computing, Universiti Sultan Zainal Abidin, Malaysia

Article Info

Article history:

Received Apr 1, 2020

Revised Oct 6, 2020

Accepted Dec 15, 2020

Keywords:

Feature extraction

Handwriting analysis

Pattern recognition

Personality trait

Pre-processing

ABSTRACT

Handwriting analysis has wide scopes include recruitment, medical diagnosis, forensic, psychology, and human-computer interaction. Computerized handwriting analysis makes it easy to recognize human personality and can help graphologists to understand and identify it. The features of handwriting use as input to classify a person's personality traits. This paper discusses a pattern recognition point of view, in which different stages are described. The stages of study are data collection and pre-processing technique, feature extraction with associated personality characteristics, and the classification model. Therefore, the purpose of this paper is to present a review of the methods and their achievements used in various stages of a pattern recognition system.

This is an open access article under the [CC BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license.



Corresponding Author:

Rudi Kurniawan

Faculty of Computer, Universitas Bina Insan

Jend. Besar H. M. Soeharto St. Km. 13

Lubuklinggau, South Sumatra, Indonesia

Email: rudi.kurniawan@univbinainsan.ac.id

1. INTRODUCTION

Every human being has a unique personality. The study of personality traits based on handwriting is called as handwriting analysis or graphology. A graphologist uses handwriting as a guidance of a person's personality traits which are representations of neurological patterns in the brain. Handwriting analysis can be done by extracting some specific features from various handwriting samples. The extracted features are analyzed using handwriting analysis rules. Automated handwriting analysis helps graphologists to understand and identify a person's personality automatically. Handwriting analysis applications have wide scopes include recruitment, medical diagnosis, forensic, psychology, and human-computer interaction.

The development of automated handwriting analysis has become an active research area at this time. Today, the role of the graphologist can be replaced by an automated handwriting analysis that can work with a very fast, accurate, inexpensive, and easy-to-use method for identifying and predicting human personality. The problem of handwriting analysis based on the pattern recognition approach can be solved by the following three general aspects: 1) data collection and pre-processing technique, 2) data representation (feature extraction or feature selection), and 3) decision making (classification). Several approaches of pattern recognition have been used in handwriting analysis like template matching, syntactic pattern recognition, statistical pattern recognition, and artificial neural networks [1].

One of the earliest studies defined for automated handwriting analysis is called computer-aided graphology which applies the principles of pattern recognition included data acquisition, pre-processing technique, feature extraction of handwriting, and feature analysis [2]. After that, automated handwriting analysis was developed rapidly and it became an area of research in determining human personality through handwriting. Several handwritings analysis studies that refer to pattern recognition methods are explained below.

Template matching, the simplest classification technique with the concept of similarity: the same patterns can be grouped into the same class. Some letters like “t” and “i” are analyzed and detect personality traits [3-8]. A template matching algorithm is used to measure the correlation between the height of the t-bar on the stem of the letter ‘t’ and the title over ‘i’ letters to determine a person’s personality traits. To maximize the correlation of measurement, the availability of a dataset containing the templates is essential. The larger dataset is used, made the greater computation process in training the dataset. Faster processors and GPU technology made this method more easily.

In a statistical approach, each pattern is represented in terms of d features or measurements and is viewed as a point in a d-dimensional space [1]. Nonlinear discriminant analysis is used to analyze the main features: time, pressure, acceleration, velocity, energy, and complexity [9]. Meanwhile, logistic regression is used to determine Alzheimer’s disease from healthy individuals through handwriting analysis with measure on-surface time, in-air time, and total time features [10, 11]. Naïve Bayes is used to determining Parkinson’s disease through handwriting analysis with measure displacement, pressure, average speed, maximal acceleration [12, 13]. Other statistical approaches such as the KNN classification method are used to measure the personality traits with the similarity matrix method revealed by extracting handwritten analysis features such as baseline, slant, margin, and height of t-bar [3, 14].

Artificial neural networks can be described as a non-linear classification algorithm that models complex relationships between input and output to find patterns in data. This algorithm maps the input data in the input layer to the target in the output layer via neurons in the hidden layer. Personality classification based on the features of handwriting analysis through unique letters using neural networks has been carried out by Multilayer Perceptron with backpropagation algorithm [15], and neural network architecture to determining personality traits from handwriting features such as baseline, pen pressure, slant, strokes, letter ‘t’ and ‘f’ [4, 5]. In several works, convolutional neural networks are used to analyze the handwriting analysis feature through baseline, spacing, slant, pressure, size, and margin [16-18]. The advantages of the neural networks are suits for nonlinear solutions, flexible procedures for finding good, and unified approaches for feature extraction and classification. This paper aims to study several approaches in handwriting analysis based on pattern recognition and each stage of the pattern recognition systems. The block diagram of handwriting analysis based on the pattern recognition approach is shown in Figure 1.

This study is organized as follows: Section 2 contains data collection and pre-processing stages. Section 3 explains the feature extraction of handwriting analysis. Section 4 presents the classification stages of several automated handwriting analysis studies and a summary of their research. Section 5 describes promising research directions, and section 6 contains the conclusion and future work.

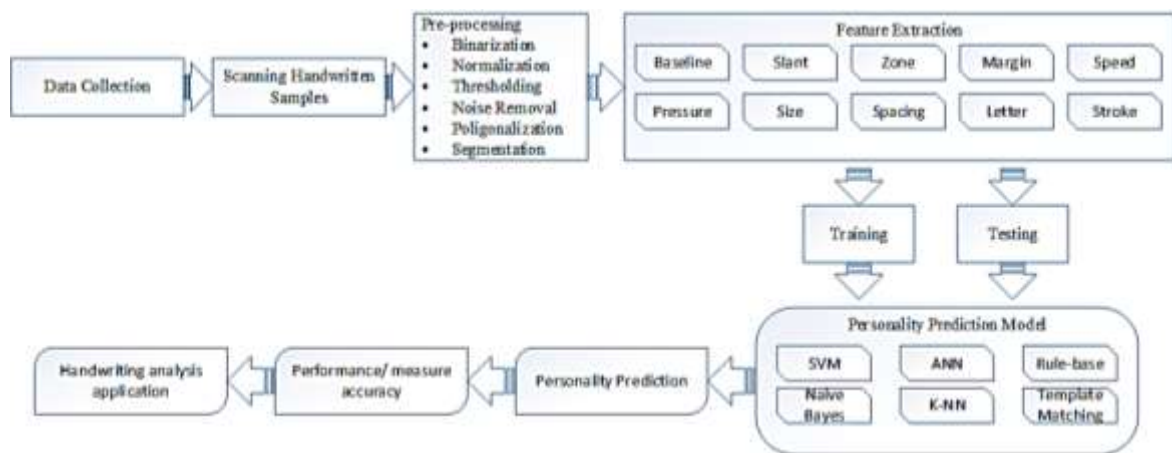


Figure 1. Block diagram of handwriting analysis based on pattern recognition approach

2. DATA COLLECTION AND PRE-PROCESSING

Many researchers have worked to make the dataset. Several factors must be considered as follows: defining a group of the respondent that might be included the ratio of male-female [19], group of age [6, 7, 19], specification of paper size [15, 20], type of pen (ballpoint or ink pen) and ink colors [20]. After the data is taken, then the data acquisition process is carried out in digital form using a scanner. Keep in mind, the quality of the scanner used affects the quality of the digital data [19, 21]. The handwriting samples are scanned and converted to JPEG format images and become a dataset of handwriting. Mostly, the dataset used by many researchers is private and unpublished. Even though, some researchers have done their studies with open access datasets and freely available such as IAM handwriting dataset English text [22-26].

The raw dataset produced by the scanning process must be improved to get a better quality image. The most common pre-processing techniques that can be used in image processing include thresholding, noise removal, and segmentation [27]. The thresholding process or what is often referred to as image binarization is the process of converting a grayscale image into a black and white image. This technique separates the foreground layer from an image that contains information (handwritten text) from the background layer that contains noise (salt and pepper noise). In other words, noise removal removes the unwanted object (interfering strokes) from the handwritten text. Segmentation in handwritten images is divided into three types: line segmentation, word segmentation, and character (letter) segmentation. The process of separating the image of a handwritten text line: word segmentation, the process of separating words from the text line image; and character segmentation, the process of separating characters (letters) from the word text image.

3. FEATURE EXTRACTION ON HANDWRITING ANALYSIS

Feature extraction is a process of dimensionality reduction (extraction data) from high dimensional input data [28]. The output data is used for analyzing human personality. Neuroscientists confirm that handwriting comes from existing minds and ideas in the human brain, so that handwriting can be made as a measure of mood, physical condition, health emotional, and mental the author. Other characteristic traits are linked to important behavioral personality traits such as concentration, emotional steadiness, motivation, intelligence, adaptability, honesty, fear, energy, and defense. Table 1 shows handwriting analysis features and associated personality characteristics.

The most common features in handwriting analysis are baseline, size, pressure, stroke, slant, spacing, speed, margin, and letters. Most of the researchers use baseline, slant, and pressure as features to predict human behavior [4, 5, 8, 18]. It is not surprising because the three features mentioned above describes the emotional stability of the writer. The size of handwriting is among the most essential factors and it reflects how an individual feels about the adaptability of a person, concentration, and nature. The size of handwriting can help to discover an individual's social aptitude [6, 18, 29-32]. The margin of handwriting could be useful in handwriting analysis and the researchers use it to indicate personality characteristics like adjustment, intelligence, past and future, truthfulness, and fastness [3, 18, 33-35]. In research [29, 31], the authors use the spacing of handwriting that contained three spacing types: spacing between lines, the spacing between words, and spacing between letters. The spacing between lines on the page refers to the clarity and the orderliness of the writer's philosophy and reasoning. The spacing between words describes the emotional comfort of the writer with their social environment. Whereas the spacing between letters reflects how the writer relates to people on a personal level. In research [4, 5, 8], the authors use the connecting strokes of handwriting to find out the information of a person's ability adaptation to changing environments. On the other hand, the speed of handwriting has been considered as one of the feature amongst baseline and lower letter case to determine personality traits [15, 29].

Table 1. Handwriting analysis features and associated personality characteristics

Feature	Type	Personality Characteristics
Baseline [3-5, 20, 22, 24, 29, 30, 33]	Normal Straight lines	Mind disciplines emotions; emotional stability
	Ascending	Continuous check on own impulse to become overly optimistic
	Descending	Fighting against depressive moods
Size [6, 18, 29-32]	Normal or average size	balance of mind, realistic, practical
	Larger than average size	Acts with boldness, enthusiasm, optimism, boastful and restless
	Smaller than average size	Not very communicative except with close friends
Pressure [5, 8, 18, 24, 29, 31, 36]	Heavy pressure	Strong-willed, firm, can get easily excited; stubborn, inclined to depression
	Medium pressure	Healthy vitality and willpower
	Light pressure	Sensitive, impressionable
Connecting Strokes [4, 5, 8]	Non-connected	Monotonous
	Medium connected	Like to change environments
	Connected	Easily adaptable to change

Table 1. Handwriting analysis features and associated personality characteristics (continue)

Feature	Type	Personality Characteristics
Slant [3-8, 16, 18, 30-32, 36]	Vertical	Head over heart emotional attitude, cautious and consider responses
	Inclined	Emotions influence decisions. Ability to express emotional self
	Reclined	Independent, completely self-interested
Letter spacing [29, 31]	Normal spacing	Balanced and flexible relationship
	Narrow	Introvert, narrow-minded, judgemental
	Wide	Cautious with own feelings
Word spacing [29, 31]	Normal spacing	Socially mature, intelligent, ability to deal flexibly and objectively
	Narrow	Craving constant contact and closeness with others; selfishness in demands
	Wide	Preferably maintaining distance from social contact, need for privacy
Line spacing [29, 31]	Normal spacing	Harmony and flexibility
	Narrow	forceful, lively, and often creative; suffer from a lack of clarity of purpose
	Wide	Isolated, fear contact and closeness
Letter 'i' [6, 15, 21]	Title is a dot	Detail-oriented, organized, and emphatic
	Title is circle	Visionary and child-like
	Title is slash	Overly self-critical
Margin [3, 18, 33-35]	Balanced	Awareness of social boundaries, poise, order, control, aesthetic sense
	Wide left margin	Avoidance of the past, sense of culture, vitality, communicative
	Wide right margin	Fear of the future, over sensitivity, self-consciousness, reserve
	Wide margin all over	Withdrawn and aloof, sensitive in color and form in surroundings, artistic
	Left margin widening	Eager to move away from the past into the world, optimistic, impatient
	Left margin narrowing	Depression or inner fatigue caused by overwork or haste
	Narrow on both sides	Acquisitiveness or stinginess, lack of consideration and reserve
	Uneven left margin	Rebellion and defiance against the rules of society
	Uneven right margin	Impulsive moods act, and reactions unreliable
	No margins anywhere	The writer eliminates all barriers between himself and other
	Wide upper margin	Modesty and formality
	Narrow upper margin	Informality, the directness of approach, lack of respect, indifference
	Wide lower margin	Idealism, aloofness, losing interest in one's environment, reserve
	Narrow lower margin	Desire to communicate, materialism, sentimental, sometimes depressed
	Letter 't' [3-5, 7, 8, 15]	Short length
Average length		Healthy, balanced: calm, self-controlled
Long length		Energetic, bold; unstoppable ambition
Lighter than stem pressure		Extremely sensitive; resignation or timidity
Heavier than stem pressure		Capable of being selfish in pursuing goals
Letter 'f' [4, 5, 8]	Big upper loop	Many theories: less concluding actions
	Big lower loop	Practical
	No loop	Austerity
Speed [15, 29]	Slow writing	A tendency toward calculation; self-conscious, possibility of dishonestly
	Fast writing	Natural and Spontaneous

4. CLASSIFICATION IN HANDWRITING ANALYSIS

Many classifiers have been used to reveal the character of human beings: artificial neural networks (ANN), support vector machine (SVM), rule-based system, naïve Bayes, and K-NN. Several researchers combine several methods to reach maximum accuracy. In this section, we discuss several studies using classifiers to determine human personality. K-NN classifier has been applied to identify the class which is most appropriate for the handwriting sample, based on the similarity matrix [3]. The similarity matrix method has been utilized to calculate the similarity of the training dataset with the feature vector matrix. Three years later, the authors had been comparing random forest, naïve Bayes, and SVM classifiers to find maximum accuracy in classification [7]. By applying the synthetic minority oversampling technique (SMOTE) algorithm, SVM achieved superior accuracy with 97%, random forest with 94%, and naïve Bayes with 90%.

SVM classifier has been revealed the character of the individual writer. In research [29], the hyper-parameters and the kernel function of SVM have been influenced to find the maximum accuracy and Radial Basis Function (RBF) kernel function archived better accuracy around 90% than linear and polynomial kernel function. In research [35], SVM has been applied to analyze psychological behavior with margin as a basic feature and the result showed an average accuracy of 82.73%. A different approach by [33], it has been Farsi's handwriting to analyze human behavior through handwriting with SVM classifier has been considered different features as an input to analyze personality traits from handwriting and the system showed promising results. Another research has been proved that the SVM classifier can perform better to reach maximum accuracy and SVM showed superior accuracy with 98% and ANN with 70% [20]. Two years later, the author's had been studying to analyze handwriting using cursive O letter (FCC and zoning features) with trained by SVM classifier and gave accuracy with 86.66% [19].

Artificial neural networks have been applied to recognize unique letters and find out human personality. Multilayer layer perceptron (MLP) has been used to identify letters a, d, i, m, and t as features with wavelet transform has been given a special feature for noise removal and it gave identification accuracy with 74% average while identification of unique letter gave accuracy with 81% [15]. In research [4], the authors have been used 3 layer neural networks architecture to analyze handwriting with Myer-Briggs type indicators (MBTI) parameter to measure human personality traits and it showed an accuracy of 86.7% that the highest accuracy is achieved for the primitive personality analysis extrovert vs introvert (E/I) and thinking vs feeling (T/F). They had been improving the previous research with combine neural network and SVM method to their classifiers and it gave identification accuracy of 88.6% [5]. Same feature but different measuring type, five-factor model (FFM) has been used to measure personality traits [7] with feedforward neural network classifier and it gave accuracy around 84.4% [8]. In another research, Promising results were also constructed to identify a person's personality traits through a deep learning approach using the convolutional neural network (CNN) method [16-18].

A rule-based approach has been used for classification. An algorithm to identify human characteristics using space has been analyzed; the system achieved the accuracy to detect skew with 96% and character analysis with 63% [22]. In research [24], the different feature has been used to detect personality traits and it gave the accuracy rate of lines segmentation with 95.65%, word segmentation with 92.56% and respectively 96% offline and words were normalized perfectly with tiny error rate. In research [6], rule base algorithm has been applied to analyze features of handwriting and it gave accuracy with 95% accuracy in identifying the handwriting features and assigning the correct trait according to principles of Graphology. The authors also propose a rule-based algorithm based on image processing to extract handwriting features like size and title over 'i' using MATLAB [21]. In research [34], the authors have been proposed a model for determining personality and gave accuracy for the left margin with 95%, the right margin with 90%, and word spacing with 85%. In research [26], the rule-based system has been used to determine personality identification using space in a handwriting image. The authors have been proposed characteristic analysis with a single feature like spacing to determine specialization in business and gave accuracy with 96% lines and words were segmented perfectly with a very small error rate and the character analysis based on space calculation accuracy with 63%.

Several researchers have been used the fuzzy system in their research. In research [30], the authors have been proposed fuzzy C-means as a classifier and the psychological method that used a series of questions to determine a human personality called enneagram and it archived the accuracy with 81.6%. In research [32], a fuzzy membership classifier is has been used to identify writer identification from handwriting Devanagari script and it gave accuracy with 97% on the test set. Another research has been used the fuzzy Sugeno model and proposed a promising framework in handwriting analysis [36]. In research [37], the authors have been applied to fuzzy rule models called the fuzzy rule-based classification system (FRCS). By applying chi's algorithm as the learning method, FRCS achieved an accuracy of around 76%. The simple overall discussion for handwriting analysis based on pattern recognition is summarized in Table 2.

Table 2. The summary of handwriting analysis based on pattern recognition

Reference	Pre-processing	Feature	Classifier	Dataset	Result
Joshi P., <i>et al.</i> 2015 [3]	Poligonalization, thresholding	Baseline, slant, Letter "t", margin	K-NN Method, template matching	100 samples of handwriting	Proposed an algorithm in handwriting analysis
Bobade, Ankur, M, <i>et al.</i> 2015 [29]	Noise removal, segmentation (letter, word, and line)	Pressure, baseline, size, spacing, margins, slant, and speed	SVM (RBF Kernel)	Unspecified; the dataset takes from a different person (write 50-60 words on a plain paper)	ACC=90%
Hashemi. S, <i>et al.</i> 2015 [33]	Pen width extraction, noise and scratch removal	Margin, size, spacing, slant	SVM	120 samples of Farsi handwriting	Proposed an algorithm in handwriting analysis
Asra. S, <i>et al.</i> 2015 [20]	Cropping (de- noising & resized), thresholding	Baseline	Comparing SVM and ANN	500 samples of handwriting; A4 paper with black ballpoint pen	SVM=98% ANN=70%
Djamal, Esmeralda C., <i>et al.</i> 2015 [15]	Grayscale, thresholding, segmentation	Speed, letter "a, d, i, m, t", wavelet	Multilayer Perceptron (MLP)	125 samples of handwriting	MLP=74%. Unique letter=81%.
Gavrilescu. M. 2015 [4]	Segmentation, thresholding, noise removal	Baseline, slant, stroke, letter "t" and "f", pressure	ANN architecture	64 samples of handwriting	ACC=86%

Table 2. The summary of handwriting analysis based on pattern recognition (continue)

Reference	Pre-processing	Feature	Classifier	Dataset	Result
Pratiwi D., <i>et al.</i> , 2016 [30]	Grayscale, thresholding	Baseline, slant, break, size	Fuzzy C-Means, Enneagram method	50 data collected; 1 data not valid	ACC=81.6%
Nagar S., <i>et al.</i> 2016 [22]	Noise removal, thresholding,	Spacing	Rule Base System	IAM Database	Normalized line and word acc.=6%, and character analysis acc.=63% ACC=96%
Bal A., <i>et al.</i> 2016 [24]	Noise removal, thresholding,	Baseline (line and word), pressure	Rule Base System	IAM Database	ACC=88.6%.
Gavrilescu. M. 2017 [5]	Segmentation, thresholding, noise removal	Baseline, slant, stroke, letter “t” and “f”, pressure	ANN, Multi-class SVM (RBF kernel), Template Matching, K-NN	64 samples of handwriting	ACC=86.66%
Asra S., <i>et al.</i> 2017 [19]	Resize, grayscale, segmentation (drop fall algorithm)	Cursive O	SVM	500 samples of handwriting; different educations, genders, ages	ACC=95%
Sen A., <i>et al.</i> 2017 [6]	Binarization	Baseline, margin, slant, size, word spacing, and title over i	Rule Base System	75 handwriting samples: the age of correspondent between 20-40 year	ACC=97%
Kumar R., <i>et al.</i> 2017 [32]	Noise removal, thresholding	Slant, baseline, and size of the letter	Fuzzy System	CPAR-2012 dataset	Proposed a framework with fuzzy Sugeno model
Lakshmi K., Nithya, <i>et al.</i> 2017 [36]	Noise removal, segmentation, thresholding	Slant, size, pressure, spacing	Fuzzy System (Sugeno Fuzzy Model)	The handwriting samples are taken as input which is taken on a plain A4 sheet	
Garoot A. H, <i>et al.</i> 2017 [38]	A survey paper that presented different methodologies that are implemented for automated graphology. This survey also presented various features on handwriting				
Varshney A, <i>et Al.</i> 2017 [39]	A survey paper on human personality identification which is used for automated graphology. The crucial part of this survey is a different classification based on Artificial Neural Network				
Joshi P., <i>et al.</i> 2018 [7]	Grayscale, thresholding	Baseline, slant, Letter “t”, margin	Naive Bayes, Random Forest and Multi-Class SVM	1890 samples of handwriting; different ages, genders	SVM=97%, RF=94%, and NB=90%
Wijaya W., <i>et al.</i> 2018 [35]	Gray Scale, Thresholding	Margins	SVM	42 samples of handwriting	ACC=82.73%
Nag S., <i>et al.</i> 2018 [40]	Segmentation (Horizontal Projection Profile)	Baseline with COLD (Cloud of Line Distribution) method	Multi-class SVM	100 writers per each class (nation);	Classification Rate (CR)=75%
Gavrilescu M., <i>et al.</i> 2018 [8]	Segmentation, thresholding, noise removal (Gabor filter)	Baseline, slant, pressure, stroke, the letter “t”, letter “f”, spacing	Feed-Forward Neural Network, Template Matching	London Letter, and 300 words texts that subjects could write freely and randomly	ACC=84.4%
Lemos N., <i>et al.</i> 2018 [16]	Noise removal, thresholding,	Baseline, Spacing, Slant	CNN	Module image of handwriting (the samples is taken through a website)	Proposed an algorithm in handwriting analysis
Sen A., <i>et al.</i> 2018 [21]	Noise removal	Size and title over “i”	Rule Base System, Image Processing	Handwriting samples with the corresponding writer	The proposed algorithm implemented to MATLAB application ACC=90%
Bhade V., <i>et al.</i> 2018 [34]	Noise removal, Grayscale, Thresholding	Margin, spacing	Rule Base System	11 different handwriting paragraphs have been taken.	ACC=76%
Riza L. S., <i>et al.</i> 2018 [37]	Segmentation, thresholding	Size, pressure, margin, baseline	Fuzzy Rule Base Classification System (FRBCS)	75 handwriting sample; 36 Males and 39 females)	AUC = up to 0.5314
Valdez-Rodríguez J. E., <i>et al.</i> 2019 [17]	Grayscale	Baseline	CNN (five convolutional layers)	2018 ICPR	
Sony D., <i>et al.</i> 2019 [18]	Noise removal, segmentation, thresholding	Baseline, slant, pressure, size, margin, zone	CNN	Handwriting samples with the corresponding writer	Proposed a framework with CNN

Table 2. The summary of handwriting analysis based on pattern recognition (continue)

Reference	Pre-processing	Feature	Classifier	Dataset	Result
Chitlangia A, <i>et al.</i> 2019 [41]	Histogram of Oriented Gradient (HOG)	Size, slant, pressure, spacing, baseline	Multi-Class SVM (polynomial kernel)	50 different writers have been asked to write several texts with the same content	ACC=80%
Chakraborty S., <i>et al.</i> 2019 [26]	Grayscale, normalization, segmentation	Spacing	Rule Base System	IAM Database	Normalized line and word acc.=96%, and character analysis acc.=63%
Ghosh S., <i>et al.</i> 2020 [42]	Grayscale, thresholding	Lower letter a to z	SVM	5300 samples of handwriting	ACC=86.70%

5. DISCUSSION

Personality traits assessment based on handwriting analysis has become a widely used benchmark in forensics, employee recruitment, and even in the medical world. The relationship between personality traits and handwriting analysis has been a long debate about the validity of the results obtained. The different opinions expressed by [43], they conclude nothing characteristics of handwriting were specific to human personality traits and there is no evidence for assessment of personality based on handwriting analysis with measured by the NEO-FFI (big five model of personality) and EPQ-R. In the pro-graphology case, it can be a relationship between handwriting analysis features and personality traits assessment (16PF-R measure assessment with zoning feature) [44]. Despite these contradictions, the studies of graphology have become very intensively as evidenced by the large number of journals related to the research area. A lot of evidence linking human personality traits based on pattern recognition to handwriting is obtainable. Although various specific issues have been already shown, in the following the most applicable are shortly discussed.

5.1. Data collection and pre-processing

Many research workers have taken database make by gathering data themselves. The tendency of researchers to use their dataset is to determine the relationship between personality traits through handwriting. To find out this relationship, some researchers compare the results with various psychological questionnaire assessment methods that can be applied to handwriting analysis [4, 8, 30, 44]. These datasets are dissimilar in sizes, age of participants, type of papers, male-female ratio, type of pen (ink pen, ballpoint pen, and ink colors), and many more. The lack of a large-scale database involving a significant amount of participants, as well as, a set of important tasks, very restricts the physical process of research. It should be noted that there is a lack of research using databases of non-western scripts. Besides that, this would be of great interest since scripts have many symbolic elements that could produce useful information [32, 33]. Taking handwriting samples from participants should be repeated much time to study human psychological change and its effects on writing (several times a month or even a year to determine personality shifts and hence conclude in mental disorders that affect the subject or psychological symptoms of a physical disease) [4]. Unfortunately, developing such a benchmark dataset is an expensive process and time-consuming.

Noise removal and thresholding techniques are generally used to refine and smooth images from the results of scanning. The median filter can be used to remove Salt and paper noise [45, 46] and the Otsu thresholding for image binarization [47, 48]. The handwriting pressure feature can be extracted by the grayscale value of the handwriting portion and discards the background portion. For line segmentation, the horizontal projection method can be applied to separate text lines from the handwritten image text [25]. After line segmentation, different lines may occur with different skew angles. The orthogonal projection method can be implemented to normalize the skew. For word segmentation, to calculate the distance between words, a vertical projection histogram approach is used to measure the threshold value between the width of the distance of each word.

Several image processing techniques are considered to apply: HOG technique [41] and two-stage filtering technique [49]. HOG technique works by calculating the appearance of gradient orientation in a localized part of sample images and the important idea behind HOG descriptors is that local object appearances and shape within an image can be depicted by edge directions or distribution of intensity gradients. HOG technique converts the digital handwriting sample into square grids. After that, based on the central difference, the edge or histogram of gradient direction is calculated. In the two-stage filtering technique, there are two steps: detection technique and filtering technique. Some algorithms in detection techniques that can be used include rank order absolute difference (ROAD), rank order logarithmic

difference (ROLD), adaptive switching median (ASM), measures of dispersions (MOD), and triangle-based linear interpolation (TBLI). The algorithms in filtering techniques that can be used include median filter (MF), fuzzy switching median filter (FSMF), and fuzzy switching weighted median filter (FSWMF).

5.2. Features in handwriting analysis

Many features have been considered to evaluate personality traits. The common features are used by many researchers are baseline, slant, margin, pressure, spacing, and size of the letter. For global feature extraction, the input is an analysis of the entire handwritten text. For local feature extraction, the page is partitioned into lines of text and every line is partitioned into several connected components. The components are used as units of local feature extraction. Several algorithms are needed for the extraction of various features [31]. Variations of size may vary from each writer. The writers with large size letters need width area and hence more number of black pixels than small-sized letters. It can be calculated by the measured area and the number of black pixels. Template matching algorithms can be applied to analyze personality traits from the lower case letter 't' and 'f'.

Unfortunately, many researchers use handwriting analysis features independently. Combining two or more features to create psychological traits would be a great interest and could convey useful information to determine personality traits. Some methods can be used: myer-briggs type indicators (MBTI) [4], five-factor models (FFM) [8], and enneagram [30] to measure the personality traits of an individual. Another measure of personality traits can be used for furthermore like 16PF questionnaire method and the Birkman Method.

5.3. Classification in handwriting analysis

Classification in handwriting analysis problems can be solved by using a pattern recognition approach. In template matching, a good distance measurement must be determined to find similarity between objects to be very important to the success of this approach. In determining the closest distance, some distance algorithm can be used like Manhattan distance, Euclidean distance, Minkowski distance, and Chebychev distance. If in the pre-processing stage there is no normalization of the object image, the same character must be represented in various positions of the object in its feature space [1]. This condition will require a large image template database and it can be affected by the computational process.

The SPR emphasizes the classification of features in the recognition process. The feature extraction process is very important in achieving maximum accuracy. SVM has generalization capabilities where it can identify patterns that do not belong to existing classes and can prevent the curse of dimensionality. SVM is only able to handle the classification of two classes, but there has been a lot of research in handwriting pattern recognition using multiclass SVM [7, 40, 41]. SVM has proven the highest achievement accuracy of classification [7, 20, 41]. It must be underlined that the accuracy achieved by SVM is only in the classification of one feature [20]. In several cases, determining the classification of personality traits using complex features, the results obtained are not too significant [19, 29, 33, 35]. Determining the right kernel also provides maximum accuracy for SVM [5].

The fundamental issue of neural networks is how to integrate the characteristics of existing features with the classification process. Choosing a good feature extraction method can improve accuracy and minimize the resulting error. FFNN with backpropagation algorithm provides an integrated procedure for feature extraction and classification [4, 5, 8]. In determining the good feature extraction must be accompanied by good pre-processing techniques, such as segmentation, noise removal, and thresholding on the image object. In several cases, deep learning approaches also use to identify human personality behavior [16-18]. Besides, combining neural network architecture with other classification is done to reach maximum accuracy [5]. A hybrid approach using deep learning and other machine learning would be a good interest to reach better accuracy [50-54].

5.4. Handwriting analysis applications

Handwriting analysis has a wide scope of application. Handwriting analysis applications are used in employee recruitment, job applications, and marriage compatibility, motivate workers, career guidance, student's exam anxiety, and child development [55]. Handwriting is considered as a kind of biometric behavior that can be used to identify someone through his/her handwriting [56]. Many researchers have worked with handwriting analysis as the object of their research to identify personality traits of a person and have studied how it can be useful in various aspects of life such as work profile, forensics, healthcare, and others [57].

6. CONCLUSION AND FUTURE WORK

These optimistic studies suggest a methodology in automated handwriting analysis based on the pattern recognition approach. The selection of good and appropriate techniques in the pre-processing stage is crucial to achieving maximum results in identifying personality traits through handwriting. We have defined several handwriting features that are used to determine a person's personality in automated handwriting analysis such as baseline, slant, margin, spacing, letter size, pen pressure, and speed of writing. These features will serve as input in determining human personality through handwriting. It has been observed that different analysis methods such as artificial neural networks, support vector machine, rule-based system, K-NN, fuzzy model, and naïve Bayes have been used widely in handwriting analysis and it shows promising results. Applying a deep learning (LSTM, RNN, CNN, auto-encoder) method for designing a fully automated handwriting analysis system to reach better accuracy on classification should be considered as future work. Otherwise, combining one classifier and others to analyze the features of handwriting, applying the feature of signature, and modified the quantity of training data could be an interesting area of future research in the relationship between personality traits and handwriting analysis.

REFERENCES

- [1] A. K. Jain, R. P. W. Duin, and J. Mao, "Statistical pattern recognition: A review," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 22, no. 1, pp. 4-37, 2000, doi: 10.1109/34.824819.
- [2] G. Sheikholeslami, S. N. Srihari, and V. Govindaraju, "Computer Aided Graphology," State University of New York at Buffalo, 1995.
- [3] P. Joshi, A. Agarwal, A. Dhavale, R. Suryavansi, and S. Kodollikar, "Handwriting Analysis for Detection of Personality Traits using Machine Learning Approach," *Int. J. Comput. Appl.*, vol. 130, no. 15, pp. 40-45, 2015, doi: 10.5120/ijca2015907189.
- [4] M. Gavrilescu, "Study on determining the Myers-Briggs personality type based on individual's handwriting," in *The 5th IEEE International Conference on E-Health and Bioengineering - EHB 2015*, 2015, doi: 10.1109/EHB.2015.7391603.
- [5] M. Gavrilescu, "3-Layer architecture for determining the personality type from handwriting analysis by combining neural networks and Support Vector Machines," *UPB Sci. Bull. Ser. C Electr. Eng. Comput. Sci.*, vol. 79, no. 4, pp. 135-152, 2017.
- [6] A. Sen and H. Shah, "Automated Handwriting Analysis System using Principles of Graphology and Image Processing," in *ICIIECS*, 2017, doi: 10.1109/ICIIECS.2017.8276061.
- [7] P. Joshi, P. Ghaskadbi, and S. Tendulkar, "A Machine Learning Approach to Employability Evaluation Using Handwriting Analysis," in *ICAICR*, pp. 253-263, 2018, doi: 10.1007/978-981-13-3140-4_23.
- [8] M. Gavrilescu and N. Vizireanu, "Predicting the Big Five personality traits from handwriting," *EURASIP J. Image Video Process.*, 2018, doi: 10.1186/s13640-018-0297-3.
- [9] J. Garre-Olmo, M. Faúndez-Zanuy, K. López-de-Ipiña, L. Calvó-Perxas, and O. Turró-Garriga, "Kinematic and Pressure Features of Handwriting and Drawing: Preliminary Results Between Patients with Mild Cognitive Impairment, Alzheimer Disease and Healthy Controls," *Curr. Alzheimer Res.*, vol. 14, no. 9, pp. 960-968, 2017, doi: 10.2174/1567205014666170309120708.
- [10] S. Müller, O. Preische, P. Heymann, U. Elbing, and C. Laske, "Diagnostic Value of a Tablet-Based Drawing Task for Discrimination of Patients in the Early Course of Alzheimer's Disease from Healthy Individuals," *J. Alzheimer's Dis.*, vol. 55, no. 4, pp. 1463-1469, 2017, doi: 10.3233/JAD-160921.
- [11] S. Müller, O. Preische, P. Heymann, U. Elbing, and C. Laske, "Increased diagnostic accuracy of digital vs. conventional clock drawing test for discrimination of patients in the early course of Alzheimer's disease from cognitively healthy individuals," *Front. Aging Neurosci.*, vol. 9, no. APR, pp. 1-10, 2017, doi: 10.3389/fnagi.2017.00101.
- [12] P. Zham, S. P. Arjunan, S. Raghav, and D. K. Kumar, "Efficacy of Guided Spiral Drawing in the Classification of Parkinson's Disease," *IEEE J. Biomed. Heal. Informatics*, vol. 22, no. 5, pp. 1648-1652, 2018, doi: 10.1109/JBHI.2017.2762008.
- [13] C. Kotsavasiloglou, N. Kostikis, D. Hristu-Varsakelis, and M. Arnaoutoglou, "Machine learning-based classification of simple drawing movements in Parkinson's disease," *Biomed. Signal Process. Control*, vol. 31, pp. 174-180, 2017, doi: 10.1016/j.bspc.2016.08.003.
- [14] H. Mohammed, V. Maergner, T. Konidaris, and H. S. Stiehl, "Normalised Local Naïve Bayes Nearest-Neighbour Classifier for Offline Writer Identification," in *Proceedings of the International Conference on Document Analysis and Recognition, ICDAR*, vol. 1, 2017, pp. 1013-1018, doi: 10.1109/ICDAR.2017.168.
- [15] E. C. Djamal and Febriyanti, "Identification of speed and unique letter of handwriting using wavelet and neural networks," in *International Conference on Electrical Engineering, Computer Science and Informatics (EECSI)*, vol. 2, no. August, 2015, pp. 74-78.
- [16] N. Lemos, K. Shah, R. Rade, and D. Shah, "Personality Prediction based on Handwriting using Machine Learning," in *International Conference on Computational Techniques, Electronics and Mechanical Systems*, 2018, pp. 110-113, doi: 10.1109/CTEMS.2018.8769221.
- [17] J. E. Valdez-Rodriguez, H. Calvo, and E. M. Felipe-Riveron, "Handwritten Texts for Personality Identification

- Using Convolutional Neural Networks,” in *Int. Conf. on Pattern Recognition*, 2019, pp. 140-145, doi: 10.1007/978-3-030-05792-3_13.
- [18] D. Sony and R. Sawant, “Identifying Human Behavior Characteristics using Handwriting Analysis,” *Int. Res. J. Eng. Technol.*, vol. 06, no. 04, pp. 4436-4439, 2019, doi: 10.1109/ICIIC.2010.29.
- [19] S. Asra and S. D.C, “Human Behavior Recognition based on Hand Written Cursives by SVM Classifier,” in *2017 International Conference on Electrical, Electronics, Communication, Computer and Optimization Techniques (ICEECCOT) Human*, 2017, pp. 260-268, doi: 10.1109/ICEECCOT.2017.8284679.
- [20] S. Asra and S. D.C, “Personality Trait Identification Using Unconstrained Cursive and Mood Invariant Handwritten Text,” *Int. J. Educ. Manag. Eng.*, vol. 5, no. 5, pp. 20-31, 2015, doi: 10.5815/ijeme.2015.05.03.
- [21] A. Sen, H. Shah, J. Lemos, and S. Bhattacharjee, “An Algorithm to Extract Handwriting Feature for Personality Analysis,” in *Proceedings of International Conference on Wireless Communication*, 2018, pp. 323-329, doi: 10.1007/978-981-10-8339-6_35.
- [22] S. Nagar, S. Chakraborty, A. Sengupta, J. Maji, and R. Saha, “An efficient method for character analysis using space in handwriting image,” in *Proceedings - 2016 6th International Symposium on Embedded Computing and System Design, ISED 2016*, pp. 210-216, 2016, doi: 10.1109/ISED.2016.7977084.
- [23] A. Bal and R. Saha, “An Efficient Method for Skew Normalization of Handwriting Image,” in *In 6th IEEE International Conference on Communication Systems and Network Technologies*, 2016, pp. 222-228.
- [24] A. Bal and R. Saha, “An Improved Method for Handwritten Document Analysis using Segmentation, Baseline Recognition and Writing Pressure Detection,” in *Procedia Computer Science*, vol. 93, pp. 403-415, 2016, doi: 10.1016/j.procs.2016.07.227.
- [25] A. Bal and R. Saha, “An improved method for text segmentation and skew normalization of handwriting image,” *Adv. Intell. Syst. Comput.*, vol. 518, pp. 181-196, 2018, doi: 10.1007/978-981-10-3373-5_18.
- [26] S. Chakraborty and J. Majumder, “Character Analysis Using Space in Handwriting Image to Determine Specialization in Business,” *Int. J. Recent Trends Bus. Tour.*, vol. 3, no. July, pp. 108-116, 2019.
- [27] R. Plamondon and S. N. Srihari, “On-Line and Off-Line Handwriting Recognition: A Comprehensive Survey,” *IEEE Trans. PATTERN Anal. Mach. Intell.*, vol. 22, no. 1, pp. 63-84, 2000, doi: 10.1109/34.824821.
- [28] M. Sachan and S. K. Singh, “Personality Detection using Handwriting Analysis: Review,” in *The Seventh International Conference on Advances in Computing, Electronics and Communication - ACEC 2018*, 2018, doi: 10.15224/978-1-63248-157-3-33
- [29] A. M. Bobade and P. N. N. Khalsa, “Character Revealing Handwriting Analysis based on Segmentation method using Support Vector Machine,” *Int. J. Electron. Commun. Soft Comput. Sci. Eng.*, pp. 203-207, 2015, doi: 10.5120/1256-1758.
- [30] D. Pratiwi, G. Budi, and F. Hana, “Personality Type Assessment System by using Enneagram-Graphology Techniques on Digital Handwriting,” *Int. J. Comput. Appl.*, vol. 147, no. 11, pp. 9-13, 2016, doi: 10.5120/ijca2016911181.
- [31] S. Mukherjee and I. De, “Feature extraction from handwritten documents for personality analysis,” in *2016 International Conference on Computer, Electrical and Communication Engineering, ICCECE 2016*, 2016, doi: 10.1109/ICCECE.2016.8009580.
- [32] R. Kumar, K. K. Ravulakollu, and R. Bhat, “Fuzzy-Membership based writer identification from handwritten devnagari script,” *J. Inf. Process. Syst.*, vol. 13, no. 4, pp. 893-913, 2017, doi: 10.3745/JIPS.02.0018.
- [33] S. Hashemi, B. Vaseghi, and F. Torgheh, “Graphology for Farsi Handwriting Using Image Processing Techniques,” *IOSR J. Electron. Commun. Eng. Ver. I*, vol. 10, no. 3, pp. 2278-2834, 2015.
- [34] V. Bhade and T. Baraskar, “A model for determining personality by analyzing off-line handwriting,” in *Advances in Intelligent Systems and Computing*, vol. 705, pp. 345-354, 2018, doi: 10.1007/978-981-10-8569-7_35.
- [35] W. Wijaya, H. Tolle, and F. Utaminigrum, “Personality Analysis through Handwriting Detection Using Android Based Mobile Device,” *J. Inf. Technol. Comput. Sci.*, vol. 2, no. 2, pp. 114-128, 2018, doi: 10.25126/jitecs.20172237.
- [36] K. N. Lakshmi, A. Keerthana, and P. R. Lakshmi, “Handwriting Analysis Based Human Personality Prediction Using Sugeno Fuzzy Model,” *Int. J. Sci. Eng. Res.*, vol. 8, no. 5, pp. 105-110, 2017.
- [37] L. S. Riza, A. Zainafif, Rasim, and S. Nazir, “Fuzzy rule-based classification systems for the gender prediction from handwriting,” *TELKOMNIKA Telecommunication Computing Electronics and Control*, vol. 16, no. 6, pp. 2725-2732, 2018, doi: 10.12928/telkomnika.v16i6.9478.
- [38] A. H. Garoot, M. Safar, and C. Y. Suen, “A Comprehensive Survey on Handwriting and Computerized Graphology,” in *Proceedings of the International Conference on Document Analysis and Recognition, ICDAR*, vol. 1, 2017, pp. 621-626, doi: 10.1109/ICDAR.2017.107.
- [39] A. Varshney and S. Puri, “A Survey on human personality identification on the basis of Handwriting using ANN,” in *International Conference on Inventive Systems and Control (ICISC-2017) A*, 2017, pp. 1-6, doi: 10.1109/ICISC.2017.8068634.
- [40] S. Nag, P. Shivakumara, Y. Wu, U. Pal, and T. Lu, “New COLD feature based handwriting analysis for ethnicity/nationality identification,” *Proc. Int. Conf. Front. Handwrit. Recognition, ICFHR*, vol. 2018-Augus, 2018, pp. 523-527.
- [41] A. Chitlangia and G. Malathi, “Handwriting Analysis based on Histogram of Oriented Gradient for Predicting Personality traits using SVM,” *Procedia Comput. Sci.*, vol. 165, no. 2019, pp. 384-390, 2019, doi: 10.1016/j.procs.2020.01.034.
- [42] S. Ghosh, P. Shivakumara, P. Roy, U. Pal, and T. Lu, “Graphology based Handwritten Character Analysis for

- Human Behavior Identification,” *CAAI Trans. Intell. Technol.*, vol. 5, pp. 55-65, 2020, doi: 10.1049/trit.2019.0051.
- [43] B. Gawda, “Lack of evidence for the assessment of personality traits using handwriting analysis,” *Polish Psychol. Bull.*, vol. 45, no. 1, pp. 73-79, 2014, doi: 10.2478/ppb-2014-0011.
- [44] Y. Chernov and C. Caspers, “Formalized computer-aided handwriting psychology: Validation and integration into psychological assessment,” *Behav. Sci. (Basel)*, vol. 10, no. 1, 2020, doi: 10.3390/bs10010027.
- [45] M. H. Suid, M. F. M. Jusof, and M. A. Ahmad, “Dual sliding statistics switching median filter for the removal of low level random-valued impulse noise,” *J. Electr. Eng. Technol.*, vol. 13, no. 3, pp. 1383-1391, 2018, doi: 10.5370/JEET.2018.13.3.1383.
- [46] B. Karthik, T. Krishna Kumar, S. P. Vijayaragavan, and M. Sriram, “Removal of high density salt and pepper noise in color image through modified cascaded filter,” *J. Ambient Intell. Humaniz. Comput.*, pp. 1-8, 2020, doi: 10.5829/idosi.mejsr.2014.20.10.114063.
- [47] N. Otsu, “Threshold Selection Method From Gray-Level Histograms,” *IEEE Trans Syst Man Cybern*, vol. SMC-9, no. 1, pp. 62-66, 1979, doi: 10.1109/TSMC.1979.4310076.
- [48] N. Paul and H. Tunga, “An Improved Method for Document Image Binarization,” *Natl. Conf. Recent Innov. Comput. Sci. Commun.*, no. July, pp. 1-6, 2016.
- [49] N. Singh, T. Thilagavathy, R. T. Lakshmi Priya, and O. Umamaheswari, “Some studies on detection and filtering algorithms for the removal of random valued impulse noise,” *IET Image Process.*, vol. 11, no. 11, pp. 953-963, 2017, doi: 10.1049/iet-ipr.2017.0346.
- [50] K. Dutta, P. Krishnan, M. Mathew, and C. V. Jawahar, “Improving CNN-RNN hybrid networks for handwriting recognition,” *Proc. Int. Conf. Front. Handwrit. Recognition, ICFHR*, vol. 2018-Augus, 2018, pp. 80-85, doi: 10.1109/ICFHR-2018.2018.00023.
- [51] A. Kumarbhunia *et al.*, “Handwriting Trajectory Recovery using End-to-End Deep Encoder-Decoder Network,” *Proc. - Int. Conf. Pattern Recognit.*, vol. 2018-Augus, 2018, pp. 3639-3644, doi: 10.1109/ICPR.2018.8546093.
- [52] Y. Su, Y. Huang, and C. C. J. Kuo, “Efficient Text Classification Using Tree-structured Multi-linear Principal Component Analysis,” *Proc.-Int. Conf. Pattern Recognit.*, vol. 2018-Augus, 2018, pp. 585-590, doi: 10.1109/ICPR.2018.8545832.
- [53] Y. Su and C. C. J. Kuo, “On extended long short-term memory and dependent bidirectional recurrent neural network,” *Neurocomputing*, vol. 356, pp. 151-161, 2019.
- [54] Y. Su, R. Lin, and C. C. Jay Kuo, “Tree-structured multi-stage principal component analysis (TMPCA): Theory and applications,” *Expert Syst. Appl.*, vol. 118, pp. 355-364, 2019.
- [55] Y. Bay Azyeren, M. Erbilek, and E. Celebi, “Emotional State Prediction from Online Handwriting and Signature Biometrics,” *IEEE Access*, vol. 7, pp. 164759-164774, 2019, doi: 10.1109/ACCESS.2019.2952313.
- [56] A. Rehman, S. Naz, M. I. Razzak, and I. A. Hameed, “Automatic Visual Features for Writer Identification: A Deep Learning Approach,” *IEEE Access*, vol. 7, no. c, pp. 17149-17157, 2019, doi: 10.1109/ACCESS.2018.2890810.
- [57] K. Chaudhari and A. Thakkar, “Survey on handwriting-based personality trait identification,” *Expert Syst. Appl.*, vol. 124, pp. 282-308, 2019.

BIOGRAPHIES OF AUTHORS



Dr. Samsuryadi Sahmin is currently a lecturer at the Department of Informatics Engineering Faculty of Computer Science, Universitas Sriwijaya, Indonesia. He received a Bachelor of Science in Mathematics from Universitas Sriwijaya, Indonesia, Magister of Computer Science from Gadjah Mada University, Indonesia, and Ph.D. in Computer Science, Universiti Teknologi Malaysia, Malaysia. His research interest is image processing, machine learning, and pattern recognition.



Rudi Kurniawan is currently a lecturer at the Department of Computer System Engineering Faculty of Computer Universitas Bina Insan, Indonesia. He received a Bachelor of Engineering in Electrical Engineering from Gadjah Mada University, Indonesia; a Magister of Computer in Informatic Engineering from Bina Darma University, Indonesia. He is a student of Doctoral Engineering Science, Informatics Engineering from Universitas Sriwijaya, Indonesia. His research interest is image processing, machine learning, and pattern recognition.



Dr. Fatma Susilawati Mohamad is currently a lecturer at the Faculty of Informatics and Computing, Universiti Sultan Zainal Abidin (UniSZA), Malaysia. She received a Bachelor of Science in Information System Management from Oklahoma, U.S.A., MIT Computer Science from Universiti Kebangsaan Malaysia, and Ph.D. in Computer Science, Universiti Teknologi Malaysia. Her research interests are feature extraction methods, pattern mining, pattern analysis, and object classification.