

Application Image Processing to Predict Personality Based on Structure of Handwriting and Signature

Esmeralda Contessa Djamil
Department of Informatics
Universitas Jenderal Achmad Yani
Cimahi, Indonesia
E-mail: esmeraldacd@yahoo.com

Risna Darmawati, Sheldy Nur Ramdhan
Department of Informatics
Universitas Jenderal Achmad Yani
Cimahi, Indonesia

Abstract—Handwriting stroke reflects how the author faced his world and the emotional honesty. By examining all elements of handwriting and interpreting them separately or integrated, we could generate a sketch of the writer's character traits, emotional disposition and social style using standard of graphology. As image, the analysis of graphology is divided into two approaches that graphics features and segmentation digit each character. In this research, using graphical approach based on a combination of signature and handwriting to predict the more personality using structure algorithms and multiple artificial neural networks (ANN). The image in A4 paper split into two areas: Signature area which nine features and handwriting based on five features. Each area had pre-processing performed to improve the recognition accuracy. Signature area is classified using ANN based on five features and using multi structure algorithms based on four features. While the handwriting area is classified using multi structure algorithm based on four features (margins, spacing between words and lines, and zone domination) and using ANN after hill valley extraction based on baseline features. Eight features are processed using multi-structure algorithms that provide 87-100% accuracy. In the meantime, six features are classified using an ANN which result an accuracy of 52-100%. It used 100 sets of data testing after training using back propagation with 25-75 data. The system has been implemented with the software so that it can be used for classification of personality from handwriting scanned automatically.

Keywords—*handwriting and signature analysis; predict personality; Graphology; structure algorithms; multiple multilayer perceptron*

I. INTRODUCTION

Handwriting and signature are image that have certain pattern reflecting anything such personality prediction. Handwriting analysis is also known as graphology which is pseudoscientific study of handwriting in relation to human mind or called the personality. One of methods in analyzing structural graphic element of handwriting is graphology. Graphology can do to identify the qualities, traits, attitudes, sentiments or postures that seem indicated in the handwriting; they further seek insight into how these aspects of selfhood may integrate together to constitute the dynamic organization

that we recognize as the personality of that writer[1]. Some of guidelines for handwriting analysis are seven basic elements: speed, pressure, shape, dimensions, continuity, direction, and order. Personality overview was obtained from the research on handwriting psychiatric patients [2]. Handwriting can be classified on the various aspects of personality. Meanwhile the use of signatures is usually used to identify certain personality as with appearance of dots, streaks, shapes or shell bottom line [3],[4]. There are two approaches in graphology i.e. graphical analysis of the structure type of writing and analysis of the type of symbol or letter. Signature analysis includes first approach.

If the graphology test is still done manually it takes a long time considering the aspects reviewed in graphology very much. Besides, accuracy of handwriting analysis depends on how skilled the analyst is. Development in image processing and pattern recognition lead to analyzing of handwriting can be done automatically. So it can be used by society at large. Handwriting is an image, so that recognition can be done through the stages of conversion of images into numerical vector, image processing for quality improvement, followed by feature extraction and pattern recognition.

Methods proposed in literature involve the preliminary process of segmentation from the sample and then application of various algorithms / techniques to determine the characteristic traits. Some of them, using baseline, the pen pressure and the height of the T-bar on the stem of the letter 't' are considered for predicting the personality of the writer[5]. Other research about personality analysis considered six main different types of features i.e. size of letters, slant of letters and words, baseline, pen pressure, spacing between letters and spacing between words in a document to identify the personality of the writer. It used SVM that generated various parameters are then calculated by simple use of trigonometry and threshold technique [6]. While other research using the polygon method for pattern recognition based handwriting baseline, slant letter and pen pressure, and the identification of letters i and f using template matching as the input of neural networks [7]. There is also research that uses holistic paradigm in handwriting recognition with whole word approach [8], using back propagation for handwriting analysis in the selection of employees with respect to the size of 'a' letters [9].

This research is a continuation of previous research that only using handwriting based on five features [10]. To provide an overview of the various sides of the personalities, handwriting recognition performed on handwriting and signature. Image was scanned and converted into bmp or jpg format. It was split into two areas i.e. handwriting area and signature area. Handwriting area was reviewed five features whereas signature area was reviewed nine features. Pattern recognition system is divided into two processes namely preprocessing and pattern recognition by feature extraction. Testing the accuracy of the system is only performed on the second stage of testing accuracy for recognizing handwriting or signature pattern of each feature are reviewed. Handwriting and signature is classified ANN and analysis of graphic structure that namely multi structure algorithm. Recognition of signature area based on appearance of five features (curve start, end streaks shell in the middle, middle streaks, underline) using multiple ANN and using multi structure algorithms based on four features (extreme margin, dot, separate signature, and signature streaks disconnected). While recognition of handwriting area based on four features (margins, spacing between words and lines, and zone domination) using multi structure algorithm and using ANN after hill valley extraction based on baseline features.

Each features recognized in parallel that indicates the distinct personality, in order to obtain a review of certain personalities. The success of the system is determined recognition accuracy against graphology test manually.

II. HANDWRITING AND SIGNATURE FEATURES

In this research handwriting analysis carried out on five features that page margin, spacing between words, spacing between lines, dominance of vertical zones, and baselines. Meanwhile signature analysis using nine features that curved start, end streak, shell, middle streaks, underline, extreme margin, dot structure, separate, and streak disconnected. It is depicted in Table I. Meanwhile, signature analysis using nine features to identify certain personality, as in Table II.

Piece of paper is a space, and how the individual stroke it with writing indicating how the author faced his world and the emotional honesty. Each feature reflects personality review as Table I [2].

Of nine features, graphically allowing appears over a signature feature of the structure. If it does not appear, attached personality of features not found on the author. Initial curve of possible signatures are divided into three types, namely backward curved, sharp and curved seamless arch. The second feature that identified the presence of streaks end which consists of a type of increased or decreased. Features to 3-9 are a feature that may appear or not appear from strokes signature that describes a particular personality.

TABLE I. DESCRIPTION FIVE FEATURES OF HANDWRITING

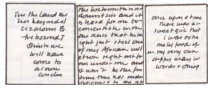
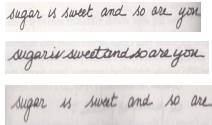
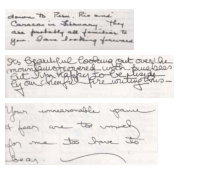
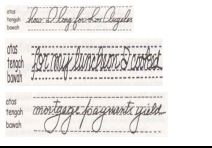

No	Features/ personality review	Type	Examples of Features each class
1	Page margin: Relationship with space and time	15	
2	Spacing between words: emotional comfort with others	3	
3	Spacing between lines: clarity of thought and the desired number of interactions	3	
4	Dominance of vertical zones: personality completely	3	
5	Baseline: how successful a person overcome the influence of external drives	5	

TABLE II. DESCRIPTION NINE FEATURES OF SIGNATURES
















No	Feature	Image	Type	Personality
1	Curved start		Curved backwards	Comfortable going past
			Curved sharply	To formulate a sharp mind
			Curved smoothly	Be careful, friendly, diplomatic
2	End Streak		Increase	Open, foresight, desires ahead, confident
			Down	Lacking spirit, realistic thinking, lack of confidence, easily discouraged.
3	Shell		Shell	Excessive fear, introvert, do not care are approximately, not sociable and do not like to work together
4	middle streaks		Middle streaks	Possessive
5	Underline		Underline	Have unique ideas and thinking, need support to make decisions, and have reliability in the lead.

TABLE II. DESCRIPTION NINE FEATURES OF SIGNATURES (CONT)

6	Extreme margin		Tends to right side	Careless, inattentive
			Tends to left side	Fear of failure, fear of others, lack of confidence, pessimistic
			Tends to top side	Respect yourself, reflect personal happy
			Tends to bottom side	Depressed, shy, feel foreign
7	Dot structure		Dot	Establishment of stable, has a suspicion, not always easy to keep believing
8	Separate		Separate	Had a less pleasant experience in the past
9	Streaks disconnected		Streaks disconnected	Limiting desires, not taking any risks, often discouraged and hesitated to take decisions.

III. METHODS AND DATA DESCRIPTION

Pattern recognition performed on a sheet of A4 handwritten scanned using a scanner in jpg or bmp format as Fig 1. The image is divided into two areas including handwriting and signature area.

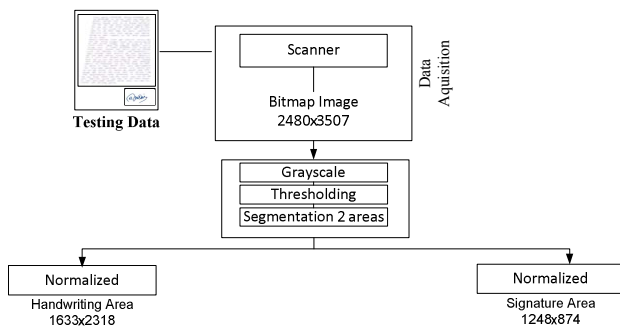


Fig. 1. Design of handwriting and signature area of image of testing data

Pattern recognition and handwritten signatures are used to predict personalities. The system is developed consists of two phases: pre-processing and pattern recognition of each feature stage.

A. Data Aquisition and Image Pre Processing

Handwriting and signature image samples of different individuals are used in this research which is digitally collected by scanning the handwriting of 25 different writers (training data) and 100 different writers (testing data). Each of them was asked to write a text document of simple in running hand and give signature in the box. Most of the handwritings are printed but few of them are cursive handwriting. The samples were

written on A4 size paper without any lines. In pre-processing stage, image processing was done with gray scale and threshold so the handwriting image in a bit then divided by two areas, yield handwriting area and signature area.

B. Recognition of Handwriting System

Handwriting or signature classified in two ways. First, identify certain pixels that can inform structure pattern of handwriting or signature. This method is used when the features have to be reviewed have simple structure pattern. Therefore this method is called multi-structure algorithm. When the feature has a complex pattern, recognition using second methods that is ANN with multi-layer perceptron (MLP) architecture.

The image that has been split over two areas, namely handwriting and signature area, each is done pre-processing stage. After that, the image is segmented correspond to the features that were reviewed. For handwriting has five features, that is page margin, spacing between word, spacing between lines, dominance vertical zones and baseline. Of the five features, four features that is page margin, spacing between words, spacing between lines, dominance vertical zones identified using multi-structure algorithm. While the baseline features were identified using ANN. Segmentation and identification process as Fig 2.

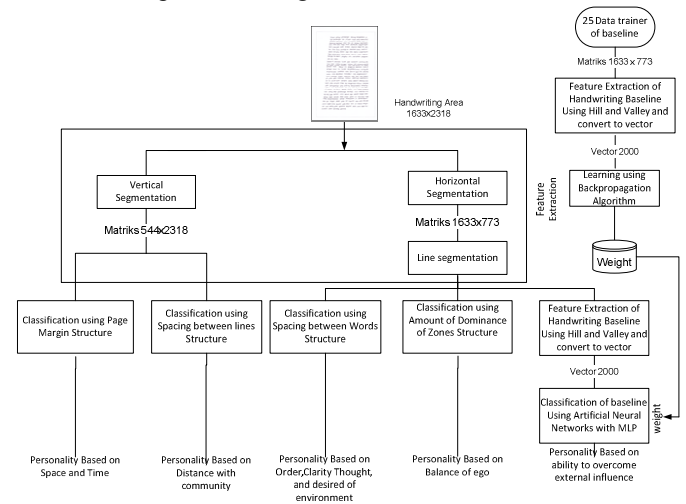


Fig. 2. Recognition of hadwriting system consist 4 features using multi structure algorithm and baseline feature using ANN

The process of feature extraction is done by saving the memory segmentation for classification. Segmentation performed on three stages, ie. vertical segmentation, horizontal segmentation and lines segmentation as Fig 3.

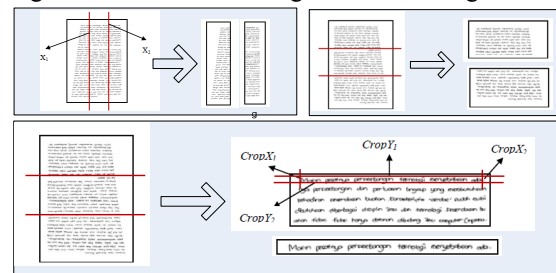


Fig. 3. Three step segmentation

Vertical segmentation divide the image into three sections that right and left side of the image used to analyze page margin. Whereas spacing between lines analysis using left side. Horizontal segmentation to divide into three parts which middle segment are processed into line segments. The process begins by taking the coordinates of x in the upper left corner and y coordinates of the lower left corner. Having found the black pixel value of x will be stored as the value CropX₁. Then look for the value of the value of x CropX₂ started last form to the right y-axis until no black pixels are found. Then look CropY₁ and CropY₂ the same process starting from the bottom point. It used to classify dominance zone, baseline patterns and spaces between words features.

To identify baseline pattern, result of line segmentation process was extracted by Hill Valley feature as shown in Fig 4. The output becomes input for identification system using artificial neural network with Multilayer Perceptron (MLP) architecture.

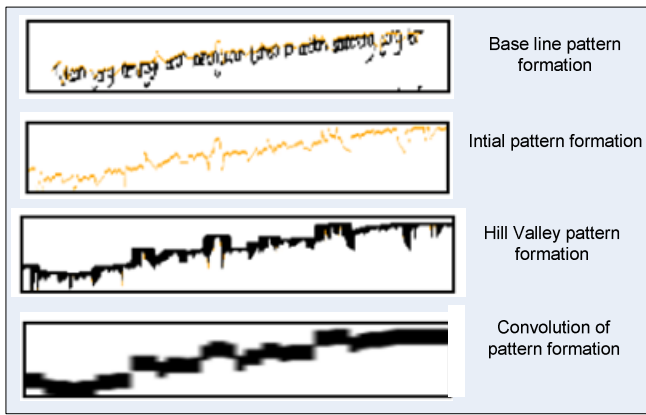


Fig. 4. Feature extraction of baseline feature using Hill Valley

Classification of human personality based on page margin, space between word or line and the dominance of zone feature used multi structures.

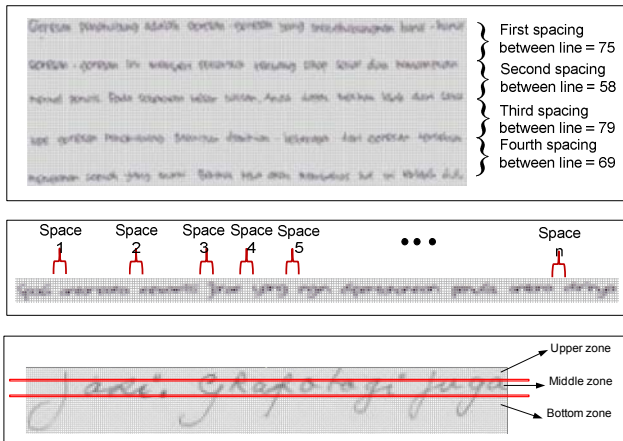


Fig. 5. Classification based on space between lines, space between words and dominance of vertical zones

Analysis of the structure of the page is done by identifying the edge of black pixels in the upper left and upper right. The

second classification is based on the spacing between lines is done by identifying the writing and spaces of color pixels after noise cancelling. Line spacing is calculated of the vertical segmentation as Fig 5. Meanwhile, classification based on space between word and dominance of vertical zones after line segmentation. For classification based on spacing between words, compute average of writing and space, as shown in Fig 5. Classification based on dominance of vertical zones was proceed starting from the left edge. There will be a row of the array, where is searched for the highest series. Highest Array will be divided into three based on the number of existing zones, with the goal to find out where the time of writing is at its decline. From these comparisons it will produce the dominant type of zone as Fig 5.

The last classification is based on baseline structure performed after feature extraction process in Fig. 4, using artificial neural network with as shown in Fig. 6.

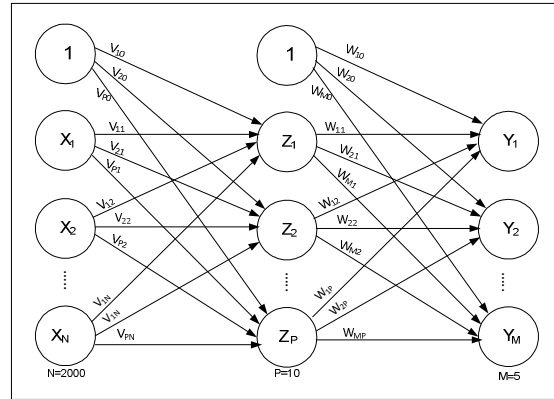


Fig. 6. Architecture of Classification System Using MLP

The system consists of 2000 neurons of input layer, ten neurons of hidden layer and five neurons of output layer. Each input neuron xi is connected to hidden neurons by each weight v_{ij} (1). Then each hidden neuron z_j, is connected to output neurons y_k by w_{jk} (2). Fungsi f(.) is sigmoid biner.

$$z_j = f\left(v_{0j} + \sum_{i=1}^n x_i v_{ij}\right) \quad (1)$$

$$y_k = f\left(w_{0k} + \sum_{j=1}^p z_j w_{jk}\right) \quad (2)$$

The error between the output of the feed-forward network and the target (e_k) correct weight by back propagation algorithm. This research used five output neurons or five classes of the total pattern of the baseline. They are pattern of straight lines, ascending base line, declining baseline, baseline convex and concave bottom line.

$$e_k = t_k - y_k \quad (3)$$

C. Recognition of Signature System

Recognition of signature area based on the nine features, i.e. five features using ANN and four features multi-structure algorithm, as shown in Fig 7. It was processed starting from the signature area.

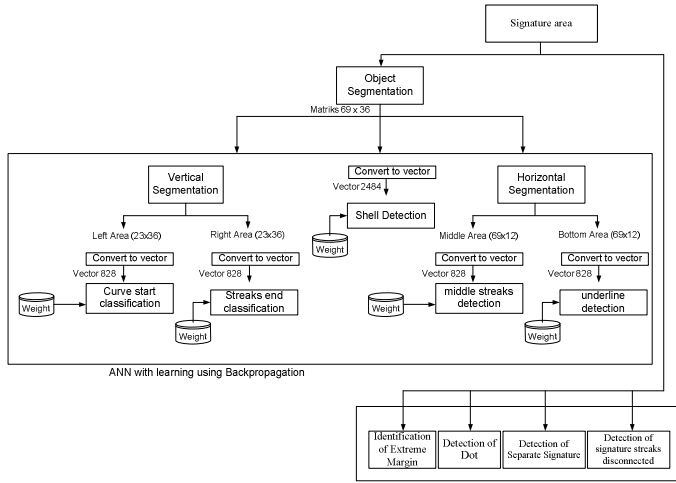


Fig. 7. Recognition of signature system consist 5 features using ANN and 4 features using multi structure algorithm

Signature identification based on four features was done by computing black pixels in signature matrix. Identification extreme margin features by comparing the distance between left and right margin, between top and bottom margin. Detection of dot structure by assuming the dot is a separate structure with a size of 5-17 and 5-15 pixels tall. Detection separate signature was done by finding from top to bottom white pixels more than 15 series. Streak disconnected structure detection was done with the assumption that are on the right, a length of 20 to 38 pixels and a width of 5 to 15 pixels. The detection is same with dot detection. Meanwhile, identification based on five features using ANN after object segmentation as Fig. 8.

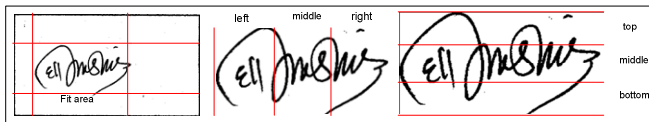


Fig. 8. Segmentation process of signature area

Fit area of signature as Fig. 8, vertically divided into three parts, the left side is used for the identification of curve start and the right part is used to identify the type of end streaks end, declining or ascending. Then the fit area of signature horizontally divided into three parts, the center is used for detection streaks middle and bottom are used for detection of the end the under line signatures. The identification and detection system used Multilayer Perceptron architecture (MLP) in parallel, so has obtained five networks. It consists of input layer with 828 neurons (except for shell, 2484), hidden layer with 10 neurons and the output layer with two neurons (except three for curved start), which are connected via weights. The weights are obtained from the training which is a generalization training data using backpropagation algorithm as Fig 6.

IV. RESULT

Testing of identification system performed comparing visual observation. Tests carried out on a sheet of A4 paper which is divided into two areas including handwriting and signature area. So testing is done separately. Training of ANN identification result 97% accuracy of page margin and space between line features, 94% accuracy of space between word feature, and exactly recognition of dominance of vertical zone feature. Meanwhile, it result 78% accuracy of base line feature using ANN. Hidden layer used 10 neurons. The training with 25 data took 10 hours 52 minutes to achieve maximum MSE for 0.014.

Whereas, testing of signature recognition system using multi structure algorithm result exactly identification of the extreme margin and streak disconnected, 87% accuracy for separate signature detection and 99% for the dot detection of signature end. Using of ANN identification result after training with 150 epoch and parameter as Table III. Accuracy of the test is not always influenced MSE training. As an example of middle streaks features result well recognized that 75%, whereas the resulting MSE of training just 0.01. This is possible given quite training data varies made great MSE, but large variations in the training data made the system better able to recognize a variety of test data. Training with a maximum of 150 epochs was provide MSE is small enough that 0.009 – 0.015. Therefore increasing the accuracy of the test is less affected by the increase in the number of epoch.

TABLE III. TRAINING PARAMETERS AND IDENTIFICATION OF SIGNATURES PATTERN USING

Feature	Training			Accuracy
	Data	MSE	Time	
Curved start	75	0.015	6 hours 30 minutes	63%
End Streak	52	0.009	2 hours 45 minutes	58%
Shell	52	0.010	13 hours a minute	56%
Middle streaks	48	0.010	2 hours 19 minutes	75%
Underline	47	0.009	an hour	70%

Using of five ANN designed in parallel and called multiple ANN allows detecting the presence of more than one feature. Pattern recognition handwriting and signature will result 7-14 different aspect of personalities. It gives more of an overview of the author's personality.

Weakness of using ANN to recognize shell and end streak are less variation of the data training and the number of neurons in the hidden layer is only 10 less than comparable to the number of input neurons of 2000 pieces. The magnitude of the size of the image does not affect the accuracy of pattern recognition significantly, but the effect increase number of

memory and processing time. If using ANN, it would take more training time.

V. CONCLUSION

This research proposes a system that can identify the type of handwriting from a variety of features personalities using multi structure and ANN algorithm. It is useful to obtain an overall picture of the personality. The system using combination of identification of signature style and handwriting using multiple network, so can predict more aspect of personality reviewed. Eight features are processed using multi-structure algorithms that provide 87-100% accuracy. In the meantime, six features are classified using an ANN which result an accuracy of 52-100%. It used 100 sets of data testing after training using back propagation with 25-75 data.

Using ANN to recognize handwriting pattern depend on variation of the data training and the number of neurons in the hidden layer comparable to the number of input neurons. In meanwhile, magnitude of the size of the image and epoch maximum are less affect the accuracy of pattern recognition significantly.

This system has been implemented in software to provide convenience to the public in identifying personality easily and quickly. It can be used in the selection of the employee or job application letter. Besides the system can apply in determination of an appropriate field of study.

For the future, handwriting pattern recognition system can increase the number of features in order to provide a more complete picture of personality. It should also consider variation of training data and more optimal parameters of training in order to provide better accuracy. In addition, it is

necessary to find the relation between personality pictures of the handwritten signatures for more accurate prediction of personality

REFERENCES

- [1] A. McNichol, "Handwriting Analysis Putting It to Work for You", Contemporary Books, 1994.
- [2] B. Ludvianto, "Handwriting Analysis", Gramedia Pustaka Utama, 2011.
- [3] P.D. Sunar, "Read the full review Graphology Personality Through Writing His handwriting", Yogyakarta : Diva Press, 2010.
- [4] Taufik, M., Rosette, "Step By Step to analyzed personality trait through handwriting", Jakarta, 2010.
- [5] H.N Champa, K.R AnandaKumar, "Artificial Neural Network for Human Behavior Prediction through Handwriting Analysis", International Journal of Computer Applications (0975 – 8887) Volume 2 – No.2, May 2010.
- [6] S. Prasad, V.K. Singh, A. Sapre, "Handwriting Analysis based on Segmentation Method for Prediction of Human Personality using Support Vector Machine", International Journal of Computer Applications (0975 – 8887) Volume 8– No.12, October 2010.
- [7] P.K.Grewal, D Prashar, "Behavior Prediction Through Handwriting Analysis", IJCST Vol. 3, Issue 2, April - June 2012
- [8] S. Madhvanath, V. Govindaraju, "The Role of Holistic Paradigms in Handwritten Word Recognition", IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE, VOL. 23, NO. 2, FEBRUARY 2001.
- [9] Z. Mohd Zam, "Handwriting Analysis for Employee Selection Using Neural Network", Thesis of Intelligent System Faculty of information Technology and Quantitative Science of University Technology MARA, 2006.
- [10] EC. Djamal, "Recognition of Human Personality Based on Handwriting Using Multi Structures Algorithm and Artificial Neural Networks", 2nd IEEE Conference on Control, Systems & Industrial Informatics Bandung, Indonesia June 23-26, 2013.