Detecting features of human personality based on handwriting using learning algorithms

Behnam Fallah[1], Hassan Khotanlou[2]

[1]Department of Computer, Qazvin Branch, Islamic Azad University, Qazvin, Iran.

bmfallah@yahoo.com

[2]Department of Computer, Bu-Ali Sina University, Hamedan, Iran.

Hassan.khotanlou@gmail.com

Abstract

Handwriting analysis is useful for understanding the personality characteristics through the patterns created by the handwriting and can reveal features such as mental and emotional instability. On the other hand, it is difficult to determine the personality, especially when it is associated with the law because there is no threshold or scale being able to make detailed results of the analysis. This thesis aims to provide an automated solution to recognize the personality of the author by combining image processing and pattern recognition techniques. The personality recognition system proposed in this project is composed of two main parts: training and testing. In the training part, after feature extraction from all image patterns of the input text, a proportional output is created through the MMPI personality test. Then these inputs are trained to the neural network as a pattern. As a result of this training, a comprehensive database will be formed. In the testing part, the database is used as a main comparison reference. After feature extraction, the input text image is compared with all patterns in the database to find the closest image to the input text image. Finally, the MMPI personality test output for the proposed text image is introduced as the output personality parameters.

Keywords: *handwriting recognition, neural network, MMPI personality test, graphology*

1. Introduction

Handwriting analysis is an act that has been performed for many years. However, when we analyze the behavior and personality of an individual, its effects are still discussable. Each of these neural brain patterns lead to a unique neural and muscular movement for an individual; therefore, this small subconscious movement occurs for each person, who has a certain personality feature while writing. the problem of recognizing the writer`s personality from his handwritten texts aims to specify the personality of the writer given a handwritten text, which of course this personality elicitation is performed based on samples of different individuals` handwritings. Professional handwriting researchers investigate and analyze the sample handwriting.

Writing can indicate personality features like feelings, fear, honesty, etc. identifying the personality of a human being by his handwriting is an old technique. Before, the nature of an individual was predicted manually, which took a long time. Recognizing a writer`s personality from his handwriting has recently become a considerable and interesting subject in psychology.

2. Previous Works

This section discusses a number of feature extraction methods from Persian and Arabic alphabets, words, and handwritten texts.

Reference [1] reviews identity recognition methods from handwriting and graphology. Moreover, a review of handwriting analysis computer systems in the market is presented and compared for better understanding.

Reference [2] points out and uses salient and important features of using handwriting in graphology analysis, including curves and figures in paper margins, line spaces, line tilts, word tilts, sharpness of edges, character sizes, text density, writing speed, and order point. This paper proposed some methods for the first time. Moreover, for automatic feature extraction from Persian handwritten texts, 24 training samples and 118 test samples are used for the experiments.

Reference [3] considers 6 different feature types for computerizing handwritten graphology: 1- character sizes, 2- word tilts, 3- baseline, 4- pen pressure, 5- space between characters, and 6-space between words in the document that is used to recognize the writer`s personality.

Reference [4] claims that potential behavioral and personal deviations of an individual are possible by analyzing his handwriting. In this paper, two methods are proposed for handwriting analysis:

1- Graphology that is a psychological analysis method.
2- Graphology that is used to identify the writer.

Reference [5] discusses the effect of brain neural patterns on micro-movements of the muscles, such that because of these micro-movements, personal parameters are emerged in human daily behaviors like writing a text. All hits, patterns, and the pressure that is applied when writing a word can express personal behaviors of an individual. In this reference, features like the tilt of the baseline, pen pressure, tilt and size of characters are used to extract the personal parameters of an individual. This reference used the linear regression method at the classification stage.

Reference [6] extracts personal parameters like extrinsic emotions, fear, trustworthiness, etc. from the graphology of handwriting. Professional handwriting tests, which are called graphology, are mostly recognized by part of the handwritten text. The handwriting analysis accuracy depends on the skills of the expert. In this paper, a method was proposed to investigate the personality based on features like baseline, pen pressure, and character t in a separate handwritten text. These parameters and features are defined as the input of a neural network, whose output is the personal behavior of the writer. The performance of this system is measured by different experiments.

Reference [7] uses computers to accelerate the image processing of Persian handwritten texts. More specifically, important features are extracted from the handwritten text to understand the psychological state of the writer. In order to do so, features like boldness, compression, and two word space measures are achieved. The tilt angle of the text is then obtained and after removing it, the main points corresponding to the right, left, and up margins, as well as the shape of the margins are achieved. Finally, the line vibration is obtained. Extracted features are divided into four groups, which are sent to three fuzzy systems and a non-fuzzy system to recognize the personality of the writer.

3. General Principles of Personality Recognition System from Persian Handwritten Texts

Figure 1 presents the block diagram of a personality recognition system from the handwritten texts. The innovative methods in personality recognition from handwritten texts are ensued by changes in running one or several blocks of the following diagram. The general name of the system is in fact changed to the block performance. For instance, recognizing personality from handwritten texts based on neural networks is in the classification section.

3.1. Pre-Processing

All operations performed on texts, which facilitate the process of following phases, e.g. removing noise, smoothing, thinning, language recognition, word fonts, etc. the set of these processes have the following goals [8]:

1- Reducing noise

2- Contour smoothing

3- Storing information that should be protected (thinning)

Activate binarizing for facilitating working with edit, ruler, and paragraph markers. By activating “Ruler” in the “View" section, you can see the settings of specific distances, columns, and margins. In order to find the undesired parts of paragraphs, like spaces between alphabets, pages, and lines, specific spaces, and chapter headers, activate paragraph markers (¶) in the “Paragraph” toolset.

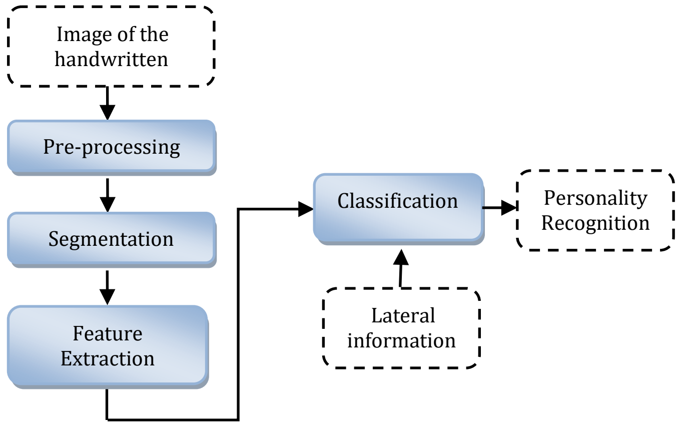

Figure 1. The block diagram of a personality recognition system

3.2 Segmentation (Separation)

The segmentation stage is very important for feature extraction, and calculating the sizes of characters and words in personality recognition systems from handwritten texts [9].

3.3 Feature Extraction

Feature extraction aim is to reduce data to a set of features, such that they are constant in the handwritten texts of a certain individual and independent of the other person's handwriting.

3.3.1 Text Independent Features

Independent text features include the margin value from the beginning of the page, word expansion, characters sizes, line spaces, word spaces, word tilts, horizontal to vertical ratio of characters, and lie tilts.

3.3.2 High-Order Local Autocorrelation as a Text Dependent Feature

Autocorrelation is one of the most well0-known functions insensitive to shift [10]. In what follows, a special type of autocorrelation is introduced:

Autocorrelation function is known as a shift insensitive function. The N-th order autocorrelation function is defined by N location changes$(a_1, a_2, a_3, \dots, a_N)$, from reference point r as equation (1).

$$x^N(a_1, a_2, a_3, \dots, a_N) = \int I(r)I(r + a_1) \dots I(r + a_N)dr \quad (1)$$

Where, I(r) is the text image and $r = z = (\log(p)\,, \boldsymbol{P})$. There are many autocorrelation functions, which are resulted from different combinations of location changes on the text image. Figure 3 presents different masks with location change patterns obtained by the autocorrelation function. In this research, the number of location changes is reduced by removing location changes that are equal due to the same shift. In other words, the value of N is limited to 3 (N= 0, 1, 2, 3). Moreover, the ratio of location changes and displacements are limited in a 3×3 position window [10]. In order to obtain the values of HLAC, the text image is first scanned by 70 local 3×3 masks. The sum of the multiplication of corresponding pixels` values by one (pixels of black mask) is then computed for each mask [11]. This method is known as HLAC feature (figure 2).

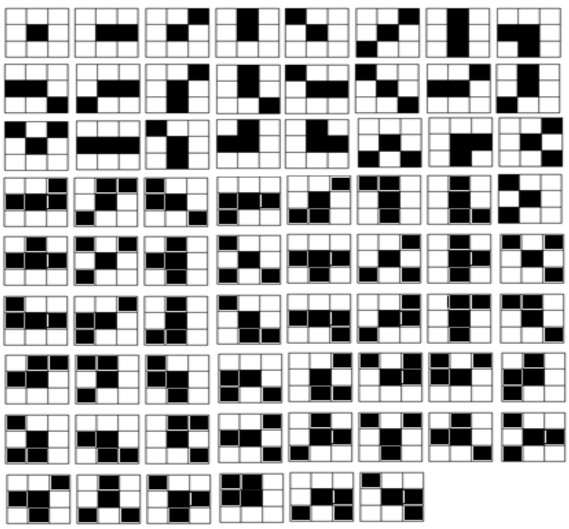

Figure (2). 70 masks with different patterns to extract HLAC [8].

3.4 Feature Vector and Generalized Discriminant Analysis (GDA)

In the proposed personality recognition system, when the number of training patters is greatly increased, the resolution of HLAC is reduced, which causes class interference. In this research, generalized discriminant analysis (GDA) is used to increase the space and resolution of different classes [12]. GDA is resulted from the non-linear extension of the linear discriminant analysis and it is successfully used for many applications. This method can be used to overcome classification problems. GDA helps us to combine features and increase the resolution of the classes.

3.5 Classification and Recognition (With One or Several Classifiers)

This stage includes methods to map each pattern extracted from the feature extraction stage with one of the classes of the corresponding pattern space. This is performed through minimizing the feature vector of each input pattern in proportion to one of the reference vectors. Reference vectors are vectors that are derived from the training samples a priori. The proposed techniques for this stage can be searched in one of the general pattern recognition discussion groups [12]:

- Pattern matching
- Statistical techniques
- Neural networks

The three groups above are not necessary performed separately and they may be found among the techniques of other groups.

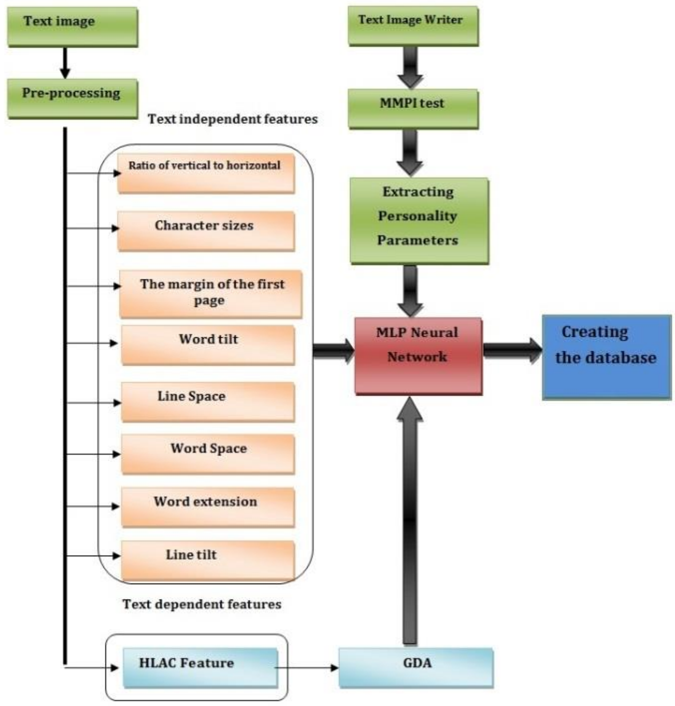

Figure 3(a). the block diagram of the personality recognition Training system

4. The Structure of the Handwriting Personality Recognition System

The proposed personality recognition system consists of two main training and test sections (figure 3). In the training section, after extracting features from all patterns of the input text image, the corresponding output is created using a MMPI personality test [13]. These input-outputs are then trained as a pattern in a neural network and finally, a comprehensive database is created as a result of training.

In the test section, this database is used as the main comparison reference. At this stage, after feature extraction, the input text image is compared with all patterns in the database to find the closest image. Finally, the output of the MMPI personality test corresponding to the selected text image is introduced as output personality parameters.

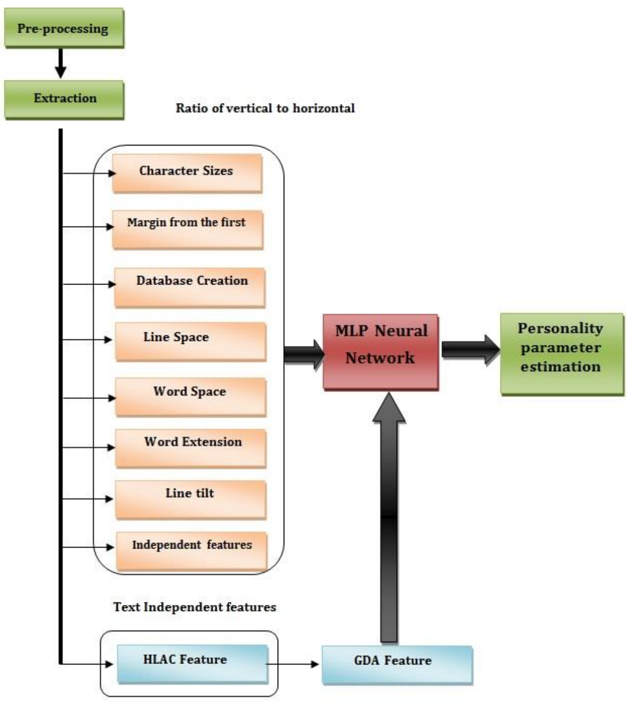

Figure (3b). the block diagram of the personality recognition Test system

5. Experimental Conditions and Database Type

In order to evaluate the proposed method, 70 individuals with different educations, ages, and genders were asked to fill the designed forms with their ordinary handwriting. These individuals were selected from ordinary people (college students, employees, etc.). Table (1) presents the characteristics of these individuals, including gender, education, and age.

Table 1. The characteristics of the individuals who filled the database's forms

Gender		Education			Age	
Female	Male	Elementary school to diploma	Diploma to undergraduate	Undergraduate to doctorate	7 to 40	Higher than 40
32	38	18	36	16	37	33

In order to perform the experiments and evaluations, these individuals were asked to write a constant text in one paragraph, which contained various words.

Subsequently, each writer took the MMPI personality test, which was designed in 1940s by two researchers of the University of Minnesota, United States. This personality test consists of 71 fundamental questions and the writer should select from true or false options. Finally, a score is given to the test that is in fact the character profile based on eleven clinical scales. These results are stored together with the ID that is assigned to each writer before writing the corresponding paragraphs.

The following points are considered when filling the forms:

- There is no limitation regarding the handwriting type. The collected samples include various handwritings. Of course, in some cases of writings filled by female individuals, the handwritings are very similar.
- The individuals were asked to fill the forms in a specific time duration and write using their ordinary handwriting without trying to alter or improve it.
- The forms do not have lines and only the paragraph area is specified.

Since the handwriting of each individual is changed by his/her mental and environmental conditions, the subjects were asked to fill the forms with patience and tranquility. After collecting the forms, they were all scanned as gray-scale images in 300dpi resolution.

6. The Environment of the Simulated Software

In this dissertation, all simulation are performed using MATLAB 2013a and in some cases, Matlab instructions are used to transform gray-scale images into binary ones and label groups. Moreover, DRTools[1] toolbox is used to apply the GDA algorithm, whose reception and route should be separately defined in the Matlab toolbox.

7. Results of the Proposed Algorithm

In this section, the output of the proposed personality recognition system is compared with those of MMPI test (i.e. the 11 scales of the MMPI test). Figure 4 presents the HLAC feature before and after using the GDA algorithm:

[1] Dimensionality Reduction

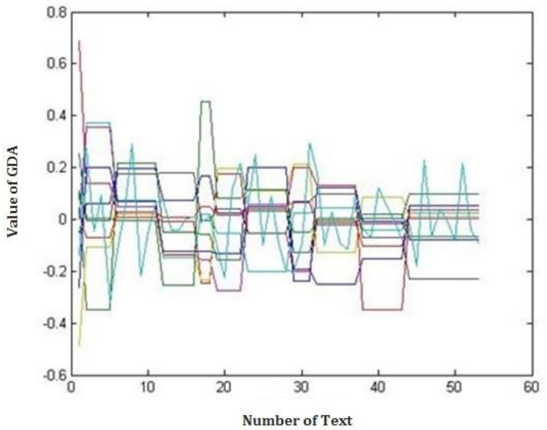

Figure 4. The extracted features for 53 handwritings without using GDA

In figure 4, the horizontal axis represents the number of handwritings and the vertical axis shows the value of the extracted feature. As we can see in the figure, after applying GDA, the resolution is considerably increased. In these experiments, from the total 70 samples, 50 were selected for training and the rest was used for tests. This comparison is performed for different training conditions, including the main parameter of the MLP neural network, e.g. the number of neurons in the middle and input layers, training duration, etc. table 2 presents the results of this comparison.

As we can see, the highest recognition rate is resulted when the number of neurons in the input layer is 18 and the number of neurons in the hidden layer is 10. Results of the neural network are presented in figure 5.

Table 2. Comparison of the results for different neural network training conditions

The average personality recognition rate in comparison to MMPI test		Training duration (seconds)	Number of trainings	Number of neurons in the hidden layer	Number of neurons in the input layer
Test	**Training**				
46%	69%	500	70	5	10
59%	72%	500	70	7	15
61%	76%	700	70	10	18
52%	74%	1100	70	15	20
57%	71%	2000	70	19	25

As we can see, the highest recognition rate is resulted when the number of neurons in the input layer is 18 and the number of neurons in the hidden layer is 10. Results of the neural network are presented in figure 5.

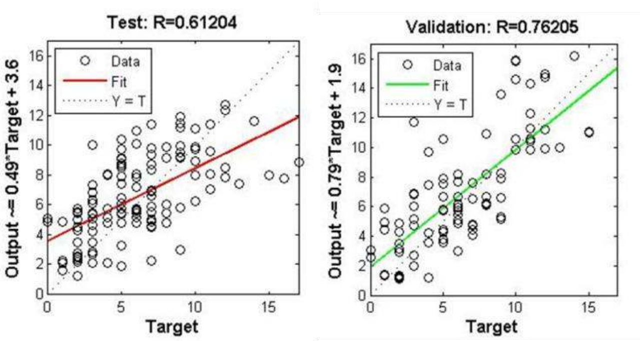

Figure 5. (a) Regression of the validation data (b) regression of test data

Figure (5a) presents the regression of validation data and figure (5b) presents the regression of the test data. Moreover, table 3 is presented to show the role of each feature extraction stage in improving the final output of the personality recognition system.

In this table, the input feature vector of the personality recognition system only consists of one feature extraction method mentioned in each row. The same feature vector is used to estimate the clinical and validity scales of the neural network. Therefore, according to the recognition rate of the personality recognition system in each row, we can find the effect of the feature extraction method on estimating the clinical and validity scales parameters.

According to table 3, we can find some important points. The sum of the values in each row of table 3 and normalizing it, shows the effect of each feature extraction stage on the final output of the personality recognition system.

As we can see, the HLAC feature has the highest effect (29.08%) and the ratio of vertical to horizontal words has the lowest effect (8.56%) on the accuracy of the propose system`s output. Moreover, the sum of the values of each column in table 3, whose results are presented in figure 6, shows the average estimated scales for each test of the proposed personality recognition system.

Table 3: The role of each feature extraction stage in improving the final output of the personality recognition system

Scale \ Feature	*Ratio of vertical to horizontal*	*Character sizes*	*Word tilt*	*Line spaces*	*Word spaces*	*Margin value from top of page*	*Word extension*	*Line tilt*	*HLAC feature*
L	5%	6.5%	6.8%	8.5%	4.5%	3.1%	6.4%	6.5%	17.3%
K	4.5%	5.8%	6.2%	7.6%	5.1%	3.8%	5.6%	6.7%	16.8%
F	4.8%	6.1%	5.6%	6.4%	3.1%	4.2%	6.1%	5.9%	16.2%
Ma	5.1%	5.4%	5.9%	6.8%	4.7%	4.5%	4.9%	6.1%	17.5%
Sc	5.7%	4.6%	6.1%	5.2%	3.4%	4.1%	5.7%	5.3%	16.5%
Pt	6.1%	6.8%	6.4%	7.5%	4.6%	3.2%	4.9%	5.6%	16.4%
Pa	4.7%	5%	5.2%	6.9%	4.5%	4.5%	5.2%	6%	16.4%
Pd	4.2%	5.3%	5.5%	7.2%	5.1%	4.3%	5.8%	6.5%	17.1%
Hy	5.8%	4.9%	6.4%	6.9%	5.7%	3.5%	4.5%	5.9%	17.6%
D	5%	5%	5%	5%	5.1%	4.9%	4.6%	5.6%	16.5%
Hs	5.7%	3.4%	5.9%	6.3%	4.4%	5.4%	5.4%	5.8%	17.4%

In this figure, the clinical scale Hs has the highest value (64.60%) and the clinical scale Pa has the lowest estimation value (56.6%).

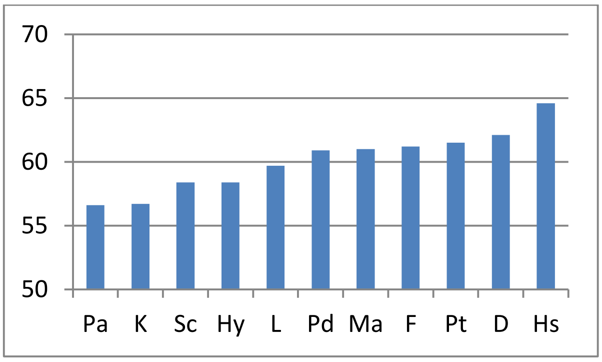

Figure 6. The average estimated scales for each test

Finally, the result of the proposed algorithm is compared with conventional methods using one database and table 4 presents the results.

Table 4. Evaluation of the proposed algorithm

Test	Efficiency	measure
13%	20%	K-Means
46%	58%	Fuzzy C means
61%	76%	Proposed method

8. Conclusions

This research employs a personality recognition system which automatically extracts the character parameters from Persian handwritten texts. Well as due to the use of valid personality test (MMPI) in the training system, the extracted personality parameters on the test step was pretty standard, in result the proposed method has superior finality to other methods. The proposed system create feature vector using HLAC independent feature, context features such as value of margin from the top, word extraction, character sizes, line space, word space, word tilts, horizontal to vertical ratio of characters, and lie tilts. The MLP neural network is used for classification; Such that the output of this network will be the parameters of the author characters. Training and evaluation prepared for the database system is used by 70 different writers.

considering the results, the advantage of the proposed algorithm is as follows:

1- Using dependent and Independent features of text in the process of feature extraction.

2- The proposed personality recognition system is automated, particularly in the process of feature extraction (None automated systems are very unendurable and time-consuming)

3- Enhance accuracy and reliability of the personality recognition system due to the use of MMPI test on training step

4- No need to segmentation on feature extraction phase

5- Using GDA to increase the space between classes

References

[1] S. Hock, K. Siang Teh , L.Y. Yee, "An Overview on the Use of Graphology as a Tool for Career Guidance", CMU. Journal 2005, Vol. 4(1).

[2] A.B.Sharif, E. Kabir, "Computer Aided Graphology For Farsi Handwriting ", Computer and electronic engineering journal (Persian), 2005 , Vol 3 , Num 2; Page(s) 73-79.

[3] Sh. Prasad, V. Kumar, A. Sapre, "Handwriting Analysis based on Segmentation Method for Prediction of Human Personality using Support Vector Machine", International Journal of Computer Applications (0975 – 8887) , October 2010,Vol 8– No.12.

[4] J. Fisher, A. Maredia, A. Nixon, N. Williams, J. Leet, "Identifying Personality Traits, and Especially Traits Resulting in Violent Behavior through Automatic Handwriting Analysis", Proceedings of Student-Faculty Research Day, CSIS, Pace University, May 4th, 2012.

[5] A. Rahiman, D. Varghese , M. Kumar," Handwritten Analysis Based Individualistic Traits Prediction", International Journal of Image Processing (IJIP),2013, Volume (7) : Issue (2) , 2013.

[6] H.N. Champa , K.R.AnandaKumar , "Artificial Neural Network for Human Behavior Prediction through Handwriting Analysis", International Journal of Computer Applications (0975 – 8887), May 2010, Vol 2 – No.2.

[7] Y.Norouz Zade, H.NezamAbadi,"Handwriting Graphology using image processing and fuzzy inference system",8th Intelligent Systems Conference, Ferdowsi University of Mashhad, Iran. 2007.

[8] I. A. Jannoud, "Automatic Arabic Hand Written Text Recognition System," *American Journal of Applied Sciences*, vol. 4, pp. 857-864, 2007.

[9] Tsai C-F, Wu J-W, (2007). "Using neural network ensembles for bankruptcy prediction and credit scoring. " Expert Systems with Applications,in press.164-173.

[10] T. Kurita, K. Hotta, and T. Mishima, “Scale and rotation invariant recognition method using higher-order local autocorrelation features of Log-Polar image,” Proc. Third Asian Conference on Computer Vision, 1998, Vol 2, pp.89-96.

[11] S. M. Lajevardi, Zahir M. Hussain, “Facial Using GA to increase the distance between classes Expression Recognition: Gabor Filters versus Higher-Order Correlators,” International Conference on Communication, Computer and Power (ICCCP’09), Muscat, Oman, 15-18 Feb. 2009.

[12] Arica Nazif ,Yamin-Vural Fatos T., “An overview of character recognition based focused on off-line handwriting”, IEEE Transactions on Systems, Man, and Cybernetics-Part C: Applications and Reviews, May 2001, Vol. 31, No. 2.

[13] Perssman, Lyons , Lavson & Strain, “Religions belief depression , and ambulation Status in elderly woman with broken hips”. American Journal of psychiatry , 1990, 147, 758-66.

Behnam Fallah received his B.Sc. in computer engineering from department of computer Engineering, Hamedan Branch, Islamic Azad University, in 2008, and her M.Sc. in computer software engineering from department of computer, Qazvin Branch, Islamic Azad University, in 2015.His research interests includes Image processing, Wireless Networks and Data Mining.

Hassan Khotanlou is an associate professor in department of computer at Bu-Ali Sina University, Hamedan, Iran. He received his B.Sc. degree in computer engineering from IUST University in 1995, and his MSc. degrees in artificial intelligence engineering from Shiraz University in 1997 and his Ph.D. degrees in artificial intelligence engineering from Pierre & Marie Curie University (Paris VI – TELECO) in 2007.His main research interests are Image processing, Fuzzy Systems, Acoustic, Data Mining, Computer Networks, Medical Image processing and Artificial Intelligence.