An Expert System for Determining the Emotional Change on a Critical Event Using Handwriting Features

Bora Uğurlu [1], Rembiye Kandemir [2], Aydın Carus [2], Ercan Abay [3]

[1]*Canakkale Onsekiz Mart University, Department of Computer Engineering, Canakkale, Turkey*
[2] *Trakya University, Department of Computer Engineering, Ahmet Karadeniz Campus, Edirne, Turkey*
[3]*Istanbul Kemerburgaz University, Department of Psychiatry, Istanbul, Turkey*

***Abstract* –An individual may sometimes feel anxious when a critical event happens. Job interview, wedding, moving in a new city/country can result this occurrence. Examinations taken in school are also that kind of events. Since our handwriting is controlled by brain, it is possible to see clear changes in handwriting style during examinations. In our study, an expert system is developed which considers handwriting features to predict student's exam anxiety state. 210 handwriting samples are collected and classification is made by using J48 decision tree algorithm. The average of Precision, Recall and F-Measure metrics are 71%, 66% and 67%, respectively.**

***Keywords* – Information system, emotional state, anxiety, decision tree**

1. Introduction

We all take some notes and sign papers in our daily life. Frequently, there are also strange symbols with notes. All of these have some deep meanings. All of these are thought to be done unconsciously. However, they are not. Like our retina and finger print, handwriting is also unique for the individual. Handwriting is under the control of brain [1]. Each personal trait is represented by neurological brain patterns and each of these patterns form unique neuromuscular movement. The movement produced is all the same for people sharing same personal traits [2].

Our emotional states are taught to have an impact on our handwritings. So much knowledge about an individual mood can be acquired when his/her handwritings are analyzed. It contains many personal characteristics such as emotional transitions, fears, honesty and anxiety, etc. Whether he/she is outspoken or able to keep secret can be inferred through a person's handwriting [3].

The discipline focused on handwriting is called "Graphology". Person who is interested in graphology is called graphologist. The research area establishes some connections between handwriting and person's psychological profile [4]. Studies about handwriting features and people's interests about this area have a long history. For Emperor Augustus in Roman age, some oddities in the handwriting are detected and reported by an ancient biographer Seutonius Tranquillus [5]. All analyses of handwriting are made in visual manner in the Roman age, which depends on the ability of graphologists to highlight emotional state of a person. Graphologists' experience and knowledge about handwriting features are determinant.

In the literature, handwriting analyses is no limited to graphology. In diagnostic perspective, Kucera and Havigerova [6] also investigated handwriting. Graphology or handwriting analysis looks through some characters and patterns in documents [7,8]. For detecting people's lies, Li and Tang [9] recommended eight features in the handwriting. The deviation from normal handwriting of people begins with telling a lie [9]. For the assessment of neuropsychological drawing tasks, Glenat et al. [10] introduced how handwriting can be used.

The recruitment, signature verification, truth tests and medicine are application areas of handwriting analyses. Anxiety is one of our main feelings. When we are afraid and worried we often show similar reactions, which are difficult to define for most of us.

DOI: 10.18421/TEM54-11
https://dx.doi.org/10.18421/TEM54-11

Corresponding author: Bora Uğurlu,
Canakkale Onsekiz Mart University, Department of Computer Engineering, Canakkale, Turkey
Email: boraugurlu@comu.edu.tr

Panic attacks occurred when a person is distressed, such difficulties may cause anxiety. When we feel anxious, the most common symptoms appear as changes in blood pressure and body temperature, rapid breathing, perspiration. In their lifetime, students experience some situations which cause anxiety. For example, mid-term and final examinations are the reasons of anxiety.

Despite the fact that there are different techniques and mechanisms used for measuring anxiety level [11], State Trait Anxiety Inventory (STAI) is widely used scale for measuring the anxiety level. It is developed by Charles Spielberger et al. [12]. This anxiety inventory is capable of measuring various types of anxiety and can be applied to great range of socio-economic segments of society. There are two types of anxiety which can be assessed by STAI. The first is state anxiety. Inconvenience, fear and some other similar feelings can be covered by state anxiety. Person feels this kind of anxiety when the threat is approaching. Therefore, it is rather temporary. In the opposite, person feels trait anxiety in habitual daily life [12].

In this study, making predictions about students' exam anxiety levels using their handwriting features is our main concern. Therefore, we have developed an expert system which is capable of making anxiety prediction based on handwriting features. As a result of the study, expert system can indicate whether an individual is anxious or not based on handwriting features.

This paper is organized as follows. Section 2 describes the related works, which are previously published papers in the literature. Section 3 provides architectural design and section 4 gives the details of implementation of the proposed expert system. The experimental results are presented, and statistical results are given in Section 5. Experimental results are based on classification results obtained by a decision tree, and classification accuracy and performance measurements and evaluation metrics are also presented. Section 6 concludes the paper.

2. Related Work

Different implementations of handwriting can be found in medicine. Dyspraxia and Parkinson are some significant illnesses and their diagnosis is important. Therefore, impacts of drug usage and diseases on handwriting are very important for the researches, which exhibit pathological aspects of these illnesses. Some serious disorders such as ataxia and tremor may show extraordinary characteristics and other findings for handwriting analyses [6].

Detecting anxiety level within offline handwriting features in a computerized way is fairly new in the literature. Discovering the relationship between anxiety and handwriting features is the key factor for a system that makes prediction about person's anxiety level within handwriting features. Whereas many debates have been going on about this issue [13], some argue that there is a relationship between students' anxiety levels and their handwriting features. Since each person's anxiety threshold on a critical event is dissimilar, the relationship mentioned above is not the same for all people [14].

Macaulay [15] investigated the relationship between stress and motor activities involved in the operation of the fingers. He indicated no obvious correlation between the speed of mouse-click and the state of anxiety. Desment and Hoste [16] described a machine learning methodology for fine grained emotion detection in suicide notes. 15 different emotions are taken into account. These emotions are taught to have some relationship between suicidal behaviors. In this study, emotional behaviors within handwritten suicidal note is obtained by natural language processing. Djamel, Darmawati, and Ramdlan [17] present a system which is able to predict personality. Their system is based on the structure of handwriting and signature. Page margins, spacing between words and lines are some of the chosen handwriting features for their prediction mechanism.

Machine learning and data mining is also another common tool used in medicine. Suhasini, Palanivel, and Ramalingam [18] show how decision support systems can be used for psychiatry problems such as depression and anxiety. Their proposed method has an accuracy of 98.75% for identifying the psychiatric problem. Lu et al. [19] developed a personalized reading anxiety model to reduce the reading anxiety of learners while reading. C4.5 algorithm runs in their model. There is another study which aims at an improvement for text based emotion classification and prediction by using a customized decision tree algorithm [20].

As seen in the above given studies, many different algorithms and approaches such as decision trees and neural networks are studied. In our study, J48 decision tree is used to classify the emotional states using handwriting features. Extracted handwriting features are located into the nodes of J48 tree. In this frame, the inference process is operated from root to leaves on the decision tree which are based on classification rules. On the contrary, the other studies did not use the handwriting features in this manner. In Section 3, decision tree operations are explained in details.

3. Proposed System

In our previous work, we developed a novel model which is able to give us statistical outputs about handwriting features and Spielberger Trait and Anxiety Inventory (STAI) scores [14]. The presented model could also mark the students if they were anxious or not. In this work, we have added a prediction mechanism into our previous model and we developed an expert system. The prediction system's details are shown as architectural design, and how it is implemented follows.

The general architectural design of the proposed system is given in Figure 1. The main input of this handwriting are samples of the students. All handwriting documents of the selected students are obtained with pen and paper style in classroom.

In Figure 1., there are three different time periods considered, which are named Part A, B and C, respectively. The first time period was chosen a month ago (Part A) before the mid-term exams. In

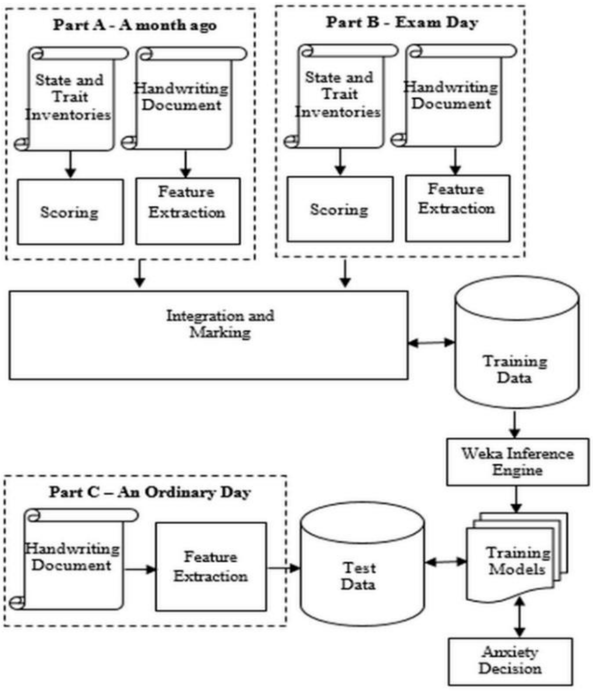

Figure 1. Architectural design of prediction system

that time period, selected students were thought not to feel anxious about exam. The second time period (Part B) was chosen just before the beginning of the mid-term exam. In this time period, the selected students were thought to have maximum anxiety level. The last time period (Part C) is an ordinary day. In Part A and B time periods, both

Assign the ***average*** to calculated average of state anxiety scores of students.
If (a ***student's state score*** is greater or equal to the ***average)*** then
    Label the owner of handwriting "***anxious"***
else
    Label the owner of handwriting "***not anxious***"

Figure 2. Pseudo code for labelling training data

handwriting samples are collected and STAI is performed. The only exception is that there is no STAI implementation in Part C. Outputs of Part C are used in test stage.

This system composes of two main databases which are used for training and testing purposes. A spreadsheet file is used for this purpose. Before the training stage, each student's handwriting features have to be marked depending on a given algorithm in Figure 2. This algorithm was also given and applied in our previous study [14].

J48 decision tree algorithm is used in the WEKA inference engine [21]. Inference engine extracts the classification rules which are totally depending on the facts and handwriting features. Rule are used to build the training model. The test data are provided by Part C. in Figure 1. According to the generated training model, decision making is realized.

4. Implementation

This expert system considers three handwriting features, which are line slant, character pixel height and width. These features are obtained from handwriting samples. As above mentioned, these samples are in pen and paper style. For this reason, some well-known image processing techniques are used. The followed steps, traced in order to get the above mentioned handwriting features, are given below.

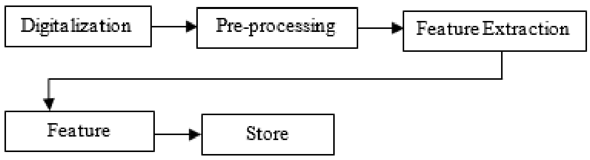

Figure 3. Feature extraction steps

In Figure 3., feature extraction process covers pre-processed images and extracted segmented characters that indicate features, then they are properly stored. Handwriting features are obtained from pre-processed handwriting images which are type of 24-bit BMP images. To acquire character pixel height and width, handwriting, character segmentation is carried out. Each properly segmented character is considered for height and width calculation.

Obtaining handwriting features form off-line documents has some difficulties [22]. The main difficulty is removing the noises and distortions which reside in document images. There could be many various reasons of this trouble. Digitization device, pencil, other impacts that are out of control may result towards this problem.

In order to predict students' anxiety levels within their handwriting features, J48 decision tree algorithm in Weka [21] framework is chosen for this purpose. After the training model is built, it is possible to make a prediction about a student's anxiety level using handwriting features.

Both collected handwriting features and STAI scores are stored into training database. Data input is done via several graphical user interfaces. As 210 handwriting instances in the training database are collected during exam time, which the students are thought to reflect through exam anxiety.

5. Experimental Results

The main input data of developed expert system consist of some selected students' handwritings. They are selected from both department of Computer Engineering at Trakya University and Canakkale Onsekiz Mart University. Their ages are between 18 and 24. The females of this group are 37% of the total group, and the rest of the students are male. The left handed persons of male group are 6%. It is also the same for female group.

The values of handwriting features for training dataset which are also part of Part A and B in Figure 1 are given in Table 1. These are the inputs for integration and marking process, which were given in our previous study [14], in order to build a prediction model for inference engine of the expert system. In the current study, we have categorized handwriting feature numerical values also seen below, in order to make our decision tree branches equally distributed.

Table 1. Categorization table

Handwriting Feature	Categories	Begins	Ends
Line Slope (LS)	**N**egative Slope **1** *N1*)	-5	-8
	Negative Slope **2** (*N2*)	0	-4
	Positive Slope **1** (*P1*)	0	4
	Positive Slope **2** (*P2*)	5	8
Character Height (CH)	Low	10	16
	Medium High	15	22
	High	23	28
Character Width (CW)	Narrow	9	13
	Medium Large	14	17
	Large	18	21

Student is labelled "anxious" due to the algorithm previously mentioned in Figure 2. J48 decision tree algorithm, which is also used by inference engine. The attributes of each instance in training data set are built as seen in Figure 4. Categorization in Table 1. is built by expert in medicine area, which depends on the proposed algorithm in Figure 2. The main reason of this qualitative categorization is that three visual handwriting features are located in document image.

Line Slant : *String {N1,N2,P1,P2}*

Character Height : *String {Low, Medium High, High}*

Character Width : *String { Narrow, Medium Large, Large}*

Anxiety Label: *Boolean*

Figure 4. Attributes of an instance in training dataset

In Figure 4., anxiety attributes are also the class attributes that have two possible values. There are some predefined symbolic expressions: True and False. True value represents "Anxious" class, whereas false value represents "Non-Anxious". The input dimensions are line slant, character height and width, whose discrete representations are given in Table 1. Since these dimensions are discrete, the decision mechanism checks the value in each node and takes appropriate branches. On the other hand, the computed handwriting features have numerical values. So, these numerical values are categorized based on the selected threshold values in Table 1.

The related node about line slant divides the input space into four subcategories, whereas both character height and width nodes divides into three subcategories. For operating the classification rules, a J48 decision tree is constructed by training data set as shown in Figure 5.

This decision tree is also known as classification tree. In experiments, we chose 90% of 210 instances used to train the decision tree and the remainder part is used to test this tree. In this way, maximum number of instances are considered for the training. This built model may be used for new students' anxiety states using handwriting features after its performance is evaluated, which is presented in the next section.

5.1. Analysis

In this section, we evaluate the classification results. Balanced accuracy is used in order to assess the results. The harmonic mean of Precision and Recall is calculated as F-measure [23]. The ground truth annotation is based on the marking made considering the proposed algorithm in Figure 2. In addition, classification results are compared to the above mentioned ground truth annotations. Therefore, we interpreted the ROC curves and Precision-Recall graphics.

Classification process needs less effort for human being, but in some circumstances such as classifying anxiety level, it would be troublesome. Binary classifiers are widely used in automatically learnt

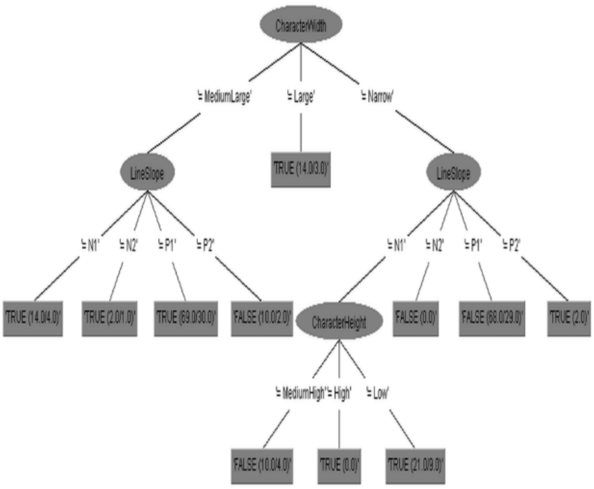

Figure 5. Anxiety decision tree

classification systems in Bayesian networks, support vector machines and decision trees also [24,25]. Classification results have two groups which are not symmetric. Classification accuracy covers different types of errors.

The binary classification results are measured by four basic metrics, which are true positive (TP), true negatives (TN), false positives (FP) and false negatives (FN). The J.48 decision tree classifies the students into “Anxious”, which are represented by “TRUE” label, and “Non-anxious”, which is represented by “FALSE” label. These result are represented in Table 2. While the table columns correspond to classification values which are the predicted classes, the rows correspond to the actual values. To evaluate the results in a more precise way, 210 instances are dived into 10 groups, and then the training set is built gradually. Classification results are given below.

Table 2. Confusion matrix

		Predicted Class	
		Anxious	Not Anxious
Actual Class	Anxious	9 (*tp*)	5 (*fn*)
	Not Anxious	2 (*fp*)	5 (*tn*)

By using values from Table 2., true positive rate and false positive rate can be calculated by using values as shown below [26].

$$TP\ Rate = TP/(TP + FN) \tag{1}$$

$$FP\ Rate = FP/(FP + TN) \tag{2}$$

In the scope of information retrieval, there are two measures about classifications defined [21]. These are recall and precision, which are formulated as below.

$$Precision = TP/(TP + FP) \tag{3}$$

$$Recall = TP/(TP + FN) \tag{4}$$

The harmonic mean of Precision and Recall values are F-Measure. This value is calculated as follows.

$$F\text{-}Measure = 2TP/(2TP + FP + FN) \tag{5}$$

In Table 3., true positive rates, false positive rates, Precision, Recall, and F-Measure values belonging to the classifier in our study are given below.

Table 3. Classifier’s performance evaluation

	TP	FP	Precision	Recall	F-Measure
Anxious	0.64	0.28	0.81	0.64	0.72
Non Anxious	0.71	0.35	0.5	0.71	0.58
Average	0.66	0.31	0.71	0.66	0.67

Average results of performance are computed as 66% for TP rate, 31% for FP rate, 71% for Precision, 66% for Recall, and 67% for F-Measure, respectively.

F-Measure reflects the quality of results. 0 indicates worst score and 1 indicates the best score. F-measure does not consider true negative values [27]. Kappa coefficient is a statistical measure for qualitative items that is calculated as follows [28].

$$\kappa = (\Pr(a) - \Pr(e))/(1 - \Pr(e)) \tag{6}$$

In Eq. 6, Pr(a) represents relative agreement among raters and Pr(e) is the hypothetical probability of chance agreement [29]. Performance measures reflect the quality of query results by considering the kappa statics range [0, 1]. In this range, values below the zero show there is no match which is also the worst case, the one value show the complete match which is the best case for the performance. In Table 4., statistical assessment of the performance of the J48 decision tree is presented.

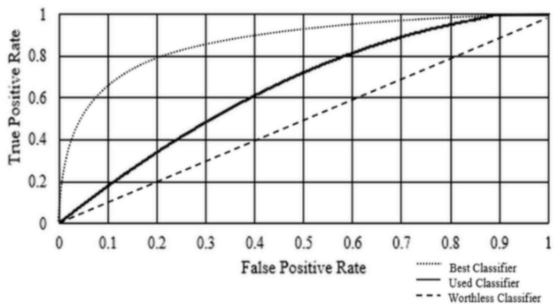

Figure 6. ROC curve of the classifier.

Table 4. Statistical results

Correctly classified instances	66.66 %
Incorrectly classified instances	33.33 %
Kappa statistics	0.3226

Receiver Operating Characteristics (ROC) is graphical technique used for evaluating the binary classifier's performance. This representation shown in Figure 6 illustrates the performance of the classifier without regard to class distribution or error costs. The horizontal and vertical axis depend on the values given in Table 3.

Accuracy is measured by the area under the ROC curve. If the ROC area is under 0.5 value, it indicates a worthless classifier. The best result is 1, which indicates a perfect classifier. In our study, the area under the ROC curve is 0.663 (for "Anxious" values), which is represented black line above. In Table 5, classification results of ten above mentioned groups of instances are shown.

Table 5. Classifications results of groups

Groups	Instance Count	Correct Classifications (%)	Incorrect Classifications (%)
1	21	0	100
2	42	25	75
3	63	66.66	33.33
4	84	62.5	37.5
5	105	60	40
6	126	69.23	30.76
7	147	53.33	46.66
8	168	47.05	52.94
9	189	31.57	68.42
10	210	66.66	33.33

Weighted F-Measure graphic which is basically depending on this ten groups illustrated in Figure 7.

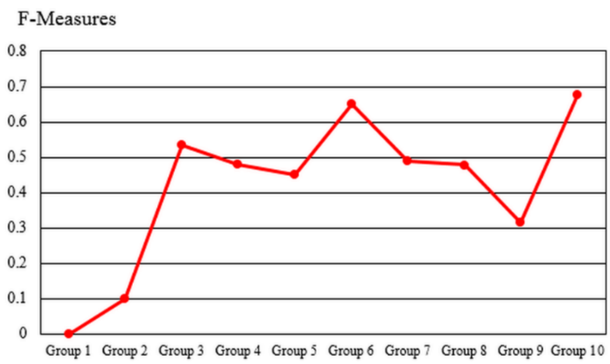

Figure 7. Weighted F-Measures of groups

When Table 5. and Figure 7. are both considered, classifier accuracy is getting better as the groups are getting more instances. However, correct classification results are so low in group 8 and 9. This situation also reflects itself in Figure 7.

6. Conclusion

The main contribution of our study is that we put a new perspective for emotional state changes on a critical event. We gave some proofs about the relationship between handwriting and person's emotional state in a computerized way. We developed an expert system to make anxiety prediction based on handwriting features. According to the performance measurements, it is possible to make a prediction depending on handwriting features, which are character's height, width and line slant, respectively. The performance results of classifier are little weak due to off-line handwriting. Off-line handwriting has own limitations such as leading to errors. On the other hand, all handwriting features may not reflect the anxiety. Since each person has different anxiety threshold, handwriting features may not always represent anxiety level.

There are continuing debates about the relationship between handwriting and personality. While many argue that it is possible to find some significant relationship between these, others completely disagree. All of these debates are around the term "Graphology". Analyzing process mostly depends on knowledge and experience of the graphologist, which is rather relative. Hence, the concrete facts stay hidden in most of times.

The proposed expert system capable of making predictions about emotional states such as anxiety can be used in industry. When a critical event occurs such as job interview or signing, signatures on checks or contracts may give us enough information whether person is under pressure or not.

In conclusion, it is possible to make an estimation about emotional states using handwriting features. Since collecting more and precise handwriting process has some difficulties, the online systems will be more helpful for researches. We plan to investigate some new aspects in other emotional state changes using handwriting features.

References

[1]. Kamath, V. ,N. Ramaswamy, N. Karanth, V. Desai, and S. Kulkarni. (2011). *Development of An Automated Handwriting Analysis System*, ARPN Journal of Engineering and Applied Sciece 6(9): pp.135-140.

[2]. Champa, H.N., and K.R. Anandakumar. (2010a). *Artificial Neural Network for Human Behavior Prediction through Handwriting Analysis*, International Journal of Computer Applications, 2(2), pp 36-41.

[3]. Roberts, P. (2002). Love Letters:*The Romantic Secrets Hidden in Our Handwriting*, The Carerr Press.

[4]. Coll, R., A. Fornes, and J. Llados. (2009). *Graphological Analysis of Handwritten Text Documents for Human Resources Recruitment*, Paper presented at the 10th International Conference on Document Analysis and Recognation, Barcelona, Spain, July, pp 1081-1085.

[5]. Klimoski, R., and A. Rafaeli. (1983). *Inferring Personal Qualities Through Handwriting Analysis*, Journal of Occupational Psychology, 56(3), pp 191-201.

[6]. Kucera, D. and J. Havigerovaç (2011). *Handwriting in Diagnostic Perspective*, E-psychologie, 5(2/3), pp 1-8.

[7]. Champa, H.N. and K.R. Anandakumar. (2010b). *Automated Human Behaviour Prediction through Handwriting Analysis*, Presented Papet at International Conference on Integrated Intelligent Computing.

[8]. Prasad, S., V. Singh, and A. Sapre. (2010). *Handwriting Analysis Based on Segmentation Method for Prediction of Human Personality using Support Vector Machine*, International Journal of Computer Applications, 8(12), pp 25-29.

[9]. Li, T. and P. Tang. (2012). *Detecting Honest People's Lie in Handwriting*, Journal of Bussiness Ethics, 106(4), pp 389-400.

[10]. Glenat, S., L. Heutte, T. Paquet, R. Guest, M. Fairhurst, T. Linnell. (2008). *The Development of A Computer Assissted Tool For The Assessment of Neuropsychological Drawing Tasks*, International Journal of Information Technology & Decision Making, 7(4), pp 751-767.

[11]. Pedrotti, M., A. Mirzaei, Tadesco M., Chardonnet A., Mérienne R.J, and Benedetto F.(2014). *Automatic Stress Classification with Pupil Diameter Analysis*, International Journal of Human-Computer Interaction,30:3,220-236, DOI:10.1080/10447318.2013.848320.

[12]. Spielberger, C.D., R.L.Gorssuch, P.R. Lushene, P.R. Vagg, and G.A Jacobs.(1983). *Manual for the State-Trait Anxiety Inventory*, Consulting Psychologists Press, Inc.

[13]. Bengerter, A., C. König, S. Blatti, and A. Salvisberg. (2009). How Widespread is Graphology in Personnel Selection Practice?. *A case study of a job market myth*, International Journal of Selection and Assessment, 17(2), pp 219-230.

[14]. Uğurlu, B., R. Kandemir, A. Carus, and E.Abay. (2014). *A Novel Model For Mood Changes On a Critical Event Using Emotional State and Trait Analysis and Handwriting Features*, 44th International Conference on Computers & Industrial Engineering, 14-16 October 2014, Istanbul, Turkey.

[15]. Macaulay, M. (2004). *The Speed of Mouse-Click As a Measure of Anxiety During Human-Computer Interaction*, Behaviour & Information Technology,23:6,427-433, DOI:10.1080/01449290412331294651

[16]. Desment, B. and V. Hoste. (2013). *Emotion Detection in Suicide Notes*, Expert Systems with Applications, 40 (16), pp. 6351-6358.

[17]. Djamal, C.E., R. Darmawati, and N.S. Ramdlan. (2013). *Application Image Processing to Predict Personality Based on Sturucture of Handwriting and Signature*, Paper Presented at the International Conference on Computer, Control, Informatics and Applications, pp. 163-168.

[18]. Suhasini, A., S. Palanivel, and V. Ramalingam. (2011). *Multimodel Decision Support System for Psychiatry Problem*, Expert Systems with Applications, 38(6), pp. 4990-4997.

[19]. Lu, Y.T., M. Lin, M.C. Chen, and H.J. Wu. (2013). *Forecasting Reading Anxiety to Promote Reading Performance Based on Annotation Behavior*, Paper Presented at IEEE 37th Annual Computer Software and Applications Conference Workshops, pp. 427- 432.

[20]. Sriram, S. and X. Yuan. (2012). *An Enhanced Approach for Classifying Emotions Using Customized Decision Tree Algorithm*, Southeastcon 2012 Proceedings, DOI: 10.1109/SECon.2012.6196948

[21]. Hall, M., E. Frank, G. Holmes, B. Pfahringer, P. Reutemann, I. H. Witten. (2009). *The Weka Data Mining Software: An Update*; SIGKDD Explorations, 11:1.

[22]. Arica, N. and F. Yarman. (2001). *An overview of Character Recognition Focused on Off-line Handwriting*, IEEE Transactions on Systems, Man and Cybernetics, Part C, 33(2), pp. 216-233.

[23]. Manning, C.D., P. Raghavan, and H. Schütze. (2009). *Introduction to Information Retrieval*. Cambridge University Press, New York, NY, 2009, Draft Online Copy (2009.04.01). Retrieved March 13, 2014 from http://nlp.stanford.edu/IR-book/pdf/irbookprint.pdf

[24]. Alpaydin, E. (2009). *Introduction to Machine Learning (Second Edition)*. MIT Press. ISBN: 978-0-262-01243-0.

[25]. Han, J., M. Kamber, and J. Pei. (2011). *Data Mining Concepts and Techniques (Third Edition)*. Morgan Kaufmann, ISBN: 978-0-12-381479-1.

[26]. Witten, H.I. and E. Frank. (2005). *Data Mining: Practical Machine Learning Tools and Techniques (Second Edition)*. Morgan Kaufmann, ISBN: 0-12-088407-0

[27]. Powers, D. M.W.(2011). *Evaluation: From Precision, Recall and F-Measure to ROC, Informedness*, Markedness & Correlation", Journal of Machine Learning Technologies, 2:1, 37-63.

[28]. Banerjee, M., Michelle C., Laura M.; Debajyoti S. (1999). *Beyond Kappa: A Review of Interrater Agreement Measures*. The Canadian Journal of Statistics / La Revue Canadienne de Statistique 27 (1): 3–23. JSTOR 3315487.

[29]. Smeeton, N.C.(1985). *Early History of the Kappa Statics*, Biometrics 41:795.