Handwriting Analysis for Detection of Personality Traits using Machine Learning Approach

Prachi Joshi
Associate Professor
Department of Computer Engineering
MIT College of Engineering
Pune, India

Aayush Agarwal
Department of Computer Engineering
MIT College of Engineering
Pune, India

Ajinkya Dhavale
Department of Computer Engineering
MIT College of Engineering
Pune, India

Rajani Suryavanshi
Department of Computer Engineering
MIT College of Engineering
Pune, India

Shreya Kodolikar
Department of Computer Engineering
MIT College of Engineering
Pune, India

ABSTRACT
Among all the unique characteristics of a human being, handwriting carries the richest information to gain the insights into the physical, mental and emotional state of the writer. Graphology is the art of studying and analysing handwriting, a scientific method used to determine a person's personality by evaluating various features from the handwriting. The prime features of handwriting such as the page margins, the slant of the alphabets, the baseline etc. can tell a lot about the individual. To make this method more efficient and reliable, introduction of machines to perform the feature extraction and mapping to various personality traits can be done. This compliments the graphologists, and also increases the speed of analysing handwritten samples. Various approaches can be used for this type of computer aided graphology. In this paper, a novel approach of machine learning technique to implement the automated handwriting analysis tool is discussed.

Keywords
Graphology, Personality Traits, Handwriting, Feature Extraction, Margins, Baselines, T-bar.

1. INTRODUCTION
Handwriting also termed as brain-writing is a useful measure in identifying the characteristic personality traits of an individual. Handwriting analysts also known as graphologists can examine an individual's handwriting to predict the personality traits of the writer. Automated handwriting analysis can be used to examine personal traits of candidates during interviews accurately as the accuracy of an analyst highly depends on his skill set. Also hiring a graphologist to analyze hundreds of samples for recruitment purpose will be time consuming and not be feasible economically.

This work discusses about a method for analysing real world handwritten text samples with the aid of technology. The analysis is done for specific features of the sample for determining various characteristic behavioural traits of the person. Various parameters of the handwritten sample like Margin, Baseline, T-bar and Slant will be taken into consideration to determine corresponding traits [1]. The proposed tool will compliment the graphologists to increase their speed and efficiency in the analysis process. Machine learning approach like KNN with incremental learning, will be implemented to improve the efficiency of the tool.

2. LITERATURE SURVEY
The study of handwriting is quite an old concept tracing back to the seventeenth century. The first book to document these methods was written almost 400 years ago by Camilo Baldi. Known as the father of graphology, Camilo Baldi, who was an Italian doctor of medicine and philosophy, performed systematic observations on handwriting samples in the year 1622. Since then, very detailed and extensive studies have been performed in this field. There are more than 2200 documented studies of handwriting analysis till date. [2]

A study conducted by the American Psychological Association's annual convention acknowledges that the use of computer technology in the field of handwriting analysis can be considered as a reliable tool for determining various traits like honesty, emotional stability, substance abuse risk and judgment.

A paper published in the year 1995 at the SUNY, Buffalo by Prof S. N. Shrihari and two others from the Center of Excellence for document analysis and recognition gave a lot of insights which led to the progress in the field of Computer Aided Graphology [3]. This paper focuses only at the prime features of a page of handwritten sample that are page margins, line spacing, line direction, slant and zone ratios. To do this, the methodology used is scanning, preprocessing, feature extraction, analysis and finally, trait determination. This system was mainly designed to prove the validity of the graphology rules that were applied in the implementation of the system. This paper restricts its scope to macro analysis of the handwriting sample. There are no micro features like alphabet, loops etc. taken into consideration.

Another paper discusses about computer aided graphology which intends to reduce the human intervention needed to perform trained analysis on the handwriting input sample. The features like margin, baseline, size and zones have been extracted through image processing and an approximate analysis of the personality trait has been given. These traits are then mapped to existing theories to determine a final personality type and generate a report for the same. This work

includes limited physical features of handwriting. Additional perspective can be provided by using approaches like Natural Language Processing which can improve the efficiency of the system [4].

A paper based on artificial neural network explores the implementation of a machine learning approach in the field of handwriting analysis. This paper proposed a method to predict the personality traits of a person by analyzing the baseline, pen pressure and the letter 't' as found in the individual's handwriting sample. These extracted features are then given as an input to the artificial neural network which in turn gives output as personality trait to the user. The future work discussed in this paper are, including more features of the handwriting like the size of the letters and the margins as inputs for personality trait determination to improve the system output[5].

The various algorithms and techniques used for the analysis have been discussed below:

2.1 Polygonalization

Polygonalization is a method of subdividing the plane into polygons. This is the main technique used to find the slant of the baseline. In this technique of Polygonalization, a closed polygon is drawn around a single line of the scanned handwriting sample. The slope of the polygon in Fig. 1 is found using the coordinates of the polygon. This slope corresponds to the slope of the baseline. [5]
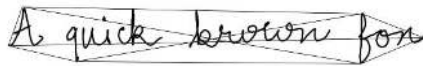

Fig. 1: Polygonalization

2.2 Thresholding Algorithm

Image thresholding is an uncomplicated and effective way of partitioning. This image analysis algorithm is a type of image segmentation that segments objects by transforming grayscale images into binary images. Image thresholding is most effective in images with high levels of contrast such like handwritten sample on white sheet using a black pen [6].

2.3 Template Matching

Template matching is a technique in image processing used for finding small parts of an image which match a given image that is template image. From the scanned handwritten sample, individual lines are isolated and individual characters are identified. Finally, the identified characters are compared with all the template images by correlation and a match is found. [7]
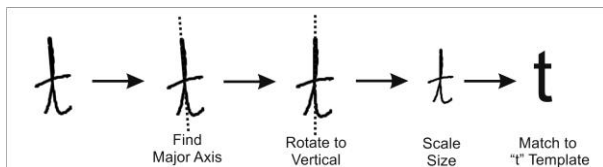

Fig. 2:Template Matching

2.4 Artificial Neural Network

When large sample of data is analysed, different combinations of traits are derived. These combinations of traits are analyzed to predict the personality trait using artificial neural network. Back propagation is the abstraction of the Widrow-Hoff learning rule to multiple-layer networks and nonlinear differentiable transfer functions. Input vectors and the

analogous target vectors are used to train a network until it can approximate function is determined which can associate input vectors to output vectors, or classify input vectors in an appropriate way defined by default [5]. Fig 3 [8] shows a neural network active node. This is a flow diagram of the active nodes used in the hidden network. Each input is multiplied by weight (the w_N values), and then summed. This produces a single value that is passed through an "s" shaped nonlinear function called sigmoid.
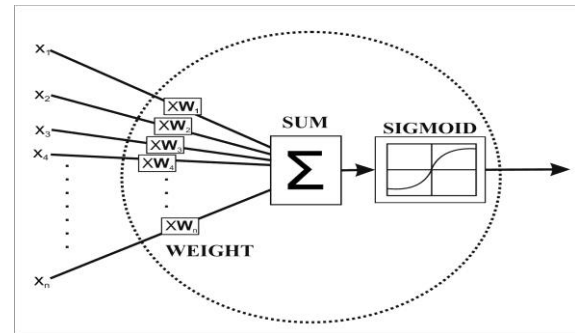

Fig. 3: Artificial Neural Network

3. PARAMETERS FOR ANALYSING HANDWRITING

Graphologists are the handwriting analysts who identify the characteristics traits of a person merely by examining the handwritten samples of the individual. As this is a manual process, the skilfulness of the examiner defines the accuracy. However this is a time consuming and costly affair solely because of the human intervention. Thus, in the proposed methodology the focus would be on the development of a computer aided tool with minimum human intervention that would be able to predict the characteristic traits of a person intelligently. Baseline, writing pressure, spacing between letters, words and lines, size of letters, strokes connecting the letters, width of margins, starting strokes, ending-strokes, slant of word etc. are the most common parameters that help in identifying the personality traits of an individual through handwriting analysis. This paper would be focusing on the following four parameters: Baseline, Letter-slant, Height of the T-bar and Width of Margins.

3.1 Baseline

The baseline is the feature which reveals a great deal of information as far as the personality of the writer is concerned. Baseline in a person's handwriting is the imaginary line along which the writer aligns the bottoms of the middle zone letters, when asked to write on a blank paper. Slanting downwards, rising upwards and level are the three most common baselines found in handwriting [5]. Different personality traits are associated with each of the above types of baselines. These characteristic traits are mentioned in the table below.

Table 1: Baseline Characteristics

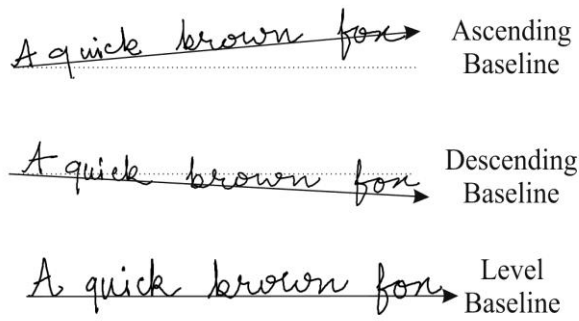| Baselines	Corresponding traits
Ascending(rising upwards)	Optimistic
Descending(slanting downwards)	Pessimistic
Level(Straight)	Balanced

Fig. 4: Types of Baseline

The technique that would be used to determine baseline, in this paper is polygonolization. In this method the smallest possible polygon around the line to be examined is drawn. This polygon should cover every point on the line. The slope of various sides of this polygon will help us identify the baseline.

3.2 Letter Slant

The slant of the letter is used to understand whether an individual's handwriting is inclined towards the right or towards the left or is it vertical. It has been observed that, around 77 percent of individuals write with a right slant, 15 percent with left slant and remaining 8 percent write vertically [9]. The slant of an handwriting is in relation to an individual's emotional direction and degree of sentimental control. The writer's connection between the inner and outer world is indicated by means of the slant of the letters.

Table 2: Slant Characteristics

Slant	Corresponding trait
Extreme left	Fear of the future, defensive, early rejection.
Extreme right	Lack of self-control, impulsive, unrestrained, intense, very expressive, low frustration tolerance.
Left	Reflective, independent, not sympathetic, difficulty in adapting and expressing emotions.
Right	Expressive, confidence in convictions, freedom of thought, extrovert, future orientation.
Vertical	Head controls over heart, independent emotional nature, and works independently.

To determine the slant, the following method is proposed. In the grid of every individual letter that would be obtained after template matching, a line will be drawn between the lowermost and the uppermost points of that letter [10]. Slope of this line will calculate and similarly it would be calculated for all other letters. The average of all slopes calculated will give the slant of letters [7].

3.3 Height of 't' bar

Another important feature that reveals lot of accurate information about the writer is the lower-case letter 't'. There are different ways in which a lowercase 't' can be written. This paper would be specifically focusing on the height of the t-bar. The t-bar can cut the stem of the letter 't' at different positions like the lower-portion, mid-portion, upper-portion of the stem or not cut the stem at all. Different characteristics traits of a writer can be identified depending on where the writer crosses the t-bar [11]. These personality traits are mentioned in the table below.

Table 3: 't' Bar Characteristics

| Position of t-bar	Corresponding trait
Crossed very high(but not above the stem)	**High self-esteem**: This reveals confidence, ambition, and the ability to plan ahead, high goals, high personal expectations and an overall good self-image.
Crossed just above the middle zone	**Moderate self-esteem**: It indicates a practical and successful person.
Crossed very low on the stem	**Low self-esteem**: This person fears failure and resists change. He is rarely successful enough in his own eyes despite his accomplishments and resists change.
Crossed above the stem	**Dreamer**: This person's goals and dreams lose touch with reality.
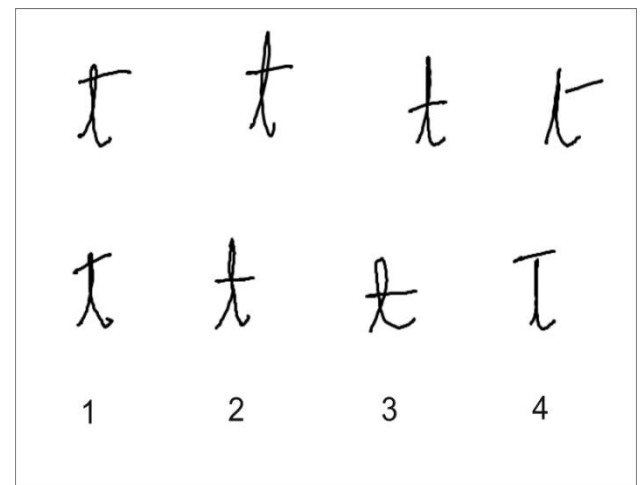
Fig. 5: Types of 't' Bar

Through template matching all the lowercase 't' letters will be singled out and compared with the predefined templates [10]. 4 templates of letter 't' will be predefined each consisting of t-bar at a height very low on the stem, at mid level on the stem, at very high height on the stem and out of the stem respectively. The given template will be matched with the predefined ones using Hamming distance [12]. Hamming distance will measure the minimum number of substitutions required to change one into the other. The predefined template which is achieved with minimum number of substitutions on the input template will be the matching template.

3.4 Margin

The margin is generally considered to define the layout of the page. Any individual assumes a particular margin while writing on a blank paper. The blank spaces on the left, right, top and bottom on a page comprises the margin. In this paper, different types of left and right margins namely- wide right, wide left, no margin at all and evenly spaced margin will be considered. Following are the characteristic personality traits of individuals associated with the above mentioned types of margin [1].

Table 4: Margin Characteristics

Margin Orientation	Corresponding trait
Wide left	Courageous.
Wide right	Avoids future and a reserved person.
No margin	Insecure and devotes oneself completely.
Even margin	Self disciplined and balanced.

To obtain the margin, virtual vertical lines will be drawn from the top left on the scanned image. The line that would encounter a good number of pixels with a low numerical value on the gray scale will be considered as the left margin. Similarly to obtain the right margin vertical lines will be drawn from the top right

4. PROPOSED APPROACH

This work aims at obtaining personal characteristic traits of individuals especially in the age group of 20-35 years when they face many interviews.

4.1 Generation of Training Dataset

Sample Training set will be generated using 100 samples of handwriting, which will be examined by a professional

Graphologist. Using Image Processing tools the feature vectors matrix of the same handwriting samples will be generated. The Feature Vector Matrix is the mathematical representation of the handwritten text in form of feature vectors. Based on the characteristics corresponding to particular trait, as identified by the graphologist, classes will be created. Initially, each FV Matrix, created for a particular handwriting sample, will be manually mapped into its corresponding class as suggested by the graphologist [13]. In this way a dataset will be generated that contains individual classes corresponding to each trait. Later a supervised machine learning algorithm would be used for mapping new samples into previously identified classes.
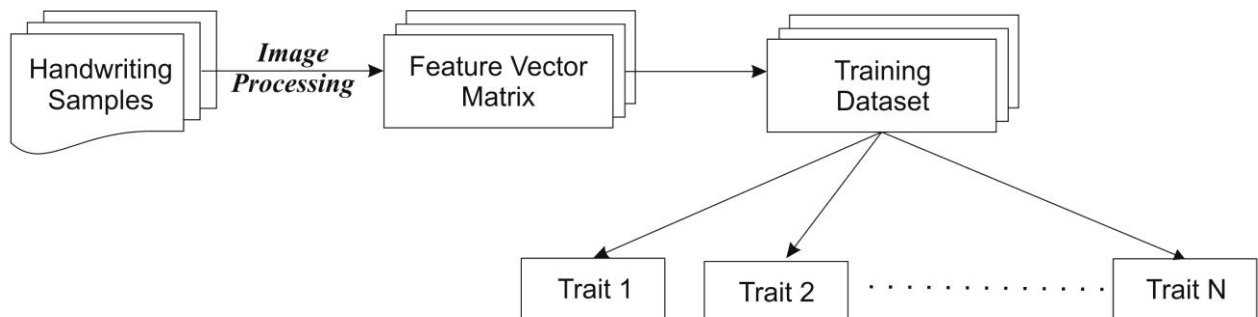

Fig. 6: Process Flow Diagram for Generation of Training Dataset

4.2 Trait Identification Based on Trained Data

Once the dataset is trained, it will be easy to identify the traits corresponding to a new sample of handwriting. The Feature Vector Matrix for a new handwriting sample will be created and then its similarity with the trained dataset would be calculated using the similarity matrix method. K-NN classifier will be used to identify the class which is most appropriate for the handwriting sample, based on the similarity matrix. In this
.

way the handwriting would be identified in a class that will map the handwriting to its corresponding trait [14]. After the results are produced, it would be stored in the trained dataset. By storing the new sample into the class to which it is mapped will be able to implement the incremental machine learning, increasing the accuracy of the future results that will be produced.
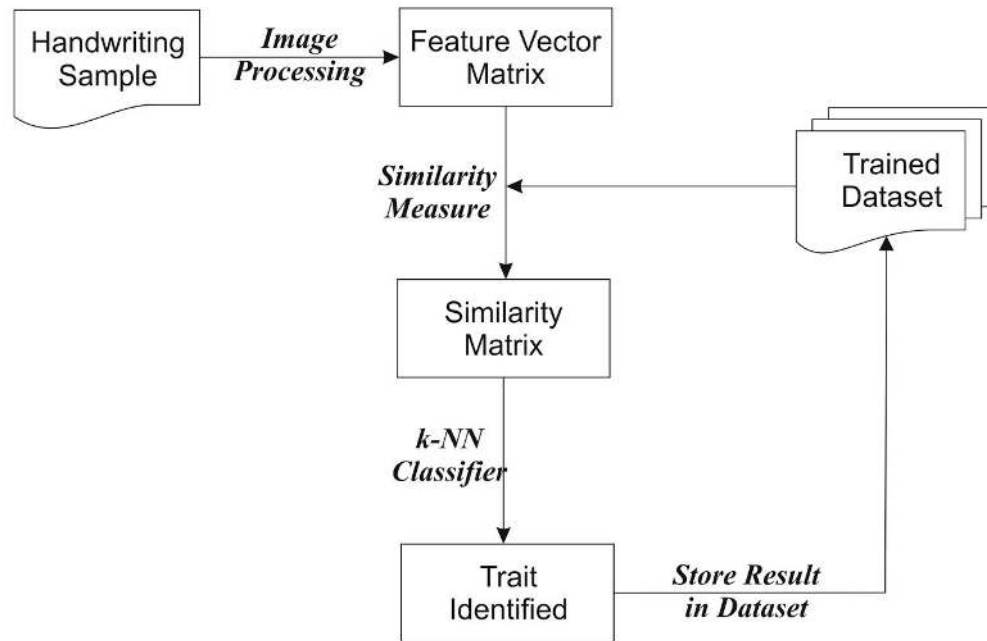
Fig. 7: Process Flow Diagram for Classification of Traits

4.3 Mathematical Model

1) Building the training set

For TD_1 to $TD_n \in C_1$ to C_m

where,

n is no. of training documents

m is no of classes

(i) Extract f_1 to f_y features [Selection based on discussed parameters]

(ii) Build FV_i for every TD_i

where,

$$FV_i = \{f_1 \ldots f_y\}$$

2) Classification of unknown sample x

For TD_1 to TD_n

(i) Calculate similarity(TD_i,x)

(ii) Determine value of k

(iii) For C_1 to C_p obtained from k;

Classify $x \rightarrow C_f | C_f$ is majority class

(iv) Determine $C_f \rightarrow$ trait identified for x

5. CONCLUSION

This paper has proposed a methodology to predict the accurate personality traits of an individual from the features extracted from handwriting using a machine learning approach. This paper explores the personality traits revealed by baseline, margin, slant of the words and height of t-bar of a person's handwriting. These features will be extracted from the handwriting samples into feature vectors which would be compared with an initially trained data set; and then mapped to the class with corresponding personality trait. The baseline would be evaluated using the method of Polygonalization while margin will be calculated using the method of vertical scanning. The height of the t-bar on the stem of the alphabet 't' and word-slant would be calculated using template matching.

The proposed system can be used as a complementary tool by the graphologist to improve the accuracy of handwriting analysis and also make the process fast. It will also assist the HR/company employer in decision making regarding the suitability of an employee for the specific job and improving the retention of an employee. The future work can be to include more features from the micro approach of handwriting analysis like the loops of alphabet 'f' and 'l', gradient, concavity of letters and so on in order to predict more accurate results.

6. REFERENCES

[1] Champa, H.N. and Anandakumar, K.R,"Automated human behavior prediction through handwriting analysis", Integrated Intelligent Computing (ICIIC), 2010 First International Conference

[2] "Handwriting Research Corporation" http://www.handwriting.com/facts/history.html

[3] G. Sheikholeslami, S. N. Srihari, V. Govindaraju, "COMPUTER AIDED GRAPHOLOGY", Center of Excel lence for Document Analysis and Recognition.Ding, W. and Marchionini, G. 1997 A Study on Video Browsing Strategies. Technical Report. University of Maryland at College Park.

[4] Rashi Kacker, Hima Bindu Maringanti, "Personality analysis through handwriting" , GSTF Journal on Computing (JoC) Vol.2 No.1, April 2012.

[5] Champa H N and Dr. 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] H. Devi,"Thresholding: A Pixel-Level Image Processing Methodology Preprocessing Technique for an OCR System for the Brahmi Script", Ancient Asia Journal Vol. 1, Dec 2006.

[7] Mr.Danish Nadeem & Miss.Saleha Rizvi, "CHARACTER RECOGNITION USING TEMPLATE MATCHING", https://www.cs.uic.edu/~srizvi/BIT_Thesis.pdf

[8] "Artificial Neural Network " ,Available at: http://cogsci.stackexchange.com/questions/8509/dynamic al-systems-theory-as-a-metaphor-in-psychology-is-it-useful-or-not Spector, A. Z. 1989. Achieving application requirements. In Distributed Systems, S. Mullender

[9] Ruth Gardner,"The Truth About Graphology",Page 19.

[10] Anlong Ming, et, al, "A Grid-based Face Recognition Approach Using General Template Matching", Proceedings of the First International Conference on Semantics,Knowledge, and Grid ,IEEE 2006.

[11] Sofianita Mutalib, Shuzlina Abdul Rahman, Marina Yusoff1, Azlinah Mohamed"Personality Analysis Based On Letter 't' Using Back Propagation Neural Network", Proceedings of the International Conference on Electrical Engineering and Informatics Institut Teknologi Bandung, Indonesia June 17-19, 2007.

[12] Hui Tian and Hong Shen, "Hamming Distance and Hop Count Based Classification for Multicast Network Topology Inference", Proceedings of the 19th International Conference on Advanced Information Networking and Applications (AINA'05).

[13] Vladimir Pervouchine and Graham Leedham,"Extraction and analysis of document examiner features from vector skeletons of Grapheme 'th'",Springer-Verlag Berline Heidelberg 2006 , pp. 196-207.

[14] Cleber Zanchettin, Byron Leite Dantas Bezerra and Washington W. Azevedo,"A KNN-SVM hybrid model for cursive handwriting recognition",Neural Networks (IJCNN), The 2012 International Joint Conference.

[15] Seema Kedar and Ms Vaishnavi Nair, Ms Shweta Kulkarni, "Personality Identification through handwriting analysis: A review ",Volume 5, Issue 1,January 2015 ISSN: 2277 128X,International Journal of Advanced Research in Computer Science and Software Engineering.

[16] Vasantha Kalyani David, S. Rajasekaran, "Applications of MicroARTMAP",Pattern Recognition Using Neural and Functional Networks.

[17] Srihari S.N., Sung-Hyuk Cha and Sangjik Lee,"Establishing handwriting Individuality using pattern recognition techniques", Proceedings of the Sixth International Conference on Document Analysis and Recognition, 2001, pp. 1195 – 1204.