

# Personality Prediction Using Handwritten Characters

Patil Abhijit <sup>[1]</sup>, Rupale Mayuresh <sup>[2]</sup>, Singh Archana <sup>[3]</sup>, Tonge Saheel <sup>[4]</sup>

Department of Computer Engineering  
MGM CET, Kamothe  
Navi Mumbai-India

## ABSTRACT

Graphology is a very old and respected science-the study of hand-writing and its analysis is used to determine the personality traits of a person. It works on the principle that while writing our hand is controlled by our subconscious mind. The graphic movements generated by every person's sub-conscious mind are very variable and unique. A writer does not write each letter by his or her hand consciously. Thus, writings can be governed by brain, so with the help of the person's writing we can predict his/her personality. In this model a raw image is taken as an input and a dataset containing large no of real-world handwritten samples will be kept up. There are different features of handwriting like Margin, Baseline, and Size, Spacing etc. these features are outlined in each of samples and extracted through image processing techniques. Hence, a rough investigation of the writer's personality trait will be done. Therefore, the proposed system centers around building up a framework that predicts the personality of the writer with the help of machine learning.

**Keywords**—Graphology, Human Personality, Psychological Analysis, Image Processing, Machine Learning, Support Vector Machine.

## I. INTRODUCTION

Graphology is a study of analysing, assessing and understanding personality traits of a person. This can be performed by identifying the physical characteristics and patterns of an individual. It helps to discover psychological nature of that individual. Handwriting finds the character of a person including feelings, fears, modesty, honesty clarity of thinking and many others. Well-versed handwriting inspectors are known as graphologists. They distinguish the writers on the basis of their handwriting. Accuracy of this study relies upon the skill and competency of the analyst. Human intercession in this analysis is expensive and prone to errors hence, we need computerized software which reduces the human intervention.

We consider seven handwriting features: (a) Letter size, (b) Slant of letters, (c) Baseline, (d) Pen pressure, (e) Line spacing, (f) Word Spacing and (g) Top margin of the image predict eight personality traits of a writer. After extracting all these features, the image which contains the handwriting, is trained by eight support vector machines which outputs personality trait of the writer

### A. Objective

The objective of this project is to develop software which takes an image as input and outputs his/her personality traits based on the handwriting features.

### B. Motivation

There are several methods which are used in identifying psychology of person.

1) **Psychological analysis:** Graphology is used clinically by psychotherapists and counsellors. However, they use it for assessment tools and not for confinement.

2) **Employment profile:** Many companies use handwriting analysis for recruitment. They used this report with other tools for background check and record working skills.

3) **Mindset of terrorism:** Various organizations use handwriting samples of terrorists to identify the tendency of destruction, bombings, killings etc. Their writing consists of inconsistent textures, blurred strokes etc which are useful for discovering their personality.

## II. REVIEW OF LITERATURE

The authors Anamika Sen UG Student, EXTC Department Dwarkadas J. Sanghvi College of Engineering, created a feature vector matrix using Image Processing which served as the training dataset. Using kNN classification, traits were later classified. [1]

In the existing system by P.S.Dandannavar, S.R.Mangalwede, P.M.Kulkarni used social media text for personality prediction. They used dataset from the Myers-Briggs Type Indicator (MBTI) and used five broad categories. The "Big Five" are broad categories of personality traits acronymed as OCEAN (Openness, Conscientiousness, Extraversion, Agreeableness, and Neuroticism). [2]

Another similar work in a research named A Local Thresholding Algorithm for Images of Handwritten Historical Documents by Renata F.P.Neves and Carlos

A.B.Mello. They have provided an approach to threshold historical images. They have proposed an algorithm that divides images into three parts and identifies main objects of the image. [12]

An existing system [14] have used features like number of breaks, baseline, margins, speed of writing, slant and pressure and spacing between the words and have used MATLAB for image processing.

Another similar work was done by N Mogharreban, S Rahimi, M Sabharwal; they have used both fuzzy and crisp methods to predict the personality traits. Parameters used by them were baseline and slant of letters. [8]

### **C. Comparative study**

In the study “Automated Human Behavior Prediction through Handwriting Analysis” by Champ H N and K R Ananda Kumar have worked on the five features of such as baseline, pen pressure, the letter ‘t’, lower loop of ‘y’ and the slant of letters. Extraction of letter t and slant of writing was done by template matching, pen pressure using grey threshold value, baseline using polygonization and lower loop of y using Generalized Hough Transform (GHT). Implementation was done using MATLAB. Behaviours predicted were Optimistic, Pessimistic, Dreamer, Self-Centered etc. [4]

Another study [5] by Rajani Suryavanshi, Shreya Kodoliar Prachi Joshi, Aayush Agarwal and Ajinkya Dhavale. They have worked in the 3 similar features mentioned above including an additional feature named width of margin. The extraction of margin width is mentioned in Classification was done using k-NN classifier. Detected traits include Courageous, Dreamer, High Self-esteem, Impulsive etc.

In,[3] the authors Shitala Prasad, Akshay Sapre and Vivek Kumar Singh, have worked on six major features such as Size of Letters, Slant of Words and Letters, Baseline, Pen Pressure, Spacing between Letters and Words. Segmentation was widely used for preprocessing techniques. Further, Support Vector Machine (SVM) and Polygonization was used to predict the psychological behaviour. Multifarious traits were predicted such as Tired, Introspective, Determined, Social Index (Sociable), Practical, Independent etc.

Another recent study [6] by the authors Abdul Rahiman M, Diana Varghese and Manoj Kumar G in their research named HABIT; have considered 5 features viz. Pen pressure, Slant of letters and baseline, Size of letters and Spacing between words. Entire implementation was done using simple Linear Regression. The platform used by them was Java (hibernate). They have extracted the features and resulted in traits such Closeness, Tired, Endure feelings, Modest etc. Based on the study, it can be concluded that there are some common features which are found in every handwriting.

Although, we have used some similar features described above; to maintain integrity, the implementation and personality traits predicted are different. We have considered seven features namely Baseline, Letter size, Line spacing, Word spacing, Pen pressure and Slant of Letters. Preprocessing methods include Image Cropping, Removal of noise, Grayscale and Binarization. Extraction methods are

briefly described in the later section. Therefore, after precisely extracting the features exactly eight traits are determined. The degree or the depth in each of the traits can be determined using range 0 – 10 as seen in the results (refer section VIII). Implementation of the system is done in PYTHON3 and libraries include Sci-Kit learn and Pandas. Hopefully, our system has achieved 90% accuracy.

### **III. PROPOSED WORK**

The proposed system is designed to develop an efficient platform to analyse the handwriting of a person. Graphologists break down the handwriting physically which is composed on the paper. The graphologist's understanding ability relies on the mental ability to identify various types of handwriting. But, it is costlier and delayed process. The proposed procedure focuses on building up a framework with the base guide of human traits by examining both miniaturized scale and full scale highlights of handwriting.

This system requires minimum human intervention and is less expensive. Also, this system includes a centralized website where the traits are easily identified within seconds after the handwritten image is uploaded.

### **IV. PRESENT INVESTIGATION**

This section will give the detailed explanation of image processing methods used which is then followed by explanation of seven selected handwriting features. Further, the personality traits will be classified by giving combinations of these seven features.

#### **D. Pre-processing**

The image uploaded contains unwanted noise etc. they also have a big resolution. Aim of is to make the image suitable for feature extraction.

1) **Image cropping and Image resolution:** Adobe photoshop is used to crop the left and right margins of the image. The size of the image is also reduced, and then the image is converted to PNG format.

2) **Removal of noise:** Gaussian Noise-This is an electronic noise which is caused by thermal vibration of atoms and nature of warm objects.

**Gamma Noise:** This is seen in laser based images.

3) **Grayscale and Binarization:** We created two-valued binary image and convert this image plane into foreground and background pixels can take either 0 or 255 (black or white), an inverted binary function can be created so that pixels are converted to either 0 or 255.

#### **E. Handwriting Features**

There are large numbers of features for handwriting of a person. Only seven significant features are extracted, which are popular.

1) **Baseline:** It is the imaginary line where the writer tends to write. The writer may write above, below or at the middle of the line. It brings out the emotional control of the writer.

2) **Letter Size:** It determines the size of the letter. It may be tall or short. Average height of word is considered. It tells whether the writer is introvert or extrovert.

3) **Line Spacing:** It is the spacing between two consecutive lines. It gives clues of thinking clarity and the interaction with the environment

4) **Word Spacing:** The distance that is present between two words. This feature counts the pixels between two words. It determines the emotional comfort of a person with other people.

5) **Top Margin:** It is simply the space left from the top of a page. If the top margin is narrow it shows lack of respect for others, if it is broad it shows modesty and formality.

6) **Pen Pressure:** It determines whether the user presses more on upstroke or downstroke (light or heavy). If it is heavy it determines strong willed and if it is light it determines sensitive, not strong willed.

7) **Slant of Letters:** It looks whether the writing is slant to the left or right. It can also be straight. The average slant is right. If left or right it determines goal oriented and sensitive. If it is straight the person might be independent and good decision maker.

**V. EXPERIMENTAL STUDY**

**F. STEP 1 : Baseline Extraction**

In order to extract Baseline inverted binary thresholding and dilation is done.

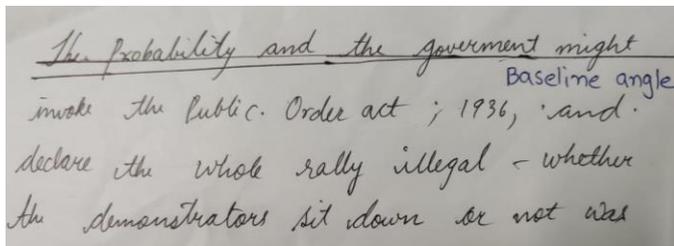


Fig. 1 Extraction of Baseline

**G. STEP 2 : Individual Line Extraction**

The resulting image is taken from STEP 1 and further horizontal projection is done.

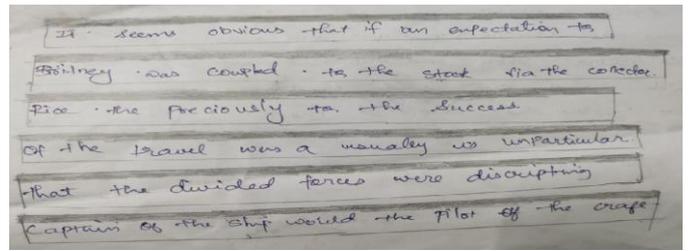


Fig. 2 Extraction of Individual Line

**H. STEP 3 : Letter Size Extraction**

After STEP 2 the letter size can be extracted by scanning the horizontal projection and counting the projection value greater than threshold.

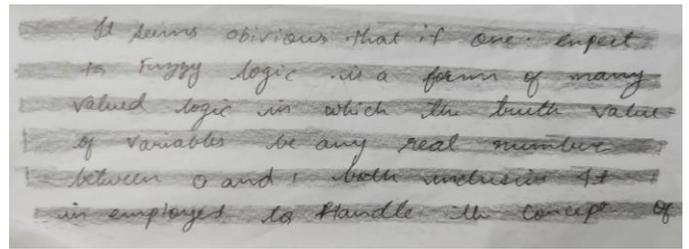


Fig. 3 Extraction of Letter Size

**I. STEP 4 : Line spacing Extraction**

It is then calculated by taking the sum of total number of rows with horizontal projection less than zero is and horizontal projection equal to zero. Further, it is then divided by the number of lines in the image.

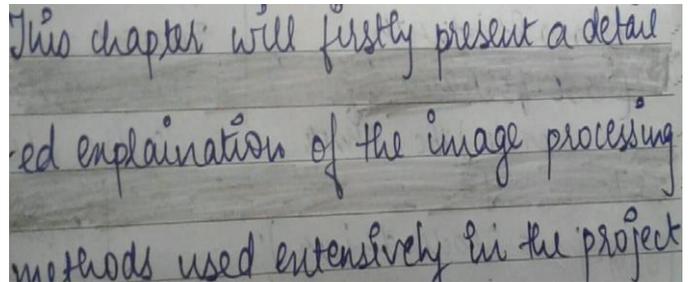


Fig. 4 Extraction of Line Spacing

**J. STEP 5 : Word Spacing Extraction**

Vertical projection is computed and the no of columns with pixel value 0 is counted. This step can be further computed similar to STEP 4.

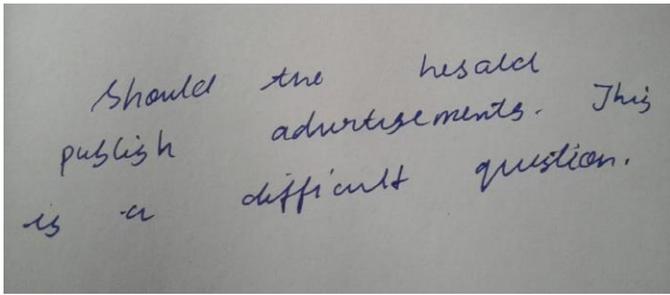


Fig. 5 Extraction of Word Spacing

**K. STEP 6 : Pen Pressure Extraction**

This step is expensive and time consuming step. In this step the image is inverted using formula:  $dst[x][y] = 255src[x][y]$ . If it is lower than threshold value, the new pixel value is set to zero.

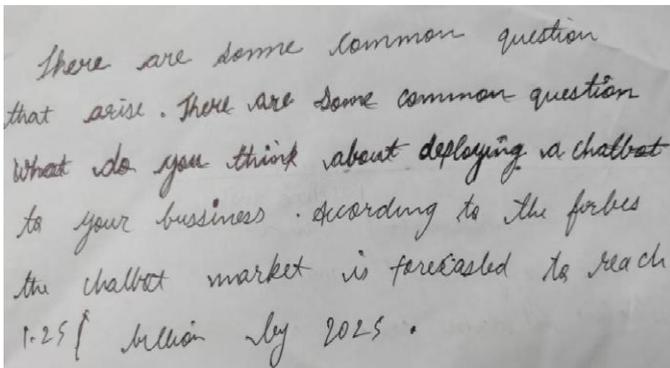


Fig. 6 Extraction of Pen Pressure

**L. STEP 7 : Slant Letter Extraction**

The slant of letters can be found by 9 different angles (-45, -30, -15, -5, 0, 5, 15, 30 and 45 degree), Then a shear transformation is applied and the histogram is calculated.

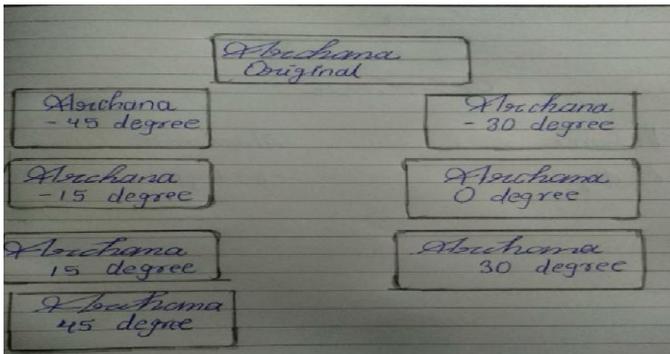


Fig. 7 Extraction of Slant Letters

**M. Personality Traits**

These seven handwriting features are combined and used to find eight personality traits given as follows:

Features	Traits
Emotional Stability	Slant angle and Baseline
Mental Energy	Pen pressure and Letter size
Modesty	Letter size and Top margin
Personal Harmony	Word spacing and Line spacing
Lack of Discipline	Slant angle and Top margin
Poor concentration	Line spacing and Letter size
Non-communicativeness	Letter size and Word spacing
Social Isolation	Word spacing and Line spacing

**VI. DATA FLOW DIAGRAM**

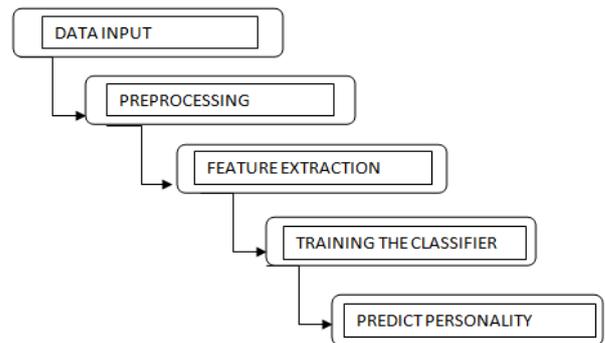


Fig. 8 Data Flow Diagram

The data input takes the image as an input in which the image is pre-processed based on the input given. The image gets pre-processed then the feature is extracted of the pre-processed image. Then classifier classifies the image and predicts the personality. There are seven traits on the basis of which the classifier classifies the image, the classifiers used in the system is SVM which further classifies the image based on the feature extracted.

The process of the feature extraction consist of various parameter, the output of the classified image predicts the personality of the person based on the features we have defined above.

## VII. ALGORITHM

In this project we've used SVM (SUPPORT VECTOR MACHINE) algorithm. SVM algorithm is an algorithm which is used for the classification as well as regression. SVM algorithm produces significant accuracy with less computation power. The SVM is to find a hyperplane in an N-dimensional space to classify the data point distinctly. Hyperplanes are lines or decision boundaries which are used to separate the data classes in a dimensional space. Support vectors are the data points which are closest to the hyper plane and affect the hyper plane.

We can choose many possible number of hyperplane to separate the two classes of data points. The dimension of hyperplane depends on number of input feature. If number of input feature is 2 then the hyperplane is just a line. If input feature is increased (more than 3) than it is hard to find hyperplanes. SVM are of two types, Linear SVM and non linear SVM. Linear SVM is used to separate the linear separable data by a straight line; on-linear data can't be classified by a straight line so here we use a non-linear SVM. Distance between hyperplane and vector is called the margin. So the main aim of the SVM algorithm is to maximize the margin. The maximized margin is known as the optimal hyperplane. When the data is linear we can use straight line to separate them by using only two dimensions. But when the data is non-linear we can't separate it by only using straight line here we have to add 1 more dimension.

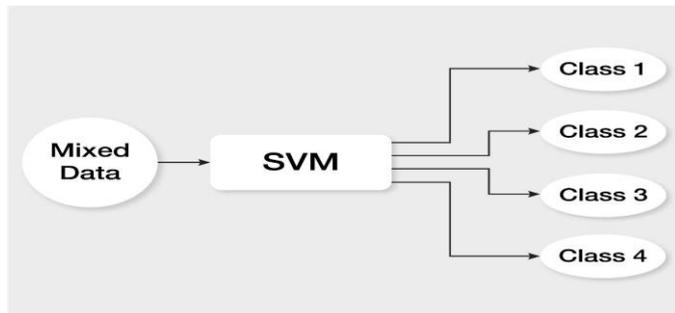


Fig. 9 Support Vector Machine [9]

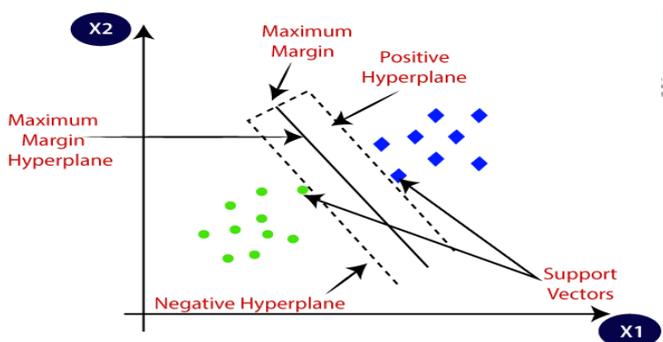


Fig. 10 Support Vector Machine [10]

SVM is used for text categorization, image classification etc. SVM implementation of Sci-kit Learn Library is done. Further, the eight SVM classifiers are trained using RBF Kernel.

## N. STEPS of SVM:

- 1) Load pandas library and data set using pandas.
- 2) Now define feature and target.
- 3) Using sklearn split the train and test dataset.
- 4) Import the SVC function from sklearn of SVM module and build the SVM.
- 5) Values will be predicted by SVM model.
- 6) Evaluate the SVM module.

## VIII. EXPERIMENTAL RESULTS

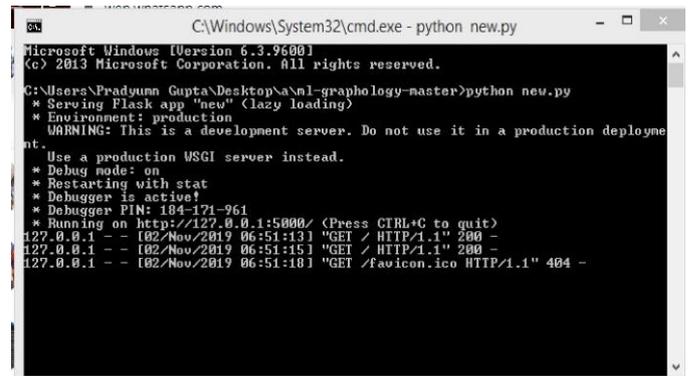


Fig. 10 Running the python file

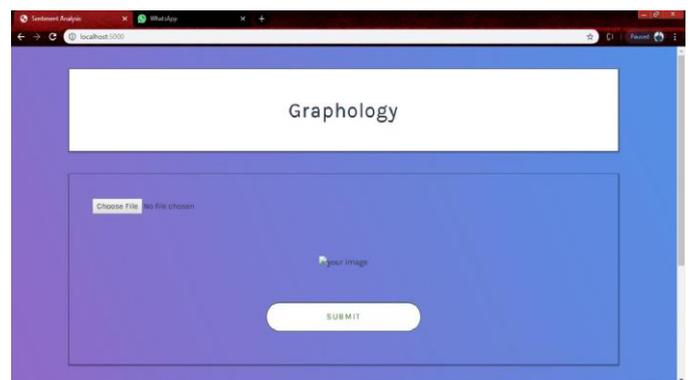


Fig. 11 Home Page

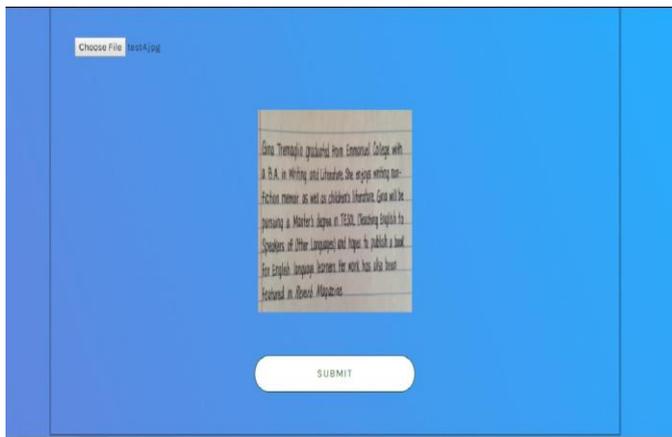


Fig. 12 Uploaded sample image of handwriting

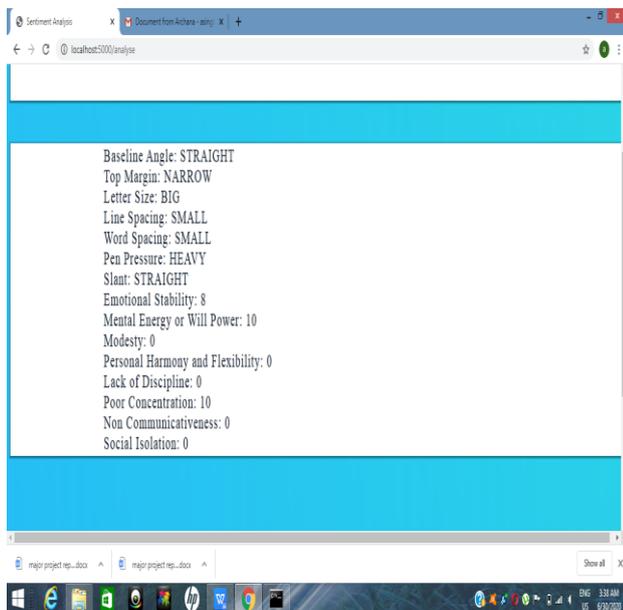


Fig. 13 Personality traits based on the image

## IX. CONCLUSION

By analyzing an individual handwriting patterns we proposed a method to predict some personality traits of an individual. We considered seven handwriting feature to be extracted and eight personality traits to be predicted by different combination of them. For each of the personality traits, each SVM classifier is trained. After sufficient of training we are able to predict personality traits with great accuracy and efficiency. We used Open CV Library for image processing Sci-kit learn Library for using the standard implementation of SVM with RBF kernel. Nevertheless, feature extraction technique may not able to handle every extreme case of handwriting styles and such cases may yield accurate result. Overall, it detects your personality on the basis of your handwriting in a very accurate manner. We are able to achieve ninety percent accuracy in the SVM classifier. It should be noted that extraction of raw features from handwriting might also introduce little inaccuracy.

## X. FUTURE SCOPE

There are spaces for future update in the undertaking. A segment of the redesigns that can be made are recorded as follows:

- O. Future plan is adding more handwriting characteristics like Diacritic placements, Pen lifts and separations and Line quality.
- P. Further, we aim to get better accuracy using different algorithms for image processing techniques.
- Q. Collecting more handwriting samples from writers more the data better the training of machine learning algorithm more accurate the result.
- R. This system can be trained to provide an insight into one's personality such as character, temperature, the present state of mind etc. Even if only a few words are scribbled hurriedly on a piece of paper.
- S. This is measure applicable in areas like human resource development, consultant services, teaching etc. where screening of people is essential.
- T. Enhancing the algorithms to handle extreme cases of handwriting for the more accuracy.

## REFERENCES

- [1] A. Sen and H. Shah, "Automated handwriting analysis system using principles of graphology and image processing," 2017 International Conference on Innovations in Information, Embedded and Communication Systems (ICIIECS), Coimbatore, 2017, pp. 1-6, doi: 10.1109/ICIIECS.2017.8276061.
- [2] P.S.Dandannavar, S.R.Mangalwede, P.M.Kulkarni THE BIG FIVE Personality traits. <https://ieeexplore.ieee.org/abstract/document/8769304>.
- [3] Shitala Prasad, Vivek Kumar Singh, Akshay Sapre "Handwriting Analysis based on Segmentation Method for Prediction of Human Personality using Support Vector Machine", International Journal of Computer Applications, Volume 8– No.12, October 2010.
- [4] Champa H N, K R Ananda Kumar, "Automated Human Behavior Prediction through Handwriting Analysis", IEEE First International Conference on Integrated Intelligent Computing, 2010.
- [5] Joshi, Prachi M. & Agarwal, Aayush & Dhavale, Ajinkya & Suryavanshi, Rajani & Kodoliar, Shreya. (2015). "Handwriting Analysis for Detection of Personality Traits using Machine Learning Approach." International Journal of Computer Applications. 130. 40-45. 10.5120/ijca2015907189.
- [6] "HABIT: Handwritten Analysis based Individualistic Traits Prediction" Abdul Rahiman, Diana Varghese, Manoj Kumar G International Journal of Image Processing (IJIP) Vol. 7 Issue 2 doi:30.04.2013
- [7] Chaudhari K & Thakkar (2019), Survey on Handwriting-Based Personality Trait Identification". Expert System with Applications.doi:10.10.16/j.eswa.2019.01.028.
- [8] Ball, Gregory & Srihari, Sargur & Stritmatter, Roger. (2010). "Writer Verification of Historical Documents among Cohort Writers." Proceedings - 12th International Conference on Frontiers

- in Handwriting Recognition, ICFHR 2010. 314-319. 10.1109/ICFHR.2010.55.
- [9] Mogharreban, N., Rahimi, S., & Sabharwal, M. (2004). "A combined crisp and fuzzy approach for handwriting analysis". IEEE Annual Meeting of the Fuzzy Information Processing Society, 2004. NAFIPS '04, 351-356. doi: 10.1109/NAFIPS.2004.1336307 ©2004 IEEE.
- [10] Plamondon R. and Srihari S.N. 2000. "On-line and off-line handwriting recognition": A comprehensive survey. IEEE Trans. PAMI. 22(1): 63-84.
- [11] H.E.S.Said, T.N.Tan and K.D.Baker "Writer Identification Based on Handwriting", IEE Third European workshop on Handwriting Analysis and Recognition, vol.33, no.1, 2000, pp 133-148.
- [12] Mishra A. Forensic Graphology: "Assessment of Personality". Forensic Res Criminol Int J. 2017; 4(1):9–12. DOI: 10.15406/frcij.2017.04.0009.
- [13] R. F. P. Neves and C. A. B. Mello, "A local thresholding algorithm for images of handwritten historical documents," 2011 IEEE International Conference on Systems, Man, and Cybernetics, Anchorage, AK, 2011, pp. 2934-2939, doi: 10.1109/ICSMC.2011.6084150.
- [14] Bobade, A.M., Khalsa, N.N., Deshmukh, S.M.: "Prediction of human character through automated script analysis." Int. J. Sci. Eng. Res. 5(10), 1157–1161 (2014).
- [15] Flow of SVM <https://images.app.goo.gl/XbVHzfWwo1v3Z8kr7>
- [16] SVM Classification <https://images.app.goo.gl/RmZ7fS2rP4qbvy3w7>