ISSN: 0974-5823 Vol. 6 No. 3 December, 2021

International Journal of Mechanical Engineering

Baseline and its Slant Based Personality Assessment from Handwritten Documents

Pradeep Kumar Mishra[1], Ali Imam Abidi[1*] and Gouri Shankar Mishra[1]

[1] Department of Computer Science & Engineering, School of Engineering & Technology, Sharda University, Greater Noida, India

Abstract: Personality assessment based on handwriting is a difficult and challenging task. However, with the advancement of technology in intelligent systems, it is now possible to determine a person's personality by examining many aspects. The purpose of this paper is to propose a personality assessment system that is based on a baseline and the slant of handwritten words. The system accepts handwritten samples as input and determines the slant of each line statement and determines the personality of a person based on this particular feature. The Center for Pattern Analysis and Recognition (CPAR) dataset is used to evaluate the system's results. The system can be utilized to figure out who you are. Radon transform is used for estimating the slant from handwritten samples. . The system uses set of Structural linguistic variable Slant = {acutely reclined, very reclined, reclined, vertical, inclined, very inclined and acutely inclined}. These slant factors are used to measure mood, self-confidence, coherence of thought, continuity, and hostility using fuzzy logic. After analyzing feedback from users the acceptance percentages lies between 60 to 85 percentages.

Keywords: Slant Estimation, Personality Assessment, Handwrittten document, CPAR dataset.

1. Introduction

The writer's emotive or dramatic expressions are depicted in a handwriting sample. All hand movements are visible in handwriting, and the writer's tendencies give handwritings their character. Individuality also stems from the handwriting learning process, which includes arm, wrist, and finger movements, the usage of writing equipment, the quantity of practice, professional needs, social imitation, and other factors. Some of these variables, such as the writer's arms, wrist, and finger, are influenced by the writer's physical and/or mental state, and thus influence the thickness, length, stiffness, and suppleness of the writer's handwriting strokes. The probability of two handwritings being identical is one in sixty-eight trillions, according to one estimate [1]. As a result, handwriting analysis results are being studied in order to learn more about the connections between handwriting samples and the writer's physical and mental health, as well as individuality. However, a quick review of recent achievements in several crucial areas is provided below for a better understanding of sources supplying knowledge for creating a handwriting processing system. Handwriting analysis is used in criminology to solve crimes by identifying the author of handwritten notes obtained during a criminal investigation [2]. Handwriting analysis [3] is being used extremely well in child development to evaluate children's examination performance [4] handwriting assess handwriting assessment [5], attention deficit hyperactivity disorder and estimation [6]. The detailed application of personality development system (Graphology) is shown in Figure 1.

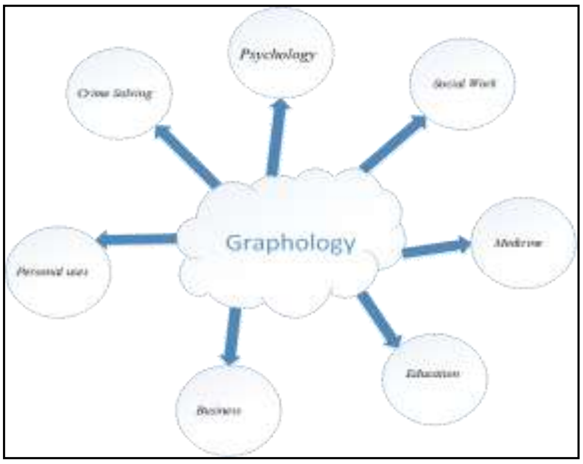


Figure 1: applications of Graphology.

The rest of paper organized as section 2 discuss literature review, Section 3 explains methodology , Section 4 details experimental results and analysis finally section 4 concludes the paper.

2. Literaure Review

Personality evaluation from handwritten materials is a difficult task. Everyone has a unique writing style. The nature of these applications reveals that developing handwriting analysis tools necessitates a thorough understanding of handwriting structure as well as its relationship to features that link handwriting constituents to personality traits, illnesses, and disabilities. There are ample amount of observational data on the relationship between handwriting constituent structures and personality traits, medical or mental diseases, and other behavioral flaws. Conclusions based on such subjective information, however, may not be scientifically acceptable. Although, decades of research efforts, indicate that systems that are being crafted for handwriting analysis applications are expected to touch the acceptable recognition performance level, but their acceptance will depend on the system's ability to provide the scientific explanation and justification of its decision(s) with reasoning, and embedding such explanation ability in systems is a challenging research issue. This research is an effort in this direction. Handwriting analysis based on individual traits Handwriting Analysis based Invidualistic

Traits Predictions [8] HABIT has been proposed. In this slant of baseline, pen pressure , slant of letters and size of writing were used for personality analysis. The roles and age and education be related with handwriting analysis [9]. To analyze data descriptive statistics were used along with person correlation and multiple regression. Results shows that there is a strong correlation between handwriting and age and education. Robust soft-biometrics prediction from handwritten documents [10] for a robust prediction of the writer's gender [11], age range and handedness. First, three prediction systems using SVM classifier and different features, that are pixel density, pixel distribution and gradient local binary patterns, are proposed. Since each system performs differently to the others, a combination method that aggregates a robust prediction from individual systems, is proposed. This combination uses Fuzzy MIN and MAX rules to combine membership degrees derived from predictor outputs according to their performances, which are modeled by Fuzzy measures. Experiments are conducted on two Arabic and English public handwriting datasets. A study based on identifying personality traits have been discussed in [12]. The study differentiates violent and non-violent behavior from multiple samples of criminals. The process shows the various similar and different behaviors amongst the criminal. Personality traits from handwritten documents have been proposed in [13]. The recognition accuracy is between 55 and 70% with different parameters have been reported. A machine learning based approach for analyzing personality has been proposed in [14]. The baseline, margin, slant of the words and height of t-bar of a person's handwritings are used as features for computing personality. A writer identification techniques based on graphologist traits have been proposed in [15], [17]. Using these features 88 to 89 % recognition accuracy has been reported. The progress on personality Assessment has been highlighted in Table 1.

Table 1: Progress on Personality Assesment

Author	**Feature Extraction**	**Data**	**Classification**	**Traits**	**Result**
Liwicky *et. al*, 2010	GMM	IAM Database	SVM	Speed, acceleration, writing direction, Normal x, Y co-ordinate	84.66%
Bouadjenek, *et. al.* 2016	LBP	NA	SVM	English	74%
Kinjal Chaudhari, Ankit Thakkar 2019	Graphology based features	NA	Prediction model	Graphology based traits	NA
Suja Sreeith Panicker, P. Gayathri , 2018 [12]	EEG based feature	NA	SVM	mental stress detection systems, Emotion detection	NA
Mekhaznia, Tahar, Chawki Djeddi, and Sobhan Sarkar [13]	textural features	TxPI-u database	artificial neural networks	Personality traits	55 % and 70%
Fisher, Janet, Anish Maredia , Anita Nixon, Nerissa Williams, and Jonathan Leet. [12]	Graphology based features	Own dataset	NEURO SCRIPT, WANDA , CEDAR-FOX, and Gaussian Mixture Model	violent and non-violent behavior	NA
P. Joshi, A. Agarwal, A. Dhavale , R. Suryavansi, and S. Kodolikar [14]	baseline, margin, slant of the words and height of t-bar	100 samples	Machine learning	Personality traits	NA
Rajiv Kumar, Kiran Kumar Ravulakollu, and Rajesh Bhat [15]	Baseline, slant, margin	CPAR dataset	KNN	Writer identification	88% and 89%

3. Methodology

The experiments starts with input images , pre-process the image, further slants are calculated and using structural linguistic variable traits are estimated and finally personality of person can be evaluated. The detailed methodology used in developing graphology system is shown in Fig. 2.

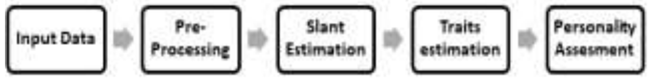


Figure 2: Methodology used in Personality Assessment

3.1 Input data

The handwriting analysis methods are generally trained and tested on non-standard in-house collected datasets. However, there some datasets that are available for benchmark studies for handwritten character, digit and handwriting recognition, and writer identification but, to the best of our information, no benchmark dataset is available for medical, forensic and personality assessment applications. To evaluate the performance of methods for these applications standard benchmark datasets is needed. In these applications, such a dataset will not only help in discovering clues to personality traits or illnesses for medical applications from the handwriting samples but also the discovered ground truth would help in authenticating the subjectivity of the handwriting related myths that are widely accepted. Therefore, without discovering the ground truth, embedded in handwriting it would be unrealistic to devise good handwriting processing methods. Thus, the system depicted in Figure 1 must have module to collect, compile, and store handwriting samples in an organized manner so that the data can be used and shared among researchers efficiently. To standardize the dataset all the writers must be asked to write a same piece of text that contains all possible combinations of alphabet. We are collecting dataset for Hindi, and English. For English no fixed test is designed. Experiments are done on any handwritten text documents in English.

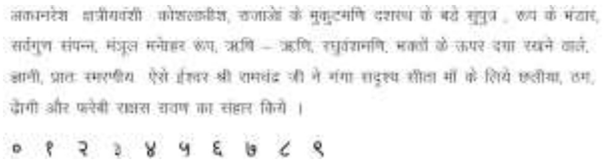


Figure 3 Hindi Text used in CPAR dataset

For Hindi language 13 most commonly used vowels, 14 modifiers and 36 consonants were used. The text was written by subjects of different **age** groups: ranging from 6 to 77 years; **genders**: male and female; **education** backgrounds: ranging from grade 3rd to post graduate levels; **professions** like: software engineer, professor, assistant professor, students, accountants, house wives, and retired persons; **regions**: states of Bihar, Uttar Pradesh, Haryana, Punjab, National Capital Region(NCR), Madhya Pradesh, Karnataka, Kerala, Rajasthan, and countries Nigeria, China and Nepal. Two thousand (2000) writers from these groups participated in this experiment. Similarly, for English & Arabic text 1000 writers from different educational and ethnic backgrounds, age groups, professions, nationalities and linguistics background participated in the experiments. The collected data is being prepared to be placed on a website, so that it can be shared among the researchers of this field. For personality assessment the handwriting samples must be collected from the writers having known personality traits. Such data is difficult to obtain. However, we are trying to create such a dataset from the handwriting samples of known personalities that are available through Internet.

3.2 Pre-Processing:

The input images were resized during the pre-processing step. The smoothing filter was used to reduce noise from the images. A low-pass filter (Gaussian filters) is used to smooth the images during pre-processing. Figure 3 depicts the original and pre-processed images.

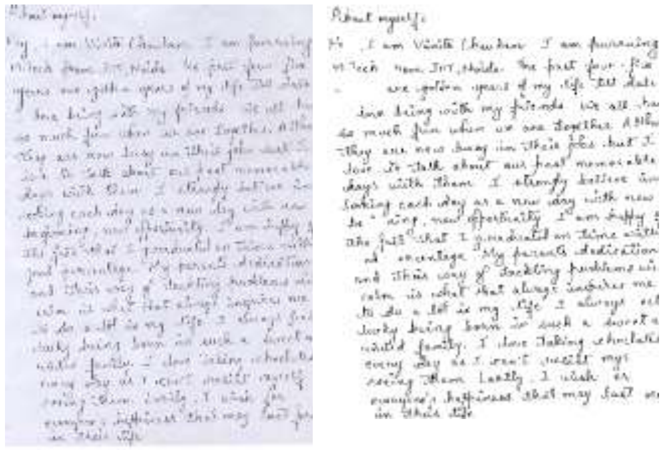


Figure 3: (a) Original Image (b) Pre-Processed Image.

During data collections different types of noises are observed in the dataset. These noises are removed by using wavelet based noise removal techniques mentioned in [16]. The processed images are shown in Figure 4.

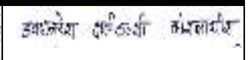
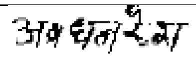
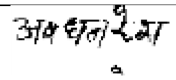
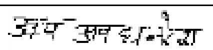
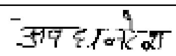
Sl.No.	Noisy Image	Processed Image
1		
2		
3		

Figure 4: Preprocessed Image.

3.3 Slant Estimation:

In this experiment features that graphologist and expert handwriting analyzer uses for writer's personality assessments are used. The advantage of these features is that their definition is based on sound historical knowledge of handwriting analysis. Slant is used to find emotional feeling of a person. This can become a feature for writer identification. The various slant used for emotions are shown in Table 1. In handwriting, as said before, a slant structural attribute signifies emotions. It takes values in set of structural linguistic variable Slant = {acutely reclined, very reclined, reclined, vertical, inclined, very inclined and acutely inclined}. In this implementation, slant values of handwriting words by Radon transform on word images are computed. The Radon transform of an image f(x, y) is the projection on x' of the rotated image f(x',y') at an angle θ as shown in Equation (1-4) where,

$$x^{'} = -x\sin(\theta) + y\cos(\theta) \quad \ldots\ldots\ldots (1)$$

$$y^{'} = x\cos(\theta) + y\sin(\theta) \quad \ldots\ldots\ldots (2)$$

At a given angle θ the projection can be computed by Equation (3)

$$R_\theta(x') = \sum_y f(x', y') \ldots\ldots\ldots\ldots\ldots (3)$$

the slant value of a word w_i in the direction θ at which R_θ (x') attains the maximum value for 0 ≤θ≤ 2π, and fuzzify it as in Equation (4)

$$\widetilde{Slant} = (s, \mu_{\widetilde{Slant}}(s)) \ldots\ldots\ldots\ldots\ldots (4)$$

where $s \in \text{Slant}$. The membership $\mu_{\widetilde{\text{Slant}}}(s)$ is obtained from the membership function of the structural linguistic variables of *Slant* defined over the universe of discourse 0 ≤θ≤ 2π. The fuzzy membership values for inclined slants (acutely reclined, very reclined, reclined, vertical, inclined, very inclined and acutely inclined) are computed by the relationship μ(θ) = (β-θ) / (β-α) where α≤β and α≤θ≤β by setting appropriate α and β values as mentioned in Figure 5.

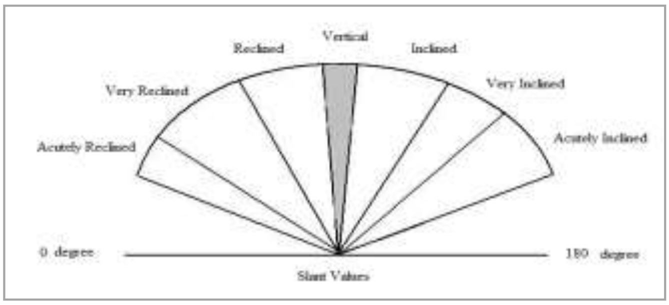


Figure 5: Slant values used in personality assessment.

The frequency of occurrences of every $s \in Slant$ is estimated and stored it in a bin b_s corresponding to each s=1,2,…7.

The distribution of slant feature across 0 to 180 degree in reference to the degree of membership over seven variables of single slant structural linguistic are demonstrated. However finding the membership function is equally effective when it comes to an implementation. In the following a slant estimation algorithm is provided using the function connected component (w, U) that scans image U from top to bottom and line by line, and returns a list of connected components, also referred to as words, in the handwriting sample image U. The functions Radon(w_i,,0..180) returns Radom transform of a word w_i in R(Θ), and the slant fuzzification ($\widetilde{\text{Slant}}_i$, max(R(Θ)), b) creates fuzzy subset $\widetilde{\text{Slant}}_i$ by fuzzifying the maximum value of *R*(Θ), and increments the respective slant type counters in *b*. The algorithm is formally described below.

Algorithm Slant

Get image U;

{

Find connected_component (w, U);

While w is not empty for each word w_i do

{

R(Θ) ← Radon(w_i,,0..180);

slant_fuzzification ($\widetilde{Slant}_i$, max(R(Θ)), b);

}

$$\widetilde{\text{Slant}} = \bigcup_{i=1}^{n} \widetilde{\text{Slant}}_i$$

}

There are two outputs of the algorithm: slant type counter array *b* and the fuzzy subset shown in Equation (5).

$$\widetilde{\text{Slant}} = \{(s, \mu_{\widetilde{\text{Slant}}}(s))\} \ldots\ldots\ldots\ldots (5)$$

where $s \in$ Slant represents a slant type observed in all the words and $\mu_{\widetilde{Slant}}(s)$ is computed from the membership values obtained in all the words. In this experiment all ten writers wrote the same text of Devnagari pangram.

Baseline direction estimation

In psychological prospect, according to E. Downey, baseline direction of handwritten text relates to the writer's mood. This relation on computational line can be described using a series of structural linguistic variables such as leveled, ascending, descending, wavy, sinuous, convex, concave erratic, word rising, word falling. The various baseline direction in handwriting samples are shown in Table 3.

In Devnagari script text is written as group of words where each words contain Sirorekha (a bridging line connecting individual characters). In order to estimate the alignment of baseline it is essential to identify all the Sirorekha of available words of that line. Horizontal projection defines the intensity of orientation to a given direction. This signifies horizontal projection can be used to estimate the orientation of all the Sirorekha of a given line.

3.4 Personality Assessment

The personality assessment depends on various factors. However in this paper personality assessment from slant based feature are considered. Table 2 shows the different highest personality traits associated with handwritten documents on slant feature [19].

Table 2: Slant and personality traits

Sl.No.	Slant	Personality
1	Vertical	Self control
2	Moderate right	Mood
3	Extreme right	Coherence of thought
4	Moderate left	Continuity
5	Extreme Left	Aggression

Similarly, handwriting experts have noted the characteristics of a large number of handwriting symbols like capital letters in a word and loops and extensions and their relationships with personality traits. A discussion on all of them is beyond the scope of this paper. Interested readers are may refer to [17-22].

4. Experimental details:

In this experiment we have implemented *fundamental feature* based components based on slants. These features are essentials for personality prediction while accessories and other features would be required for improving the prediction strength. We describe below methods to estimate the values of features: slant, baseline direction, letter size, continuity, form, arrangement, pressure and speed. As can be accessed from their definitions, the estimation of these feature values is a challenging task. However, we have tried to emulate the process, as realistic as possible, as described by graphology experts. Our system stores the values obtained from all handwriting samples. Its' advantage is that the feature values can be scientifically quantized and that may provide better personality trait estimation.

4.1 Personality Trait Estimation

We have created the fact base of the system. It consists of feature values along with associated personality traits. Both are represented as a vector pair (V_f, T_p), where V_f is a set of *m* values ($v_{f,1}, v_{f,2}, v_{f,3}, \ldots, v_{f,m}$) of feature *f* and associated with it are set of *n* personality traits $T_p=(t_{p,1}, t_{p,2}, t_{p,3}, \ldots, t_{p,n})$. To improve the prediction veracity, the system keeps on updating, regularly, the personality trait list T_p from experts observations and graphology literature. To discover a set of personality traits in an unknown writer we have implemented a nearest neighbor based procedure which is describe below.

1. Extract N feature vectors $V_{f,i}^{u}$ $\forall i = 1, N$ from an unknown handwriting image U

2. $\forall i = 1, N$, compute $\mathrm{T}_{\mathrm{p,i}}^{\mathrm{u}} \leftarrow \Phi\left(V_{f,i}^{u}, V_f\right)$, where $\Phi\left(V_{f,i}^{u}, V_f\right)$ returns $\mathrm{T}_{\mathrm{p,i}}^{\mathrm{u}}$—the traits of the nearest neighbor of the feature vector of the unknown handwriting $V_{f,i}^{u}$ in V_f .

3. Observed personality traits $T_p^u = \sum_{i=1}^{N} T_{p,i}^{u}$ in image U

A comparative analysis of each trait can be seen by clicking the trait. After the click, the current trait values are displayed along with its minimum, average, and maximum values that have been observed by the system. In this case the value of the mood trait is displayed. It is above minimum value but less than the population average.

4. 2Results and discussion

The performance evaluation of the system is difficult because of the involved subjectivity. However, we have used a pretest and posttests agreement measure. In pretest, we asked an individual to assess his/her traits as low, medium and high. After the test, we asked the individual to write their agreement with the result on the scale of 0 to 5 where indicating no to total agreements. In these experiments 500 number of people participated. Their average agreement with prediction ranges between 70 to 95 percent. To check the prediction consistency, a single writer was asked to write different text over a period of time. The chart below shows the analysis of 15 samples. Only the sincerity trait showed high variation in that chart. To assess the prediction variability among writers, we analyzed handwritings of writers were. The chart below shows the trait prediction results for 7 writers. These results illustrate that the personality traits can be used to assess the handwriting individuality. The obtained results are verified by the writer. The writers are agree with approximately 50-60 percentage. From figure 6 it is clear that

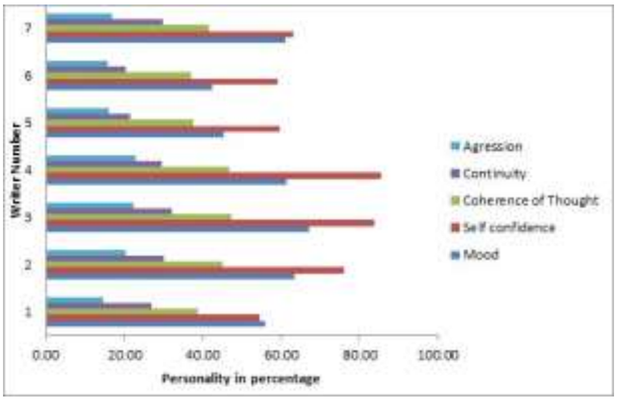


Figure 6: personality characters.

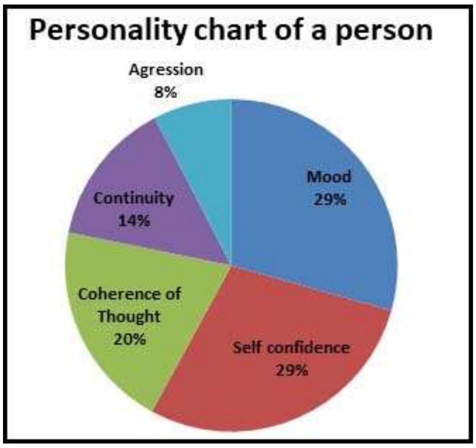


Figure 7: Personality chart of one person.

Personality chart of a person is shown in Figure 7. From figure it is clear that self confidence and mood of person is high, whereas coherence of thought and continuity is medium and aggression is low. To check the prediction consistency, a single writer was asked to write different text over a period of time. The chart below shows the analysis of 15 samples. Only the sincerity trait showed high variation in Figure 8.

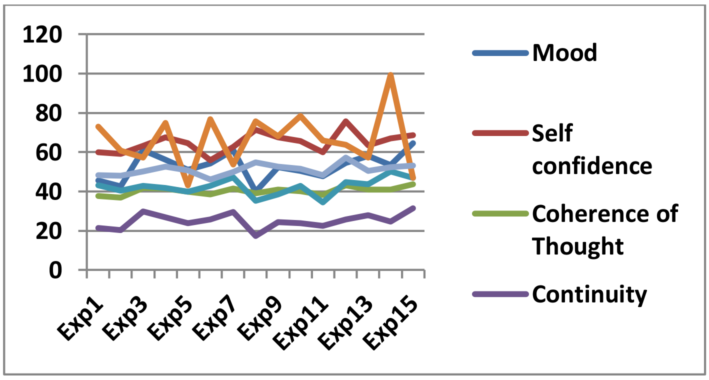


Figure 8: Consistency of single writer over period of time.

To assess the prediction variability among writers, we analyzed handwritings of writers were. The chart below shows the trait prediction results for 10 writers. Figure 3 illustrate that the personality traits can be used to assess the handwriting individuality.

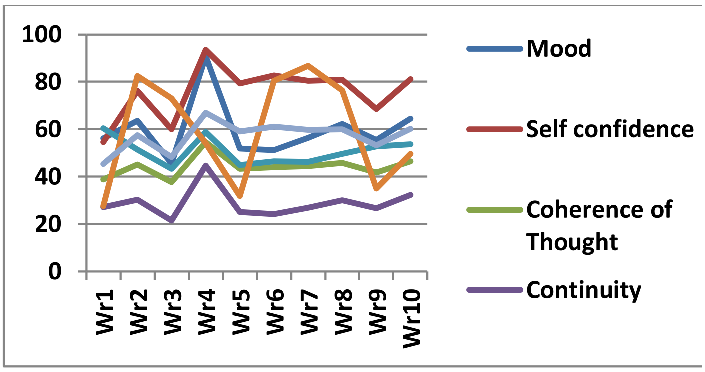


Figure 9: Variations among different writers.

5. Conclusion and future research

Handwriting analysis applications in diverse areas show its importance in exploring the hidden phenomena by a noninvasive means, but, the knowledge discovered from handwriting analysis need to be authenticated to provide the scientific basis. In this paper, we have presented an architecture of a system that allows continuous collection of handwriting samples from subjects having different personality traits, illnesses, learning disorders, criminal records, and other attributes that might help in exploring the writers mindset, compilation, store, retrieve and organize the samples for long term studies, discover knowledge from a large set of samples collected over a long period of time, and to make in scientifically viable authenticate the discovered knowledge. In addition to these, the system can provide justification of its decision along with the explanation of the process through which it reaches to a decision. These two characteristics can help greatly in knowledge justification process and it may guide in learning about the knowledge. To assess the viability of the system we have implemented a personality assessment module using the general and fundamental features. For this experiment, we have implemented simple and straight, forward algorithms to extract the speed, continuity, form, arrangement and pressure features. It gives the values of personality traits: emotions, mood, self-confidence, coherence of thought, and Aggression. The result obtained is also mapped with the users view. Almost we observe 50 to 60 percentage agreement with these results. In future we will develop a process of grouping these values so that a personality assessment model can be formed.

References

1. Downey, June Etta, "*Graphology and the Psychology of Handwriting*", No. 24. Baltimore: Warwick & York, 1919.
2. Hepler, Amanda B., Christopher P. Saunders, Linda J. Davis, and JoAnn Buscaglia. "Score-based likelihood ratios for handwriting evidence", *Forensic science international* 219, no. 1-3 (2012): 129-140.
3. Jonason, Peter K., Virgil Zeigler-Hill, and Jake Baldacchino. "Before and after: Personality pathology, childhood conditions, and life history outcomes", *Personality and Individual Differences* 116 (2017): 38-43.
4. Lowis, Michael J., and Sarah Mooney. "Examination performance and graphological analysis of students' handwriting", *Perceptual and motor skills* 93, no. 2 (2001): 367-381.
5. Falk, Tiago H., Cynthia Tam, Heidi Schellnus, and Tom Chau. "On the development of a computer-based handwriting assessment tool to objectively quantify handwriting proficiency in children." *Computer methods and programs in biomedicine* 104, no. 3 (2011): e102-e111.
6. Shen, I-Hsuan, Tsai-Yun Lee, and Chia-Ling Chen. "Handwriting performance and underlying factors in children with Attention Deficit Hyperactivity Disorder." *Research in developmental disabilities* 33, no. 4 (2012): 1301-1309.
7. Kushnir, T., Y. Arzouan, A. Karni, and D. Manor. "Brain activation associated with practiced left hand mirror writing." *Brain and language* 125, no. 1 (2013): 38-46.
8. Rahiman, Abdul, Diana Varghese, and Manoj Kumar. "Habit: Handwritten analysis based individualistic traits prediction." *International Journal of Image Processing (IJIP)* 7, no. 2 (2013): 209.
9. Holekian, Farnaz, Masoud Gholami Lavasani, and Yaser Madani. "Handwriting analysis: the role of age and education." *International Journal of Modern Management and Foresight* 1, no. 6 (2014): 208-221.
10. Bouadjenek, Nesrine, Hassiba Nemmour, and Youcef Chibani. "Robust soft-biometrics prediction from off-line handwriting analysis." *Applied Soft Computing* 46 (2016): 980-990.
11. Topaloglu, Murat, and Seher Ekmekci. "Gender detection and identifying one's handwriting with handwriting analysis." *Expert Systems with Applications* 79 (2017): 236-243.
12. Fisher, Janet, Anish Maredia, Anita Nixon, Nerissa Williams, and Jonathan Leet. "Identifying personality traits, and especially traits resulting in violent behavior through automatic handwriting analysis." *Proceedings of Student-Faculty Research Day, CSIS, Pace University* (2012).
13. Mekhaznia, Tahar, Chawki Djeddi, and Sobhan Sarkar. "Personality Traits Identification Through Handwriting Analysis." *Pattern Recognition and Artificial Intelligence* 1322 (2021): 155.
14. P. Joshi, A. Agarwal, A. Dhavale, R. Suryavansi, and S. Kodolikar, "Handwriting Analysis for Detection of Personality Traits using Machine Learning Approach," Int. J. Comput. Appl., vol. 130, no. 15, pp. 40-45, 2015, doi: 10.5120/ijca2015907189.
15. Kumar, Rajiv, Kiran Kumar Ravulakollu, and Rajesh Bhat. "Fuzzy-Membership Based Writer Identification from Handwritten Devnagari Script." *JIPS* 13, no. 4 (2017): 893-913.
16. Ruikar, Sachin D., and Dharmpal D. Doye. "Wavelet based image denoising technique." *IJACSA) International Journal of Advanced Computer Science and Applications* 2, no. 3 (2011).
17. Panicker, Suja Sreeith, and Prakasam Gayathri. "A survey of machine learning techniques in physiology based mental stress detection systems." *Biocybernetics and Biomedical Engineering* 39, no. 2 (2019): 444-469.
18. Kumar, Rajiv, Amresh Kumar, and Pervez Ahmed. "A benchmark dataset for devnagari document recognition research." In *6th International Conference on Visualization, Imaging and Simulation (VIS'13), Lemesos, Cyprus*, pp. 258-263. 2013.
19. Chaudhari, Kinjal, and Ankit Thakkar. "Survey on handwriting-based personality trait identification." *Expert Systems with Applications* 124 (2019): 282-308.

20. Schomaker, Lambert. "From handwriting analysis to pen-computer applications." *Electronics & Communication Engineering Journal* 10, no. 3 (1998): 93-102.
21. Beyerstein, Barry L., and Dale F. Beyerstein. *The write stuff: Evaluations of graphology, the study of handwriting analysis*. Prometheus Books, 1992.
22. Amend, Karen Kristin, and Mary Stanbury Ruiz. *Handwriting analysis: The complete basic book*. Red Wheel/Weiser, 2000.
23. https://github.com/rajivbec1/CPAR-DATASET [Center for Pattern Analysis and Recognition dataset]