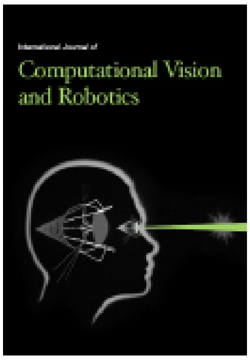


International Journal of Computational Vision and Robotics

ISSN online: 1752-914X - ISSN print: 1752-9131
https://www.inderscience.com/ijcvr

Identification of personality traits from handwritten text documents using multi-label classification models

Salankara Mukherjee, Ishita De Ghosh

DOI: 10.1504/IJCVR.2022.10049835

Article History:

Received:	25 January 2022
Accepted:	13 June 2022
Published online:	01 December 2023

Identification of personality traits from handwritten text documents using multi-label classification models

Salankara Mukherjee*

West Bengal State University,
Malikapur Berunanpukuria, Barasat, North 24 Parganas,
Kolkata-700126, West Bengal, India
Email: getsalu@gmail.com
*Corresponding author

Ishita De Ghosh

Department of Computer Science,
Barrackpore Rastraguru Surendranath College,
West Bengal, India
Email: ishitade.ghosh@gmail.com

Abstract: Handwriting is widely investigated to mark emotional states and personality. However, the majority of the studies are based on graphology, and do not utilise personality factor models. We use the well-known five-factor model which says that people possess five basic traits, together known as big-five. Hence the problem of personality prediction from handwriting is essentially a multi-label problem. In addition to that, the predicted values should be non-binary decimal numbers since the model says people possess the traits in various degrees. Multi-label classifiers have not been explored for personality assessment using handwriting features. The current work aims to bridge the gap. Multi-label classifiers are trained by trait scores obtained by big-five inventory as well as handwriting features. A number of classifiers including classifier chain, binary relevance and label power-set are employed in the work. Best accuracies of 95.9% with non-binary label values and 97.9% with binary label values are achieved.

Keywords: multi-label classification; personality assessment; big-five traits; handwriting features; non-binary label values.

Reference to this paper should be made as follows: Mukherjee, S. and De Ghosh, I. (2024) 'Identification of personality traits from handwritten text documents using multi-label classification models', *Int. J. Computational Vision and Robotics*, Vol. 14, No. 1, pp.18–41.

Biographical notes: Salankara Mukherjee received her MSc and MTech both in Computer Science from University of Calcutta. She has eight years of teaching experience in higher school level. Currently she is pursuing her PhD degree in the Department of Computer Science, West Bengal State University. Her research interest includes image processing and pattern recognition.

Ishita De Ghosh is a senior member of IEEE and ACM since 2013. She received her MTech and PhD degrees both in Computer Science from Indian Statistical Institute, Kolkata. Currently, she is an Associate Professor in the Department of Computer Science, Barrackpore Rastraguru Surendranath College, West Bengal, India. Her current research interests include pattern recognition, artificial intelligence, and their applications. She has published research papers in several international journals, and contributed chapters in books published by reputed publishers.

This paper is a revised and expanded version of a paper entitled 'Big-five personality prediction from handwritten character features and word 'of' using multi-label classification' presented at Int. Conference on Mathematics and Computing (ICMC 2021), Indian Institute of Engineering Science and Technology (IIEST), Shibpur, India, 2–5 March 2021.

1 Introduction

Personality assessment is helpful, if not necessary, in many practical situations. For example, a job aspirant who is stressed easily and disagrees with most of the people most of the time should not be expected to perform well in a team. Hence prior personality assessment would help the recruiters to take the appropriate decision like not assigning him or her to a job that requires teamwork. Consider the example of a boy who is flamboyant, outgoing, and enjoys new experiences. He may not be a good choice as a partner for a girl who is timid, shy, and keeps to herself. Here also, prior personality assessment of the couple may save a future break-up. In other situations, like medical assistance to a depressive or an addicted person, the assessment becomes a compulsion to understand the changes in personality and behaviour of him or her.

Handwriting is a complex human activity that involves an intricate blend of components like cognitive skills (perception, memory, reasoning and concentration), perceptual-motor skills and kinaesthetic abilities (Srihari et al., 2002; Plamondon, 2010; Kim and Lee, 2021). It is also influenced by the training received during childhood and, occupational experiences (Srihari et al., 2016). Being a habitual act, it is difficult to change or fake instantaneously. Moreover, handwriting samples are collected easily by using paper and a pen. Hence people thought of using handwriting as a tool for personality analysis for a long time. The discipline known as graphology must be mentioned in this context. The term is derived from two Greek words, 'graph' and 'logos', meaning respectively 'to write' and 'study'. So it essentially means 'study of writing' (Roman, 1962). The discipline tries to understand the psycho-physical behaviour of individuals, via meticulous observation, analysis and interpretation of the graphic signs detected in handwriting. There are many notable work on *automated personality analysis from handwriting* based on computer-aided graphology (Sheikholeslami et al., 1996) and machine learning (Champa and Kumar, 2010b; Ghosh et al., 2020). The discipline is inherently experimental in nature and there are different schools of thoughts in across the continents (Seifer, 2009). In general, the analysis is done manually by expert graphologists and subjective variations may creep in. Hence graphology is less acceptable as a standard theory for personality analysis (Bushnell, 1996; Tett and Palmer, 1997; Dazzi and Pedrabissi, 2009).

Among the personality theories, factor theory has a descriptive structure which utilises natural language terms people use to describe themselves and others. There are several factor models in personality psychology like Myers-Briggs type indicator (MBTI), Minnesota multiphasic personality inventory (MMPI), five factor model (FFM), etc. There are many published work on automated personality analysis from handwriting using these models (Gavrilescu, 2015; Gavrilescu and Vizireanu, 2018; Mekhaznia et al., 2021). Raymond Cattell's 16 fundamental factors of personality were reduced to five primary factors in FFM (Goldberg, 1990). FFM serves as an integrative system which can represent the diverse system of personality description in a common framework (John et al., 2008). It has accumulated thousands of explorations within its framework, across multiple cultures and diverse populations (Rothmann and Elize, 2003; Arora and Rangnekar, 2016; Ortet et al., 2017). Current work is based on FFM.

According to FFM, every person possesses five basic traits, viz. openness to experiences, conscientiousness, extraversion, agreeableness and neuroticism, together known as *OCEAN*, also as *big-five*. So big-five personality prediction is essentially a multi-label problem, one label for each trait. In addition to that, the predicted values should be non-binary numbers in nature, since the model specifies a method to estimate the degree of presence by calculating score values which come as decimal numbers.

Multi-label classification (MLC) is a relatively new technique which is used to solve problems in which the underlying data contains multiple labels (Gibaja and Ventura, 2010). To solve such problems, MLC is more cost effective than single-label classification (SLC) in respect of time and space because all the labels are considered simultaneously in MLC (Pushpa and Karpagavalli, 2017). Finally, inter-connectivity of the data-points is considered in MLC (Sadhukhan and Palit, 2019). Use of MLC to find out the interrelationship between personality traits and handwriting is not available in literature except one introductory work (Mukherjee et al., 2021).

Objective of the current work is to find out the interrelationship between handwriting and personality by employing MLC and to predict big-five trait scores from handwriting. A system is built to predict the scores without using graphology, thereby eliminating the subjectivity of the prediction method. Handwriting samples for a pre-defined English text are collected from a set of subjects. Various features are extracted from the samples. Personality data is collected from the subjects using big-five inventory (BFI), which is a standard questionnaire to assess the five traits (John and Srivastava, 1999). Personality scores are calculated according to the BFI-scale. After creation of the dataset with handwriting features and personality scores, multi-label classifiers like binary relevance (BR), label power set (LP) and classifier chain (CC) are employed for prediction of personality traits from handwriting using various feature combinations. Two separate test options, namely 10-fold cross validation and percentage split (66% and 80% of total data are used for training purpose and the rest 34% and 20% are used for testing purpose) are used in this work.

Contribution of the work is summarised below:

- A scheme to predict personality traits from handwriting using MLC is proposed. Most of the earlier work has used SLC (Gavrilescu, 2015; Gavrilescu and Vizireanu, 2018). Moreover, many of them have relied upon graphological models (Ghosh et al., 2020). In contrast to that, the current work is accomplished using five factor model (FFM) of psychology

- According to FFM, each of the five traits is present in various degrees in every individual. In earlier work, researchers predicted binary label values (high and low) for the traits (Gavrilescu and Vizireanu, 2018; Lima and de Castro, 2014). In this work, to tackle the multi-valued nature of the traits, multi-label classifiers are trained and tested successfully with non-binary, namely decimal label values. For the sake of completeness, classification is done using binary label values as well.
- Performance of the proposed scheme is good. The best accuracies of 95.9% and 97.9% are achieved by classifier chain with non-binary and binary label values respectively.
- Since no public dataset in English script is available for this type of work, a multi-label dataset comprising of personality trait values as well as handwriting feature values has been built by us. It is one of the few datasets with non-binary label values to be used in multi-label classification.

Organisation of the work is given as follows: a brief survey of related work is given in Section 2; personality assessment technique is discussed in Section 3; classification techniques along with evaluation metrics are discussed in Section 4; proposed method along with dataset creation is presented in Section 5; experimental results and discussion are presented in Section 6 and concluding remarks are given in Section 7.

2 Related work

A brief survey of related work is given now. We emphasise on three components:

a handwriting features

b personality model

c classification techniques.

In most of the work, offline English handwritten text is used. When nothing is mentioned about the mode and the script, it is to be assumed that the mode is offline and the script is English. In some of the related works, online text samples and non-English script is used.

Computer aided graphology was first explored more than 25 years ago (Sheikholeslami et al., 1996). Page level features like margin and line level features like *slope*, *slant*, *ratio of zones* and *spacing between lines* were used by the authors. They analysed handwriting by syntactic pattern recognition using a graphological model. Though the work presents the topic in a lucid and interesting manner, neither the classification techniques nor the validation results were discussed in the work.

Srihari et al. (2002) studied individuality of handwriting as an evidence in court proceedings. Two types of features, viz. *conventional* and *computational*, were used in the paper. Conventional features are generally examined by forensic document experts, whereas computational features are automatically extracted by using computer programs, though with the advent of technology, conventional features may become computational ones. In the work, extracted features were analysed by a 3-layer artificial neural network (ANN). Though the study was based on a large dataset collected from a representative US population of 1,500 people and individuality of handwriting was validated with a

95% confidence, limitation of the work is that no personality model was used for analysis of handwriting.

In 2010, Champa and Kumar (2010b) analysed handwriting using line features, e.g., *baseline*, *pen pressure*, *slant* and features of small letters '*t*' and '*y*' for behaviour prediction using graphology. They proposed a rule-based method to analyse 120 handwritten text samples collected from 120 individuals. In the same year, the authors used a back propagation ANN model using the similar set of features for behaviour prediction (Champa and Kumar, 2010a). The best performance in respect of time was achieved when the number of hidden layer nodes and the number of epochs were 8 and 4,500, respectively. There was no mention of dataset used in the experiments. Moreover, prediction accuracy was reported in neither of the work.

Prasad et al. (2010) used a support vector machine (SVM) to predict personality traits using the similar features employed by Champa and Kumar as given above. For experiments, 100 handwritten text samples containing 70–80 words were collected from 100 individuals, and personality analysis was done manually. Although accuracy of 90.3% was achieved using SVM with RBF kernel, disadvantage of the work is that manual personality analysis may suffer from subjectivity. In 2015, Kedar et al. (2015b) presented a graphology-based study using various features like, *margins*, *zones* and *baseline*; *slant*, *size* and *spacing*; *pressure*, *speed* and *connecting strokes*; *print* and *cursive writing*; and *signature of the writer*. In another work published in the same year, the authors described a subset of the above features and their relationship with the emotional stability and well-being of people (Kedar et al., 2015a). Both the papers describe the relationship between personality and handwriting characteristics in a clear and lucid manner, but the details of feature extraction, classification and evaluation techniques are given in neither of the work.

Deep learning techniques are recently being used for graphology-based personality prediction. Lemos et al. (2018) analysed characteristics like pessimism, optimism, balance, shyness, etc. by convolution neural network (CNN). They did not specify the accuracy obtained. Fatimah et al. (2019) used CNN for prediction of personality characteristics like honesty, sincerity, generosity, diligence, self-esteem, etc. They obtained 82.5% to 100% accuracy for various categories of handwriting features. Ghosh et al. (2020) published a work on analysis of lowercase English letters employing VGG-16 model. Correctness of prediction was evaluated by the subjects themselves by choosing agree or disagree options. The accuracy of the system was reported to be 86.7%. Advantage of the work is that almost all the lower case English letters were used in the study, but the disadvantage is that correctness of the prediction was evaluated by the subjects themselves.

Though all the work mentioned above claimed a close link between handwriting and personality, the relationship was not explored using standard personality assessment models, rather graphology-based models were relied upon. Problems of these models have been mentioned in the previous section. Moreover, in most of these works, very little has been written about the procedure of data collection, background of the subjects, and the specific personality model used in the work. The first work that used a personality factor assessment model to analyze handwriting was done by Gavrilescu (2015). He proposed a method to predict the traits characterised by MBTI model using the handwriting features, *baseline*, *pen pressure*, *connecting strokes*, *slant of words* and features of *lowercase letters* '*t*' and '*f*'. He employed a 3-layer neural network architecture for the work. Average accuracy of 86.7% was achieved, with highest accuracies achieved for dichotomies like extrovert versus introvert, and thinking versus feeling. In a related work, similar type of handwriting features and neural network architecture were employed, but five factors of FFM model were predicted (Gavrilescu and Vizireanu, 2018). Responses of personality assessment questionnaire and handwriting samples were collected from 128 subjects. The highest prediction accuracy was obtained for openness, extraversion, and neuroticism (over 84%), whereas, it was about 77% for conscientiousness and agreeableness.

There are other groups working in the field. Fallah and Khotanlou (2016) proposed a method for personality prediction using MMPI. They employed text independent features like margin; length of words; size of characters; spacing between words and lines; tilt of words and lines and ratio of horizontal to vertical length of characters. They also used text dependent features like *high-order local auto correlation values computed from text images*. Limitation of the work is that data from only 70 individuals was utilised for the experiments and only 76% accuracy was achieved by a neural network-based classification. In 2017, Chen and Lin proposed a method to measure seven dimensions of personality using Chinese personality scale (QZPS) questionnaire. In addition to offline handwriting features, e.g., *size*, *spacing* and *height-to-width ratio of letters*, online handwriting features, e.g., *velocity*, *acceleration*, *duration of pen-tip in the air*, *pause between two successive moves*, were used in the work. Binary classification was done by four methods, support vector machine (SVM), k-nearest neighbour (KNN), AdaBoost, and artificial neural network (ANN). The results were promising with accuracy ranging from 62.5% to 83.9%. Limitation of the work is that, handwritten text samples of only one sentence per subject (collected from 56 subjects) were used for the study.

Recently, Mekhaznia et al. (2021) published a work to predict big five personality traits using statistical features of handwriting. Experiments were done using a database named TxPIu which consists of handwriting in Spanish language written by 418 writers (Ramírez-de-la-Rosa et al., 2018). Classification was done by a three-layer-ANN with feed-forward architecture, and the results showed prediction accuracy of 70% for the first two features and 55% for the other features. Mukherjee et al. (2021) presented a work to predict big-five personality traits using features of the word ‘of’ and four lowercase letters. For experimentation, a dataset was built using BFI responses. MLC with both non-binary and binary values were used in the work. Accuracies of 94.1% for non-binary label values and 98.1% for binary label values were achieved by the authors. Limitation is that data from only 50 writers were used. Details of personality assessment method used in this work are discussed now.

3 Personality assessment

As mentioned already in introduction, *big-five* are the basic five traits: openness to experience, conscientiousness, extraversion, agreeableness and neuroticism, collectively known as OCEAN (John and Srivastava, 1999). A short description of each one is given now.

- *Openness to experience:* people high on this trait express their emotions easily; have interest in diverse topics and a desire for adventure. They are interested in art forms, and in general novelty seeking. Low levels indicate a preference for familiarity and conventionality. Hence for this trait, people are rated based on the dichotomy: curious versus consistent.

- *Conscientiousness:* people high on this trait are accurate, careful, on time, thorough, and, organised. Lack of conscientiousness is related to disorganised, frivolous and irresponsible people. On this scale, people are rated based on the dichotomy: careful versus careless.
- *Extraversion:* people high on this trait are sociable, talkative, and adventurous. It is also related to people who express positive emotions easily and, enjoy other people's company. While introversion relates to people who are generally reserved and fearful. On this scale, people are rated based on the dichotomy: extrovert versus introvert.
- *Agreeableness:* it is a measure of how cordial a person is towards other people. People high on this trait are compassionate and helpful instead of being suspicious. On this scale, people are rated based on the dichotomy: compassionate versus detached.
- *Neuroticism:* it is a measure of emotional control. Low levels of neuroticism indicate steadiness and more noteworthy control over feelings, whereas high levels demonstrate more affectability, uneasiness and less control over feelings.

The five traits remain relatively stable throughout a large span of one's lifetime. They are also used to predict certain important life outcomes such as education and health.

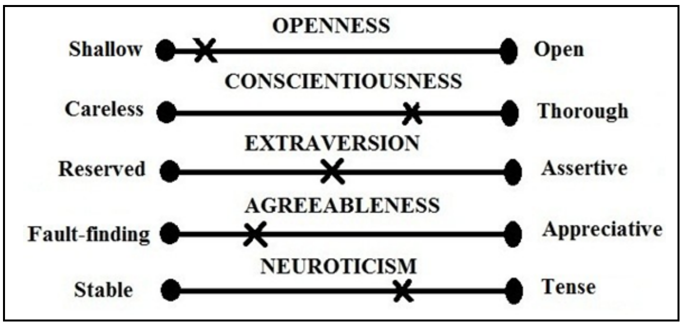
3.1 Assessment method

Methods commonly used to assess the five factors are given now. The NEO PI-R is a 240-item inventory developed by Costa and McCrae (1992). It measures not only the five factors, but also six subordinate dimensions of each one. NEO-FFI is a 60-item truncated version of the above which measures the five factors only. Both NEO-PI-R and NEO-FFI are commercial products. International personality item pool (IPIP), developed and maintained by Goldberg et al. (2006), provides scales similar to NEO PI-R and NEO-FFI scales.

Big-five inventory (BFI) is a common inventory used by research community (John and Srivastava, 1999). Both BFI and IPIP are freely available for non-commercial research purposes, but IPIP is comparatively bigger with 120 items, whereas BFI has 44 items consisting of short phrases and simple vocabulary. Honesty, sincerity and spontaneity of the subjects in responding to the items are essential for our work and boredom of a long process may affect these conditions. Moreover, BFI scaling system enables us to determine not only the presence or absence of a particular trait, but also its gradation in a spectrum of values. Therefore, BFI is chosen for our work.

In BFI system, subjects have to give responses for each item in a five-point scale as following: 1 for disagree strongly, 2 for disagree a little, 3 for neither agree nor disagree, 4 for agree a little, 5 for agree strongly. Then score values for five factors are calculated according to BFI scale. Value for each trait is obtained as a floating point number in the range of [1, 5]. In almost all major work linking Big-Five traits and handwriting, trait scores are converted to binary numbers by using some threshold (Gavrilescu and Vizireanu, 2018; Mekhaznia et al., 2021). So for each trait, the subjects are evaluated for one of the two extreme ends, either lowest or highest. As this does not represent the real scenario, non-binary integer values are utilised in this work. For this, each trait value is scaled up in the range of [1, 10] because the increased spectrum disperses the values leading to more effective training. Then it is converted to nearest integer. Figure 1 illustrates the gradation of traits for a subject in BFI system. Left end, middle point and right end of each scale respectively show the lowest tendency, neutral tendency and the highest tendency. So a particular individual may have lower tendency in one trait, higher in a second trait, and neutral in a third trait.

Figure 1 Plotting of big-five trait scores on scales: cross-mark on each scale denotes the score values for a particular person



4 Classification techniques

SLC techniques are employed to solve learning problems for which each instance is associated with a single target variable that describes its property. However, there is an increase in the number of real-life applications involving data with multiple target variables. MLC algorithms are used effectively to learn from this type of data. In MLC, the training set consists of instances each of which is associated with a set of class labels, and the task is to predict the label-set of an unknown instance by analysing training instances with known label-sets (Cherman et al., 2011). Problem transformation (PT) and algorithm adaptation (AA) are two main approaches of MLC. Various problem transformation techniques for MLC are discussed by Modi and Panchal (2012). MLC is employed to classify Tamil phonemes by Pushpa and Karpagavalli (2017). A comprehensive review of the currently available multi-label learning software like MEKA, MULAN, CLUS, etc. are provided in Charte (2020). In PT approach, a multi-label learning task is mapped into one or more single-label tasks. If multiple single-label tasks are generated, then each of them is treated as an independent task (Cherman et al., 2011). BR, LP and CC, which is an improvement of BR, are common PT techniques used in our work (Pushpa and Karpagavalli, 2017; Multi-Label Classification with Scikit-Multilearn, 2018). Details of classification techniques are discussed in Section 5.3.

4.1 Evaluation metrics

In SLC, classification result of a given instance is either correct or incorrect. However in MLC, an instance can be classified as partially correct and hence different metrics are required. Some of them are adaptations of the ones used in SLC, while others are specially designed (Pushpa and Karpagavalli, 2017; Cherman et al., 2011). Following evaluation metrics are used in our work to measure the classification accuracy.

- *Accuracy* is the average ratio of correctly predicted labels to total number of labels for all instances. Greater accuracy indicates better classification.
- *Hamming_score* is the average accuracy value of the labels in a non-binary multi-label classification.
- *Hamming_loss* is the average ratio of incorrectly predicted labels to total number of labels for all instances. It is simply the difference of hamming score from unity. For hamming loss, smaller value indicates better performance.
- *Exact_Match* is the number of instances for which the predicted set of labels is equal to the true set of labels, divided by the total number of instances. This is a strict measure as it requires the predicted set of labels to be an exact match of the true set of labels.
- *Average_precision* is the proportion of labels ranked ahead of a certain relevant label. The goal is to establish how many positions have to be traversed until this label is found. The bigger the value of average precision is, better is the performance. The best performance is reached when average precision is equal to 1.
- *F-measure* or *F1_score* is a measure of a model's accuracy on a dataset. It is the harmonic mean of precision and recall.

Mathematical formulas of the above metrics are given in equations (1)–(5). Let N be the number of instances in a multi-label dataset and L be the set of all labels defined for the dataset. For i^{th} instance, let Y_i is the set of predicted labels, and, Z_i be the set of true labels, $Y_i, Z_i \subseteq L$.

$$\text{Accuracy} = \frac{1}{N}\sum_{i=1}^{N}\frac{|Y_i \cap Z_i|}{|Y_i \cup Z_i|} \tag{1}$$

$$\text{Hamming\_Loss} = \frac{1}{N}\sum_{i=1}^{N}\frac{|Y_i \,\Delta\, Z_i|}{|L|} \tag{2}$$

where Δ is the symmetric difference between Y_i and Z_i

$$\text{Exact\_match} = \frac{1}{N}\sum_{i=1}^{N} I\left(Y_i = Z_i\right) \tag{3}$$

where $I(\text{true}) = 1$ and $I(\text{false}) = 0$.

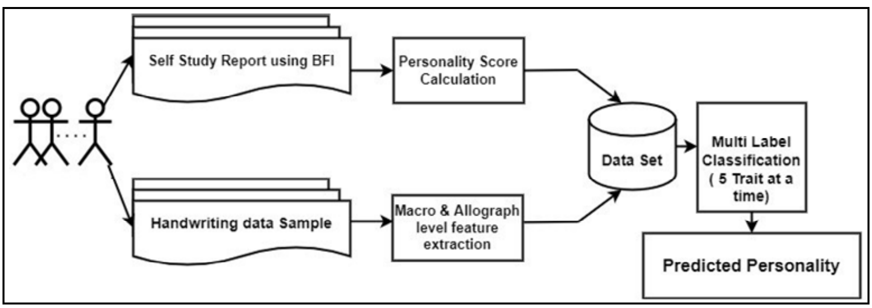
$$\text{Avg\_precision} = \frac{1}{N}\sum_{i=1}^{N}\frac{1}{Y_i}\sum_{y \in Y_i}\frac{\left|\left\{y' \in Y_i : rank\left(x_i, y'\right) \leq rank\left(x_i, y\right)\right\}\right|}{rank\left(x_i, y\right)} \tag{4}$$

$$\text{F1\_score} = \frac{1}{N}\sum_{i=1}^{N}\frac{2\left|Y_i \cap Z_i\right|}{\left|Y_i\right| + \left|Z_i\right|} \tag{5}$$

5 Proposed work

Schema diagram of proposed work is given in Figure 2. Creation of the dataset is an important component of the work and it is described below.

Figure 2 Schema diagram of proposed work



5.1 Data collection and dataset creation

Two types of data, viz. self-study reports according to BFI (as discussed in Section 3) and handwriting samples in English script were collected from total 120 (62 males and 58 females) inhabitants from West Bengal, India. Though their mother-tongue is Bengali, they are proficient in English. Their ages vary from 25 to 70 years and each of them possesses at least one higher educational degree from a college/university.

To discuss the purpose and procedure of the study, a session was initiated with each subject individually or in a group. Moreover, the subjects were requested to provide honest self-study reports and their natural handwriting. For handwriting data collection, each subject was requested to write a predefined text consisting of five sentences (62 words and 309 characters in total) thrice in separate A4-size pages with ball-point pens with black/blue colour ink. The pages are then scanned with a document scanner at 300 dpi. Feature extraction from handwriting is an important part of our work. It is elaborated in the next subsection. Personality score values for a subject are calculated for the five traits according to BFI scoring system (John et al., 2008). Finally, the dataset is created by storing personality trait values as well as handwriting feature values in one row for each subject in data files.

5.2 Handwriting feature extraction

For this work, lines, words, and, specific characters are segmented from the scanned document images. A total of 360 (= 120 × 3 × 1) single lines, 2,520 (= 120 × 3 × 7) words (considering 7 words per line and per set) and 5,760 (= 120 × 3 × 4 × 4) characters or allographs (considering average four samples of each type of four letters from each set) are segmented. Then features are extracted from individual text lines, words, or characters. Different types of features used in the work along with their description are presented in Table 1 (X-axis = horizontal axis, Y-axis = vertical axis). Line and word level macro features and different types of allograph features are discussed now.

- *Macro features:* macro-features are extracted at document level (entire written manuscript) or at page, paragraph, line, and word levels (Srihari et al., 2002). For this work, the first three features, 'baseline direction and pattern', 'word spacing' and 'pen pressure' are extracted at line level. The next three, viz. 'character connection within word', 'zone width ratio' and 'word slant' are extracted at word level.
- *Baseline direction and pattern:* direction of baseline is the angle between the X-axis and the baseline. It is determined by using the horizontal projection profile of a text-line (Mukherjee and De, 2016). Pattern of baseline (levelled, ascending, or descending) is determined by using thresholds on the direction angle of baseline. Different types of baseline are illustrated in Figure 3.
- *Word spacing:* number of white columns between the word segments within a text-line gives the word spacing value. Different types of word spacing are illustrated in Figure 4.
- *Pen pressure:* greyscale intensity values as well as corresponding binary values are used for pressure calculation (Mukherjee and Ghosh, 2020).
- *Character connection within word:* number of connected components within a word is divided by total number of characters in the word to get the connectivity ratio. Connectedness (loosely connected or strongly connected) is determined by thresholding the ratio. Different types of connectedness are illustrated in Figure 5.
- *Zone width ratio:* to balance the size variation of the writing, ratio of middle-zone-width to three-zones-width is considered in this work (Mukherjee and Ghosh, 2020).
- *Word slant:* for calculating the angle between the Y-axis and the vertical axis of the word we implement the method mentioned in Mahanta and Deka (2013).
- *Allograph features:* special attention has been given to allograph or shapes of characters because individuality of handwriting is revealed greatly in strokes used to write characters. Each stroke has a specific direction, length and curvature relative to the other strokes present in a character (Srihari et al., 2016; https://skillsforaction.com). Minute inspection of them reveal the design of allograph, their construction, dimensions (vertical and horizontal), slant or slope, and initial or terminal strokes pattern.

Table 1 Macro and allograph features used in current work

Nature of feature	*Feature name*	*Description*
Macro features line level:	Baseline direction and pattern	Direction is angle between line of writing and X-axis, calculated anti-clockwise
	Word spacing	White column count between the words along the baseline
	Pressure of writing	Pressure exerted by the writer during writing
Macro features word level:	Character connection within word	Character connections within a word due to variation of stroke
	Zone width	Gap between three zones (upper, middle, lower) of writing
	Word slant	Angle in degrees (measured clockwise) from the vertical line to the axes of letters
Allograph level features	Letter 'a' orientation and ellipticity	Orientation, angle between the x-axis and the major axis of the conic. Ellipticity, the angle θ of a conic; where $\theta = \tan^{-1}(minor\_axis/major\_axis)$
	'g' lower stem loop enclosure	Whether the lower zone loop of 'g' is closed or not
	'g' lower stem loop structure	Description of loop structure with horizontal width and height to width ratio
	'g' lower stem slant	Inclination of 'g' stick with respect to Y-axis
	'n' hump's eccentricity	Deviation of a conic section from being circular
	'n' hump curvature	Curvature of hump resulted from strokes, rounded, at rounded or sharp
	Letter 't' connectedness	't' bar is stuck with the 't' stem or not
	Height of 't' bar	In which place the 't' bar is intersecting with 't' stem
	't' stick structure	't' stick upper part looped or not, 't' stick lower part umbrella shaped or blunt
	't' stick slant	Inclination of 't' stem with respect to Y-axis

Figure 3 Types of baseline, (a) levelled (b) ascending (c) descending

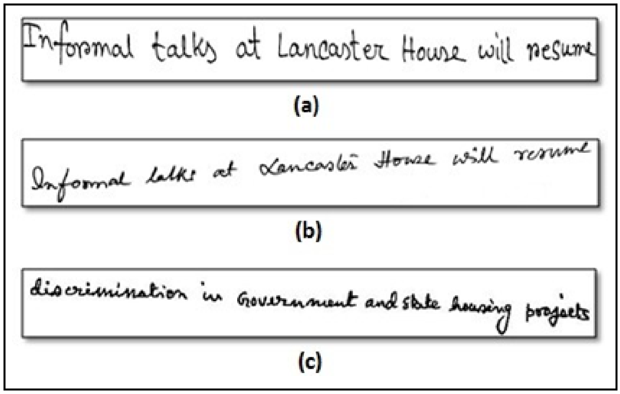


Figure 4 Types of word spacing, (a) wide spacing (b) narrow spacing

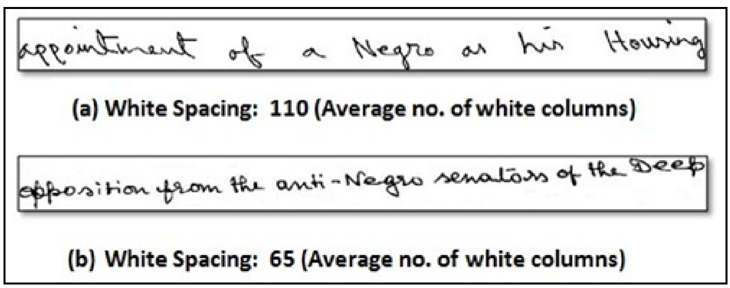
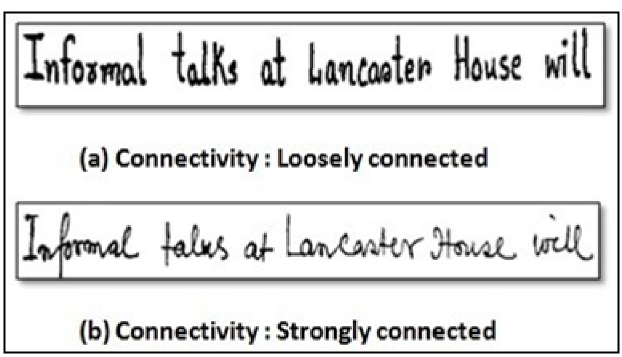


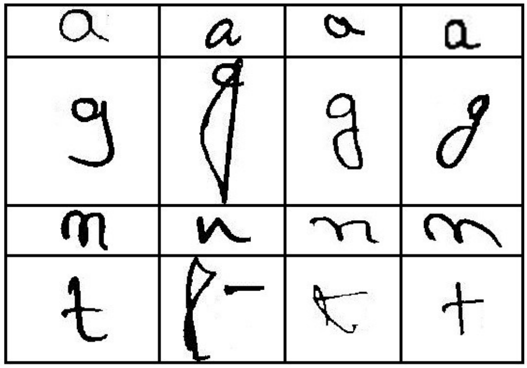
Figure 5 Types of character connectivity within a word, (a) loosely connected (b) tightly connected



In our work, we have selected some common lowercase letters from English alphabet, viz. ‘a’, ‘g’, ‘n’ and ‘t’. Among them, letters ‘t’, ‘a’ and ‘n’ occupy higher position in frequency table (with 9.1%, 8.2%, and 6.7% respective occurrences) available for English alphabet (https://en.wikipedia.org/wiki/Letter frequency). Letter ‘g’ is also an important letter as it covers the full lower zone part of the three writing zones. Moreover, it is observed that writers unconsciously reveal high amount of individuality during construction of letters ‘g’ and ‘t’ (Mukherjee and Ghosh, 2020). For each letter, a region of interest (ROI) or a particular portion of a stroke having sufficient amount of handwritten ink is selected (Ratha and Govindaraju, 2008). Different forms of the four characters are illustrated in Figure 6. Allograph feature extraction process is elaborated now.

- *Doughnut part of letter 'a'*
  - a *Orientation and ellipticity:* the doughnut shape region of letter 'a' is examined for two features, orientation and ellipticity (Mukherjee and Ghosh, 2020). Orientation is positive if the letter has right slant, whereas it is negative if the letter has left slant, and, finally it is zero if the letter has zero slant. Ellipticity is calculated by the ratio between the lengths of minor and major axes. Its value is 1 if the region is of circular shape.
- *Lower stem loop part of letter 'g'*
  - a *Lower stem loop closed or not:* lower stem part of 'g' is thinned to one pixel, then for each pixel number of neighbours is counted. If every pixel has at most two neighbours then it is decided that the loop is not closed, otherwise the loop is closed.
  - b *Lower stem loop structure:* within closed loop part horizontal gap, height of loop and height-to-width ratio is calculated to represent the structure of stem loop.
  - c *Lower stem slant:* slant angle is calculated by vertical projection profile (Mahanta and Deka, 2013).
- *Arch of letter 'n'*
  - a *Eccentricity :* it is a measure of curvature of the arch part or hump part of letter 'n' (Mukherjee et al., 2021). It tells us whether the hump is of arcade shape or angular shape. For determination, the hump part of letter 'n' is thinned to one pixel. From top to bottom manner consecutive black pixels at each row are calculated. Depending upon a threshold value the nature of the hump is detected, viz. whether it is angular, arcade or flat arcade.
- *Stick part of letter 't'*
  - a *Connectedness:* calculate number of connected components. Count value two represents 't' bar is not connected with the stem. Count value one represents 't' structure is connected.
  - b *Height of 't' bar:* count number of consecutive black pixels in each row. Row number having highest value present the height of 't' bar.
  - c *Nature of 't' stick:* presence of the upper stem loop and lower part umbrella structure is identified in the stick part of letter 't' (Mukherjee and Ghosh, 2020).
  - d *Nature of 't' slant:* slant of 't' stick is calculated in the same way as the 'g' stem slant (Mahanta and Deka, 2013).

Figure 6 Examples of different shapes and types of four characters ‘a’, ‘g’, ‘n’ and ‘t’

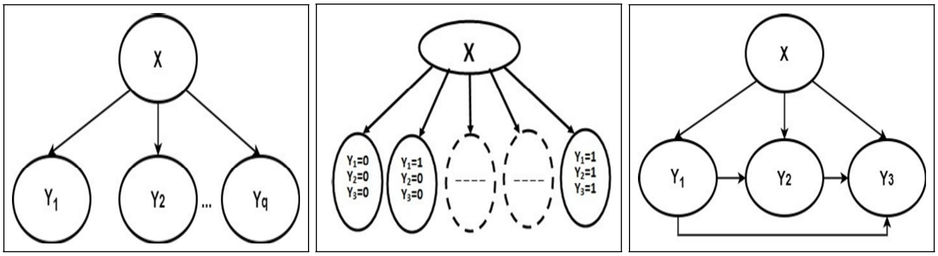


5.3 Classification

As mentioned already, five different personality traits are predicted considering each trait as a class by using MLC. Experiments are carried out using BR, LP and BR’s improved version CC and then results are analysed. The three techniques are discussed in brief now and illustrated in Figure 7.

- *Binary relevance (BR):* in this technique, an MLC task with $q = |L|$ possible label values are converted into q independent SLC-tasks with binary label values. Each SLC-task is addressed independently to get the final result. Essentially, BR transforms the original dataset D into q datasets, each of which contains all the instances of D. If a particular instance contains label y_j, $1 \leq j \leq q$, then it is labelled positively otherwise labelled negatively. Training is done separately for each of the q datasets.
- *Label power-set (LP):* in BR, relationship between labels is not considered, but it is considered in LP which generates a new class for every combination of labels and then solves the problem using multi-class classification approaches. Its drawback is that exponential number of classes may be generated, leading to very few instances in some classes.
- *Classifier chain (CC):* CC involves q binary classifiers as in BR, but it achieves higher performance by overcoming the disadvantages of BR. Each classifier deals with the binary relevance problem associated with label $l_j \in L$ and the classifiers are linked along a chain. The feature space of each link in the chain is associated with the previous links by 0/1 label extensions. Hence target variables $(y_1, y_2, y_3, \ldots, y_q)$ are not fully independent. The features $(x_1, x_2, x_3, \ldots, x_m)$ are initially used to predict y_1. Next, $(x_1, x_2, x_3, \ldots, x_m; y_1)$ are used to predict y_2. At q^{th} step, $(x_1, x_2, x_3, \ldots, x_m; y_1, \ldots, y_{q-1})$ are used to predict y_q. Prediction order of the labels which can be specified by the user, influence the result greatly. Time complexity of CC is smaller than other two methods.

Figure 7 Problem transformation methods, (a) binary relevance (BR) (b) label power-set (LP) (c) classifier chain (CC)



(a) (b) (c)

After splitting the multi-label problem into several single-label problems by problem transformation method, base classifiers are used to handle them separately. Base classifiers are the popular classification algorithms generally used in single-label classification. In our work we have used K-nearest neighbour (KNN), KSTAR and multi-layer perceptron (MLP) as base classifiers. A brief discussion on them is given below:

- *K-nearest neighbour (KNN):* KNN is a lazy learner which stores all instances of training data in an n-dimensional space. Given an unknown instance, it finds the closest K instances in stored data and returns the most common class as the prediction result. Popular distance measures used by KNN are Euclidean distance (ED) and Manhattan distance (MD) (https://towardsdatascience.com).
- *KSTAR:* KSTAR is an instance-based lazy learner which uses an entropy-based distance measure. It classifies by calculating the complexity of transforming an unknown instance into every member of a known class. For this, the probabilities of the transformations occurring in a random walk manner are considered (Cleary and Trigg, 1995; Hernández, 2015).
- *Multi-layer perceptron (MLP):* MLP is an eager learner based on artificial neural network (ANN). There is a set of connected input/output units and it learns by adjusting the weights to predict the correct class label of unknown instances (Haykin, 1998).

6 Experimental results

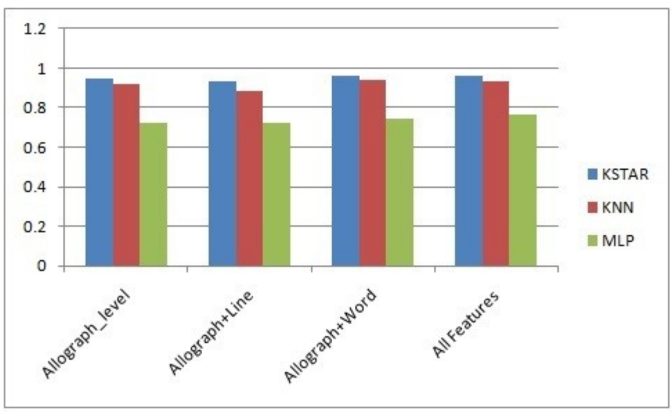
Methods described above are implemented using MATLAB version R2014a (32 bit) in a machine with Intel Core-i5 processor, 4-GB RAM and 64-bit operating system. For classification process Java-based data mining software MEKA (version 1.9.0) is used (Read et al., 2016). Two sets of experiments are conducted. In the first set, decimal numbers in the range [1, 10] are used as personality trait values as discussed in Section 3.1. In our second set of experiments, we have used binary-label values.

Table 2 Output values using classifier chain as multi-label classifier with decimal label values

Features	*Test option*	*Base classifier*	*Label wise accuracy*					*Hamming score*	*Hamming loss*	*Exact match*
			O	*C*	*E*	*A*	*N*			
Allograph level features	Cross validation	KSTAR	0.939	0.951	0.951	0.951	0.934	0.945	0.055	0.931
		KNN (MD)	0.928	0.916	0.919	0.916	0.902	0.916	0.084	0.839
		MLP	0.818	0.715	0.715	0.720	0.643	0.722	0.278	0.507
	Train/test	KSTAR (66:34)	0.864	0.881	0.907	0.924	0.881	0.892	0.108	0.864
		KSTAR (80:20)	0.929	0.929	0.957	0.943	0.929	0.937	0.063	0.929
Allograph level and macro features (line level)	Cross validation	KSTAR	0.939	0.928	0.922	0.942	0.919	0.930	0.070	0.911
		KNN (MD)	0.896	0.867	0.879	0.896	0.867	0.881	0.118	0.847
		MLP	0.801	0.677	0.732	0.746	0.669	0.725	0.275	0.487
	Train/test	KSTAR (66:34)	0.890	0.856	0.873	0.881	0.873	0.875	0.125	0.847
		KSTAR (80:20)	0.929	0.871	0.886	0.857	0.871	0.883	0.117	0.857
Allograph level and macro features (word level)	Cross validation	KSTAR	0.963	0.954	0.948	0.968	0.963	0.959	0.041	0.945
		KNN (MD)	0.939	0.934	0.937	0.942	0.942	0.939	0.061	0.919
		MLP	0.833	0.729	0.706	0.769	0.683	0.744	0.256	0.501
	Train/test	KSTAR (66:34)	0.914	0.915	0.898	0.924	0.915	0.914	0.086	0.890
		KSTAR (80:20)	0.914	0.914	0.914	0.929	0.929	0.920	0.080	0.914
All features combination (allograph + macro)	Cross validation	KSTAR	0.968	0.948	0.951	0.968	0.960	0.959	0.041	0.945
		KNN (MD)	0.948	0.925	0.925	0.939	0.925	0.933	0.067	0.914
		MLP	0.804	0.764	0.746	0.793	0.735	0.768	0.232	0.542
	Train/test	KSTAR (66:34)	0.915	0.881	0.890	0.949	0.890	0.905	0.095	0.873
		KSTAR (80:20)	0.929	0.900	0.929	0.929	0.914	0.920	0.080	0.900

Performance of allograph level features, macro features, and, their various combinations are explored in this work. Experimental results with non-binary label values in CC as the multi-label classifier and KSTAR, KNN (MD) and MLP as the base classifier are depicted in Table 2. Two different test options cross validation with ten folds and percentage split with (66:34 and 80:20) are used. The table contains the accuracy value for each of the five labels, and corresponding hamming score (which is same as the average accuracy of five labels), hamming loss and exact match values. It is observed that KSTAR as the base classifier gives the best average accuracy of 95.9% in two cases: combining all the features (allograph level, line-level macro features and word-level macro features) and combining allograph and word level macro features. The second best average accuracy of 94.5% is obtained only by allograph level features. It is important to note the classifiers BR and LP could not handle the non-binary label in our experiments. A comparison of Hamming score of different base classifiers with different feature combinations is depicted in Figure 8.

Figure 8 Hamming score of CC with different base classifiers and different feature combinations with non-binary label values (see online version for colours)

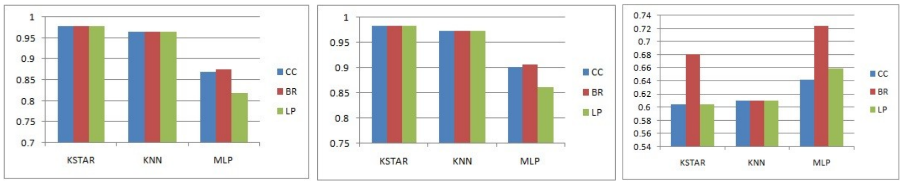


Since non-binary label values could not be used for BR and LP, we have converted the non-binary label values to binary format for these two classifiers. For this conversion the median value for each label is treated as the threshold value for that label. The values which are less than the threshold value are treated as '0' and '1' otherwise. All feature combinations are used here and 10-fold cross validation and percentage split are used as test options. Experimental results in terms of hamming score, exact match, F1-score and average precision values are displayed in Table 3. It is observed in Table 3 that all the three multi-label classifiers (BR, CC and LP) provide the highest accuracy value 97.9% by KSTAR as base classifier. Equality of accuracy value also happened for KNN (MD) and provides 96.4% accuracy. For MLP the three multi-label classifiers give different values. A comparative study of three evaluation metrics namely hamming score, F1-score and average precision with different MLC having different base classifiers are shown in Figure 9.

Table 3 Output values for multi-label classification using binary label values

Features	*Test option*	*Multi-label classifier*	*Base classifier*	*Label-wise accuracy*					*Hamming score*	*Exact match*	*F1-score*	*Avg. precision*
				O	*C*	*E*	*A*	*N*				
Allograph level and macro features	Cross validation	CC, BR, LP	KSTAR	0.988	0.965	0.983	0.983	0.974	0.979	0.957	0.984	0.630
		CC, BR, LP	KNN	0.968	0.951	0.965	0.968	0.968	0.964	0.922	0.973	0.610
		CC	MLP	0.879	0.853	0.885	0.865	0.862	0.869	0.654	0.901	0.642
		BR	MLP	0.879	0.862	0.853	0.899	0.882	0.875	0.611	0.906	0.724
		LP	MLP	0.807	0.775	0.830	0.827	0.853	0.818	0.608	0.861	0.659

Figure 9 Hamming score, F1-score and average precision of different multi-label classifiers and different base classifiers with binary label value (see online version for colours)



6.1 Discussion

We have experimented with handwriting features extracted at different levels, viz. line and word level (together termed as macro level), and their various combinations. We have also experimented with various multi-label and base classifiers. Important observations from experimental results are discussed now.

- CC with KSTAR as the base classifier shows the best accuracy values (95.9%) for non-binary label values using two feature combinations (allograph + macro (all)) and (allograph + macro (word)) with cross validation (10 fold).
- Similarly all the three MLCs CC, LP and BR provide highest accuracy of 97.9% with KSTAR as the base classifier for binary label values.
- MLP as base classifier in multi-label classification show comparatively lower accuracy values than those values for KSTAR and KNN.
- Dataset split in train and test set ratio of 66:34 shows good results for all types of feature combinations, but the performance is better when the dataset is split in 80:20 ratio.

We recorded the time taken to build and test the system is almost negligible for KSTAR and KNN (MD) but the build time for MLP is 8.30 minutes and the test time is few seconds.

6.2 Comparison with state of the art work

Comparison of the proposed work with some state-of-the-art work discussed in Section 2 is shown in Table 4. Comparison of methods in respect of efficiency is hard if the methods are evaluated on different datasets. Since no such dataset is publicly available, researchers have worked with their private datasets. So we had to compare our work with such works and we have tried to be consistent with the personality model (big-five) and the handwriting script (English) as far as practicable. All of them except one viz. (Mukherjee et al., 2021) use binary data as label values, whereas proposed method utilises both binary and non-binary data as label values for classification and achieves satisfactory performance in cases, 97.9% for binary data and 95.9% for non-binary data.

Table 4 Comparison of proposed method with state-of-the-art work

Work	*Year*	*Method*	*Personality model*	*Dataset*	*Prediction accuracy values*	
					Binary label value	*Non-binary label value*
Gavrilescu and Vizireanu	2018	Feed-forward neural networks; binary label values	Big-five	Private, 128 writers	84.4% (intra-subject), 80.5% (inter-subject)	-
Mekhaznia et al.	2021	Three-layer-ANN with feed-forward architecture	Big-five	418 writers	70% for both EDH and RLD and 55% for other features	-
Mukherjee et al.	2021	Multi-label classifier: CC (non-binary) CC, BR, LP (binary)	Big-five	Private, 50 writers	98.1%	94.1%
Current work	2022	Multi-label classification with problem transformation algorithms	Big-five	Private, 120 writers	97.9%	95.9%

7 Conclusions

In this work, a scheme for automated personality prediction from handwriting is proposed. To avoid the debatable role of graphology, personality assessment is done by Big-Five inventory which provides a means for quantitative measurement. Both macro and allograph level handwriting features and their combinations are analysed. Results obtained by extensive experiments with various combination of features show that allograph level features play the key role in prediction. It is also seen that the proposed method gives competitive results with respect to state of art results. Like ‘Solar flare’, our dataset is one of the few datasets with non-binary label values which can be used for multi-label classification. The dataset will be made available to the researchers on request by email.

There are several possible extensions of the work which we would like to work on. More critical analysis of features and their extraction may further improve the prediction accuracy. The work can be extended with bigger dataset with more number of subjects with diverse background. Online handwriting data from the same set of subjects will provide a more complete picture. Moreover, self-study reports can be collected by questionnaires with more number of items. Sub-factors of big-five can be experimented with.

There are several interesting work on the relationship between big five traits and social media footprints (Lima and de Castro, 2014; Azucar et al., 2018; Shahreki et al., 2022). If the subjects in this work have their presence in social media, then it will be interesting to compare the results obtained from media data with that obtained from handwriting data. Finally, it will be a promising work if we can use ECG, EEG or some other sensor data along with handwriting.

Acknowledgements

We gratefully acknowledge the help obtained from Prof. Bhabatosh Chanda of ISI, Kolkata, India and Prof. Kaushik Roy of WBSU, Barasat, India for their valuable advices during the work. We are thankful to Dr. Debatree Mukherjee of BRS College, Barrackpore, India and Dr. Chayan Halder of RKMVC College, Rahara, India for their useful suggestions to complete the work. We are also thankful to Dr. Ajoy Kumar Mukherjee and others who helped us in collection of data for our experiments. Finally we convey that this work could not be done without active and enthusiastic participation of the persons who provided data.

References

Arora, R. and Rangnekar, S. (2016) 'Linking the big five personality factors and career commitment dimensions: a study of the Indian organizations', *Journal of Management*, Vol. 35, No. 9, pp.1134–1148.

Azucar, D., Marengo, D. and Settanni, M. (2018) 'Predicting the big 5 personality traits from digital footprints on social media: a meta-analysis', *Personality and Individual Differences*, April, Vol. 124, pp.150–159.

Bushnell, I.W.R. (1996) 'A comparison of the validity of handwriting analysis with that of the Cattell 16PF', *Int. Journal of Selection and Assessment*, January, Vol. 4, No. 1, pp.12–17.

Champa, H.N. and Kumar, K.R.A. (2010a) 'Artificial neural network for human behavior prediction through handwriting analysis', *Int. Journal of Computer Applications*, May, Vol. 2, No. 2, pp.36–41.

Champa, H.N. and Kumar, K.R.A. (2010b) 'Automated human behavior prediction through handwriting analysis', *Proc. Int. Conf. on Integrated Intelligent Computing*, August, Bangalore, India, pp.160–165.

Charte, F. (2020) 'A comprehensive and didactic review on multilabel learning software tools', *IEEE Access*, Vol. 8, pp.50330–50354.

Chen, Z. and Lin, T. (2017) 'Automatic personality identification using writing behaviors: an exploratory study', *Behavior and Information Technology*, Vol. 36, No. 8, pp.839–845.

Cherman, E.A., Monard, M.C. and Metz, J. (2011) 'Multi-label problem transformation methods: a case study', *CLEI Electronic Journal*, Vol. 14, No. 1.

Cleary, J.G. and Trigg, L.E. (1995) 'K*: an instance-based learner using an entropic distance measure', *12th International Conference on Machine Learning*, pp.108–114.

Costa, P.T. and McCrae, R.R. (1992) 'Normal personality assessment in clinical practice: the neo personality inventory', *Psychological Assessment*, March, Vol. 4, pp.5–13.

Dazzi, C. and Pedrabissi, L. (2009) 'Graphology and personality: an empirical study on validity of handwriting analysis', *Psychological Reports*, Vol. 105, No. 3, pp.1255–1268.

Ramirez-De-La-Rosa, G., Villatoro-Tello, E. and Jiménez-Salazar, H. (2018) 'TxPI-u: a resource for personality identification of undergraduates', *Journal of Intelligent and Fuzzy Systems*, May, Vol. 34, No. 5, pp.2991–3001.

Fallah, B. and Khotanlou, H. (2016) 'Identify human personality parameters based on handwriting using neural network', *Proc. Artificial Intelligence and Robotics*, August, Qazvin, Iran, pp.120–126.

Fatimah, H., Djamal, E.C., Ilyas, R. and Renaldi, F. (2019) 'Personality features identification from handwriting using convolutional neural networks', *4th IEEE Int. Conf. on Information Technology, Information Systems and Electrical Engineering (ICITISEE)*, November, Yogyakarta, Indonesia.

Gavrilescu, M. (2015) 'Study on determining the Myers-Briggs personality type based on individual's handwriting', *Proc. 5th IEEE Int. Conf. on E-Health and Bioengineering Conference (EHB)*, November, Iasi, Romania.

Gavrilescu, M. and Vizireanu, N. (2018) 'Predicting the big five personality traits from handwriting', *EURASIP Journal on Image and Video Processing*, Vol. 57, pp.1–17.

Ghosh, S., Shivakumara, P., Roy, P., Pal, U. and Lu, T. (2020) 'Graphology based handwritten character analysis for human behavior identification', *CAAI Transactions on Intelligence Technology*, January, Vol. 5, No. 1, pp.55–65.

Gibaja, E. and Ventura, S.N. (2015) 'A tutorial on multi-label learning', *ACM Computing Surveys*, April, Vol. 47, No. 3, Article No. 52, pp.1–38.

Goldberg, L.R. (1990) 'An alternative 'description of personality': the big-five factor structure', *Journal of Personality and Social Psychology*, Vol. 59, No. 6, pp.1216–1229.

Goldberg, L.R., Johnson, J.A., Eber, H.W., Hogan, R., Ashton, M.C., Cloninger, C.R. and Gough, H.G. (2006) 'The international personality item pool and the future of public-domain personality measures', *Journal of Research in Personality*, Vol. 40, pp.84–96.

Haykin, S. (1998) *Neural Networks: A Comprehensive Foundation*, 2nd ed., Prentice Hall PTR, USA.

Hernández, D.T. (2015) 'An experimental study of K* algorithm', *International Journal of Information Engineering and Electronic Business*, March, Vol. 7, No. 2, pp.14–19.

John, O.P. and Srivastava, S. (1999) 'The big-five trait taxonomy: history, measurement, and theoretical perspectives', in Pervin, L.A. and John, O.P. (Eds.): *Handbook of Personality: Theory and Research*, 2nd ed., pp.102–138, Guilford Press, New York.

John, O.P., Naumann, L.P. and Soto, C.J. (2008) 'Paradigm shift to the integrative big-five trait taxonomy: history, measurement and conceptual issues', in John, O.P., Robins, R.W. and Pervin, L.A. (Eds.): *Handbook of Personality: Theory and Research*, 3rd ed., pp.114–158, Guilford Press, New York .

Kedar, S., Bormane, D.S., Dhadwal, A., Alone, S. and Agarwal, R. (2015) 'Automatic emotion recognition through handwriting analysis: a review', *Proc. IEEE Int. Conf. on Computing Communication Control and Automation*, February, Pune, India, pp.811–816.

Kedar, S., Nair, V. and Kulkarni, S. (2015) 'Personality identification through handwriting analysis: a review', *Journal of Advanced Research in Computer Science and Software Engineering*, January, Vol. 5, No. 1, pp.548–556.

Kim, B. and Lee, K. (2021) 'Study on hand gesture recognition with CNN-based deep learning', *International Journal of Computational Vision and Robotics*, September, Vol. 6, pp.571–579.

Lemos, N., Shah, K., Rade, R. and Shah, D. (2018) 'Personality prediction based on handwriting using machine learning', *International Conference on Computational Techniques, Electronics and Mechanical Systems (CTEMS)*, pp.110–113.

Letter Frequency [online] https://en.wikipedia.org/wiki/Letter frequency (accessed March 2021).

Lima, A.C.E. and de Castro, L.N. (2014) 'A multi-label, semi-supervised classification approach applied to personality prediction in social media', *Neural Networks*, Vol. 58, pp.122–130.

Machine Learning Classifiers [online] https://towardsdatascience.com (accessed 11 June 2018).

Mahanta, L.B. and Deka, A. (2013) 'Skew and slant angles of handwritten signature', *Int. J. of Innovative Research in Computer and Communication Engineering*, November, Vol. 1, No. 9.

Mekhaznia, T., Djeddi, C. and Sarkar, S. (2021) 'Personality traits identification through handwriting analysis', in Djeddi, C., Kessentini, Y., Siddiqi, I. and Jmaiel, M. (Eds.): *Proc. 4th Mediterranean Conf. Pattern Recognition and Artificial Intelligence (MedPRAI 2020)*, Springer Nature, Switzerland, pp.170–187.

Modi, H. and Panchal, M. (2012) 'Experimental comparison of different problem transformation methods for multi-label classification using Meka', *Int. Journal of Computer Applications*, Vol. 59, No. 15, pp.10–15, ISSN: 0975-8887.

Mukherjee, S. and De, I. (2016) 'Feature extraction from handwritten documents for personality analysis', *Proc. IEEE Int. Conf. on Computer Electrical and Communication Engineering (ICCECE)*, Kolkata, India, December.

Mukherjee, S. and Ghosh, I.D. (2020) 'Writer identification based on writing individuality and combination of features', *Proc. IEEE Int. Conf. on Applied Signal Processing (ASPCON)*, Kolkata, India, October.

Mukherjee, S., Ghosh, I.D. and Mukherjee, D. (2022) 'Big-five personality prediction from handwritten character features and word 'of' using multi-label classification', Giri, D. et al (Eds.): *Proceedings of 7th Int. Conf. on Mathematics and Computing, Advances in Intelligent Systems and Computing (AISC)*, Vol. 1412, pp.275–299, Springer Nature, Singapore.

Multi-Label Classification with Scikit-Multilearn (2018) [online] https://xang1234.github.io/multi-label (accessed 8 October 2020).

Ortet, G., Martínez, T., Mezquita, L., Morizot, J. and Ibáñez, M.I. (2017) 'Big five personality trait short questionnaire: preliminary validation with Spanish adults', *Spanish Journal of Psychology*, February, Vol. 20, pp.1–11.

Plamondon, R. (2010) 'Neuromuscular studies of handwriting generation and representation', *International Conference on Frontiers in Handwriting Recognition (ICFHR)*, November, Kolkata, India.

Prasad, S., Singh, V.K. and Sapre, A. (2010) 'Handwriting analysis based on segmentation method for prediction of human personality using support vector machine', *Int. Journal of Computer Applications*, Vol. 8, No. 12, pp.25–29.

Pushpa, M. and Karpagavalli, S. (2017) 'Multi-label classification: problem transformation methods in Tamil phoneme classification', *7th International Conference on Advances in Computing and Communications*, August, Cochin, pp.572–579.

Ratha, N.K. and Govindaraju, V. (2008) *Advances in Biometrics Sensors, Algorithms and Systems*, Springer, London.

Read, J., Reutemann, P., Pfahringer, B. and Holmes, G. (2016) 'MEKA: a multi-label/multi-target extension to WEKA', *Journal of Machine Learning Research*, February, Vol. 17, No. 21, pp.1–5.

Roman, K.G. (1962) *Handwriting a Key to Personality*, The Noonday Press, USA.

Rothmann, S. and Elize, C. (2003) 'The big five personality dimensions and job performance', *South African Journal of Industrial Psychology*, October, Vol. 29, No. 1, pp.68–74.

Sadhukhan, P. and Palit, S. (2019) 'Lattice and imbalance informed multi-label learning', *IEEE Access*, Vol. 8, pp.7394–7407.

Seifer, M. (2009) *The Definitive Book of Handwriting Analysis*, New Page Books, Career Press, New Jersey, USA.

Shahreki, J., Ling, T.S., Ibrahim, N.F., Chin, A.L.L., Jayiddin, N.F. and Ai, Y.J. (2022) 'The effect of big five personality traits on social network usage among the young generation', *International Journal of Business Excellence*, April, Vol. 26, No. 3, pp.378–397.

Sheikholeslami, G., Srihari, S.N. and Govindaraju, V. (1996) 'Computer aided graphology', *Int. Workshop on Frontiers in Handwriting Recognition*, September, Essex, England, pp.457–460.

Skills for Action [online] https://skillsforaction.com (accessed 11 October 2020).

Srihari, S.N., Cha, S.H., Arora, H. and Lee, S. (2002) 'Individuality of handwriting', *Journal of Forensic Sciences*, August, Vol. 47, No. 4, pp.856–872.

Srihari, S.N., Meng, L. and Hanson, L. (2016) 'Development of individuality in children's handwriting', *Journal of Forensic Sciences*, Vol. 61, No. 5, pp.1292–1300

Tett, R.P. and Palmer, C.A. (1997) 'The validity of handwriting elements in relation to self-report personality trait measures', *Personality and Individual Differences*, Vol. 22, No. 1, pp.11–18.