Research Article

International Journal of Advanced Technology and Engineering Exploration, Vol 9(93)
ISSN (Print): 2394-5443 ISSN (Online): 2394-7454
http://dx.doi.org/10.19101/IJATEE.2021.875577

A self adaptive cognitive deep learning framework for classifying graphology features to Big five personality traits

Lakshmi Durga[1*] and Deepu. R[2]
Department of CSE, Maharaja Institute of Technology Mysore, Mysore, India[1]
Professor, Department of CSE, ATME College of Engineering, Mysore, India[2]

Received: 02-March-2022; Revised: 28-July-2022; Accepted: 30-July-2022

Abstract
Graphology is a technique for study and analysis of the individual personality from his/her handwriting style. Most of the existing graphology-based solutions for personality detection recognize nonstandard application dependent personalities. Even the very few big five personality recognition approaches have limited accuracy and lacks adaptivity to new handwriting styles. Towards these problems, a novel self-adaptive cognitive learning framework based on deep learning convolutional neural network (CNN) features is proposed to classify the handwritten document to big five personality traits. This framework correlates the various document level and character level graphology features to big five personality traits to recognize features with a strong correlation to various big five personality traits and uses these features to classify the personality. To enhance the deep learning feature learning ability an enhanced convolution kernel is proposed for the CNN. Through testing with various handwritten documents, the proposed solution is found to provide 2.18% higher accuracy and 5% lower false positives compared to existing works on big five personality classification.

Keywords
Graphology, Big five personality, Deep learning, Cognitive learning.

1.Introduction
Graphology is a science of handwriting analysis. In this, the behaviour or personality of an individual is predicted using various characteristics like strokes, features, left margin, right margin, spacing between the words and patterns in his handwritten documents [1]. Many personality traits like anger, morality, fears, mental problems, social skills, work habits can be predicted using handwriting analysis. Computer aided analysis of personality detection using graphology is being increasingly used in many applications like crime investigations, candidate screening, personality grooming etc. The personality that can be classified based on graphology is diverse and it becomes difficult to comprehend compared to Big five models. Big five models provide a simple and effective way to comprehend the personality using five core personality traits [2] of neuroticism, conscientiousness, openness, agreeableness and extraversion. A person with the traits of creativity, willing to explore the unexplored has a higher degree of openness personality in him. A person with higher time consciousness, futuristic, pays more attention to details has a higher degree of conscientiousness on him. A person with traits of social active and free going with other has a higher degree of extraversion in him. A person with the traits of empathy for another, caring and lovable has a higher degree of agreeableness in him. A person with the traits of frequent mood shifts, feeling anxious, mostly has a higher degree of neuroticism in him.

Though there are many works on big five personality detection from body clues, questionnaire analysis etc., they are intrusive and results are subjective. The results are also prone to bias. Towards this end, there is a need to develop a personality assessment system which is non-intrusive and free from bias without any subjectivity in results. The link between the handwriting and neurological aspects of the brain has been demonstrated in many research works [3, 4]. Graphology or handwriting analysis based personality assessment satisfies the requirement of non-intrusiveness, free from bias and no subjectivity in results. Along these lines, many works have been

*Author for correspondence

proposed mapping graphology features to personality traits [5]. But in most of the works personality classes were huge and non-standard. There are not many works for mapping features in graphology domain to standard personality traits like big five model. An in depth study on existing methods are presented in section 2. The existing methods were limited in terms of features, coverage of features and lacked correlation of features to each of big five personality traits. As the result, these methods could not achieve higher accuracy in the presence of diverse handwriting patterns.

In this work, we explore the problem of mapping graphology features to big five personality model solving the problems in feature engineering to achieve higher accuracy for diverse handwriting patterns. Feature engineering techniques are the novel contribution of this work. Features are extracted in two level of: document and individual characters. A statistical correlation test is conducted to identify the features with significant influence on the big five personality. These significant features are encoded to a binary feature vector. A cognitive artificial neural network is trained to classify the binary feature word vector to big five personality traits. Following are the innovations of the proposed personality classification system.

(i) A novel two stage feature engineering for graphology based on both document level and character level features. The two-stage feature engineering selected, highly relevant features for big five personality classification. Statistical correlation of graphology features to big five personality traits.
(ii) Feedback based adaptive cognitive artificial neural network classifier optimized with Hinge loss feedback control to classify the big five personality traits using the features selected by two stage feature engineering.
(iii) A novel adaptation to the convolution operation called as Differences-Magnifier (DM) is proposed to enhance the accuracy of 't' and 'i' pattern recognition. This provides attention to subtle regions which affects the recognition of characters.

The paper is organized into sections as below. Section 2 presents the survey on various graphology features and their problem in personality traits classification. Section 3 presents the statistical correlation analysis to identify the significant features. Details are provided about novel feature vector based on significant features and the machine learning model for big five personality prediction. Section 4 presents the results of personality classification with proposed machine learning model. Section 5 provides discussion of results and Section 6 presents the concluding remarks and future work scope.

2.Related work

Fallah and Khotanlou [6] explored the problem of personality assessment from handwriting analysis. Handwritten image is processed to extract features of margin, size of character, spacing between lines &words, character tilt and classified the features to Minnesota multiphasic personality inventory (MMPI) personality scale with an accuracy of 70%. The personality model used in this work is outdated compared to big five personality assessment. Mekhaznia et al. [7] extracted textural features from handwritten image documents. Neural network classified the textural features to two different personalities of the individual. Compared to big five personality, the scope of personalities that can be detected is limited and also the accuracy is less than 70%. Mutalib et al. [8] classified the personality of the individual from his handwriting document using 't' character style. Based on the 't' pattern the method was able to classify if a person is a pessimist or optimist, which is very limited compared to Big five scales. Personality classification using 't' pattern was also explored by Gavrilescu and Vizireanu [9]. The authors classified the 't' pattern to two different personalities. Though the method was able to achieve more than 85% accuracy, the personality scope is limited and the time to classify personality is also high.

Mishra et al. [10, 11] extract line direction and line spacing features from handwritten images and classified it to personality with support vector machine. But due to limitation in feature, the accuracy was less than 60% in this method. Asra and Shubhangi [12] classified the personality using character zonal features. Handwritten document image is processed to extract character zonal features. The features are classified to two different personalities using support vector machine. The accuracy is less than 80% in this approach. Champa and AnandaKumar [13] proposed a handwriting-based personality classification model using features of 't' pattern and pen pressure and neural network classifier. The method used handwriting features to classify self-esteem scales (3 scales). Rahiman et al. [14] classified personality from handwritten documents using rule matching. Following features of baseline, pen pressure, slant of characters, size of the letter is extracted from a handwritten document to classify personality. Even though it classified more personality almost all are nonstandard. Fisher et al. [15] predicted individual tendency to commit crimes by extracting features of letter slant, shape. The features are classified using neural network.

Prasad et al. [16] used multi class support vector machine (SVM) for personality classification by extracting six different features of baseline, letter size, pen pressure, spacing between lines and spacing between letters. Though this method was able to classify 16 different personalities most of them are not standard and application dependent. Grewal and Prashar [17] classified the personality of individuals from the handwritten document. Document level features like baseline and character level features like t patterns are classified to personality scales using artificial neural network. The personalities classified by this approach are about 50 and most of them are application dependent. Coll et al. [18] proposed a solution to rank the applicants based on their attitude using handwriting analysis. Article neural network was used to predict the attitude using character features like shape and slant. A ground truth handwriting characteristics for desired attitude is established and used to train the neural network. Mukherjee and De [19] classified personality of individual from handwriting features of size of the letter, spacing between letter and slant of signature. The personality scope is limited and not standard.

Joshi et al. [20] classified the personality from handwritten document using features of letter slant, baseline etc. k-nearest neighbors (KNN) classifier was used. The number of personalities classified using this approach is huge and nonstandard. Kacker and Maringanti [21] processed then handwritten document to extract features of margins, baselines, letter size and zones. These features were classified using rules to different personalities. The method covered over 20 different personalities and most of them are not standard. Mutalib et al. [22] used fuzzy logic to classify the handwritten document to four different personality levels using the baseline document features. Compared to Big five models, the personality scale of this approach is limited. Wijaya et al. [23] used multi class SVM to classify 15 different personalities using the margin features of handwritten document. Chitlangia and Malathi [24] classified five different personalities of energetic, extrovert, introvert, sloppy and optimistic from the histogram of gradient (HOG) features of handwritten document. This method works for document with a single line and have higher false positives. Enneagram personality scale classification using graphology features was explored in Pratiwi et al. [25]. Though the accuracy is limited in this approach, it showed a positive correlation between personality and the graphology features.

Majumder et al. [26] proposed a method for big five personality classification using document level stylistic features and word semantic. Deep learning was used to extract word semantics. But the method has higher false positives as word semantics have lower correlation to big five personality traits. Lokhande and Gawali [27] proposed a big five personality classification method using rule matching on features extracted from handwritten document. Features of dot placement over letters, curve around letters, strokes and connected between letters are extracted from handwritten documents. Due to rule matching, the classification rules lack adaptation. Hashemi et al. [28] proposed a similar rule matching based personality classification using features of line skew, slant, letter size, text density extracted from the handwritten document. The method did not cover big five personalities. Though the personality scope is higher most are not standard.

Chaubey and Arjaria [29] trained CNN model to classify the big five personality. CNN is trained with an entire handwritten document image as input. The method was able to achieve only 43% accuracy. Rahman and Halim [30] extracted graph features from handwriting text and classified the graph features to big five personality traits using the general adversarial network. The feature is scale and rotation variant and the solution expects the handwriting text is a same Cartesian coordinate system for graph-based feature extraction. Also, it becomes to extract graph based features for multiline texts. Anari et al. [31] used lightweight deep convolutional neural network (DCNN) to classify the personality. The handwriting images are preprocessed and the image is passed as whole to DCNN. Personality classes are not standard and specific to applications. Samsuryadi and Mohamad [32] made a survey on graphology feature extraction methods and presented the correlation between graphology features and personalities. But the personality traits correlated in this work were non-standard. Pathak et al. [33] demonstrated that use of deep learning techniques provided higher accuracy in classifying graphology features to personality. But the work did not classify Big five personalities and used nonstandard personalities.

Nolazco-Flores et al. [34] classified the emotional state of the person based on handwriting style. But the work fixed itself to 3 specific handwriting and 4 specific drawing tasks. Elngar et al. [35] extracted various personality features from different handwriting. The extracted features were classified to Big five personality traits using ANN based models, personaNet and CNN neural network. The feature set and feature coverage are limited and the approach does not work for diverse handwritings. Impedovo and Pirlo [36] extracted handwriting features and used it to classify neurodegenerative diseases. They extracted features like velocity, acceleration of handwriting which requires special sensors. Gahmousse et al. [37] extracted textural features and used it to classify big five personality. The accuracy is less than 60% in this method. Valdez-Rodríguez et al. [38] split the images to patches and extracted CNN features for patches. The features of patches are used for personality classification. The personalities classified by this approach are not standard. Lemos et al. [39] passed the entire image to CNN to predict personality. But passing entire image was over fit and accuracy was low in this method.

Kedar and Bormane [40] extracted 18 different graphological features from handwritten images to detect hypertension. Most of the features are considered for correlation analysis in the proposed work. Fatimah et al. [41] extracted handwritten document structural features and CNN based character features and classified the features to personality. The false positives were high as the CNN was not sensitive to different character patterns. Costa et al. [42] used both text semantics and handwriting structural features to detect personality of the individual. Using text semantic made the approach specific to applications and not generic.

From the survey, it can be seen that most of the existing works don't cover standardized personality model like big five personality traits. They either classify very limited personalities or higher number of nonstandard application dependent personalities. In approaches based on big five personality traits, the features were limited and feature coverage is also limited. This necessitates better feature engineering approaches to improve the classification of the big five personality.

3.Proposed solution

The overall flow of the proposed solution is given in *Figure 1*. Character level and document level features are extracted from the input handwritten image. There are a total of six different features (F1 to F6) extracted from the handwritten image. F-score (Equation 3) is calculated for each of the feature combination and the feature combination with higher correlation to Big five personality traits are found. Each feature has multiple attributes and not all the attributes are not relevant for personality classification and the most relevant features is found using thresholding against symmetric uncertainty (SU) (Equation6) from the feature combination. The selected features are classified by the self-adaptive cognitive artificial neural network to one of big five personality classes.

From the analysis of related works, most used features for graphology based personality trait classification are in two categories

(i)Document level features

(ii)Character level features.

Document level features are baselines, margins, line space and word space. Character level features are patterns of different letter like t, i, a, s, y etc. Since the character space for personality mapping is huge, this work restricts its scope to 't' and 'i' characters.

The features and feature attributes were selected based on their correlation to various personality traits [30–42].

The features and the input values of the features were found through related works. On a sample of two hundred people, both their handwritten document and their Big five psychometric test [43] results are collected. The psychometric test provides a value of 1 to 5 to participant based on the dominant big five personalities found for the participant. The mapping is given in *Table 1.*

The collected handwritten document image from sample population is labeled with their dominant personality type in the range of one to five as given in Table 1. The handwritten images are preprocessed to make it suitable for feature extraction. Preprocessing involves steps of normalization, binarization, morphological operation and image orientation. The raw image is normalized at the first step. The object of interest is segmented out and normalized [44]. The purpose of normalization is to bring the image to a range suitable for processing. After normalization, the image is binarized using Otsu segmentation method. The binarized image is then subjected to morphological operations of open and close to remove noise. The handwritten image is then oriented to the common Cartesian system to correct for any skew angle. The total of 6 features with possible feature values is extracted. The feature vector is of dimension of 40 elements. There are well proven algorithms for document level feature extraction. In this work, those algorithms are used for extraction of document level features like baselines, margin, space between lines and the space between words [45] as in *Table 2*. Statistical correlation analysis is conducted to identify the correlation between document and character level graphology features to the big five personality traits. The input and output variables for correlation test are given in *Table 2*. For detection of 't' and 'i' types deep learning is used in this work.

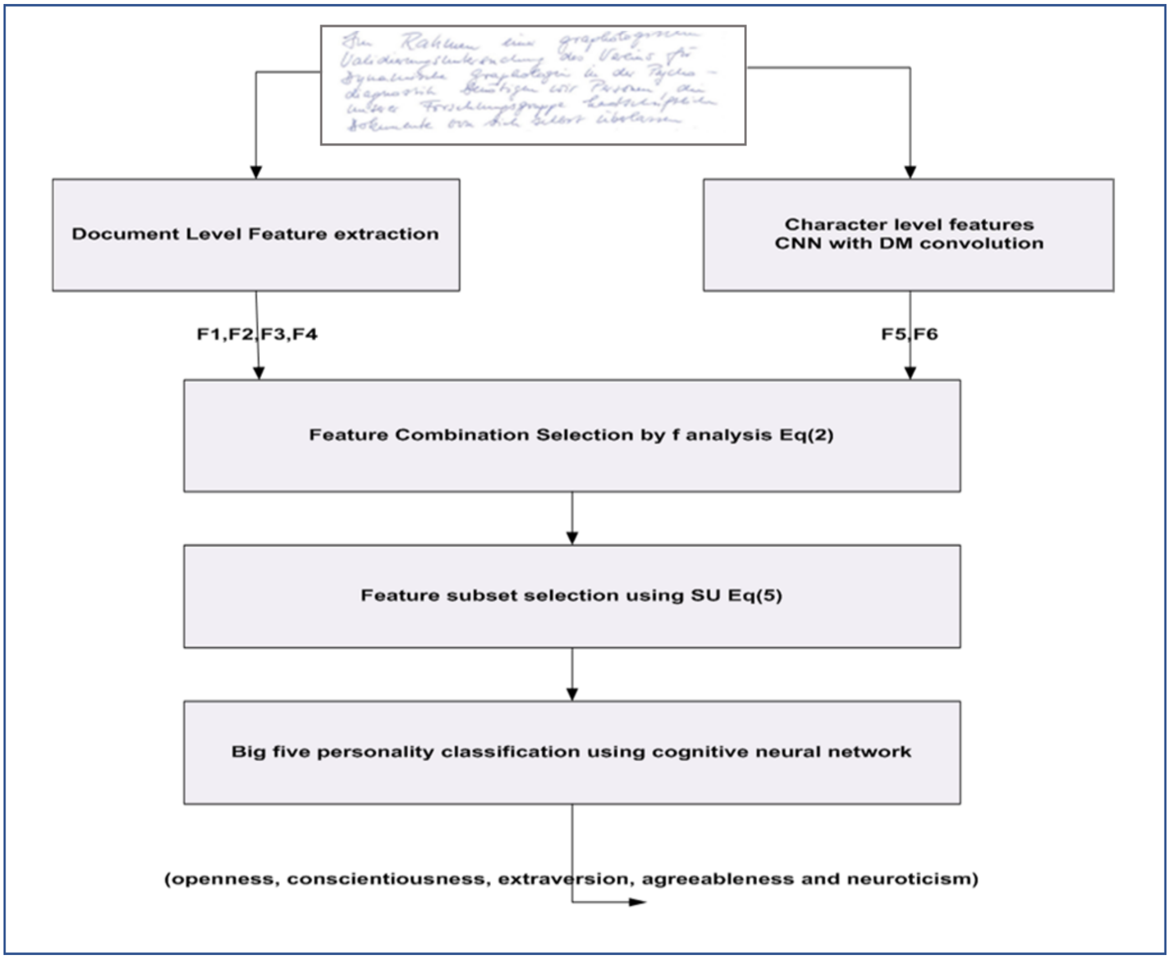

Figure 1 Proposed solution architecture

Table 1 Scoring for personality

Dominant personality	Score
Openness	1
Conscientiousness	2
Extraversion	3
Agreeableness	4
Neuroticism	5

Table 2 Input and Output variables for correlation

Features	Input values	Output variable
Baseline (F1)	Level	Openness
	Ascending	Conscientiousness
	Descending	Extraversion
	Varied	Agreeableness
	Convex	Neuroticism

Features	Input values	Output variable
	Concave	
Margin (F2)	Even margin all around	
	Margin wide all around (f8)	
	Left margin is wide (f9)	
	Right margin is wide (f10)	
	Narrowing of Left margin (f11)	
	Widening of Left margin (f12)	
	Narrow on left (f13)	
	Wide on upper (f14)	
	Narrow on upper (f15)	
	Wide on lower (f16)	
	Narrow on lower (f17)	
	Without margin (f18)	
Line spacing (F3)	Even (f19)	
	Narrow (f20)	
	Wide (f21)	
	Very wide (f22)	
	Tangled (f23)	
	Varied (f24)	
Space between words (F4)	Narrow (f25)	
	Wide (f26)	
	Even (f27)	
	Uneven (f28)	
	Very wide (f29)	
‘t’ patterns (F5)	Checkmark (f30)	
	Tapering (f31)	
	Heavy (f32)	
	Absent (f33)	
	Up-sloping (f34)	
	Down-sloping (f35)	
‘i’ patterns (F6)	Regular (f36)	
	Heavy dot (f37)	
	Light dot (f38)	
	Concave-cresent (f39)	
	Dash-dot (f40)	

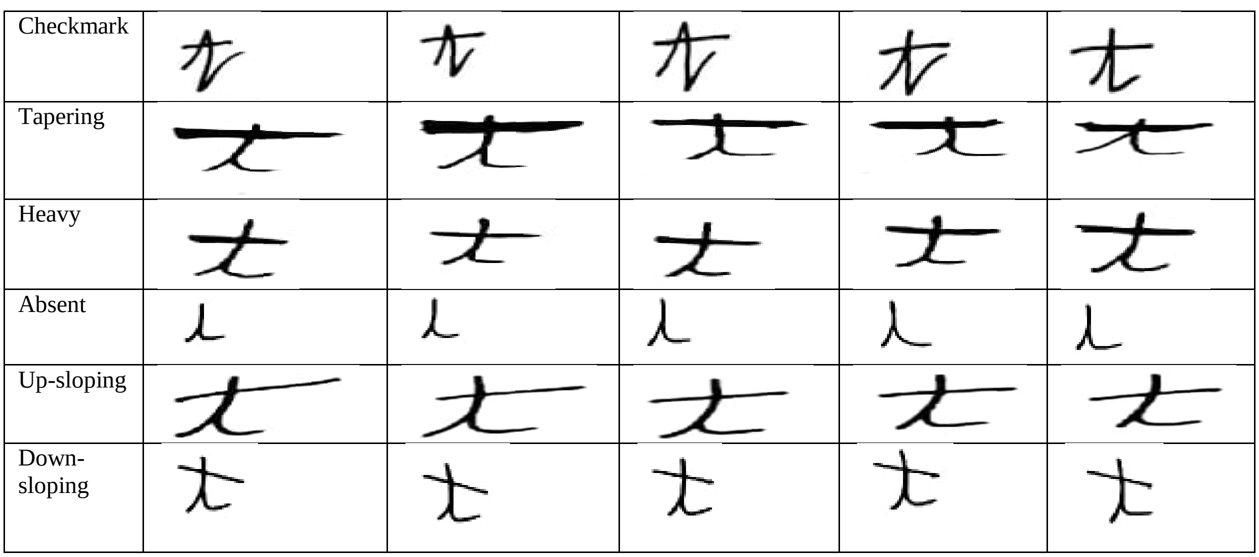

Figure 2 t patterns

The document level features (F1, F2, F3, F4) and character level features (F5, F6) are extracted from the input handwritten document image. The best feature combination with high correlation to big five personality traits is found using f analysis (Equation 2). From the feature combination, the feature subset is selected using SU analysis (Equation 5). A binary feature encoding is created from a selected subset of features and given to trained cognitive neural network model for dominant big five personality trait classification.

The patterns of ‘t’ and ‘i’ has higher similarities (*Figure 2*) and this makes a selection of suitable handcrafted features with higher accuracy a challenge. Even with deep learning convolutional neural networks, the default convolutional kernel can miss out subtle differences. To overcome this problem, this work proposes a deep learning architecture with novel convolution operation called DM convolution. DM convolution magnifies the significant areas for patterns, so that the features in those significant layers becomes amplified in subsequent feature learning. By this way, the classification accuracy is improved. Since there are 6 different types of t with each having 5 different types of occurrences, a binary mask is prepared for each occurrence. The binary mask is of size 64×64. For each occurrence, the binary mask is constructed by converting the letter pattern image to binary (by OSTU thresholding) and resizing the image to 64×64. Through this process 30 different binary masks are constructed for ‘t’ patterns and 25 different masks are constructed for ‘i’ patterns.

For an input gray scale image (q) of size 64×64 to the convolution operation, Local binary pattern (LBP) is computed. This LBP result and each of masks of letter patter are joined with logical AND operation. Each of the 30 results after AND is then convolved with 48×48 kernel and summed up to get the output feature map. The output feature map with DM convolution is given as Equation 1.

$$\mathrm{C(q)} = \sum_{\mathrm{m=1}}^{\mathrm{M}} \sum_{\mathrm{n=1}}^{\mathrm{N}} \mathrm{AND}\big(\mathrm{LBP(q)}, \mathrm{mask(m)}\big). \mathrm{K\,(j)} \quad (1)$$

Where M is the number of masks, N is the number of kernels and mask(m) is the binary mask of $\mathrm{m^{th}}$pattern. Deep learning architecture for ‘t’ and ‘i’ type recognition used in this work is given in *Figure 1* with the novel DM kernel at the first layer of convolution.

The CNN takes the image of size 64×64 pixels and activates the output neuron corresponding types. Two CNN models are trained with output layer neurons as 6 for t and 5 for i. The CNNs are trained with patterns of t and i to recognize the types. The CNN configuration for recognizing t character is given in *Table 3.*

The CNN configuration for recognizing i character is given in *Table 4*.

Table 3 CNN Configuration for t character recognition

Layer	**Size**	**Kernel**
input	64×64	-
Conv1 with DM kernel	48×48×128	5×5
Pool1	24×24×128	2×2
Conv2	20×20×192	5×5
Pool2	10×10×192	2×2
Conv3	8×8×256	3×3
Pool3	4×4×256	2×2
Conv4	2×2×128	3×3
FC	1024	-
Softmax	6	

Table 4 CNN Configuration for i character recognition

Layer	**Size**	**Kernel**
input	64×64	-
Conv1 with DM kernel	48×48×128	5×5
Pool1	24×24×128	2×2
Conv2	20×20×192	5×5
Pool2	10×10×192	2×2
Conv3	8×8×256	3×3
Pool3	4×4×256	2×2

Layer	Size	Kernel
Conv4	2×2×128	3×3
FC	1024	-
Softmax	5	

The CNN used for character type recognition has four convolutional layers. DM kernel is used in the first convolutional layer. Average pooling in done in each of the pooling layer after convolution. Feature are extracted at the fourth convolutional layer using the fully connected layer and passed to Softmax classifier. At the Softmax layer, features are classified to character types. In the CNN model for t type identification, there are 6 neurons in the output of Softmax each corresponding to the type of t pattern. In the CNN model for i type identification, there are 5 neurons in the output of Softmax each corresponding to the type of i pattern.

After the features are extracted, the subset of features and the subset of variables in each feature which are relevant for big five personality trait classification are identified.

To identify the relevancy of features: baselines(F1), margin(F2), line spacing(F3), space between words(F4), 't' patterns(F5) and 'i' patterns(F6) to the Big five personality score, multi linear regression fit is made between the features and the Big five scores in different combinations as Equation 2.

$$y = \beta_0 + \beta_1 x_1 + \ldots + \beta_n x_n \quad (2)$$

Where y is the Big five score, β_0 is the intersect and $\beta_1, \beta_2 \ldots \beta_n$ are the regression coefficients.
There are 63 combinations of features. For each of the combinations f value is calculated as Equation 3.

$$f = \frac{S_R/p}{S_E/(n-(p+1))} \quad (3)$$

Where S_R is regression sum of squares given as Equation 4.

$$S_R = \sum(\hat{y}_i - \bar{Y})^2 \quad (4)$$

S_E is error sum of squares between expected results and MLR result. It is given as Equation 5.

$$S_E = \sum(\hat{y}_i - y_i)^2 \quad (5)$$

The f values for the top five feature combinations are given in *Figure 3*.

The combination of features F1, F2, F3, F4, F5, F6 provides the highest value and this feature combination is used.

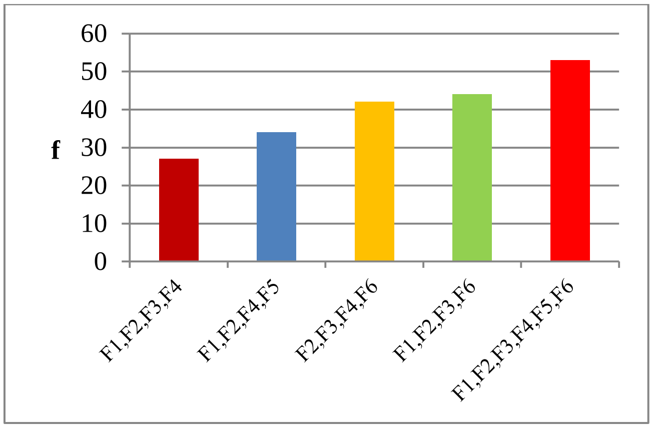

Figure 3 F-Values for feature combinations

From the identified features, to identify the relevant feature values correlation test is done. The correlation between each of the feature values to the class (Big five score) is measured using SU. The SU between the input variable aand output variable b is calculated as Equation 6.

$$SU(a,b) = \frac{2\times MI(a,b)}{H(a)+H(b)} \quad (6)$$

Where $MI(a,b)$ is the mutual information between the variable a andb. $H(a)$ is the entropy for the variable a Mutual information between variable a and b is calculated as Equation 7.

$$MI(a,b) = \sum_a \sum_b PDF(a,b) \log \frac{PDF(a,b)}{p(a)\times p(b)} \quad (7)$$

$PDF(a)$ is the probability density function for the variable a and $PDF(a,b)$ is the joint probability density function. $H(a)$ is calculated in terms of shanon's entropy as Equation 8.

$$H(a) = -\int PDF(a) \log\big(PDF(a)\big) dx \quad (8)$$

SU is calculated for each of the 40 input variables against the Big five scores. The average value of SU is calculated as Equation 9.

$$\alpha = \frac{\sum_{\forall a} SU(a,b)}{N} \quad (9)$$

Where N is the total number of input features (N=40 in this case).

The feature value a is selected as relevant feature when its $SU(a,b) \geq \alpha$. Based on the SU test, a total of 29 features were found to be relevant.

The algorithm for feature selection is given below:

Algorithm 1: Featureselection
Input: Imageset , Big five score set
Ouput: SigAttr
1. FSet ←[]
2. for img in Imageset
Fset.add(Extract F1-F6(img))
3. Score ←0
4. maxcomb ←0
5. for each icombinations of F1 to F6
f ←calcualte f using Equation.3
If f>Score
f=score
maxcomb ←i
6. SUset ←[]
6. For each attribute x in maxcob
SUset[x]=Calculate SU using Equation. 6
7. Th ←mean (SUset)
8. SigAttr ← []
9. For each attribute x in maxcomb
. if SUset[x]>Th
SigAttr.add(x)
10. return SigAttr

The relevant set of feature value forms the feature vector. The value at each place is set to 1 or 0 based on presence of that value as found from feature extraction. A training dataset mapping the binary feature word vector to the Big Five personality trait is prepared. The training dataset is used to train an artificial neural network. The configuration of the neural network for personality prediction is given in *Table 5*. It is three layer feed forward neural network. The neural network is loop controlled based on error loss functions to improve the accuracy of prediction.
The feed forward neural network is used for prediction of personality from the binary feature word vector.

Table 5 Neural network parameters

Parameters	Values
Number of layers	3
Input neurons	29
Hidden layer neurons	57
Output neurons	5
Hidden layer function	Gaussian activation
Output layer function	Linear activation

For a neural network with K neurons in the hidden layer, the output is modelled as Equation 10.

$$y_j^i = \alpha_{j0} + \sum_{k=1}^{K} \propto_{jk} \Phi_k(X^i), j = 1,2 \ldots n \qquad (10)$$

Loop control mechanism is used to optimize the performance of the ANN. Hinge error is considered for fine tuning the performance of ANN. It is calculated as absolute difference between predicted and actual result. The training process is repeated by selective removal of data items till the lowest value for Hinge error is obtained. Three different Hinge loss functions were experimented: Crammer & Singer, Westson & Walkins and Zhang quadratically smoothed.

The accuracy achieved for different Hinge loss functions is given in *Table 6.*

Table 6 Accuracy of hinge loss functions

Hinge loss functions	Accuracy
Crammer & Singer	0.85
Weston &Walkins	0.87
Zhang quadratically smootheed	0.91

The best value of accuracy is achieved with Zhang quadratically smoothed function at 0.91 which is atleast 6% higher compared to Crammer and Singer and 4% higher compared to Weston and Walkins. Thus, Zhang quadratically smoothed function is used as the hinge loss function for the cognitive learning framework.

The algorithm for big five personality classification (ClassifyPersonality) is given below:

Algorithm 2: ClassifyPersonality
Input: handwritten image-img,SigAttr
Ouput: personality class
1. Fvset ←Extract F1-F6(img)
2Fv ←extract selected features (Fvset,SigAttr)
3. PC ←Invoke on ANN(Fv)
4. return *PC*

The proposed solution was implemented in python using openCV, keras, sklearn and tensor flow modules. A total of 200 handwritten document images and their Big five score were collected. The image was first pre-processed and passed to feature extraction. Feature F1- F6 was extracted. The most relevant features and feature attributes were selected. This work identified features combination F1, F2, F3, F4, F5, F6 providing the higher f score and 29 significant attributes in the feature set. The CNN model for character type recognition was implemented using keras module. The ANN for big five personality classification was implemented using sklearn module.

4.Results

Two hundred handwritten documents belonging to different dominant big five personality traits are collected from graphologists. Each of the images was tagged with the dominant big five personality traits. The images are converted to a binary feature word vector of dimension 29, corresponding to document level and character level features. A dataset is created with feature vector vs. dominant personality trait. The dataset is split to 80:20 ratio. The proposed cognitive neural network is trained with 80% training dataset and performance is measured against 20% test set. The performance is measured in terms of personality wise accuracy. For comparing the effectiveness of proposed solution, three recent relevant solutions: Rahman and Halim (2022) [30], Mekhaznia et al. (2021) [7], Chaubey and Arjaria (2022) [29] were used. The accuracy results for different personality types are given in *Table 7*. In comparison to Rahman et al. the proposed solution has 2% higher accuracy. In comparison to Mekhaznia et al. [7] proposed solution had 8% higher accuracy. In comparison to Chaubey and Arjaria [29] proposed solution has 43% higher accuracy. The combined use of document level and character level features along with error loss minimization based loop back control in the proposed solution has increased the accuracy. The proposed solution has consistent higher accuracy for all five personality types. Chaubey and Arjaria [29] used an entire image for personality classification without selection of significant features, due to which accuracy is low. Mekhaznia et al. [7] extracted only edge based features due to which the accuracy is low.

Table 7 Personality prediction accuracy

Personality type	**Rahman and Halim (2022) [30]**	**Mekhaznia et al. (2020) [7]**	**Chaubey and Arjaria (2022) [29]**	**Proposed**
Openness	88.3	78.9	43	89.1
Conscientiousness	80	76.4	46	83
Extraversion	87.4	79	41	89.2
Agreeableness	80	80	40	82
Neuroticism	85.3	77.6	46	88.6
Average	84.2	78.38	43.2	86.38

The receiver operating characteristic curve (ROC) plot comparing the solutions is given in *Figure 4.* Higher the ROC, better is the sensitivity of the classifier. The ROC area is proposed solution is 0.907 which is higher compared to existing solutions. It demonstrates the sensitivity of the proposed solution. The sensitivity has increased in the proposed solution due to effective filtering of irrelevant features and consideration for only highly relevant features for classification.

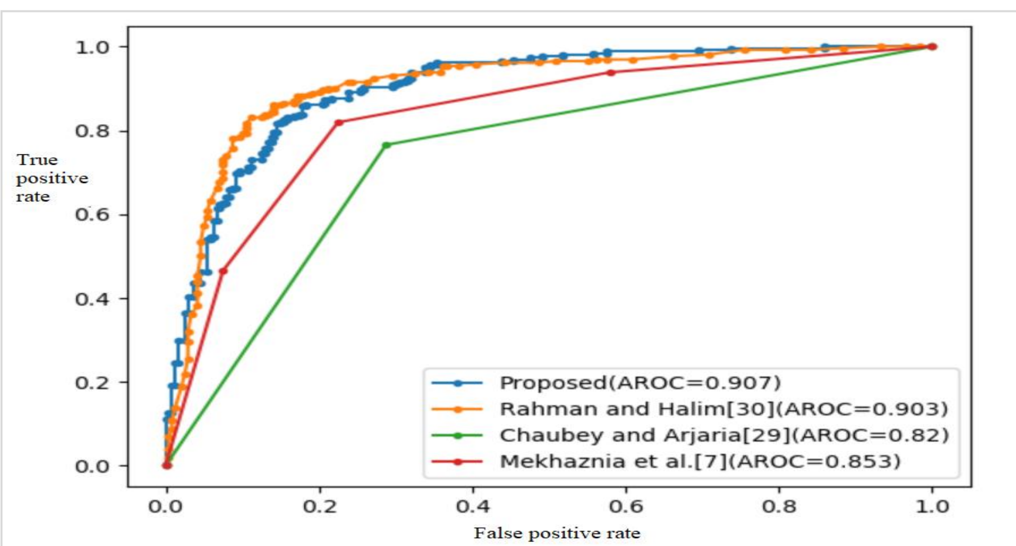

Figure 4 ROC plot

The classification error rate is compared across the solutions and the result is given in *Figure 5*. The error rate is lower in the proposed solution at 0.1 compared to 0.16 in Rahman and Halim [30], 0.21 in Mekhaznia et al. [7] and 0.54 in Chaubey and Arjaria [29]. The error rate has lowered due to the selection of the most relevant features for classification in the proposed solution.

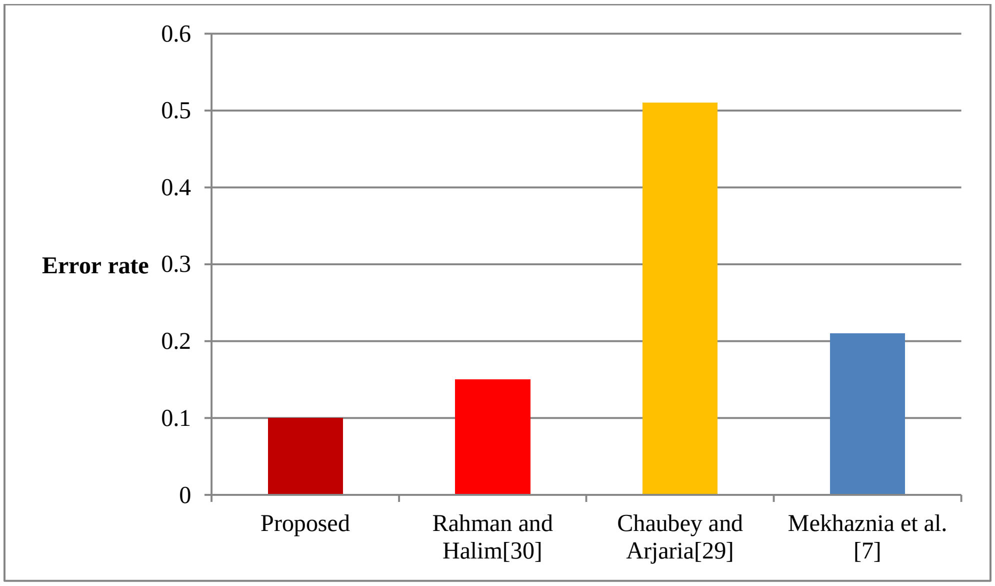

Figure 5 Error rate

The ability of features to diversity the big five personality classes is measured through clustering. The features are clustered using K-mean clustering algorithm with K value as 5 (each corresponding to the big five personality class). Across the cluster three metrics of average cohesion (AC), the average separation (AS) and silhouette coefficient (SC) are calculated. The features are effective when average cohesion value is lower, average separation value is lower and SC value is higher. The AC, AS and SC are calculated as in Equation 11 to 13.

$$AC = \sum_i \sum_{x \epsilon C_i} (x - m_i)^2 \qquad (11)$$

$$AS = \sum_i |C_i| \, (m - m_i)^2 \qquad (12)$$

$$s = \begin{cases} 1 - \frac{a}{b}, & \text{if } a < b \\ \frac{b}{a} - 1 & \text{if } a \geq b \end{cases} \qquad (13)$$

Where m_i is the feature set, x is the cluster centroid, $|C_i|$ is the size of the cluster (number of elements), a is the a is the average distance of i to the points in its cluster and b is minimum of the average distance of i to points in another cluster.

The clustering analysis results across the solutions are given in *Table 8*. The results in *Table 8*, demonstrates the clustering efficiency of features in the proposed solution. The proposed features are able to provide clusters with higher cohesion, better separation and higher SC value. Since the big five classes have higher separation, the classification accuracy is also higher in the proposed solution. A CNN with DM is used for recognizing 't', 'i' and other characters. This CNN is trained with modified national institute of standards and technology database (MINST) dataset [46]. The CNN used for recognition of characters achieved 100% accuracy at epoch of 8 seconds as shown in *Figure 6*. The character recognition has increased in the proposed solution due to use of DM convolution kernel in the first convolutional layer.

Table 8 Clustering analysis results

Metrics	Rahman and Halim (2022) [30]	Mekhaznia et al. (2020) [7]	Chaubey and Arjaria (2022) [29]	Proposed
AC	71.5	80	83	64.2
AS	122.1	112.5	119.2	140
SC	0.72	0.66	0.69	0.81

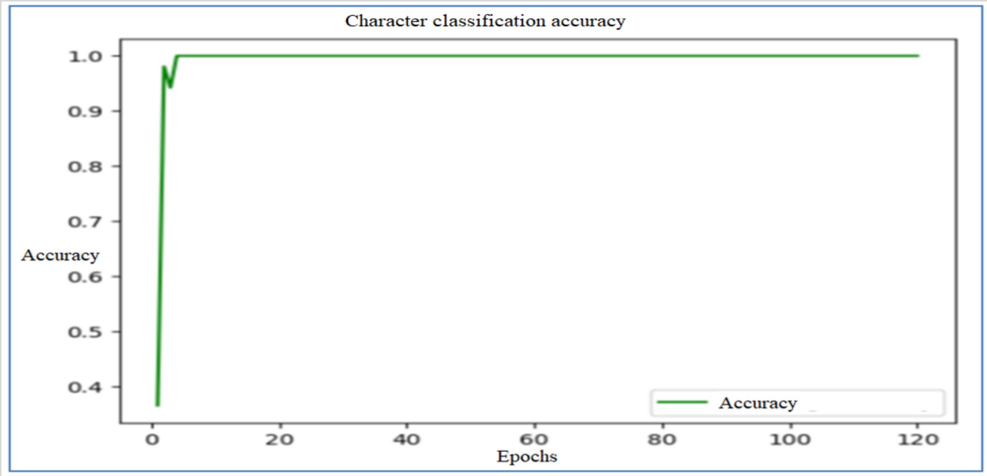

Figure 6 Char recognition accuracy

The CNN with DM convolution is trained for recognition of six different types of 't' for a total of 60 images with 10 images for each class of t. The performance of the CNN with DM convolution and default convolution is compared for 't' type identification and the result is given in *Table 9*. The performance of the CNN model for 't' type identification has increased in the proposed solution. Accuracy has increased by almost 4%, sensitivity of 7% and specificity of 7%. This is due to amplification of significant areas differentiating the 't' types in the first convolution layer with DM convolution.

Table 9 Comparison of performance with DM convolution for t pattern

Performance parameters	**CNN with DM convolution**	**CNN with default convolution**
Accuracy	96.1	92.2
Sensitivity	97.2	90.2
Specificity	95.49	87.26

The performance of the CNN with DM convolution and default convolution is compared for 'i' type identification and the result is given in *Table 10*.

The performance of the CNN model for 'i' type identification has increased in the proposed solution. Accuracy has increased by almost 3%, sensitivity of 6% and specificity of 6%.

This is due to amplification of significant areas differentiating the 'i' types in the first convolution layer with DM convolution.

Table 10 Comparison of performance with DM convolution for i pattern

Performance Parameters	**CNN with DM convolution**	**CNN with default convolution**
Accuracy	95.1	92.5
Sensitivity	96.2	90.5
Specificity	94.61	88.16

The type of 't' and 'i' pattern recognition with the proposed CNN model is tested for effectiveness in terms of: precision, recall and accuracy. For comparison of the proposed CNN model, Zonal feature method by Asra and Shubhangi (2017) [12] and template matching method by Gavrilescu and Vizireanu (2018) [9] were used, *Table 7*. The precision is measured for a total of 50 't' and 'i' pattern's images with 10 images in each type. The result is given in *Figure 7*. The proposed CNN model has at least 5.7% higher precision compared to exiting work for 't' type recognition and at least 3.5% higher precision for 'i' type recognition. Compared to template based by Asra and Shubhangi [12] zone-based features by Gavrilescu and Vizireanu [9], the proposed model was able to learn more intricate features using CNN with extended convolution function and this has increased the precision in the proposed solution.

The recall is measured for a total of 50 't' and 'i' patterns' images with 10 images in each type. The result is given in *Figure 7*. The proposed CNN model has at least 3.7% higher recall for 't' type recognition and it has at least 1.5% higher precision compared to 'i' type recognition.

The recall difference is very low for 'i' type recognition among the solutions compared. To improve it, handcrafted features can be combined with deep learning features and used in hybrid mode. The use of extended convolution operation enabled more intricate learning from 't' and 'i' patterns and this has increased the recall in the proposed solution while Asra and Shubhangi [12], Gavrilescu and Vizireanu [9] could not learn these intricate features (*Figure 8*).

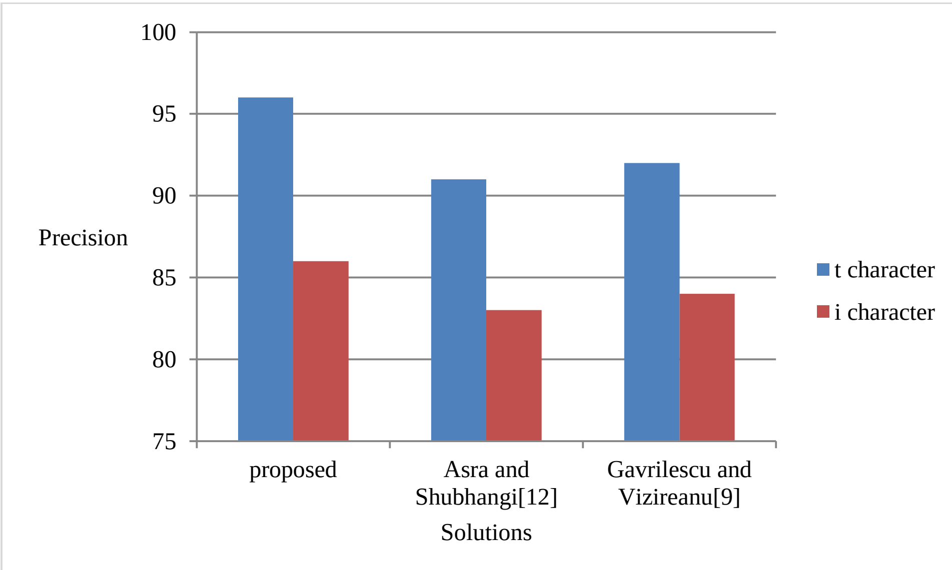

Figure 7 Comparison of precision

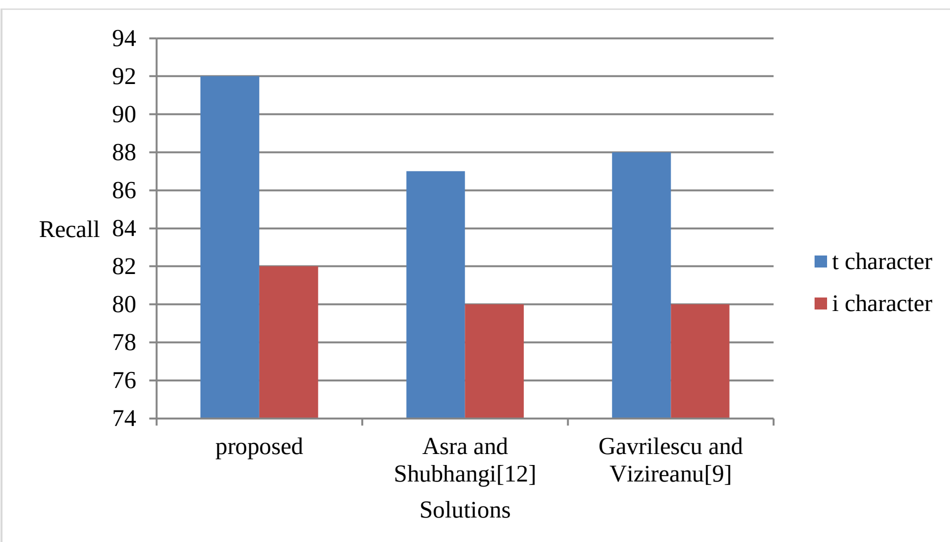

Figure 8 Comparison of recall

Accuracy is measured for a total of 50 't' and 'i' patterns' images with 10 images in each type and the result is given in *Figure 9*. Accuracy in proposed CNN model is at least 4.7% higher for 't' type recognition and at least 3.2% higher for 'i' type recognition. The accuracy is about 96% in proposed solution for 't' characters and it is only 91% in proposed solution for 'i' characters. This difference is due to higher false positives in 'i' pattern. Use of CNN with extended convolutional operation has allowed for more feature discriminating ability in the proposed solution and this has increased the accuracy compared to Asra and Shubhangi[12], Gavrilescu and Vizireanu [9].

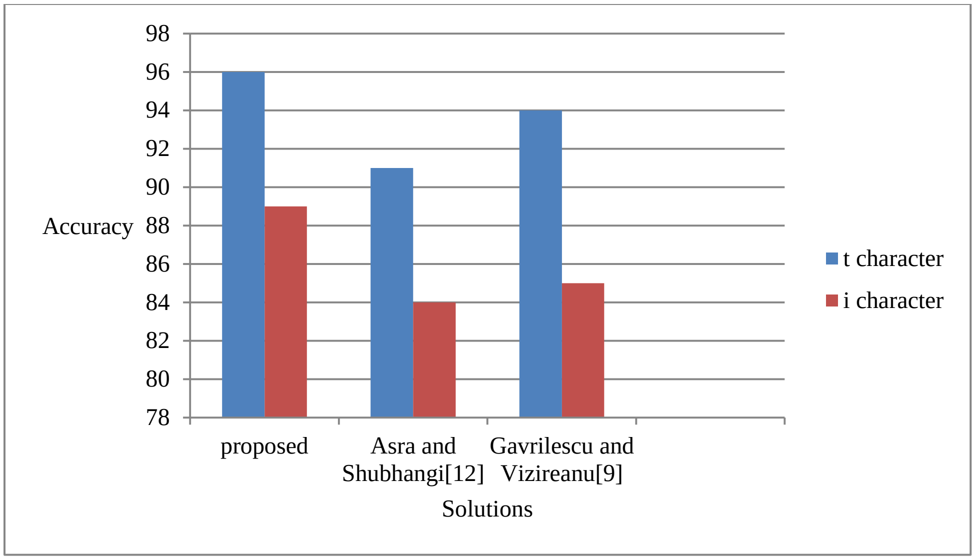

Figure 9 Comparison of accuracy

5.Discussion

The experimental results prove the effectiveness of the proposed solution in classifying the Big five personalities. Feature engineering is the contributor to better performance in the proposed solution. Features of both document and character level are used in the proposed solution. Rahman and Halim (2022) [30] constructed a graph from the entire handwritten document image and used for classification. Characters which are insignificant for personality detection also become a feature in this approach. This affected the classification accuracy. Mekhaznia et al. [7] (2021) used only edge statistical features like direction and hinge. The features are only character level and it missed document level features like spaces, margin etc. which are found to have significant influence on personality traits. Chaubey and Arjaria (2022) [29] used the entire image to learn deep features. Attention was not given to areas of an image like significant characters influencing the personality. But in the proposed solution, only highly relevant features were used for classification. High relevant features were selected in two stages. In the first stage, various features combinations were scored using f score and the feature combination with highest f score is selected. In the second stage, most important attributes in the features are selected using SU based entropy metric. By these two stages, only best set of features and their attributes are used for classification. This attention given to personality influencing factors has increased the classification accuracy in the proposed solution.

Another important contribution in the proposed solution is that DM kernel in the convolutional layer of CNN for character type (t and i) identification. There were subtle differences between different types of t and i patterns. The CNN model was adapted from works of Weng and Xia [47]. The CNN model of Weng and Xia [47] failed to provide attention to subtle changes in same character. This problem is solved in the proposed CNN model using the DM kernel mask which is constructed based on the LBP of character patterns. The mask provided attention to regions of subtle differences. This has increased the character type recognition accuracy compared to Zonal approach proposed by Asra and Shubhangi (2017) [12] and template-based matching proposed by Gavrilescu and Vizireanu (2018) [9].

Limitations

There are other personality influencing characters like f, s etc. This work has limited to only two characters (t and i) which is considered in many works. The work was tested only for 200 handwritten documents but testing against large datasets with diverse handwriting styles is needed. There were no standard datasets for graphology based big data personality classification and this limited the testing scope in this work.

A complete list of abbreviations is shown in *Appendix I.*

6.Conclusion

A big five personality classification system using graphology features has been proposed in this work.

As part of the work five different feature extraction algorithms are proposed to extract document level and character level features. Deep learning is used to extract character level features. The features are encoded into a binary feature word vector. Feedback controlled cognitive neural network is trained to predict the personality from the binary feature word vector. The proposed solution was able to classify the big five personality scales from graphology features with an average accuracy of 86%. Compared to earlier works the proposed solution was able to achieve 2% higher accuracy. In future CNN models for f and s character type recognition can also be added to the feature set. Also testing can be conducted across large and diverse dataset.

Acknowledgment
The lab facilities Department of CSE, ATME College of engineering, Mysore have been used extensively for the experiment

Conflicts of interest
The authors have no conflicts of interest to declare.

Author's contribution statement
Lakshmi Durga: Literature review, conceptualization, investigation, creation of graphology dataset, design, analysis of results. **Deepu. R:** Supervision, review and editing.

References
[1] Durga L, Deepu R. Handwriting analysis through graphology: a review. In international conference on advances in computing, communications and informatics 2018 (pp. 1160-6). IEEE.
[2] John OP, Srivastava S. The big-five trait taxonomy: history, measurement, and theoretical perspectives.
[3] Plamondon R. Neuromuscular studies of handwriting generation and representation. In 12th international conference on frontiers in handwriting recognition 2010. IEEE.
[4] Rahman AU, Halim Z. Identifying dominant emotional state using handwriting and drawing samples by fusing features. Applied Intelligence. 2022:1-17.
[5] Aggarwal A, Varma PS, Sehgal L, Shah KP, Gupta GU, Ranjan S. Detection of personality using machine learning. International Journal of Research in Engineering, Science and Management. 2022; 5(1):46-50.
[6] Fallah B, Khotanlou H. Identify human personality parameters based on handwriting using neural network. In artificial intelligence and robotics 2016 (pp. 120-6). IEEE.
[7] Mekhaznia T, Djeddi C, Sarkar S. Personality traits identification through handwriting analysis. In Mediterranean conference on pattern recognition and artificial intelligence 2020 (pp. 155-69). Springer, Cham.
[8] Mutalib S, Rahman SA, Yusoff M, Mohamed A. Personality analysis based on letter 't' using back propagation neural network. In proceedings of the international conference on electrical engineering and informatics institut teknologi bandung, Indonesia. 2007 (pp. 17-9).
[9] Gavrilescu M, Vizireanu N. Predicting the big five personality traits from handwriting. EURASIP Journal on Image and Video Processing. 2018; 2018(1):1-17.
[10] Mishra PK, Abidi AI, Mishra GS. Improved methodology for personality assessment using handwritten documents. Journal of Positive School Psychology. 2022:3263-73.
[11] Mishra A. Forensic graphology: assessment of personality. Forensic Research & Criminology International Journal. 2017; 4(1):9-12.
[12] Asra S, Shubhangi DC. Human behavior recognition based on hand written cursives by SVM classifier. In international conference on electrical, electronics, communication, computer, and optimization techniques 2017 (pp. 260-8). IEEE.
[13] Champa HN, Anandakumar KR. Artificial neural network for human behavior prediction through handwriting analysis. International Journal of Computer Applications. 2010; 2(2):36-41.
[14] Rahiman A, Varghese D, Kumar M. Habit: handwritten analysis based individualistic traits prediction. International Journal of Image Processing. 2013; 7(2):209-18.
[15] Fisher J, Maredia A, Nixon A, Williams N, Leet J. Identifying personality traits, and especially traits resulting in violent behavior through automatic handwriting analysis. Proceedings of Student-Faculty Research Day, CSIS, Pace University. 2012.
[16] Prasad S, Singh VK, Sapre A. Handwriting analysis based on segmentation method for prediction of human personality using support vector machine. International Journal of Computer Applications. 2010; 8(12):25-9.
[17] Grewal PK, Prashar D. Behavior prediction through handwriting analysis. International Journal of Computer Science and Technology. 2012; 3(2):520-3.
[18] Coll R, Fornés A, Lladós J. Graphological analysis of handwritten text documents for human resources recruitment. In international conference on document analysis and recognition 2009 (pp. 1081-5). IEEE.
[19] Mukherjee S, De I. Feature extraction from handwritten documents for personality analysis. In international conference on computer, electrical & communication engineering 2016 (pp. 1-8). IEEE.
[20] Joshi P, Agarwal A, Dhavale A, Suryavanshi R, Kodolikar S. Handwriting analysis for detection of personality traits using machine learning approach. International Journal of Computer Applications. 2015; 130(15):40-5.
[21] Kacker R, Maringanti HB. Personality analysis through handwriting. GSTF Journal on Computing. 2014; 2(1):94-7.

[22] Mutalib S, Ramli R, Rahman SA, Yusoff M, Mohamed A. Towards emotional control recognition through handwriting using fuzzy inference. In international symposium on information technology 2008 (pp. 1-5). IEEE.
[23] Wijaya W, Tolle H, Utaminingrum F. Personality analysis through handwriting detection using android based mobile device. Journal of Information Technology and Computer Science. 2017; 2(2):114-28.
[24] Chitlangia A, Malathi G. Handwriting analysis based on histogram of oriented gradient for predicting personality traits using SVM. Procedia Computer Science. 2019; 165:384-90.
[25] Pratiwi D, Santoso GB, Saputri FH. Personality type assessment system by using enneagram-graphology techniques on digital handwriting. International Journal of Computer Applications. 2016; 147(11):9-13.
[26] Majumder N, Poria S, Gelbukh A, Cambria E. Deep learning-based document modeling for personality detection from text. IEEE Intelligent Systems. 2017; 32(2):74-9.
[27] Lokhande VR, Gawali BW. Analysis of signature for the prediction of personality traits. In 1st international conference on intelligent systems and information management 2017 (pp. 44-9). IEEE.
[28] Hashemi S, Vaseghi B, Torgheh F. Graphology for Farsi handwriting using image processing techniques. IOSR Journal of Electronics and Communication Engineering. 2015; 10(3):1-7.
[29] Chaubey G, Arjaria SK. Personality prediction through handwriting analysis using convolutional neural networks. In proceedings of international conference on computational intelligence 2022 (pp. 59-70). Springer, Singapore.
[30] Rahman AU, Halim Z. Predicting the big five personality traits from hand-written text features through semi-supervised learning. Multimedia Tools and Applications. 2022;81:33671-87.
[31] Anari MS, Rezaee K, Ahmadi A. TraitLWNet: a novel predictor of personality trait by analyzing Persian handwriting based on lightweight deep convolutional neural network. Multimedia Tools and Applications. 2022; 81(8):10673-93.
[32] Samsuryadi RK, Mohamad FS. Automated handwriting analysis based on pattern recognition: a survey. Indonesian Journal of Electrical Engineering and Computer Science. 2021; 22(1):196-206.
[33] Pathak AR, Raut A, Pawar S, Nangare M, Abbott HS, Chandak P. Personality analysis through handwriting recognition. Journal of Discrete Mathematical Sciences and Cryptography. 2020; 23(1):19-33.
[34] Nolazco-flores JA, Faundez-zanuy M, Velázquez-flores OA, Cordasco G, Esposito A. Emotional state recognition performance improvement on a handwriting and drawing task. IEEE Access. 2021; 9:28496-504.
[35] Elngar AA, Jain N, Sharma D, Negi H, Trehan A, Srivastava A. A deep learning based analysis of the big five personality traits from handwriting samples using image processing. Journal of Information Technology Management. 2020:3-35.
[36] Impedovo D, Pirlo G. Dynamic handwriting analysis for the assessment of neurodegenerative diseases: a pattern recognition perspective. IEEE Reviews in Biomedical Engineering. 2018; 12:209-20.
[37] Gahmousse A, Gattal A, Djeddi C, Siddiqi I. Handwriting based personality identification using textural features. In international conference on data analytics for business and industry: way towards a sustainable economy 2020 (pp. 1-6). IEEE.
[38] Valdez-rodríguez JE, Calvo H, Felipe-riveron EM. Handwritten texts for personality identification using convolutional neural networks. In international conference on pattern recognition 2018 (pp. 140-5). Springer, Cham.
[39] Lemos N, Shah K, Rade R, Shah D. Personality prediction based on handwriting using machine learning. In international conference on computational techniques, electronics and mechanical systems 2018 (pp. 110-3). IEEE.
[40] Kedar S, Bormane DS. An approach to predict hypertension based on handwritten manuscript. Indian Journal of Public Health Research & Development. 2018; 9(11): 2235-40.
[41] Fatimah SH, Djamal EC, Ilyas R, Renaldi F. Personality features identification from handwriting using convolutional neural networks. In 4th international conference on information technology, information systems and electrical engineering 2019 (pp. 119-24). IEEE.
[42] Costa EP, Villaseñor-pienda L, Morales E, Escalante HJ. Recognition of apparent personality traits from text and handwritten images. In international conference on pattern recognition 2018 (pp. 146-52). Springer, Cham.
[43] https://openpsychometrics.org/tests/IPIP-BFFM/. Accessed 30 June 2022.
[44] Deepu R, Murali S, Raju V. A mathematical model for the determination of distance of an object in a 2D image. In proceedings of the international conference on image processing, computer vision, and pattern recognition 2013 (pp.1-5).
[45] https://books.google.co.in/books/about/The_Little_Giant_Encyclopedia_of_Handwri.html?id=ADsOAAAACAAJ&redir_esc=y. Accessed 30 June 2022.
[46] https://www.nist.gov/itl/products-and-services/emnist-dataset. Accessed 30 June 2022.
[47] Weng Y, Xia C. A new deep learning-based handwritten character recognition system on mobile computing devices. Mobile Networks and Applications. 2020; 25(2):402-11.

Lakshmi Durga is a research scholar at MITM college, Mysore. She has completed her Mtech and has around 25 years of teaching experience. She has published 5 papers in the international journals in the area of image processing and machine learning. Her expertise is Image Processing, Pattern Recognition, Machine Learning and Deep Learning.
Email: lakshmidurga0812@gmail.com

Dr. Deepu. R, is currently working as Professor in Department of Computer Science & Engineering and Head of Department of Career Guidance & Placements at ATME College of Engineering, Mysore. He has 22 Years of experience in Teaching, Research, Administration & Industry. He had completed his Ph.D in Computer Science from University of Mysore in 2016. He has done M.Tech in Computer Networks Engineering from NIE, Mysore and B.E in Computer Science and Engineering from SJCE, Mysore. He has received funding for his research proposals from various funding agencies such as VTU Belagavi, KSCST etc. He is an active member in various associations such as The Institution of Engineers (I), IEEE, Computer Society of India, ISTE, Karnataka Rajya Vignana Parishath etc. He is an elected member of The Institution of Engineers (I) Mysore centre from computer engineering division for 2018-2021. He is a senior member of IEEE, Executive Council member in Computer Society of India, Bangalore Region V and Chairman of ACM Mysuru professional Chapter, the only professional chapter of ACM in Karnataka State. He is a lifetime member of the International Association of Engineers, Hong Kong.
Email: rdeepusingh21@gmail.com

Appendix I

S. No.	Abbreviation	Description
1	AC	Average Cohesion
2	ANN	Artificial Neural Network
3	AS	Average Separation
4	CNN	Convolutional Neural Network
5	DCNN	Deep Convolutional Neural Network
6	DM	Differences-Magnifier
7	HOG	Histogram of Oriented Gradients
8	KNN	K-Nearest Neighbors
9	LBP	Local Binary Pattern
10	MINST	Modified National Institute of Standards And Technology Database
11	MMPI	Minnesota Multiphasic Personality Inventory
12	ROC	Receiver Operating Characteristic Curve
13	SC	Silhouette Coefficient
14	SU	Symmetric Uncertainty
15	SVM	Support Vector Machine