

Personality Prediction based on Handwriting using CNN & MLP

Achinthu Haridas

Computer Science & Engineering
TKM Institute of Technology (Approved by AICTE and affiliated to APJ Abdul Kalam Technological University)
Karuvellil, Kollam, Kerala, India

Sravan S

Computer Science & Engineering TKM Institute of Technology (Approved by AICTE and affiliated to APJ Abdul Kalam Technological University)
Karuvellil, Kollam, Kerala, India

Arjun R

Computer Science & Engineering TKM Institute of Technology (Approved by AICTE and affiliated to APJ Abdul Kalam Technological University)
Karuvellil, Kollam, Kerala, India

Mruthula N R

Computer Science & Engineering TKM Institute of Technology (Approved by AICTE and affiliated to APJ Abdul Kalam Technological University)
Karuvellil, Kollam, Kerala, India

Rohith Muralidharan

Computer Science & Engineering
TKM Institute of Technology (Approved by AICTE and affiliated to APJ Abdul Kalam Technological University)
Karuvellil, Kollam, Kerala, India

Abstract- Handwriting is exclusive to each person as much as fingerprint is exclusive. Someone can imitate another person’s handwriting for less than a few words creating it distinctive. Various forms of handwriting styles are taken into thought like; slope, baseline, top margin, word size, line spacing, word spacing, left or right or normal slant or irregular of the sentence, etc. The complete system evaluates the handwriting samples based on the above-mentioned handwriting styles. It is divided into three modules, i.e.; Pre-processing, Feature Extraction and CNN-MLP combined model.

Keywords— Handwriting samples, baseline, top margin, word size, line spacing, word spacing, left/right/normal slant/irregular, Pre-processing, Feature extraction, CNN-MLP combined model.

I. INTRODUCTION

The very minute patterns produced in handwriting can reveal a person’s personality traits. The study of handwriting styles where the handwriting expert examines the handwriting sample and checks for numerous trails present in the sample and predicts the personality trait present in the handwriting sample. This paper focuses on analyzing various style of handwriting like ascending, descending and straight inclination of the baseline of a sentence, Spacing between words, the right, left, irregular slants which will be used in identifying a set of traits associated with the person, if present.

II. AREA OF IMPLEMENTATION

Deep learning is a subset of machine learning where Artificial neural networks, algorithms inspired by the human brain, learn from large amounts of data. Deep learning allows machines to solve complex problems even when using a data set that is very diverse,

unstructured and inter-connected. This is an application of Graphology too. Existing method involves identifying various personality based on handwritten characters by applying CNN. In the existing method only images of baseline, slant and word space are used. The proposed work is divided into several modules. In the first module image of handwriting is taken from the Datasets. In the second module pre-processing on the image is done by using various technique to remove noise and to smoothen the image for better results. In the third module; feature extraction i.e. extraction of the seven features are performed. In the last module, for better accurate result the images along with the extracted feature is passed to a combined model consisting of CNN and MLP. By observing probabilities of different output labels, we can predict the most dominant class to which the test handwriting sample belongs or the personality traits that person has.

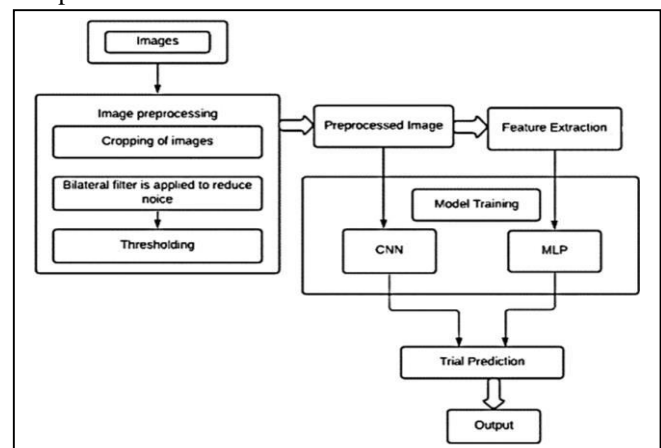


Fig. 1. Methodology for personality prediction based on hamdwriting using MLP & CNN

III. MATERIALS & METHODS

The dataset are of two types, i.e. Dataset-1 contains 1533 images of the scanned text of handwriting. These images are cropped and saved as PNG images to avoid the printed text in the images. On analysing IAM dataset it is found that slight variation of features were observed so it may affect accuracy so a new dataset is formed, Dataset-2. Dataset-2 is formed in order to compare with the standard Dataset-1. Important dataset samples are shown below;

i. Baseline

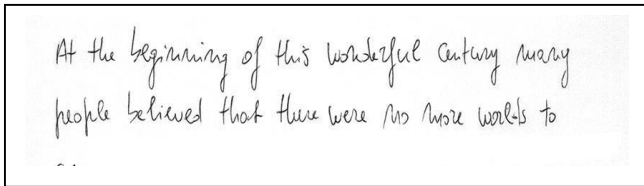


Fig. 2. Handwriting sample with same baseline angle.

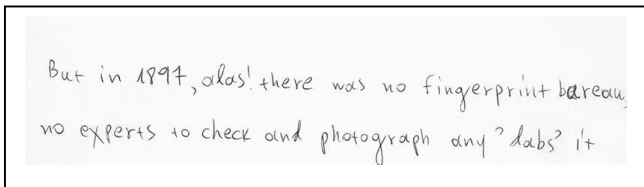


Fig. 3. Handwriting sample with falling baseline angle.

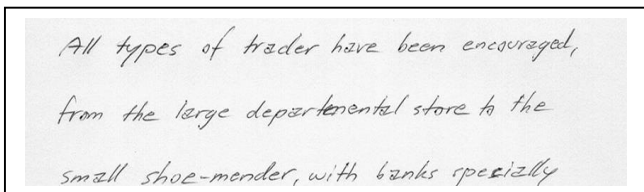


Fig. 4. Handwriting sample with rising baseline angle.

ii. Letter Size

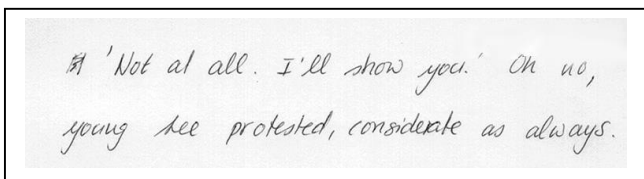


Fig. 4. Handwriting sample with average letter size.

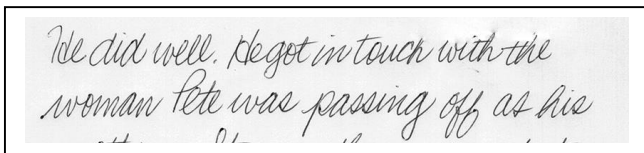


Fig. 5. Handwriting sample with large letter size.

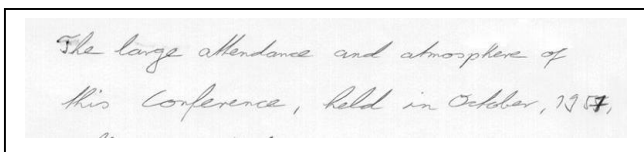


Fig. 6. Handwriting sample with small letter size.

iii. Line spacing

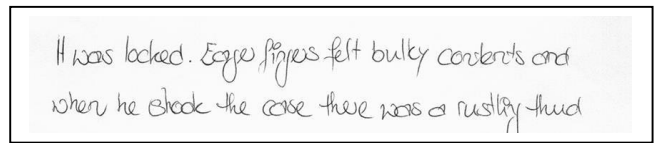


Fig. 7. Handwriting sample with normal line spacing.

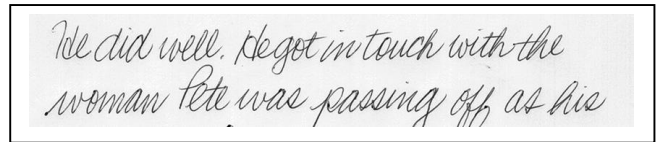


Fig. 8. Handwriting sample with small line spacing

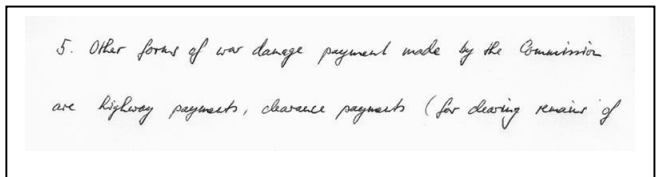


Fig. 9. Handwriting sample with large line spacing.

iv. Word spacing

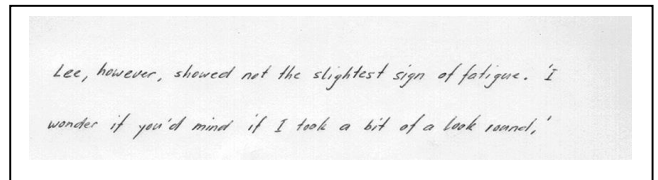


Fig. 10. Handwriting sample with normal word spacing.

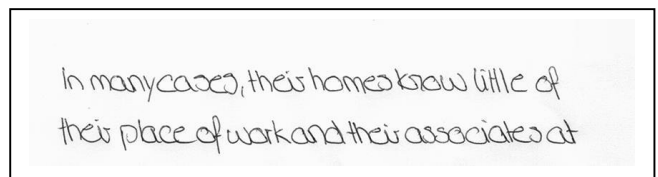


Fig. 11. Handwriting sample with narrow word spacing.

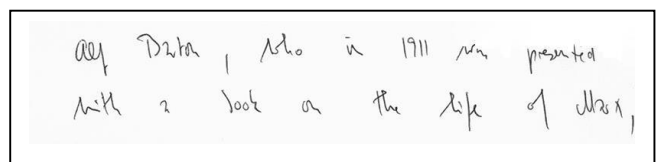


Fig. 12. Handwriting sample with wide word spacing.

v. Top margin

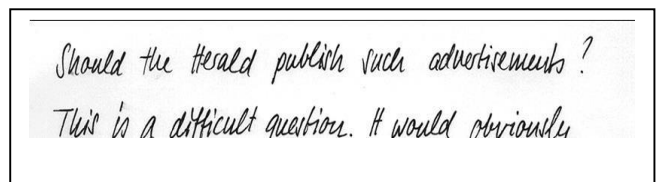


Fig. 13. Handwriting sample with narrow top margin.

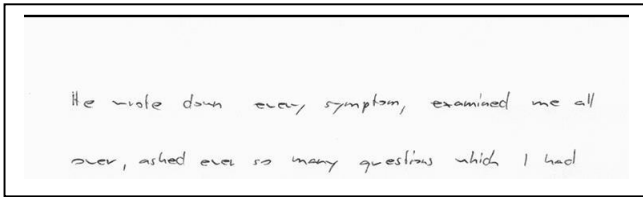


Fig. 14. Handwriting sample with wide top margin.

vi. Pen pressure

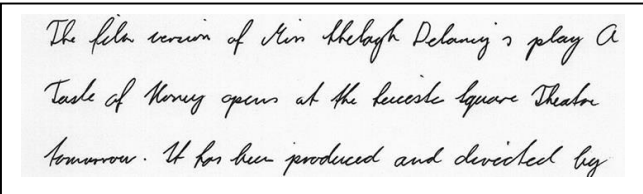


Fig. 15. Handwriting sample with heavy pen pressure.

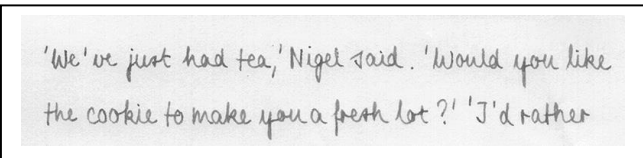


Fig. 16. Handwriting sample with light pen pressure.

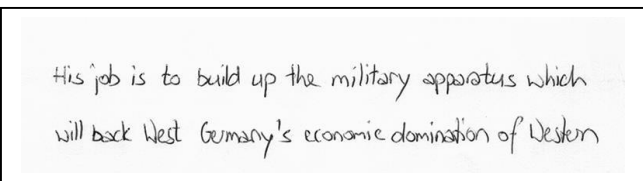


Fig. 17. Handwriting sample with medium pen pressure.

vii. Slant of letters

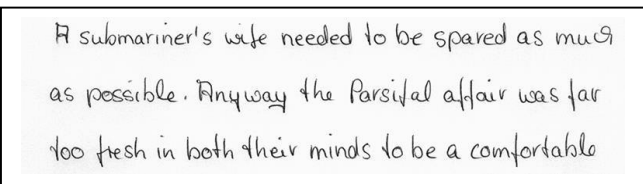


Fig. 18. Handwriting sample with vertical slant.

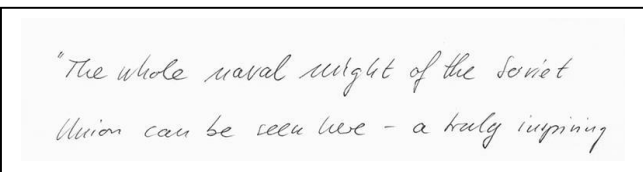


Fig. 19. Inclined or right slanted writing.

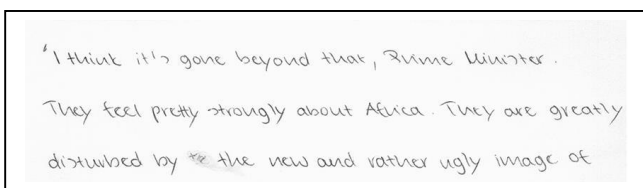


Fig. 20. Reclined or left slanted handwriting.

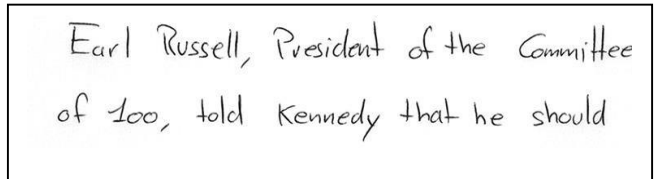


Fig. 21. Irregular or unstable slanted handwriting

IV. ALGORITHMS

CNN (Conventional Neural Network) is a powerful set of techniques used in learning of neural networks. CNN is a class of machine learning algorithms that uses multiple layers to progressively extract higher level features from the raw input. For example, in image processing, lower layers may identify the concepts relevant to a human such as digits or letter or faces. MLP (Multilevel Perceptron) is a class of feed forward artificial neural network. An MLP consist of at least three layers of nodes named; input layer, a hidden layer and an output layer. Except for the input node is a neuron that uses a nonlinear activation function. And MLP utilizes a supervised learning techniques called back propagation for training.

V. PRE-PROCESSING

- 1) Image cropping (Only for Dataset-1)
- 2) Noise removal
- 3) Gray scaling & Binarization
- 4) Contour & Warp affine transformation

Pre-processing an image is an important step in graphology, and cropping and noise removal are primary treatment. And then, Grayscale and binarization takes place, where an inverted binary image function is constructed such that those pixels above specific threshold (foreground) are converted to 255 and below the threshold are (background) converted to 0. Horizontal projection of an image is python list of sum of all the pixel values of each row of the image. Vertical projection is also a python list of sum of all the pixel values of each column of the image. Both the operations are performed on gray scaled images, and then proceeded to the next process named contour & warp affine transformation. A contour is closed curve of points or line segments, representing the boundaries of an object in an image. In other words, contours represent the shapes. Warp affine transformation is applied to rotate the contour found on an image so that the baseline of the handwriting is strictly horizontal.

VI. CNN & MLP MODELLING

Seven features are extracted from the pre-processed handwriting image, and are saved to a csv file so that we can use this file for further evaluation in the CNN model. After feature extraction, baseline, word space and slant are the three handwriting features we used to compare as the output in both models. In MLP all the features are taken to compare and in CNN only the pre-processed image are given as input and finally the output of both CNN and MLP are concatenated. Here for comparing personality of each person is matched

with their corresponding handwriting feature. Secondary personalities are predicted based on the seven extracted features and primary personalities are predicted based on combined CNN and MLP model.

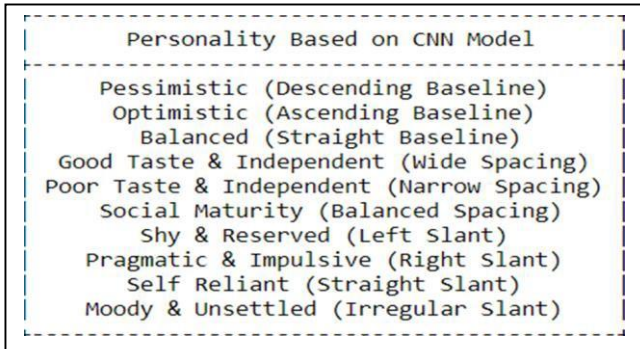


Fig. 22. Primary personality predictions based CNN and MLP models.

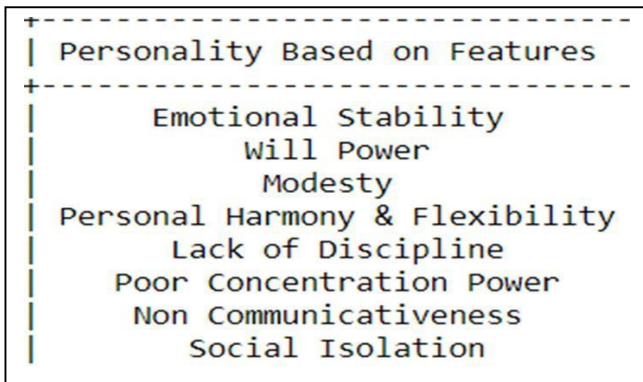


Fig. 23. Secondary personality predictions based on extracted features.

VII. FUTURE ENHANCEMENT

Personality prediction based on handwriting is an interesting topic to dig in and many more enhancement apart from this can be introduced to this particular idea. In the future Advanced CNN models like Presnet can be used, and we may be able to predict the personality of a person even by reading each letter instead of a group words or sentences.

VIII. CONCLUSION

Handwriting is exclusive to each person as much as fingerprint is exclusive. Someone can imitate another person’s handwriting for less than a few words creating it distinctive. In this project, personalities or traits are detected by performing a handwriting analysis on the input image. The system compares the input image and features with the combined model. The key feature of this project is exacting all the possible traits using combined model (CNN and MLP).

ACKNOWLEDGMENT

The above mentioned research is an outcome of a work carried by group of students of TKM Institute of Technology, Karuvellil, Kollam, Kerala, India, under the vision and guidance of Ms. Mruthula NR, Assistant Professor, TKM Institute of Technology, Karuvellil, Kollam, Kerala, India.

REFERENCES

- [1] Subham Nagar, Sudiksha Chakraborty, Arka Sengupta, Joymallya Maji, and Rajib Saha, "An efficient method for character analysis using space in handwriting image". In 2016 Sixth International Symposium on Embedded Computing and System Design (ISED), pages 210–216. IEEE, 2016.
- [2] Behnam Fallah and Hassan Khotanlou, "Identify human personality parameters based on handwriting using neural network". In 2016 Artificial Intelligence and Robotics (IRANOPEN), pages 120–126. IEEE, 2016.
- [3] Anamika Sen and Harsh Shah, "Automated handwriting analysis system using principles of graphology and image processing". In 2017 International Conference on Innovations in Information, Embedded and Communication Systems (ICIIECS), pages 1–6. IEEE, 2017.
- [4] Afnan H Garoot, Maedeh Safar, and Ching Y Suen, "A comprehensive survey on handwriting and computerized graphology". In 2017 14th IAPR International Conference on Document Analysis and Recognition (ICDAR), volume 1, pages 621–626. IEEE, 2017.
- [5] Anupam Varshney and Shalini Puri, "A survey on human personality identification on the basis of handwriting using ann". In 2017 international conference on inventive systems and control (ICISC), pages 1–6. IEEE, 2017.
- [6] Vasundhara Bhade and Trupti Baraskar, "A model for determining personality by analyzing off-line handwriting". In Advances in Machine Learning and Data Science, pages 345–354. Springer, 2018.