



# Personality Trait Analysis by Graphology Technique using Machine Learning

Ranjith R, Jothi S, Chandrasekar A

**Abstract:** Graphology is a technique for analyzing an individual's personality based on the given sample of their handwriting. Graphology considers the fact that the movement of our hand is directly related to the state of our brain. This paper implements the techniques of graphology using machine learning. The pre-defined data about the analysis of handwriting is given as learning data set for the machine. An image of a handwritten text document in jpg format is given as input to this machine. The processed output gives the personality of the individual corresponding to the input handwriting.

**Keywords:** Graphology, SVM algorithm, personality, analysis, sample of handwriting, pre-defined data set.

## I. INTRODUCTION

Every individual has a unique personality some of which are comprehended on the basis of the books they read, the songs they hear and the kind of music taste they have. It is a lesser known fact that handwriting conveys your personality. A simple evidence to this statement is that the human brain controls all activities done by a human and the same brain helps to write. Thus, handwriting is used as a medium for studying the brain's behaviour. This is called as graphology. The physical, mental and emotional state of an individual influences their handwriting. Graphology does not focus on the content that has been written but by the way they are written. There are several factors based on which graphology works. Some of them are as follows:

### A. Margin and slant

Under this category, the spacing of the paragraph along with the four margins and the slant of letters, whether it is left or right, is considered. The variation of spacing from all the marginal sides plays a major role in determining their personality.

### B. Pen Pressure

Every person has a different level of exerting pressure on a pen as they write. The pen pressure can be light, medium, heavy, distinct, lateral, pasty, sharp and so on. Thus, varying levels of darkness are obtained based on the pen pressure. This can be used to determine the energy level and ego of the person.

### C. Size and Baseline

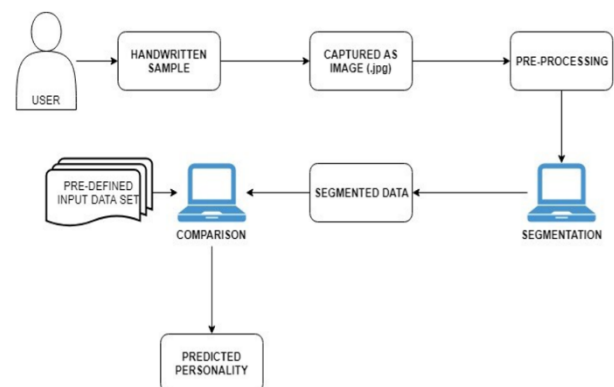
The size of the letters indicates the self-esteem that a person possesses in oneself. On the other hand, the baseline

of their handwriting indicates the disposition of the person, that is, the organizational capabilities that a person holds.

### D. Features of letters

Each letter has a particular feature such as looping, curves, arcs, strokes etc., This paper focuses on many such features of handwriting to detect the personality of the individual.

## BLOCK DIAGRAM



## ALGORITHM

The requirements for the machine are as follows,

### E. Pre-defined data set

The algorithm works on the basis of a pre-defined data set. The pre-defined data set consists of different ways of representing a letter. For each way of representation there is an associated personality trait. The sample data set is as follows,

s.no	image	shape	Personality
1		Without initial stroke	Simple personality
2		With long, straight initial stroke	Firm principles, rigor, good abilities and aggressiveness. May be inappropriate and tiring.
3		Long, straight, disconnected initial stroke	Oppositionist. Inclination towards disagreement and oratory.
4		Initial stroke begins in the superior zone	Rebels against parental authority, introversion, possessive instincts, conservative. Objections against religion.

The pre-defined data set consists of various letters and their personality traits.

Revised Manuscript Received on November 30, 2019.

\* Correspondence Author

**Ranjith R\***, Assistant Professor, St. Joseph's College of Engineering, Chennai, India,

**Jothi S**, Associate Professor, St. Joseph's College of Engineering, Chennai, India,

**Chandrasekar A**, Professor, St. Joseph's College of Engineering Chennai, India

© The Authors. Published by Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP). This is an [open access](https://creativecommons.org/licenses/by-nc-nd/4.0/) article under the CC-BY-NC-ND license <http://creativecommons.org/licenses/by-nc-nd/4.0/>

## F. Segmentation of input data

Segmentation of the input data has to be done. This is done in order to distinguish between white and black spaces present in the image. The image consists of one or more lines written by the individual whose character has to be determined. The segmentation part analyses the margin space and baseline of the given input. It helps to recognize the position of the letter in the given image and focusses on it.

## G. Python Compiler

The machine is coded in python. Image processing has to be done in order to extract the individual letters in the given word. This processed image produces multiple images each consisting of a single letter. This is used for comparison with the pre-defined input data. To do this, the system requires the following packages

- Cv2
- NumPy

The cv2 package of OpenCV helps to read an image as the input. This image is stored in an object and converted to a two-dimensional array for processing purposes. Each array element indicated the pixel density in that region. An array

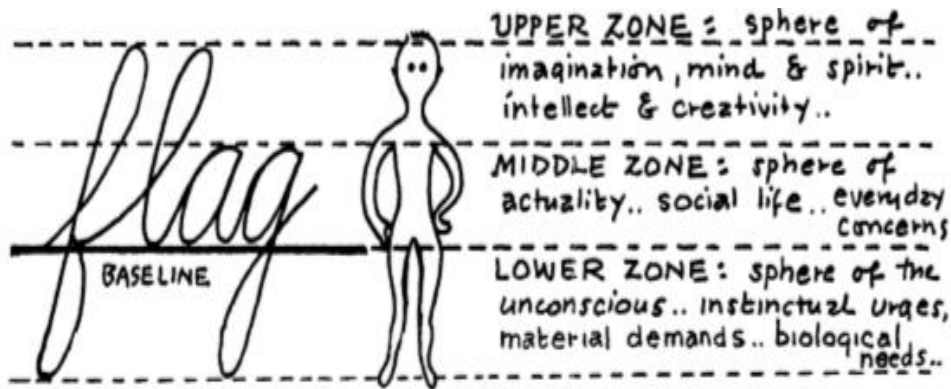
that has to be analyzed is captured as an image and given as input to the compiler.

The output of the program is a result of the analysis of the input image. It describes the personality traits of the individual whose handwriting was given as input. The accuracy of this result is dependent on the quality of the input data.

## REPRESENTATION OF CHARACTERS

The handwritten text of the individual who has to be analysed is given as input in the form of an image. The OpenCV package reads this image and converts it to greyscale. Any unnecessary backgrounds or disturbances are removed. This pre-processed image is stored in an object and analysed. The result of this analysis gives a two-dimensional array, that is, a matrix representing the colour intensity in each pixel.

For the input image as follows,



value of 0 indicates the color black and 255 indicates the color white.

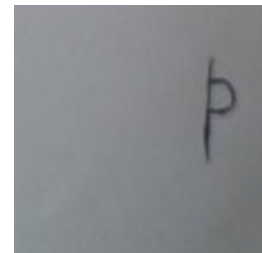
The NumPy header helps to work on two-dimensional arrays. The comparison of the matrix forms of the pre-defined data set with the given input image. This is done by accessing a particular pixel in any of the images. By accessing the required pixel of both the images, comparison of the images can be done. Using a python code, these header files are implemented.

## D. Feature Extraction

The character extracted by SkImage processing function is used as input here. This character in image format should be converted to a data format so that it is easier to compare it with another image. To do this the “NumPy” package of python is used. This NumPy package converts the image data into a multi-dimensional array. This array consists of entries 0 or 255 only. 0 represents black while 255 represents white. Thus, it considers the regions which consist of 0 to process the image. This multi-dimensional array is given to TensorFlow. Each value in the array is added to a fixed pre-determined value so as to generate a final value which can be used for comparison

## E. Input and Output

This program aims at obtaining an image as input. This input image should be in .jpg format. The handwritten text



The generated matrix is as follows,

```
[[152 152 152 ... 126 126 126]
 [152 152 152 ... 126 126 127]
 [152 152 152 ... 126 126 127]
 ...
 [145 144 143 ... 115 115 115]
 [145 144 143 ... 116 115 114]
 [144 144 143 ... 116 116 114]]
```

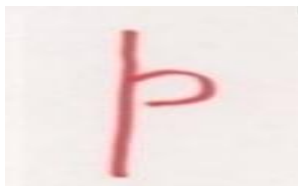
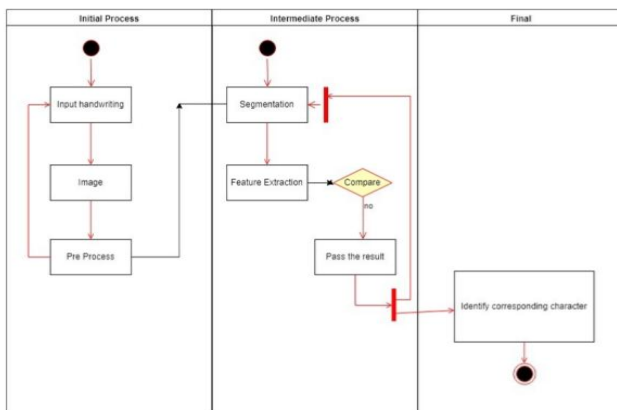
This pixel representation is compared with pre-defined input data set. This pre-defined input data set is also present in matrix format. The comparison of these two matrices provides percentage of matching. The matrix with highest matching percentage is the

# Personality Trait Analysis By Graphology Technique Using Machine Learning

required matrix. The pre-defined character associated with this matrix is the matching character. The personality trait associated with this character is the personality of the individual. This generated matrix matches with the following matrix,

[[243 242 241 ... 236 244 212]  
[242 242 241 ... 236 240 208]  
[242 242 242 ... 236 241 210]  
...  
[246 246 247 ... 241 242 242]  
[247 247 247 ... 241 242 242]  
[248 248 247 ... 241 242 242]]

It is noted that the approximate difference between each value of corresponding cells is the same. This is because of the varying colour intensity when the image is converted to a grayscale. And hence, the matching 'p' associated with this matrix is



Therefore, it is concluded that the person reflects good education, erudition, simplicity and mental activity.

Thus, each character from the given input is extracted and corresponding matrix is generated for the processed image. This process is continued until no unique characters are recognized, that is, until the full input text has been analysed.

This is the easiest method for recognition of characters using matrices.

## FLOW GRAPH SVM ALGORITHM

For cases other than comparing the features of letters, such as, the cases of checking baseline and margins the SVM algorithm can be used. After performing segmentation, the white space is calculated for each line of the given input. This is stored in percentage. The same process is repeated for each line in the given input.

After calculating the scores of all the lines in the input image it is sent as input to TensorFlow. The TensorFlow is trained with input sets which are pre-defined. It consists of

various possibilities of baselines and margins and their associated characteristics. These are numbered from 0 to 10 each number representing a particular character.

When it receives the scores of the input data from the segmentation part, it calculates the average of all these scores and finds the corresponding value from a total of 100 to a total of 10. The score obtained lies in the range of 0, 1, 2, 3, ...10. This value is compared with the input set given above and the corresponding character is printed on the screen.

## RESULT ANALYSIS

The system was tested with different possible inputs and the cases where an error might occur were analyzed. The following are the cases that were checked:

- The analysis of features of letters were checked first. All possible inputs were given to the machine and checked for accuracy. In all the cases, the matching was done and the personality trait was found to be right.
- The remaining algorithms such as baseline, margins, slant, size of letters, spacing were checked. Since, all the algorithms have to be executed for a single input they were checked simultaneously. It was found that, the baseline algorithm gives a result which is correct 9 out of 10 times. Hence, the algorithm has 90% accuracy.
- Similarly, for other features also, the accuracy was checked. It was found that there were a few deviations from the expected result. In conclusion, the error rate for the system is about 5% which can be decreased when a paid software of higher accuracy is used for the same.

## VI. CONCLUSION

This technique can be implemented in forensic departments and in psychology. In forensic departments, many handwritten evidences will be obtained. On performing the above personality analysis on these handwritten evidences, the mental state of the individual who wrote the text can be determined. This could help in further investigation of the cases. On the other hand, in psychology, especially in child psychology, where the patient is unable to convey their situation, handwriting analysis could help to a greater extent to understand the reasoning of their particular behavior. Apart from psychology, it can also be used in cancer detection and identification of heart diseases. But to do this data of at least 10 years is required.

However, the same technique cannot be applied for determining a teenager's psychology. This is because the mental state of a teenager changes frequently due to hormonal changes during this age period. Also, this does not determine the differences between a fake and a original handwriting, unlike an expert. With improved levels of accuracy these defects can be overcome easily.

## REFERENCES

1. Andrew W. Senior and Anthony J. Robinson, "An Off-Line Cursive Handwriting Recognition System", IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 20, No. 3, March 1998.
2. Bouchra El qacimy, Mounir Ait kerroum and Ahmed Hammouch, "Handwritten Digit Recognition based on DCT features and SVM Classifier", 2014 IEEE.



3. C. Hertel and H. Bunke, "A Set of Novel Features for Writer Identification," Proc. Int'l Conf. Audio- and Video-Based Biometric Person Authentication, pp. 679-687, 2003.
4. Cho, S. B. "Neural-network classifiers for recognizing totally unconstrained handwritten numerals," IEEE Transactions on Neural Networks, Volume 8, Issue 1, 1997, pp.43-53.
5. C.L. Liu, K. Nakashima, H. Sako, and H. Fujisawa, "Handwritten digit recognition: benchmarking of state-of-the-art techniques," Pattern Recognition, vol. 36, no. 10, pp. 2271-2285, 2003.
6. Dharam Veer Sharma, Geeta Saini, Mohit Joshi, "Statistical Feature Extraction Methods for Isolated Handwritten Gurumukhi Script", Vol. 2, Issue 4, July-August 2012, pp.380-384.
7. D. M. Monro, S. Rakshit, and D. Zhang, "Dct-based iris recognition," Pattern Analysis and Machine Intelligence, IEEE Transactions on, vol. 29, no. 4, pp. 586-595, 2007.
8. D.Y.C. Leung, C.H. Liu, "Limitations of the relative standard deviation of win percentages for measuring competitive balance in sports leagues" Economics Letters, Volume 109, Issue 1, October 2010, pp. 38-41.
9. Effective Multiple-features Extraction for Off-line SVM-Based Handwritten Numeral Recognition. By Shen-Wei Lee and Hsien-Chu Wu. 2012 IEEE.
10. H.Bunke, M. Roth, and E.G. Schukat-Talamazzini, "Off-Line Cursive Handwriting Recognition Using Hidden Markov Models," Pattern Recognition, vol. 28, no. 9, pp. 1399-1413, 1995.
11. Hu Guosheng, Zhang Guohong, "Comparison on neural networks and support vector machines in suppliers' selection," Journal of Systems Engineering and Electronics, Volume 19, Issue 2, April 2008, pp. 316-320.
12. Jaime Melendez, Miguel Angel Garcia, Domenec Puig, Maria Petrou. "Unsupervised texture-based image segmentation through pattern discovery," Computer Vision and Image Understanding, Volume 115, Issue 8, 2011, pp. 1121-1133.
13. J.C. Simon and O. Baret, "Regularities and Singularities in Line Pictures," Int'l J. Pattern Recognition and Artificial Intelligence, vol. 5, nos. 1-2, pp. 57-77, 1991.
14. J.F. Canny, "A Computational Approach to Edge Detection," IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 8, pp. 679-698, 1986.
15. Krishna S. Nathan, Homayoon S. M. Bezgi, Jayashree Subrahmonia, Gregory J. Clary and Hrozsh Maruyama, "Real-time online unconstrained handwriting recognition using statistical methods", Handwriting Recognition Group IBM Research, T.J. Watson Center, Yorktown Heights, NY 10598
16. L. Deng, "The mnist database of handwritten digit images for machine learning research," IEEE Signal Processing Magazine, vol. 29, no. 6, pp. 141-142, 2012.
17. Ni, J., Zhang, C., & Yang, S. X. "An adaptive approach based on KPCA and SVM for real-time fault diagnosis of HVCBs," IEEE Transactions on Power Delivery, 2011, pp. 1960-1971.
18. Richard J. Klimoski and Anat Rafael, "Inferring personal qualities through handwriting analysis", Journal of Occupational Psychology, 1983, 56, 19 1-202.
19. Prof. Seema Kedar, Ms Vaishnavi Nair and Ms. Shweta Kulkarni, "Personality Identification through Handwriting Analysis: A Review", International Journal of Advanced Research in Computer Science and Software Engineering, Volume 5, Issue 1, January 2015.
20. Shen-Wei Lee, Hsien-Chu Wu, "Effective Multiple-features Extraction for Off-line SVM-Based Handwritten Numeral Recognition", 2012 IEEE.
21. Tal Steinherz, David Doermann, Ehud Rivlin and Nathan Intrator, "Offline Loop Investigation for Handwriting Analysis", IEEE transactions on pattern analysis and machine intelligence, vol. 31, no. 2, February 2009
22. V. Govindaraju and R.K. Krishnamurthy, "Holistic Handwritten Word Recognition Using Temporal Features Derived from OffLine Images," Pattern Recognition Letters, vol. 17, no. 5, pp. 537-540, 1996.
23. V. Govindaraju, D. Wang, and S.N. Srihari, "Using Temporal Information in Off-Line Word Recognition," Proc. US Postal Service Fifth Advanced Technology Conf., pp. 529-543, 1992.
24. Vlad Atanasiu, "Surveys and projects in handwriting analysis", MIT Postdoctoral fellow, Department of Brain and Cognitive Sciences 77, Massachusetts Av., Cambridge MA 02139, USA.
25. X.-T. Yuan and B.-G. Hu, "Robust feature extraction via information theoretic learning," in Proceedings of the 26th Annual International Conference on Machine Learning. ACM, 2009, pp. 1193-1200.
26. Y. LeCun, L. Jackel, L. Bottou, A. Brunot, C. Cortes, J. Denker, H. Drucker, I. Guyon, U. Muller, E. Sackinger et al., "Comparison of

learning algorithms for handwritten digit recognition," in International conference on artificial neural networks, vol. 60, 1995.

27. [27] Ying-Ming Wang, Ying Luo, "Integration of correlations with standard deviations for determining attribute weights in multiple attribute decision making," Mathematical and Computer Modelling, Volume 51, Issues 1-2, January 2010.

## AUTHORS PROFILE



**R. Ranjith** is Assistant Professor of the Department of CSE at St. Joseph's College of Engineering, Chennai, Tamil Nadu. He has completed his Bachelor of Technology Degree in Information Technology from Anna University in the year of 2012. He has completed his Master of Technology in Software Engineering from Anna University in the year of 2014. He has published 2 technical papers in various journals. He has 5 years of teaching experience on graduate and post-graduate level. His area of interest includes Artificial Intelligence and Machine Learning.



**S. Jothi** is Associate Professor of the Department of CSE at St. Joseph's College of Engineering, Chennai, Tamil Nadu. She has completed her Bachelor of Engineering Degree in Computer Science and Engineering from Madurai Kamaraj University in the year of 2003. She has completed her Master of Engineering in Computer Science and Engineering from Annamalai University in the year of 2005. She has completed her Ph.D. in Wireless sensor networks from Anna University, Chennai, in the year of 2016. She has published 12 technical papers in various journals. She has 13 years of teaching experience on graduate and post-graduate level. Her area of interest includes Wireless Sensor Networks, Mobile ad-hoc Networks, Big Data Analysis and Image Processing.



**A. Chandrasekar** is Professor and Head of the Department of CSE at St. Joseph's College of Engineering, Chennai, Tamil Nadu. He has overall teaching experience of over 21 years in Engineering Colleges. He has guided more than 12 Research Scholars and more than 50 M.E. students. He has published over 110 research articles in refereed International and National journals and he is guiding research scholars and M.E. students in the areas of Network Security, Cloud Security, Data mining, Artificial Intelligence and Big Data Analysis.