

Identifying Cancer Characteristics Utilizing Handwriting Method

S. V. Kedar¹, Alisha Shinde², Aditya Velapurkar³, Prashant Pajiyar⁴, Shrinivas Sanglikar⁵

¹HOD of Department of Computer Engineering, JSPM's Rajarshi Shahu College of Engineering, Savitribai Phule University of Pune, India.

²Department of Computer Engineering, JSPM's Rajarshi Shahu College of Engineering, Savitribai Phule University of Pune, India.

³ Department of Computer Engineering, JSPM's Rajarshi Shahu College of Engineering, Savitribai Phule University of Pune, India.

⁴Department of Computer Engineering, JSPM's Rajarshi Shahu College of Engineering, Savitribai Phule University of Pune, India.

⁵Department of Computer Engineering, JSPM's Rajarshi Shahu College of Engineering, Savitribai Phule University of Pune, India.

Article History: Received: 10 November 2020; Revised: 12 January 2021; Accepted: 27 January 2021; Published online: 05 April 2021

Abstract: Handwriting is an action administered by the brain like each and every other action. This procedure is frequently insensible and is closely tied to instincts from brain. Any kind of sickness affects the kinetic movement and reflects in a person's handwriting. To recognize the health and mental problems, it is important to focus on how the person writes instead of what person writes. This also makes the procedure of handwriting analysis independent of all languages. Person handwriting is scientific proof that whatsoever person writes subconsciously it affects in handwriting. The structures related to motion, time and pressure have been used for analysis of person health. Cancer is the second top cause of death globally, and is accountable for an estimated 9.8 million deaths in 2019. Universally, around 1 in 6 deaths is due to cancer. On an approximation 72% of deaths due to cancer are in middle and low salaried countries. One third deaths from cancer are due to 5 foremost dietary and behavioural risks that are low fruit and vegetable intake, lack of physical activity, high body mass index, tobacco use, and consumption of alcohol. Cancer can be cured if the person gets to know as soon as possible. So, substitute method to patterned whether the person is diagnosed from a cancer or not, can be done by handwriting sample. For this testing 100 various person sample are used for diverse handwriting data samples. To find a solution to this mounting problem we propose the method of cancer characteristics detection by utilizing handwritten text by machine learning, SVM. Various machine learning methods were used to find a model, which can discriminate statistically Cancer patients with approximately 90% accuracy. The classification we use to discriminate are SVM, Naïve Bayes algorithms.

Keywords: Handwriting analysis, Handwriting features, Graphology, SVM.

1. Introduction

Recognition of handwriting analysis, various characteristics are taken into consideration to analyses a matching handwriting. The Quality of line gives an idea about the strength, thickness and the flow in which the letters are arranged. Approximate aspects are the letters are if shaky, flowing or very thick. The amount of space put between each and every letters is basically the letter spacing. The scopes of the handwriting are analyses based on the height, width and size of the letters. While writing a person may lift the pen or write with separations. They may stop before writing a new letter or connect the letters. Basically, connecting strokes are separations and the pen lifts such as the capital letters are connected to lowercase letters. Strokes aspects like how the person writing starts and ends their words. Does the writer ends the letter with a curl or an upstroke or a down stroke? The pressure of the pen is analyses. Where the pressure is more is calculated. In which manner each letter tends to slant, either in left or right direction or no slant at all. Usually people tend to write slant towards the right. A writer may tend to write above or below the line or even on the line. This is basically stated as the Baseline behaviours of a writer.

All activities including writing start in the brain. Like all other actions, the act of writing depends on central nerve system [1]. Our brain sends impulses to hand through nervous signals, achieving the motor act. Graphology is pseudo-science based on combination of psychoanalysis and neuroscience nested in subconscious mind. Though handwriting is driven through pen, its movement is administrated by the central nervous system, which is a process usually unconscious, but most revealing [2]. Handwriting is closely tied to impulses from the brain and therefore it can be reliably used to predict state of physical, emotional and mental health of individual [3]. Handwriting analysis is used to find out disturbance in the subject's handwriting.

This paper is based on the same procedure of handwriting detection and estimate the characteristic of the person whose features are comparable to cancer patients. The very first step is to make the computer understand using machine learning. The collection of digital handwriting of a person and computer prediction through that sample is economical and appropriate method. Any individual gives the digital sample of handwriting on a tab by which features are extracted from raw data. And it is uploaded on computer and various features are estimated with help of image processing techniques and characteristics of writer is predicted.

Digital handwriting analysis consist of 5 different standard cases and by this data samples are collected from individual. These 5 standard cases are important as all features are dependent, such as pressure, x coordinate, y coordinate and time spent on tab. These features help us to extract 11 different features which are important as characteristics of cancer patient lie.

These samples are considered as input to the support vector machine, naïve Bayes, KNN for classification. The algorithm planned is simple and easy to implement.

2. Proposed System

This paper proposes a system which uses digitized handwriting analysis to identify the cancer characteristics. The proposed system consists of four modules.

1. Tab analysis
2. Feature extraction
3. Classification

Digital handwriting analysis method takes the input of the handwriting samples in two sets, set A which consist of handwriting samples of Cancer diagnosed patients and set B which consist of a healthy person. Below Fig 1. Shows that proposed system for identification of cancer characteristics utilizing handwriting method.

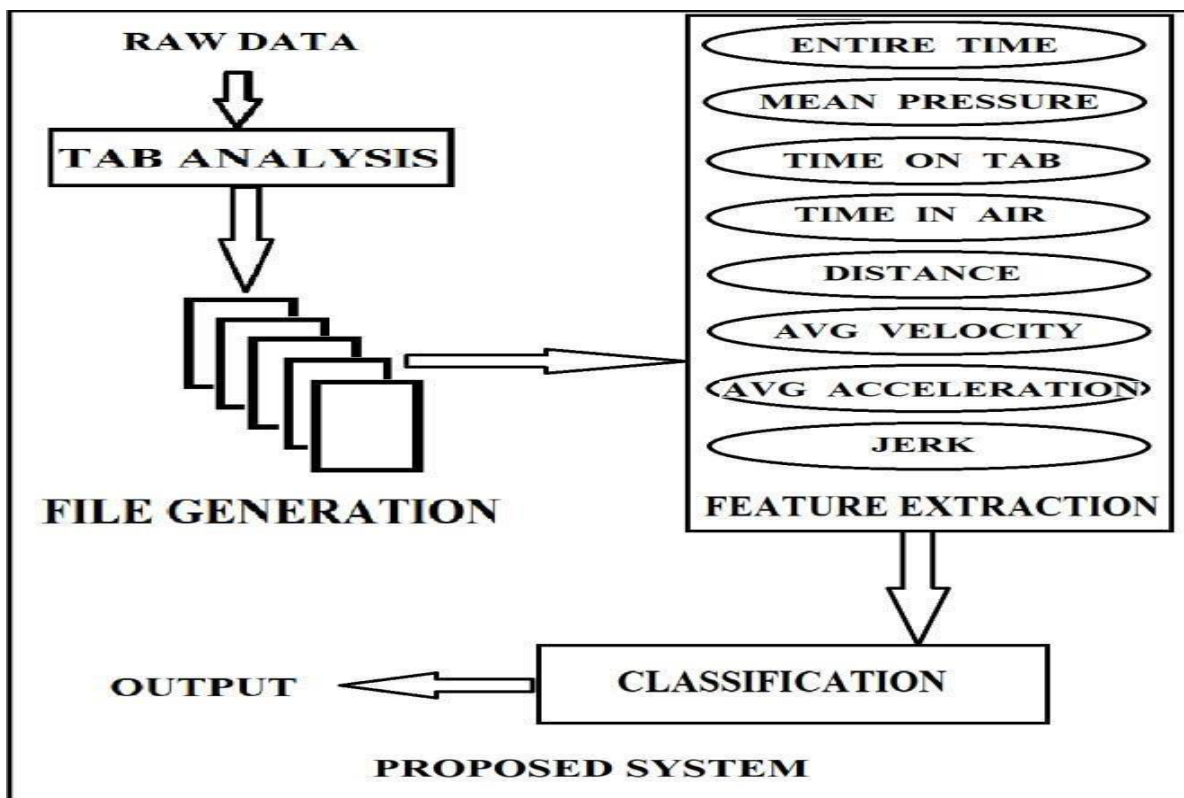


Figure 1. Proposed System.

Each individual is to complete a draw-write standard test for the handwriting analysis on Wacom Intuos tab. Each set has the 100 handwriting samples. The five files of each test from the standard test is obtained. The features are then extracted from these files. These features are then used to train and test the suitable classifier for providing the accuracy for identifying the cancer characteristics.

3. Modules

3.1. Tab Analysis

Digitally collected handwriting samples of different individuals are used in this research which uses standard test on a Wacom Tab. The handwriting of 100 individuals which were diagnosed with cancer and another of 100 handwriting samples which are weren't diagnosed with the cancer. The standard test consists five test cases which focuses on the different aspects of the handwriting. Each of them was asked to write on the tab and complete the five tests in running

hand. The first test consists of intersecting pentagon copying [4], test 2 consist of writing the three words in block letters. The test 3 is to write a sentence on the tab. The task 4 is to draw the analog clock showing the time of 5.15 without showing the numbers. As the test 5 is to draw a house with a tree and the person. These tests provide a file each contains the five parameters. The parameter is as following i. Timestamp ii. Count in microseconds iii. X co-ordinate iv. Y co-ordinate and v. Pressure. The parameters obtained in each of these files while performing the test indicates the variables obtained at that instance.

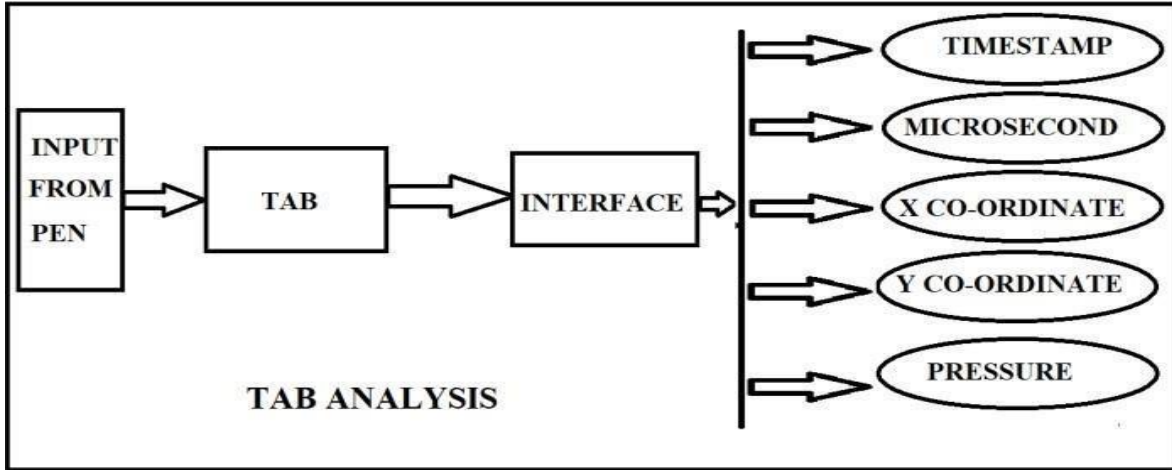


Figure 2. Tab Analysis Architecture

3.1. Feature Extraction

Feature extraction is a technique of dimensionality reduction from a high dimensional input data. This reduced output data is a transformation of the high dimension input data. represented as feature vector [5]. In our case, the 7 important factors on which the proposed system predicts the cancer characteristics are considered as the features. Below are the features extracted from the 5 files generated. These features are used for the handwriting analysis which represents a distinct characteristic developed by the cancer patient.

I. Number of on-paper and in-air strokes while completing a task:

1. On Paper Strokes:

$$pno = \text{count}(pi(t) > 0 \text{ and } pj(t) = 0) \text{ where } j = i + 1 \dots \dots \dots(1)$$

2. In Air Strokes:

$$ano = \text{count}(pi(t) = 0 \text{ and } pj(t) > 0) \text{ where } j = i + 1 \dots \dots \dots(2)$$

i) Time to complete entire task:

$$dur = dn(t) - dl(t) \dots \dots \dots(3)$$

where n is the total number of data points and dl(t) is time-stamp recorded for data frame l.

ii) Total time spent in-air for completing a task:

$$pi(t) = 0 \text{ and } pj(t) = 0 \dots \dots \dots(4)$$

iii) Total time spent on-paper for completing a task:

$$ptime = \sum_{i=1}^n (pj(t) - pi(t))$$

where $pi(t) > 0$ and $pj(t) = 0 \dots \dots \dots(5)$

iv) Mean pressure applied on pen:

$$pmean = \frac{\sum_{i=0}^n pi}{n} \dots\dots\dots(6)$$

v) Speed of trajectory calculated as length of segment divided by duration:

$$savg = \sum_{i=1}^n pi(d) / dur \dots\dots\dots(7)$$

vi) Velocity: It is the rate at which the position of a pen changes with time

$$p(a) = pi(d) / pi(t) \dots\dots\dots(8)$$

vii) Acceleration: It is the rate at which the velocity of a pen changes with time.

$$amean = savg / t \dots\dots\dots(9)$$

vii) Jerk: It is the rate at which the acceleration of a pen changes with time.

$$10) \quad asd = \sqrt{\sum_{i=1}^{n-1} (pi(a) - \alpha_{\mu sav})^2} \dots\dots\dots(10)$$

3.3. Classification

One of the main conducts for building a classifier is to train the database based on handwriting samples obtained from the cancer patients. In a primary study of numerous classification models, linear support vector machines (SVMs) and Naïve Bayes (Binary) classification raised out, even though not with a large margin. Certainly, linear SVMs and Naïve Bayes have been used broadly on a wide range of classification, regression, and ranking applications in science, medicine, and engineering and also known admirable observed performance. In all succeeding works, linear SVMs and Naïve Bayes were used to train and test the models. The classifier was set such that proficient or nonproficient products are connected with the undesirable or desirable output of classifier.

4. Experiment and Result

In demand to accurately evaluate the performance of the trained classifiers taking place handwriting samples, tenfold cross- validation was used. This procedure delivers an impartial estimator for the performance of classifiers on concealed data. In this 60 handwriting samples collected from both cancer patients as well as non-cancer patients. Data are split in 80-20 ratio for training and testing purposes. The 80% training data consist of 96 samples and are tested for accuracy. The accuracy we get after training data is 84%. Remaining 20% of dataset are used for testing. Testing dataset consist of 24 samples and are tested for accuracy. The accuracy we get after testing data is 92%. In our case where the accuracy chart jumped from 84% towards 92% which is a respectable result.

4.1. Performance Measure

The performance of the trained classifier on a database which was assessed by calculating the accuracy. Accuracy is a distinctive metric for classifier performance, measuring the fraction of correctly classified handwriting samples data and shows that better the accuracy better the output.

4.2. Results

The SVM classifier confirms precise predictions for 77 out of 100 writing samples, which deciphers to an accuracy of 76.6% whereas the Naïve Bayes classifier shows the precise prediction for 99 out of 100 writing sample, which gives 99.6% accuracy.

Table 1. Accuracy of algorithm used

Classification Algorithm	Accuracy of algorithm	
	Accuracy for Training Set	Accuracy for Testing set
Support Vector Machine	100%	76.6%
Naive Bayes	100%	99%

5. Conclusion

The procedure of evaluating the handwriting samples to recognize cancer characteristics of the writer is very considerable beneficial in current scenario. It senses the characteristics which determine the precise condition of the writer. It would be really useful to find out if the writer is suffering from cancer or not and if writer remains found to be so, essential action for treatment or provision can be on condition that with the intent that they overcome it.

References

1. Cronje, Pierre E. and Hester E Roets, "Graphology in Psychological Assessment: A Diagnosis in Writing", *Universal Journal of Psychology*, vol. 1, No. 4, pages: 163-168, 2013.
2. Bora Ugurlu, Rembiye Kandemir, Aydn Carus, Ercan Abay, "An Expert System for Determining the Emotional Change on a Critical Event Using Handwriting Features", *TEM Journal*, vol. 5, Issue 4, Pages 480-486, 2016.
3. Emotions and Feelings, Available from: <http://graphicinsight.co.za/emotions.htm>
4. S. Rosenblum, P. L. Weiss, and S. Parush, "Handwriting evaluation for developmental dysgraphia: Process versus product," *Reading Writing*, vol. 17, no. 5, pp. 433–458, 2004.
5. Seema Kedar. S. Bormane, Sandeep Joshi Online "Analysis of Handwriting for Disease Diagnosis: A Review" *International Journal of Engineering & Technology*, 7 (3.24) (2018) 505- 511
6. Naive Bayes Classifier, Available from: https://en.wikipedia.org/wiki/NaiveBayes_classifier
7. Atrayee Dutta, Aditya Dubey, "Detection of Liver Cancer using Image Processing Techniques", 978-1-5386-7595-3/19/\$31.00, ©2019IEEE
8. Matko Saric, Mladen Russo, Maja Stella, Marjan Sikora, "CNN-based Method for Lung Cancer Detection in Whole Slide Histopathology Images, University of Split, FESB, Split, Croatia Serban Radu Ștefan Jianu, Loretta Ichim , Dan Popescu, "Automatic Diagnosis of Skin Cancer Using Neural Networks", THE XIth INTERNATIONAL SYMPOSIUM ON ADVANCED TOPICS IN ELECTRICAL ENGINEERING 2019
9. Ahmed Ech-Cherif, Mohammed Misbhauddin, Mohammed Ech-Cherif, "Deep Neural Network based Mobile Dermoscopy Application for Triaging Skin Cancer Detection" In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 4510-4520, 2018.
10. Md. Sakif Rahman, Pintu Chandra Shill, Zarin Homayra, "A New Method for Lung Nodule Detection Using Deep Neural Networks for CT Images", 2019 International Conference on Electrical, Computer and Communication Engineering (ECCE), 7-9 February, 2019.
11. Xiangfeng Dai, Irena Spasić, Bradley Meyer, Samuel Chapman, Frederic Andres, "Machine Learning on Mobile: An On-device Inference App for Skin Cancer Detection", 2019 Fourth International Conference on Fog and Mobile Edge Computing (FMEC).
12. Yifan Chen, Muhammad Ali, Shaolong Shi, U Kei Chaeng, "Direct Targeting Strategy for Smart Cancer Detection as Natural Computing ", *Adv. Nat. Sci.: Nanosci. Nanotechnol.*, vol. 8, pp. 023002, 2017.
13. P. I. Khalid and R. Adnan, "Extraction of dynamic features from hand drawn data for the identification of children with handwriting difficulty," *Res. Develop. Disabilities*, vol. 31, no. 1, pp. 256–262, 2010.
14. J. Medwell and D. Wray, "Handwriting automaticity: the search for performance thresholds," *Lang. Learn.*, vol. 28, no. 1, pp. 34– 51, 2014.
- A. O'Hare, "Hands up for handwriting," *Develop. Med. Child Neurol.*, vol. 46, no. 10, pp. 651–651, 2004.
15. R. Karlsdottir and T. Stefansson, "Problems in developing functional handwriting (monograph supplement 1-V94)," *Perceptual Motor Skills*, vol. 94, no. 2, pp. 623–662, 2002.

16. Murad Badarna, Ilan Shimshoni, Gil Luria and Sara Rosenblum, "The Importance of Pen Motion Pattern Groups for Semi-Automatic Classification of Handwriting into Mental Workload Classes", *Cognitive Computation*, vol. 10(2): 215-227, 2017.
17. Emotions and Feelings, Available from: <http://graphicinsight.co.za/emotions.html>.
18. U.V. Marti, R. Messerli, H. Bunke, "Writer Identification Using Text Line Based Features", Proc. ICDAR'01, Seattle(USA), pp 101-105, 2001.
19. Champa H N, K R AnandaKumar, "A Scientific Approach to Behavior Analysis through Handwriting Analysis", National Conference on Research Trends in Information Technology, S R K R Engineering College, Bhimavaram, Andhra Pradesh, 2008.
20. Graphology-Handwriting Analysis <http://www.businessballs.com/graphologyhandwriting-analysis.htm>.