Design and Implementation of Graphology using Machine Learning

A Project Work - II Report

Submitted in partial fulfillment of requirement of the

Degree of

BACHELOR OF TECHNOLOGY in COMPUTER SCIENCE & ENGINEERING

BY
Mihir Khatri (EN18CS301140)
Milind Kaul (EN18CS301141)
Prakhar Soni (EN18CS301174)

Under the Guidance of
Dr. Kailash Chandra Bandhu

Department of Computer Science & Engineering
Faculty of Engineering
MEDI-CAPS UNIVERSITY, INDORE- 453331

Jan-April '2022'

Report Approval

The project work **"Integrating Graphology and Machine Learning for Accurate Prediction of Personality: A Novel Approach"** is hereby approved as a creditable study of an engineering/computer application subject carried out and presented in a manner satisfactory to warrant its acceptance as prerequisite for the Degree for which it has been submitted.

It is to be understood that by this approval the undersigned do not endorse or approve any statement made, opinion expressed, or conclusion drawn therein; but approve the "Project Report" only for the purpose for which it has been submitted.

Internal Examiner

Name: Dr. Kailash Chandra Bandhu

Designation: Assistant Professor

Affiliation: Medi-caps University

<u>Declaration</u>

We hereby declare that the project entitled **"Integrating Graphology and Machine Learning for Accurate Prediction of Personality: A Novel Approach"** submitted in partial fulfillment for the award of the degree of Bachelor of Technology in 'Department of Computer Science and Engineering' completed under the supervision of **Dr. Kailash Chandra Bandhu, Assistant Professor, Department of Computer Science and Engineering,** Faculty of Engineering, Medi-Caps University, Indore is an authentic work.

Further, we declare that the content of this Project work, in full or in parts, have neither been taken from any other source nor have been submitted to any other Institute or University for the award of any degree or diploma.

Mihir Khatri [EN18CS301140]

Milind Kaul [EN18CS301141]

Prakhar Soni [EN18CS301174]

Certificate

I, **Dr. Kailash Chandra Bandhu** certify that the project entitled **"Integrating Graphology and Machine Learning for Accurate Prediction of Personality: A Novel Approach"** submitted in partial fulfillment for the award of the degree of Bachelor of Technology/Master of Computer Applications by **Mihir Khatri, Milind Kaul and Prakhar Soni** is the record carried out by them under my guidance and that the work has not formed the basis of award of any other degree elsewhere.

_____

Dr. Kailash Chandra Bandhu

Assistant Professor

Department of Computer Science and Engineering

Medi-Caps University, Indore

_____

Dr. Pramod S. Nair

Head of the Department

Computer Science & Engineering

Medi-Caps University, Indore

Acknowledgements

I would like to express my deepest gratitude to Honorable Chancellor, **Shri R C Mittal,** who has provided me with every facility to successfully carry out this project, and my profound indebtedness to **Dr. Dilip K. Patnaik,** Vice Chancellor, Medi-Caps University, whose unfailing support and enthusiasm has always boosted up my morale. I also thank **Prof. (Dr.) D K Panda,** Pro Vice Chancellor, **Dr. Suresh Jain,** Dean, Faculty of Engineering, Medi-Caps University, for giving me a chance to work on this project. I would also like to thank my Head of the Department **Dr. Pramod S. Nair** for his continuous encouragement for betterment of the project.

I express my heartfelt gratitude to my Internal Guide, **Dr. Kailash Chandra Bandhu and Dr. Ratnesh Litoriya** without whose continuous help and support, this project would ever have reached completion.

It is their help and support, due to which we became able to complete the design and technical report.

Without their support this report would not have been possible.

Mihir Khatri, Milind Kaul, Prakhar Soni
B.Tech. IV Year
Department of Computer Science & Engineering
Faculty of Engineering
Medi-Caps University, Indore

Table of Contents

Chapter – 1

1.1 Introduction

Among all the unique characteristics of a human being, handwriting carries the richest information to gain the insights into the physical, mental and emotional state of the writer. Handwriting is the reflection of each individual's personality. It is the art of studying and analyzing handwriting, a scientific method used to determine a person's personality by evaluating various features from the handwriting.

To make this method more efficient and reliable, introduction of machines to perform the feature extraction and mapping to various personality traits can be done. This compliments the graphologists, and also increases the speed of analyzing handwriting samples. Proposed methodology focuses on computer assisted handwriting analysis by considering 4 features such as Baseline, slant of the words and margin. The estimation accuracy of the program is approximately 95%.

Personality identification of a human being by their handwriting is an old technique. Earlier these were done manually by spending a lot of time to predict the nature of the person. Handwriting is brain writing, representing the mental status of the person. Handwriting analysis is a projection technique as the body language that profiles human behavior in areas of social skills, achievements, thinking styles, or work habits. Handwriting also depicts the possible ways of persons transactions with stress. It is a system to study frozen graphic structures which is generated in the author's brain and is placed on the paper in a cursive or printed handwriting style which is different with other authors comparing the personalities and their potential for problem solving.

1.2 Understanding Graphology

Handwriting develops right from childhood. When you write, your pen is under the control of the muscles of your fingers, hands and arm. All these body parts are under the control of your mind. The manner in which the words are eventually formed by the pen must bear a direct relationship to the mind that guides their formation. Each vibration of movement is unconsciously directed by the brain, so we can judge the mental state of the writer. It is a guide to the will power, intellect and emotions of a person.

1.3 Motivation for Handwriting analysis

Many people want to know why they should go for handwriting analysis. Handwriting analysis is both important and necessary for us at some level. There are a variety of reasons why a person must get his handwriting analyzed at least once. The main reason is that handwriting analysis unfolds many things about your personality. It is a tool to know your strengths and weak points. Till about a few years ago, handwriting analysis was restricted to forensic experts. Not anymore. Today, several multinational corporate firms use handwriting analysis to know personality traits of their job candidates. Handwriting tells about your suitability for the job, your motivational level, creativity, leadership, teamwork etc. In fact, we would suggest you to not only get your handwriting analyzed, you should also know the art of handwriting analysis. Advantages are many like having access to the inner feelings of the person sitting next to you, your colleague. It is more productive and beneficial for us to know how to get along with other people and know yourself a little better.

1.4 Literature Survey

Literature survey chapter provides an overview of previous research on knowledge sharing and intranets. It introduces the framework for the case study that comprises the main focus of the research described in this thesis. It provides a description, summary, and critical evaluation of each work. Survey may include: scholarly journals, books, dissertations, conference proceedings, etc. It may be completed en route to an essay, thesis or dissertation and included in the final project. Or, its Materials may be conducted as its own entity. This was in order to scope out the key data collection requirements for the primary research to be conducted, and it formed part of the emergent research design process. An appreciation of previous work in this area served three further purposes. First, through providing direction in the construction of data collection tools, it guarded against the risk of overload at the primary data collection stages of the project. Second, working the endings from extant literature into a formal review helped maintain throughout the study a sense of the topic's perspective. Finally, this activity raised the opportunities for articulating a critical analysis of the actual meaning of the data collected when the data analysis stages of the research were reached.

1.5 Detailed Survey

Throughout history, scientists, philosophers, artists and others have been interested in the relationship between the handwriting and the writer. This interest appeared as early as 1622. Efforts at handwriting analysis began in 1872, with the work of the French abbe, Hypolite Michon, who gave graphology its name.

Methods proposed in literature involve the preliminary process of text extraction from the sample and then application of various algorithms/ techniques to determine the characteristic traits. Polygonalization method is one such technique which involves a closed polygon produced around a line in the scanned image of the hand written text. The slope of the text/alphabet is found using the coordinates of the polygon. Generalized Hough Transform is a second technique used to detect any arbitrary shape in an image by creating a table for storing all the edge pixels of the target shape. Template matching with certain predefined templates is also used as a technique for behavioral analysis Segmentation method which involves splitting up of the handwriting sample into individual letters is another work available in literature. Methods discussed are not

very simple to automate and therefore a need exists for a simple method which could be automated easily.

- **Development of an automated handwriting analysis system**

  This work focuses on development of an automated technique for determining the characteristic traits of a person through Image Processing called AHWAS (Automated Handwriting Analysis System). The proposed work involves less image pre-processing of the image as it crops the given sample automatically and uses a RGB filter to extract the text in the handwriting and identifies eight features in the handwriting simultaneously. The features identified are: size of the letters, baseline, pressure of the writing, slant of the handwriting, number of breaks, spacing between the words, margins and speed of the writing in the sample. The system is designed to directly indicate the behavior of the person from the above features. The system can be used in various applications such as detection of diseases like Parkinson's disease or Alzheimer's disease, forensic document examination and lie detection.

  The primary step involved in calibration of the images to extract the maximum amount of the handwriting, is to use a RGB (red green blue) filter to have a clear image of the handwriting. The width of an A4 sheet, which is 8.27 inches, is used for size calibration. This was done by determining the edges of the paper and passing the input to the program.

- **Handwriting Analysis based on Segmentation Method for Prediction of Human Personality using support vector machines**

  In this method, to predict the actual personality of the individual there are various features , such as slant, size, pressure, upper zone or case(as in i, t, h, s, etc), lower zone (as in g, q, y, z, etc), word spacing, line spacing, page margins, middle zone or case(as in a, o, c, s, e, etc), arcade, garland, angle, thread, wavy line (written by authors who are mentally mature and are skilful), and many others. But the proposed system has only six features among them discussed above. These features are size of the letters, slant of words and letters, baseline, pen pressure, spacing between letters and words as they are enough to predict the behavior of the person. The main attention of this paper is feature

extraction and its classification. All features are extracted automatically from the digital image of handwriting. These samples are then input to the support vector machine for classification.

SVM is time efficient than neural networks as tested. It analyzes the data and recognizes patterns in the data. Multiclass- SVM aims to assign labels to instances by using support vector machines, learning algorithms, where the different classes are drawn from a finite set of elements to clearly separate them. SVM takes training data x, of some space, Rd, where x=x1, x2,. xn such that x belongs to Rd. We also give their class label y= y1, y2, ..yn as to which belongs to. In binary SVM y is a subset of -1,+1while in case of multi-class SVM the y ranges 1 to N, where, N is the number of classes can belong to the linear SVM classification function is represented by f(x)=w. x+b, which relates to a separating hyper-plane. x+b=0 The hyper-planes separating the training data samples with the maximum margin such that the closest point to the plane is 1/kwk.

- **Artificial Neural Network for Human Behavior Prediction through Handwriting analysis**

In this paper a method has been proposed to predict the personality of a person from the features extracted from his handwriting using Artificial Neural Networks. The personality traits revealed by the baseline, the pen pressure and the letter t as found in an individual's handwriting are explored in this paper. Three parameters, the baseline, the pen pressure and the height of the t-bar on the stem of the letter t are the inputs to the ANN which outputs the personality trait of the writer. The evaluation of the baseline is using the polygonalization method and the evaluation of the pen pressure utilizes the gray-level threshold value. The height of the t-bar on the stem of the letter t is calculated using template matching. The baseline, the pen pressure and the letter 't' in one's handwriting reveal a lot of accurate information about the writer. MATLAB is the tool used for the purpose. The performance is measured by examining multiple samples.

The main technique for finding the baseline slant is the polygonalization of the single line in the handwritten text .The pen pressure can be calculated as a count of the number of foreground pixels in the thresholded image. The number of black pixels is indicative of the pen pressure, thickness of strokes, and the size of writing. The lowercase letter t is

one of the letters in one's handwriting that reveals a lot of accurate information about the writer. People write letters in many different ways. There are various ways to make the stem, the cross on the t-bar, and even the entry and exit to this letter t, each one of which relates to a specific personality trait of a person

- **Automated Prediction of human behavior system for Career counseling of an individual through handwriting analysis.**

  The many features in handwriting whereby behavior is usually predicted are pen pressure, baseline, slant, width of margins, spacing between letters, spacing between words, measurements writing, height of bar on letter 't', letter 'y', etc. A may be proposed to calculate the personality of the person on the feature extracted from his handwriting using Artificial Neural Networks. Most of the research was done to recognize the characters of handwriting and commonly used Artificial Neural Network (ANN) for the recognition . It is easier to apply neural networks for that purpose because ANN is known as a good method for pattern recognition. Anyhow the training process for ANN requires a lot of time and data.

  This work has been carried out to implement multi-input multi layered Neural Network(parallel distributed system) when considering recognition of Punjabi characters, that is trained using back propagation, for that final utilization of this trained network to realize the patterns trained for, and classify these under different, distinct output classes how the network was taught to group them under. This problem is divided into two phases:

  1. Reading a windows image format.
  2. Development of Artificial Neural Network model.

  Second phase is further divided into two sub-phases:

  1. Training phase of neural network
  2. Testing phase of neural network

- **Honesty and Dishonesty through Handwriting Analysis**

In order to determine dishonesty in one's handwriting, the ability to recognize honesty is a prerequisite.

Psychopathology in Handwriting

1. The Habitual Liar

The technique of lying, it seems, has at least three ways of achieving its ends. In the liar's presentation of the story

   (a) One (essential) part is simply left out.

   (b) One (essential) part is left out and a freely invented part is substituted for it.

   (c) One (essential) part is left out and the gap is filled with chitchat, or

      meaningless,vague tales.

In all three ways, the liar tries carefully not to appear as such his story and approach must not arouse suspicion.

2. The Pathological Liar

These two seemingly different handwritings were written by one person, a pathological liar. She executed this writing for the doctor who had her under his care,in order to show "how clever she was". From the standpoint of graphology, these handwritings are identical with the exception of the slant; neither contains a basic characteristic that the other lacks. The pathological liar, to be sure, is not merely a person who tells many lies. He is almost completely identified with the false roles he unconsciously assumes. Consequently, he will characteristically show two or more different styles of writing, rather than merely the slips of the "habitual liar". Such shifting of style is the clue to pathology, which the graphologist can discover.

- **A paper published in the year 1995 at the SUNY, Buffalo** by **Prof S. N. Srihari**

This paper focuses only on the prime features of a page of handwritten samples that are page margins, line spacing, line direction, slant and zone ratios. The methodology used is scanning, preprocessing, feature extraction, analysis and finally, trait determination. This system was mainly designed to prove the validity of the graphology rules that were applied in the implementation of the system. This paper restricts its scope to macro

analysis of the handwriting sample. There are no micro features like alphabet, loops etc. taken into consideration.

- **Handwriting Analysis for Detection of Personality Traits using Machine Learning Approach**

Among all the unique characteristics of a human being, handwriting carries the richest information to gain the insights into the physical, mental and emotional state of the writer. Graphology is the art of studying and analyzing handwriting, a scientific method used to determine a person's personality by evaluating various features from the handwriting. The prime features of handwriting such as the page margins, the slant of the alphabets, the baseline etc. can tell a lot about the individual. To make this method more efficient and reliable, introduction of machines to perform the feature extraction and mapping to various personality traits can be done. This compliments the graphologists, and also increases the speed of analyzing handwriting samples. Various approaches can be used for this type of computer aided graphology. In this method a novel approach of machine learning technique to implement the automated handwriting analysis tool is discussed.

This paper has proposed a methodology to predict the accurate personality traits of an individual from the features extracted from handwriting using a machine learning approach. This paper explores the personality traits revealed by baseline, margin, slant of the words and height of t-bar of a person's handwriting. These features will be extracted from the handwriting samples into feature vectors which would be compared with an initially trained data set; and then mapped to the class with corresponding personality traits. The baseline would be evaluated using the method of Polygonalization while margin will be calculated using the method of vertical scanning. The height of the t-bar on the stem of the alphabet 't' and word-slant would be calculated using template matching.

The proposed system can be used as a complementary tool by the graphologist to improve the accuracy of handwriting analysis and also make the process fast. It will also assist the HR/company employer in decision making regarding the suitability of an employee for the specific job and improving the retention of an employee. The future

work can include more features from the micro approach of handwriting analysis like the loops of alphabet 'f' and 'l', gradient, concavity of letters and so on in order to predict more accurate results.

- **A paper based on Artificial Neural Network by Dr. KR Ananda Kumar,** explores the implementation of a machine learning approach in the field of handwriting analysis. This paper proposed a method to predict the personality traits of a person by analyzing the baseline, pen pressure and the letter t' as found in the individual's handwriting sample. These extracted features are then given as an input to the artificial neural network which in turn gives output as personality traits to the user. The future work discussed in this paper are, including more features of the handwriting like the size of the letters and the margins as inputs for personality trait determination to improve the system output.

1.6 Existing Solutions

- Handwriting analysis based on Segmentation method using Support Vector Machines. Segmentation is used to calculate the features from digital handwriting and is trained to SVM which outputs the behavior of the writer.
- Hough transform is a technique used to detect any arbitrary shape in an image by creating a table for storing all the edge pixels of the target shape.
- Polygonalization method is a technique which involves a closed polygon produced around a line in the scanned image of the handwritten text. The slope of the text/alphabet is found using the coordinates of the polygon.
- Template matching with certain predefined templates is also used as a technique for behavioral analysis. Methods discussed are not simple to automate and therefore a need exists for a simple solution that could be automated easily.

- Generalized Hough Transform(GHT) is used to determine the shape of the lower loop of letter 'y' and the height of the t-bar on the stem of the letter 't' is calculated using template matching.

1.7 Objective

a) Users should be able to upload the image.
b) System should be able to pre process the given input to suppress the background.
c) System should detect text regions present in the image.
d) System should retrieve text present in the image and display them to the user.
e) The main attention of our project is feature extraction and classification.

1.8 Significance

Many people want to know why they should go for handwriting analysis. Handwriting analysis is both important and necessary for us at some level. There are varieties of reasons why a person must get his handwriting analyzed at least once. The main reason is handwriting analysis unfolds many things about your personality. It is a tool to know your strengths and weak points. Till about a few years ago, handwriting analysis was restricted to forensic experts. Not anymore. Today, several multinational corporate firms use handwriting analysis to know personality traits of their job candidates. Handwriting tells about your suitability for the job, your motivational level, creativity, leadership, teamwork etc. In fact, we would suggest you to not only get your handwriting analyzed; you should also know the art of handwriting analysis.

Chapter - 2

2.1 Specification

► **Software Specifications:**

- Operating Systems: Windows
- Language used: Python
- Platform Used : Visual Code

► **Hardware Specifications:**

- 1 PC/Laptop
- Processors: Any Intel or AMD x86 processor supporting SSE2 instruction set.
- RAM: 1024 MB (At least 2048 MB recommended).

2.2 Requirements

Requirement specification encompasses the needs of this project. It aims at providing a full description of the requirements based on the concepts defined in the Problem Domain. The system requirement specification is produced at the culmination of the analysis task. The function and performance allocated to software as part of system engineering are refined by establishing a complete information description, a detailed functional description, representation of system behavior, an indication of performance requirements and design constraints, appropriate validation criteria and other information pertinent to requirements.

Functional Requirements

- Sample Training set will be generated using 100 samples of handwriting, which will be examined by a professional Graphologist
- The Graph logical analysis will be done using Computer Aided Technology. This process will help automate the process of analyzing real world text samples as well as efficiently predicting the personality traits of the writer.

- Based on the characteristics corresponding to a particular trait, as identified by the graphologist, the Feature Vector Matrix will be created.
- The Feature Vector Matrix for a new handwriting sample will be created and then its similarity with the trained dataset would be calculated using the similarity matrix method.

Non-Functional Requirements

► **Performance:** Handwritten characters in the input image must be processed accurately.

► **Functionality:** This system will deliver on the functional requirements mentioned in this document.

► **Flexibility:** Easy to use and flexibility to upload multiple image formats.

► **Learn Ability:** Trained to recognize handwriting features to their corresponding personality traits.

Chapter - 3

3.1 Problem Statement

Automated handwriting analysis can be used to examine personal traits of candidates during interviews accurately as the accuracy of an analyst highly depends on his skill set. Also hiring a graphologist to analyze hundreds of samples for recruitment purposes will be time consuming and not feasible economically.

The scope of this project is to determine a methodology for analyzing real world handwritten text samples with the aid of technology. Various parameters of the handwritten samples like Margin, Baseline and Slant will be taken into consideration to determine corresponding traits.

The objective of this work is to also determine a proposed tool that will complement the graphologists to increase their speed and efficiency in the analysis process.

3.2 Features Recognized

Baseline

► The imaginary line along which the writer aligns the bottoms of the middle zone letters when asked to write on a blank paper is called the Baseline.

► Slanting downwards, rising upwards and level are the three most common baselines found in handwriting analysis.

► The emotional stability and disposition of a writer is judged by the baseline in the handwriting. It is obtained by finding the coordinate of all the pixels and the pixels are fitted to a line from left to right.

| BASELINE	Corresponding Traits
Ascending	Optimistic
Descending	Pessimistic
Level straight	Balanced

Margin

► Margin is the space left while writing either to the left or to the right.

► This feature reveals the exquisite temperament, adjustment and social interaction of the person with the society.

► It is obtained by applying the bounding box to whole paragraph .

► The bounding box gives us values so that we can get the margin distances from those parameters.

MARGIN	Corresponding Traits
Wide left	Courageous
Wide right	A reserved person
No magin	Insecure
Even margin	Self disciplined

Letter Slant

► The slant of the letter is used to understand whether an individual's handwriting is inclined towards the right or towards the left or is it vertical. Slant of the handwriting indicates the emotional interactions of the writer.

► There are five classes in slant. To determine the slant, the following method is proposed. In the grid of every individual letter that would be obtained after template matching, a line will be drawn between the lowermost and the uppermost points of that letter. Slope of this line will be calculated and similarly it would be calculated for all other letters. The average of all slopes calculated will give the slant of the letters.
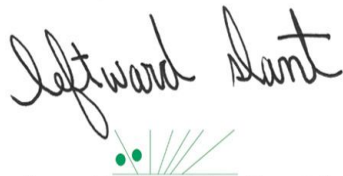
Shown by a consistent leftward slant. The upstrokes and slant are a minimum of 90 degrees from the baseline. The smaller the writing size, the more introverted. This person keeps his own counsel, rarely expresses feelings, and makes logical, unemotional decisions.

emotionally withdrawn

LETTER SLANT	Corresponding Trait
Extremely Left	Fear of the future, defensive
Extremely Right	Lack of self-control, impulsive
Left	Reflective, independent
Right	Expressive, freedom of thought
Vertical	Head controls over heart, independent emotional nature

Size of the letter

- ➤ Size has its own significance in the process of handwriting. Size is an unfixed trait; it is determined by the average height.

- ➤ To get the average height, the image is scanned across column by column to find the first black pixel and the last black pixel.

Size of the letter	Corresponding Trait
Big	High Self Esteem
Medium	Emotionally stable
Small	Generous

Chapter - 4

System Design

System design describes any constraints in the system design and includes any assumptions made by the project team in developing the system design. It is the process or art of defining the hardware and software architecture, components, modules, interfaces, and data for a computer system to satisfy specified requirements. One could see it as the application of systems theory to computing. The design of the system is essentially a blueprint, or a plan for a solution for the system. We consider a system to be a set of components with clear and defined behavior, which interacts with each other in a fixed, defined manner, to produce some behavior or services to its environment. Our system is a software, which can be used to predict the personality of the person by his handwriting
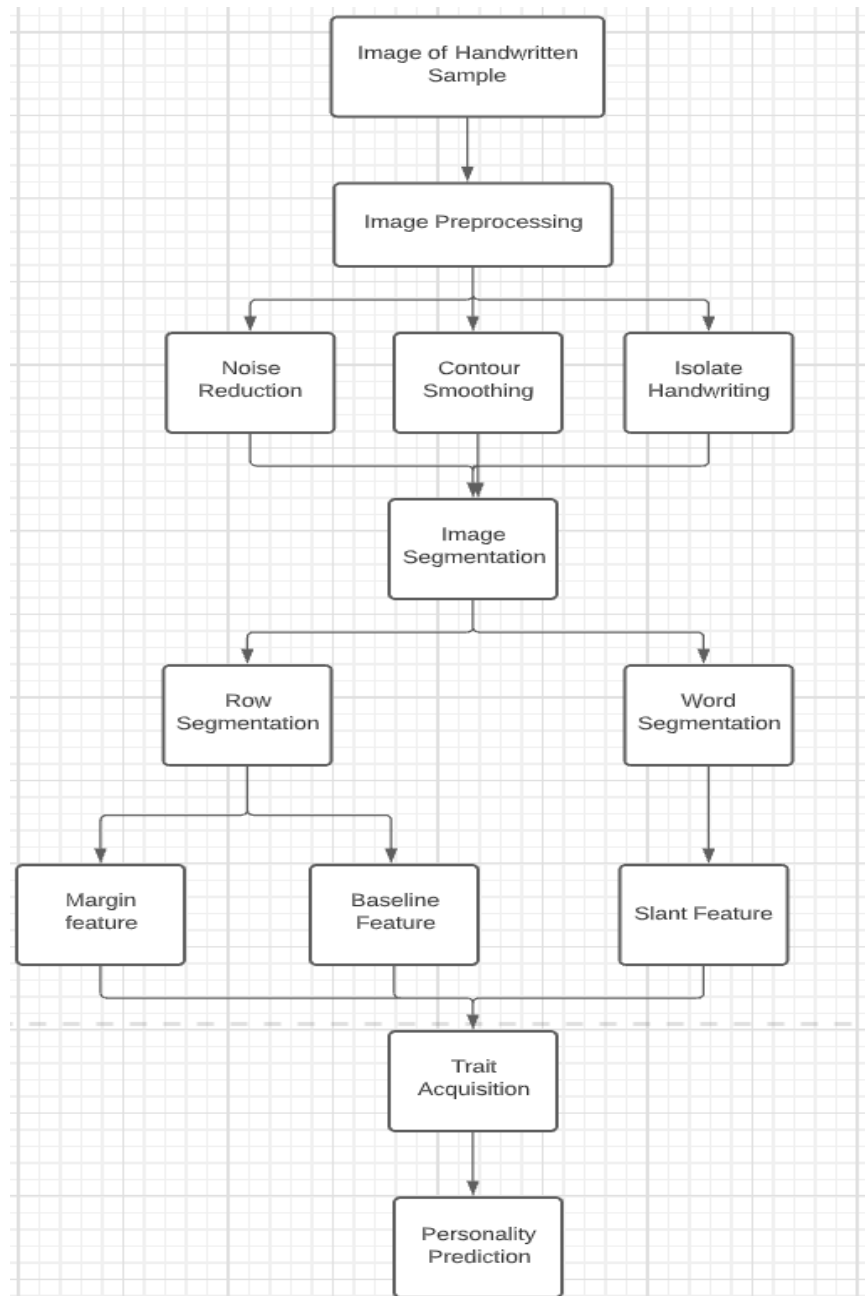
4.1 Block Diagram

Figure 1 Block Diagram

Figure 1 shows the steps used to solve personality prediction problems. It is easier to comprehend a system's functions using a block diagram, which also helps connect its components. The rectangular represents the operation performed by the system using the algorithm.

A. Handwritten Sample

Data from the IAM Handwriting Database of Research Group on Computer Vision and Artificial Intelligence INF, University of Bern, Switzerland is obtained. This database contains forms of handwritten English text suitable for training and testing personality prediction programs and identifying the behavior of writers.

B. Image Preprocessing

Data images are enhanced by preprocessing, such as noise reduction, contour smoothing, etc., before computational processing to eliminate background noise. As a result of the noise introduced in capturing or scanning the images, the quality of the input images may be degraded. The quality of the pictures necessary for experimentation should be improved by removing the background noise.

I.   Noise Reduction: The term noise refers to a random variation of image intensity that can be observed as grains in an image. It is often a result of capturing or transmitting an image. In an image with noise, the pixels have different intensity values than their actual pixel values. By smoothing the entire image with no distortion, noise removal algorithms reduce or remove the visibility of noise.

II.   Contour Smoothing: As a result of the algorithm included in the work, we have been able to eliminate possible glitches caused by unwanted hand movements during writing. Therefore, only the strokes pertinent to our analysis were kept and analyzed. The contour point positions obtained from our algorithm are essential for handwriting analysis.

III.   Isolate Handwriting: We recursively cut the page into two dimensions until only the handwritten text is left to preserve the handwritten text for the next steps of our handwriting analysis task.

C. An analysis of the handwriting sample is based on segmented segments in this process, namely line segmentation, and word segmentation.

I.   Row Segmentation: To analyze the writer's personality traits, the paragraph is divided into separate rows so that familiar features such as the baseline and margin can be extracted.

   ● Margin feature: Margin is the extra space left after writing either to the left or the right. The boundary box is applied to the entire paragraph to get it. Based on the bounding box, we can find the margins, i.e. [left, top, width, and height], to calculate margin distances.

   ● Baseline feature: The baseline refers to the imaginary line that the writer draws on the blank paper. The pixels are matched to a line from left to right once the coordinates of all the pixels are determined.

II. Word Segmentation: To determine slant, this segmentation procedure segments the words in digital handwriting documents. The image is dilated first to connect all of the letters in words, and then each word is clipped out.

   ● Slant Feature: Vertical, rightward, and leftward slant are the three types of slants. It's calculated by calculating an image's mean angle by looking for maximum projection values and then changing the picture using the geometric transformation.

III. Margin, baseline, and slant feature each gives us a unique trait of the writer, and from the traits, the overall personality of the writer is predicted.
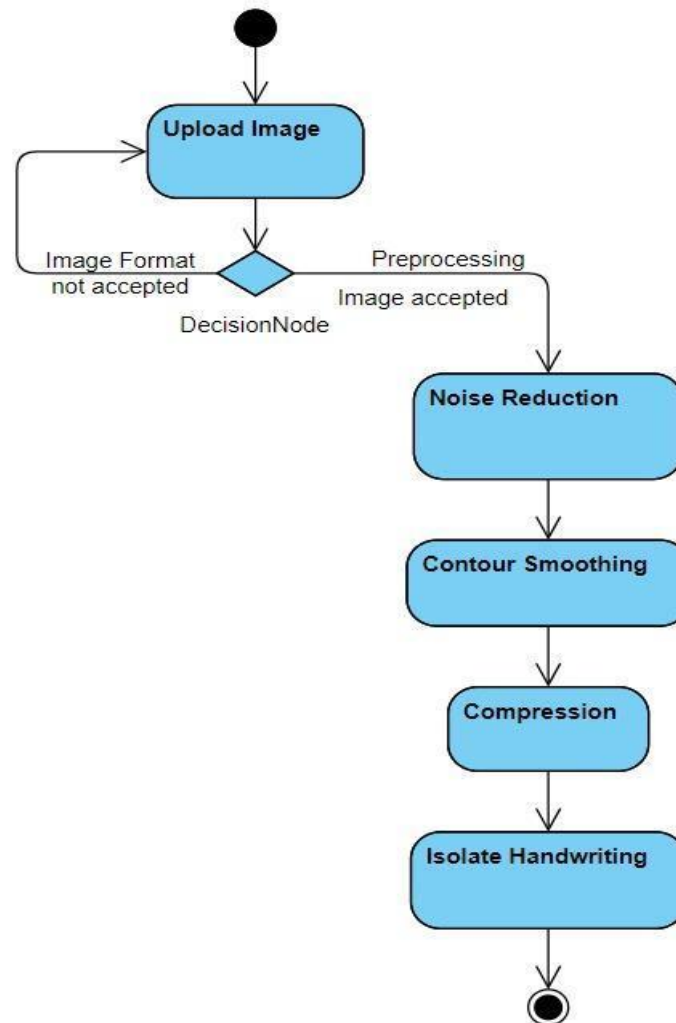
4.2 Activity Diagram

Figure 3 shows the activity diagram of the uploading the image and preprocessing the image

Row Segmentation

No — Are there unprocessed rows — yes

Fetch next row

No — Is this First row — Yes

Compute spacing between lines

Compute Margin feature

Compute Letter Size

Compute Baseline Feature

Word Segmentation

No — Is there unprocessed word — yes
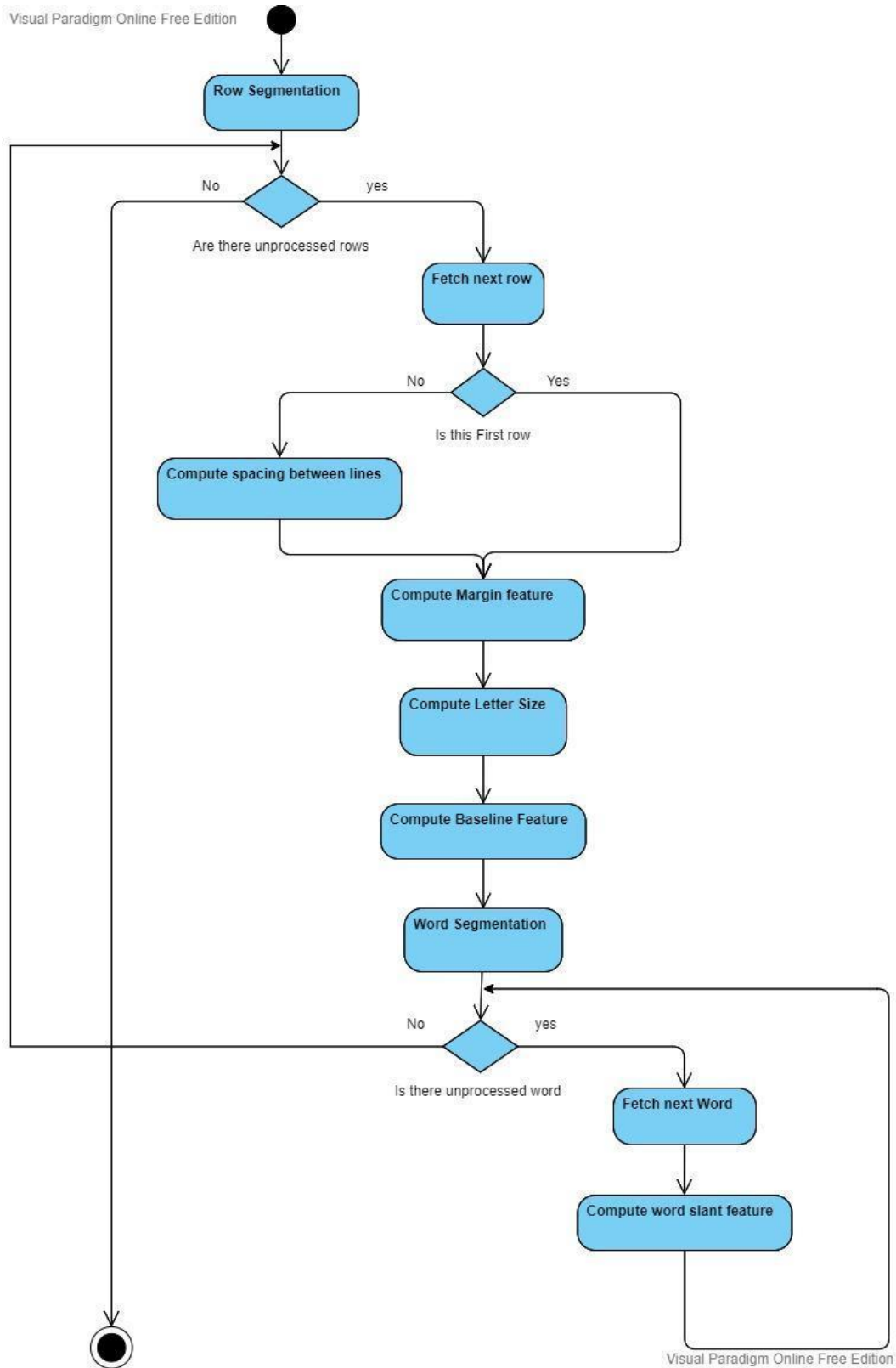
Fetch next Word

Compute word slant feature

Figure 4 This Figure shows the image segmentation and feature extraction

4.3 Class Diagram

In software engineering, a class diagram in the Unified Modeling Language is a type of static structure diagram that describes the structure of a system by showing the system's classes, their attributes, operations, and the relationships among objects.
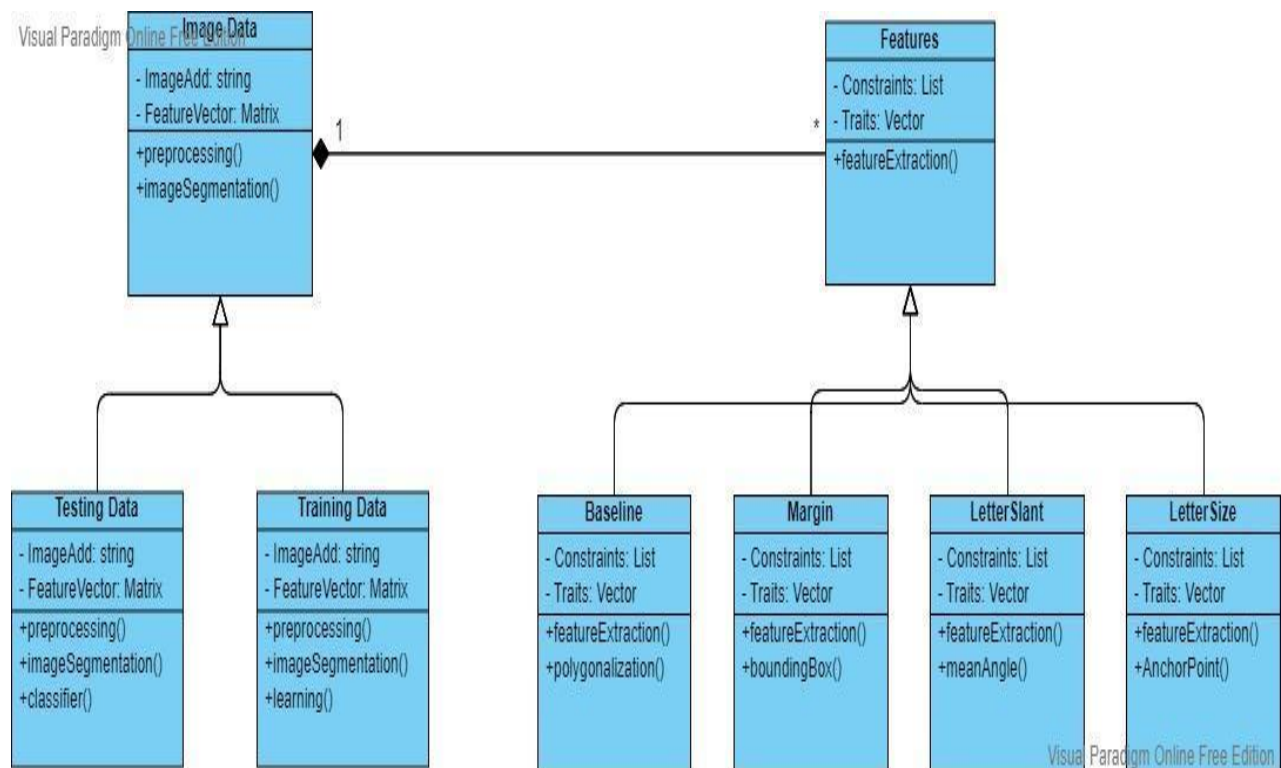
Chapter – 5

5.1 Inversion

This process is done to invert all the colors from the original image. In this case, the image will be converted into black and white. Image at threshold values 50, 100, 200, and 255. The image at threshold value 255 gives the accurate output for inversion.

The visibility of the image increases with the increase in inversion value. This is because the pixel value increases from 0 (black) to 255 (white). This is done thereby keeping in mind that the maximum value for a pixel can only be 255, which will be the brightest and thus be taken in this process.

5.2 Dilation

This process is used mainly in Contour Formation. Two consecutive lines can easily be segmented for classifications of traits by doing so. Image output at dilated value (5,50), (5,100), (5,200) and (5, 10000). The dilated value (5,100) gives more accurate results.The dilation increases the foreground object area or the image pixels. A kernel value is set to achieve this. From the results obtained, it is clear that the contour formation decreases after a certain threshold with the increase in dilation value. In this case, the kernel value (5,100) gives the best output in contour formation.

5.3 Changes with Anchor Point

In this process, the Line Segmentation is done by selecting the value of the Anchor Point. This value is taken at random, and depending on the result obtained and verifying it manually, an appropriate value is fixed upon.

For the result, multiple values of Anchor Point are taken, ranging from 1000 to 15000. Since the accuracy of this process can only be determined manually, this test is carried out on multiple images before fixing an appropriate value of Anchor Point.

5.4 SVM

SVM classifiers are used to recognize the letters and words in the handwritten samples accurately which are then used for Template Matching. This involves several steps, where each step has its benefit for the recognition process.

- Image processing
- Feature Extraction
- Features Normalization
- Classification and Recognition

Chapter – 6

6.1 Image processing

Image Preprocessing is the technique of enhancing data images in order to remove background noise. The quality of the image may get disturbed due to noise that is introduced in the process of scanning or capturing the images. It is necessary to remove the background noise to improve the quality of the image to be used in the experiment, so in these processes different techniques like contour smoothing, noise reduction techniques are used to remove background noise.

Methods of Image Processing There are two methods available in Image Processing, analog Image Processing and digital image processing. Various techniques have been developed in image processing during the last four to five decades. Most of the techniques are developed for enhancing images obtained from unmanned spacecrafts, space probes and military reconnaissance flights. Image Processing systems are becoming popular due to easy availability of powerful personal computers, large size memory devices, graphics softwares etc. The common steps in image processing are image scanning, storing, enhancing and interpretation. After converting the image into bit information, processing is performed. This processing technique may be, Image enhancement, Image restoration, and Image compression.

Image Processing Operations

This section discusses image processing operations, which are used in this project.

1. Image Segmentation

Image segmentation is the process of partitioning a digital image into multiple segments (sets of pixels, also known as super pixels). The goal of segmentation is to simplify and/or change the representation of an image into something that is more meaningful and easier to analyze. Image segmentation is typically used to locate objects and boundaries (lines, curves, etc.) in images. More precisely, image segmentation is the process of assigning a label to every pixel in an image such that pixels with the same label share certain visual characteristics.

2. Image Thresholding

In image processing, thresholding is used to split an image into smaller segments, or junks, using at least one color or grayscale value to do their boundary.

6.2 Thinning

A large amount of variability may be present among the handwritten words because writers may use different types of pens of unequal stroke width. Hence thinning of letters is done in order to remove the variability.

Thinning is an image morphological operation in which selected foreground pixels are removed by erasing an image. The gray scale images are then converted in a binary matrix format. The resultant binary images have values of 0 each for all the foreground black pixels and 1 each for all the background white pixels.

6.2 Segmentation

In this part of the process, the sample of the handwriting is segmented into three different parts for the efficient analysis, which is line segmentation, word segmentation and letter segmentation.

a) Line Segmentation: Firstly the paragraph is divided into separate lines so that features like baseline and spacing between words is extracted to examine the writer's emotional stability and dispositions in the baseline of the writing.

b) Word Segmentation: Segmentation of word, This segmentation process is used to segment the words in digital handwriting document to calculate different features like slant related to words.

6.3 Classification

The extracted features are then evaluated to find the writers personality which is classified into 34 unique personality traits, namely,

Optimism	Pessimism	Depression	Others Value	Stable
Conservative	Observant	Adjustment	Self Esteem	Joy
Concentration	Self Confidence	Goal Oriented	Management	Fear
Extrovert	Introvert	Generous	Selfish	Polite
Honesty	Emotional	Practical	Imagination	Vision
Dependent	Talkative	Relate	Expressive	Trust
Self Control	Perfectionist	Initiation	Listener	

Chapter – 7

Testing and Results

7.1 Introduction

Testing is intended to show that a system conforms to its specification and that the system meets the expectation of the user of the system. Large systems are built out of the subsystems which are built out of modules which are composed of procedures and functions. The testing process should therefore proceed in stages where testing is carried out incrementally in conjunction with system implementation. The testing of the learning tool was done along with the implementation of the various modules.

This method of testing helps to ensure the proper working of the modules at the time of their implementation. Testing involves exercising the real data processed by the program. The existence of program defects or inadequacies is inferred from unexpected system outputs. For verification and validation, we make use of the program testing technique.

7.2 Unit Testing

The primary goal of unit testing is to take the smallest piece of testable software in the application, isolate it from the remainder of the code, and determine whether it behaves exactly as you expect. Each unit is tested separately before integrating them into modules to test the interfaces between modules. Unit testing has proven its value in that a large percentage of defects are identified during its use. This type of testing is driven by the architecture and implementation teams. This focus is also called black-box testing because only the details of the interface are visible to the test. Limits that are global to a unit are tested here.

7.3 Results

The accuracy of each classifier is calculated using *the accuracy score()* function of the *sklearn* module, which compares the output of the testing dataset with the production of the training dataset and gives the accuracy between 0.0 to 1.0.

The overall accuracy of the project can be calculated as an average of these accuracy scores,

This came out to be equal to 0.950 approximately.

Chapter 8

8.1 Conclusion and future work

In this work, the unique personality is predicted using a Support Vector Machine with various parameters such as baseline angle, margin width, letter size, and slant of handwriting, and the optimal parameters are found. The model's accuracy came out to be 95.05% with the optimal parameters' values; inversion as 255 pixels, dilation at kernel value (5,100), Anchor Point value 6000, and Baseline threshold angle 45 degree, and these parameters give the best accuracy. There are still some ways in which the proposed work can be improved upon to make the prediction of personality more efficient. Furthermore, improvements can also be made using machine learning algorithms other than Support Vector Machine to get the optimal result. Apart from this, a recommender system may improve its prediction accuracy by analyzing personalities from Facebook profile statuses by recommending items, such as TV shows, music, or sports events, according to the user's personality.

8.2 Applications

1. Personnel Selection

In the selection of personnel, handwriting analysis is an invaluable tool for helping to choose the most suitable person for the job. It is also useful in existing job situations where major personality differences can have an adverse effect on employer - employee interaction as well as on the managerial staff itself. Handwriting analysis allows us to focus on these problem areas in order to correct them.

2. Psychological analysis

Graphology has been used clinically by European counselors and psychotherapists. When it is used, it is generally used alongside other projective personality assessment tools, and not in isolation. It is often used within individual psychotherapy, marital counseling, or vocational counseling.

3. Medical diagnosis

Research studies have been conducted in which a detailed examination of handwriting factors, particularly timing, fluidity, pressure, and consistency of size, form, speed, and pressure are considered in the process of evaluating patients and their response to pharmacological therapeutic agents.

The study of these phenomena is a by-product of researchers investigating motor control processes and the interaction of nervous, anatomical, and biomechanical systems of the body.

4. Investigations

When evaluating any mysterious handwritten letter or note, whether it be threatening, anonymous, suicide or unexplained, often a personality profile will reveal a great deal of insight about the type of person responsible for writing the note.When evaluating suicide notes, much can be learned about the emotional state of the writer at the time of writing, but also earlier samples written by that person can provide insight into differences between the emotional state of the writer in earlier days/weeks/months and during the writing of their last note.

5. Employment Profiling

A writing sample of an applicant is analyzed to match his profile with the ideal psychological profile of an employee required for the vacancy. A graphologist report is used for comprehensive background checks, practical demonstration or as a record of work skills. The analysis can be a complete failure and may have also proved to be highly successful in certain cases. In most cases, traditional methods are preferred in the employment process where there is an absence of evidence of a direct link between handwriting analysis and various measures of job performance. The use of graphology in the hiring process has been facing criticism since ages on ethical grounds and legal grounds.

6. Career Guidance

Handwriting analysis is also used for career guidance where knowledge about personality is essential in order to match the individual correctly to the type of work that would best suit his or her personality and interests.

7. Personal Relationships

Handwriting is analyzed to check compatibility in all types of personal relationships. Where family members are incompatible or find themselves in situations of conflict, an analysis of their handwritings will highlight the areas where personality clashes occur.In all areas of communication, and wherever people need to understand one another better, be it of personal, social or professional nature, handwriting analysis can provide deep insight and penetrating revelations.

8.3 References

[1]   V. Patil and H. Mathur, "A Survey: Machine Learning Approach for Personality Analysis and Writer Identification through Handwriting," in *2020 International Conference on Inventive Computation Technologies (ICICT)*, 2020, pp. 1–5.

[2]   L. Durga and R. Deepu, "Ensemble deep learning to classify specific types of t and i patterns in graphology," *Global Transitions Proceedings*, vol. 2, no. 2, pp. 287–293, Nov. 2021.

[3]   S. Asra and D. C. Shubhangi, "Human behavior recognition based on hand written cursives by SVM classifier," in *2017 International Conference on Electrical, Electronics, Communication, Computer, and Optimization Techniques (ICEECCOT)*, 2017, pp. 260–268.

[4]   H. N. Champa and K. R. AnandaKumar, "Automated Human Behavior Prediction through Handwriting Analysis," in *2010 First International Conference on Integrated Intelligent Computing*, 2010, pp. 160–165.

[5]   R. Litoriya, S. Soner, and M. Lovanshi, "Preventing Road Accidents Caused by Drunk Driving Using Machine Learning & Internet of ThingsNo Title," *International Journal of Advanced Science and Technology*, vol. 29, no. 3, pp. 11060–11070, 2020.

[6] S. Başaran and O. H. Ejimogu, "A Neural Network Approach for Predicting Personality From Facebook Data," *SAGE Open*, vol. 11, no. 3, p. 215824402110321, Jul. 2021.

[7] D. Niranjan, M. Kavya, K. T. Neethi, K. M. Prarthan, and B. Manjuprasad, "Machine learning based analysis of pulse rate using Panchamahabhutas and Ayurveda," *International Journal of Information Technology*, vol. 13, no. 4, pp. 1667–1670, Aug. 2021.

[8] A. M. U. D. Khanday, Q. R. Khan, and S. T. Rabani, "Identifying propaganda from online social networks during COVID-19 using machine learning techniques," *International Journal of Information Technology*, vol. 13, no. 1, pp. 115–122, Feb. 2021.

[9] S. Ghosh, "Identifying click baits using various machine learning and deep learning techniques," *International Journal of Information Technology*, vol. 13, no. 3, pp. 1235–1242, Jun. 2021.

[10] N. Sharma, R. Litoriya, and A. Sharma, "Application and Analysis of K-Means Algorithms on a Decision Support Framework for Municipal Solid Waste Management," in *Advanced Machine Learning Technologies and Applications. AMLTA 2020. Advances in Intelligent Systems and Computing*, H. A., B. R., and D. A., Eds. Springer, Singapore, 2021, pp. 267–276.

[11] P. Pandey and R. Litoriya, "Fuzzy Cognitive Mapping Analysis to Recommend Machine Learning based Effort Estimation Technique for Web Applications," *International Journal of Fuzzy Systems*, vol. 22, no. 4, pp. 1212–1223, 2020.

[12] A. R. Patil and S. Subbaraman, "Performance analysis of static hand gesture recognition approaches using artificial neural network, support vector machine and two stream based transfer learning approach," *International Journal of Information Technology*, Nov. 2021.

[13] Aulia, M. Reza, and E. C. Djamal, "Personality Identification Based on Handwritten Signature Using Convolutional Neural Networks," in *Proceedings of the 5 th NA International Conference on Industrial Engineering and Operations Management*, 2020, pp. 1761–1772.

[14] A. Chitlangia and G. Malathi, "Handwriting Analysis based on Histogram of Oriented Gradient for Predicting Personality traits using SVM," *Procedia Computer Science*, vol. 165, pp. 384–390, 2019.

[15] M. Gavrilescu and N. Vizireanu, "Predicting the Big Five personality traits from handwriting," *EURASIP Journal on Image and Video Processing*, vol. 2018, no. 1, p. 57, Dec. 2018.

[16] N. Sharma, R. Litoriya, D. Sharma, and H. P. Singh, "Designing a Decision Support Framework for Municipal Solid Waste Management," *International Journal on Emerging Technologies*, vol. 10, no. 4, pp. 374–379, 2019.

[17] P. Pandey and R. Litoriya, "An IoT Assisted System for Generating Emergency Alerts Using Routine Analysis," *Wireless Personal Communications*, vol. 112, no. 1, pp. 607–630, May 2020.

[18] S. Soner, R. Litoriya, and P. Pandey, "Exploring Blockchain and Smart Contract Technology for Reliable and Secure Land Registration and Record Management," *Wireless Personal Communications*, vol. 121, no. 1, pp. 2495–2509, Aug. 2021.

[19] M. Pandey, R. Litoriya, and P. Pandey, "Novel Approach for Mobile Based App Development Incorporating MAAF," *Wireless Personal Communications*, vol. 107, pp. 1687–1708, 2019.

[20] M. Pandey, R. Litoriya, and P. Pandey, "Validation of Existing Software Effort Estimation Techniques in Context with Mobile Software Applications," *Wireless Personal Communications*, vol. 110, pp. 1659–1677, 2019.

[21] M. Pandey, R. Litoriya, and P. Pandey, "Novel Approach for Mobile Based App Development Incorporating MAAF," *Wireless Personal Communications*, vol. 107, no. 4, pp. 1687–1708, 2019.

[22] S. A. Ayyadevara, P. N. V. S. R. Teja, and M. Rajesh Kumar, "Handwritten Character Recognition Using Unique Feature Extraction Technique," in *2018 3rd IEEE International Conference on Recent Trends in Electronics, Information & Communication Technology (RTEICT)*, 2018, pp. 1239–1243.

[23] P. Joshi, A. Agarwal, A. Dhavale, R. Suryavanshi, and S. Kodolikar, "Handwriting Analysis for Detection of Personality Traits using Machine Learning Approach," *International Journal of Computer Applications*, vol. 130, no. 15, pp. 40–45, Nov. 2015.

[24] H. N. Champa and D. K. R. AnandaKumar, "Artificial Neural Network for Human Behavior Prediction through Handwriting Analysis," *International Journal of Computer Applications*, vol. 2, no. 2, pp. 36–41, May 2010.

[25] S. Prasad, V. K. Singh, and A. Sapre, "Handwriting Analysis based on Segmentation Method for Prediction of Human Personality using Support Vector Machine," *International Journal of Computer Applications*, vol. 8, no. 12, pp. 25–29, Oct. 2010.

[26] L. S. OLIVEIRA, E. JUSTINO, C. FREITAS, and R. SABOURIN, "The Graphology Applied to Signature Verification."