**ABSTRACT**

Personality is a trait human possess to make himself identify to the community which describes the abilities of how he/she is reacting or behaving with the people around them. There are so many techniques to identify personality. One of them is Handwriting. Handwriting speaks about an individual. It is one of the unique characteristics to represent what is going in our minds, how to communicate with others? and so on. The way our words are organized on a page reflects both the state of our hearts and minds. For instance, sloppy handwriting and a variety of mistakes indicate confusion, whereas neat handwriting demonstrates excellent mental and motor coordination. Handwriting analysis is a successful and trustworthy indicator of personality and behavior. The analysis of handwriting is done by the method called as Graphology. In order to implement the automated handwriting analysis tool, a novel approach of machine learning technique is discussed i.e., Graphology, which is a study of identifying the patterns involved in handwriting sample through various characteristic’s involved such as line spacing, word spacing, slant etc. In order to complete the project and to automate or reduce the work load on Graphologist, we have used Deep Learning Technique to identify the personality of a person. Deep learning is like a child node of much bigger field i.e., Artificial Intelligence which helps us to make ideas into reality. These Deep learning filed has got the name has neural network is as they help us achieve human possible things by mimicking the way the human brain works, as in the form of how neurons interact with each other in the brain, similarly the neural network transmits information from one neuron to other neurons in the network. Deep Convolutional Neural Networks had been incorporated which performs very well in classifying images or detecting objects in the frame. The dataset consists of two classes and around 50 images in each class, these images have been loaded and trained on CNN to classify what personality a person is using his/her handwriting sample given as an input.

**PROBLEM STATEMENT**

To build a model which helps us to identify what type of personality the person is inculcated inside him with the help of his/her handwriting. As we are trying to solve the problem of identifying type of personality using his or her handwriting, so we have you used artificial intelligence to help solve this particular problem as we all are aware of the fact that how artificial intelligence has been evolved through the decades and helping large number of businesses across the world to achieve what they have dreamt of and increase their Business in terms of revenue and reputation. As we know how hectic this labelling can be especially when we think of it doing manually, scanning a document line by line, examining it word by word and letter by letter, this can take a lot of time and effort of the individual, so in order to reduce amount of workload that comes on top of a graphologist who is responsible to tell what type of personality a person is given or understanding or studying his or her handwriting this might take a lot of time when it is done manually by a graphologist which involves a lot of characteristics to be examined such as examine baseline, slant of the letters written by an individual, letter size, pen pressure, word spacing, the curve of individual letters whether they are tilting towards left side or right side as other than these there are different types of characteristics but most importantly most graphologist tend to use or examine individual letter T which plays a very significant role in any graphologist which allows the graphologist 2 exactly identify the personality of a person there are a variety of ways a person can write a **T** which are mentioned below in the data set section like how does the person cut the **T,** whether the individual cutting the letter at a root or is cutting above the root, there are like more than three different ways of writing a **T** based upon these the graphologist film able to approximately predict the personality of other than this there are also other characteristics which specifically get the graphologist observes but we have mostly concentrated on the above mentioned characteristics. Although identifying personality of a person using and rating is requires a lot of time and effort to be put in order to get the approximate personality of a person which we think also a graphologist can do this job very accurately but given enough amount of data to a machine it can predict the personality much better than the graphologist, these has got many advantages and also disadvantages such as when examining any input which has a very high pen pressure on a paper this might result in the model to not accurately get what the personality of the person is but whereas given the same input to a graphologist he or she can approximately tell the personality of a person because of that sensation is a graphologist at which he or she could know by simply moving a finger through the letters written by the individual where she could get the exact state of how much pen pressure or pencil pressure has been applied by the individual. The main questions that were answered in this particular project where hey there the processing technique that we were using is reliable and can be trusted is reliable and can be trusted, show the processing technique that we have used in order to solve this particular project was carefully processing the image by cropping and applying OpenCV filter enhance the image or samples these samples are then fed said to the model to be trained for making predictions ultimately this will produce a deep learning model which will eventually help us alright produce result in the form of the characteristic involved based on the personality and his/her handwriting style sample passed as an input. The other question other question that is will the accuracy effect by any condition of environment the picture is clicked are the paper being used in upper case as the project is still in development phase yes the condition of the environment the picture is closed does affect and also the type of paper which we are using to train must be similar to the paper which we are using while testing the model that is when the model is being trained we have used a normal A4 size sheet paper and made the individuals to write specific or random sentences on top of that paper and the images have been clicked in a well-lighted conditions during daytime. So, if we pass another type of paper is being used while testing surely the accuracy might decrease are the results might not be as accurate as we were hoping for because when we train use the different set of images and while testing we are using different type of images so this makes a lot of sense to the model so in order to avoid this in this development phase we have to use same set of papers for training and for testing moving forward the model will be train with a different set of images and different set of input data which has been written on different types of paper which can result in providing good accuracy when any input is past different kind of different type of paper as we have trained our model with variety of data we can hope that it will also give approximate or better results comparatively when it is train on one specific set of data.

**MOTIVATION**

The main motive of choosing this particular title of identifying personality of a particular individual using his/her handwriting is to get the idea of how does a graphologist manually test it, so in order to understand that we have gone through some of the papers and read through some articles of how does a graphologist does everything like scanning each and every detail of a piece of a paper to identify his or her personality, so at last, I came to a conclusion of replicating the same when given an input to a machine which can identify or detect the personality of a person when passed a piece of paper as an input to the machine. The main motivation of this particular title is that how fascinating it is to identify the personality of a person using machine, at first it was just an idea or a fake title of whether it can be solved or not but as the days progress we got to know more information about this which eventually helped us in trying and solving the problem using artificial intelligence. Moving forward this can not only help the graphologist in verifying the personality but also this will also be a very starting point to a lot of individuals would like to know and change their personality. So, that has motivated me of implementing a system which can in turn reduce the workload on a graphologist and the use of artificial intelligence which can help us detect personality of a person with minimal efforts and also saving a lot of time.

**OBJECTIVE**

The main objective of this particular project is to reduce the workload on a graphologist, to save time and efficiency, to increase the accuracy of identifying the personality trait of a person using handwriting sample. As we know traditional way of identifying personality involves in putting a lot of effort and time to yield better results but in order to reduce this issues to some extent and provide some insights to the graphologist we have incorporated to use artificial intelligence that is deep learning to solve this issue there individual handwriting samples have been collected and label to be trained with her deep learning model to generate how to create a system which can help us to identify the personality of a person when an image is passed.

**LITERATURE SURVEY**

In this Paper [1] the authors have used very complicated technique estimating the personality of the person using Hidden Markov model and multilayer perceptron which explains about classifying the input picture along with their specific properties. The office of meetings of hidden Markov and MLP which was specifically dedicated to different such as Denmark pause identify the properties of the handwriting image and MLP was used to classify those properties which Intern give us the reasons of classifying which personality the person was as this involved in using two different models the results were quite promising. The main reason for using hidden Markov model and MLP well as we know how neural networks work when given an image which calendar anything and everything when given right parameters which involves in using multiple layers such as convolutional layer which is a combination of two or more computations and activation function along with pooling layer which will give us a feature vector, this feature vector will be in the form of matrix which is then passed on to the next convolutional which then processes that matrix and flattens it based upon the output the weekend create necessary amount of dense layers in order to get the desired result for example output layer consists of as many neurons as the number of classes we are working on that is if we are working on two classes that is a binary class yes or no we can give a dense layer tumor to neurons which will give us the probability score of each class. Hidden Markov model is basically a statistic model which helps us to model times series, weather and speech recognition models and helps us to predict the states alrighty then using sequential data which were mentioned above, not only the application of hidden Markov model are limited to these but they also work very well with the image classification as the main aim of Markov model is to identify the rapidly changing states where there is a very good property of Markov model that is the future state which we are trying to predict totally depends upon the previous state/ event but not the older states. The simple explanation of this property can be as follows whenever we are trying to predict the weather off tomorrow, we need to have information of today is when we are trying to predict the property of the future state, we are making use of the current state that is today’s weather. The main terms involved in hidden Markov model R transition probabilities; these transition probabilities are then converted into matrix. The initial probability distribution D which health system start over system from a state and this is denoted by π, during implementation we can assign the same probability to all the states over here the states represent the number of classes that we are trying to predict in our case really trying to predict whether the weather is sunny rainy or windy some based upon these we can give as many initial probability distributions as the number of classes we have. The authors I finally concluded that the method they have used has given a superior finality over the others which involved in creating feature vectors music HLSC independent features such as line spacing provide spacing slant of the letters and other characteristic and then multilayer perceptron has been applied such that they output will be the personality trait of a person the authors have also concluded that this method as shown quite promising results they MMPI(Minnesota Multiphasic Personality Inventory) test on training step.

This paper [2] is based on Personality Detection using Handwriting Analysis. Handwriting analysis is a scientific tool for recognizing, assessing and to understand writer's personality with the help of shapes and word patterns present in the handwriting. Handwriting shows the true personality of the person including his behaviour, emotional outlay, anger, fears, self-esteem, honesty, imagination and many other personality traits. It is also known as Graphology. Each personality trait has some neurological brain pattern in the human brain. Each neurological brain pattern design delivers some kind of neuromuscular movement which is same for each person who has that specific personality trait. Each and every stroke or movement in handwriting reveals a particular personality trait. Graphology is the science that identifies the strokes in handwriting and describes the corresponding personality trait. The author added the related work done by different people previously in the field of handwriting analysis. He has mentioned some of the important handwriting features and personalities. Some of the features explained in the paper are zone, baseline, pressure, margin, slant, size, spacing, etc. Along with the handwriting features he has also described the personality characteristics of the person. To make the features clearer to everyone, he also put the handwriting samples added with feature types. Handwriting Analysis System includes steps such as Pre-processing, segmentation, feature extraction, classification. The author took few handwriting samples on a white A4 size paper and scanned the images of the handwritings using scanner. Those images are saved in PNG or JPEG format. The author has done pre-processing to enhance the image data as input and also done further processing to improve some image features qualities. Image pre-processing includes noise removal, binarization, and normalization. Binarization was used to convert the Gray scale images into binary images. Noise removal techniques were applied to remove the unwanted data and to improve quality of images. Normalization was used for removing some of the variations of handwriting styles and simplifying the shapes of symbols. Then, the author segmented the handwritten page into three different types of segments, i.e., line, word and character segment. Line segmentation was used to segment the handwritten image into text lines. Lines were further used for word segmentation that were used in feature extraction process. Character segmentation was used to segment the words into characters. Thereafter, feature extraction was applied which reduced the dimension or important data from a high dimensional input data were extracted. The output data was used for analysing the personality of the person. The author used classification method to recognize the personality traits of the writers. The various features extracted in the feature extraction step acted as inputs to the classifier. According to the feature values the personality of writer is identified. There are various other classification methods, classifiers or rule base system using which classification can be done. The author finally concluded that HAS identifies the personality of a writer, scanned images of handwritings of different persons are taken as input and a set of personality traits are produced as output. For getting more accurate results, more features can be added to HAS. This system can be further used to recognize the changes in the past and present personality of the writer through the change in the handwriting. A language independent human personality analysis tool can be created, which can recognize personality from different handwritings written in different languages. Using HAS, personality detection will be a useful and efficient system for personality traits classification. Handwriting analysis or graphology has wide scopes in the fields such as forensic, recruitment, medical diagnosis, psychology, human computer interaction. Handwriting represents the behaviour and personality of the writers so it can be used in recruitment and staff selection.

In this paper [3] the authors have discussed about the Personality Prediction based on Handwriting using CNN & MLP. Handwriting is exclusive to each individual as much as fingerprint is exclusive. So, the author focused on analysing various style of handwriting like ascending, descending and straight inclination of the baseline of a sentence, space between words, the right, left, irregular slants which will be used in identifying a set of traits associated with the individual, if it is present. The author divided the work into several modules: In the first module, image of handwriting is taken from the two datasets. Dataset-1 contains 1533 scanned images of handwriting. These images are cropped and saved as PNG images. On analysing Dataset-1, it is found that slight variation in features were observed that may affect accuracy, so, a new dataset ‘Dataset-2 is formed. CNN (Convolutional Neural Network) is an important set of techniques used for learning neural networks. It is a class of machine learning algorithms that uses multiple layers to extract higher level features from the raw input. Convolution is a mathematical operation that merges the two sets of information. In the case of CNN, convolution is applied to the input data for filtering the information and producing a feature map. This filter is also known as a kernel or feature detector. MLP (Multilevel Perceptron) is a class of feed forward artificial neural network. MLP consists of at least three layers of nodes namely, input layer, hidden layer and output layer. Except for the input node is a neuron which uses a non-linear activation function and for training MLP utilizes a supervised learning techniques called back propagation. In the second module the author has done pre-processing of the images. It is an important step in graphology, cropping and noise removal are of primary treatment. And then, grayscale and binarization has taken place, where an inverted binary image function is constructed such that those pixels which are above specific threshold (foreground) are converted to 255 and those which are below the threshold are (background) converted to 0. Horizontal projection of an image is the python list of sums of all the pixel values of each row of the images. Vertical projection is also a python list of sums of all the pixel values of each column of the image. Both the operations are performed on Gray scaled images, and then proceeded to contour & warp affine transformation. A contour is closed curve of points or line segments, which represents the boundaries of an object in an image. In other words, contours simply represent the shapes. Warp affine transformation is applied for rotating the contour that is found on an image so that the baseline of the handwriting is strictly horizontal. In the third module, the author extracted all the seven features from the pre-processed handwriting image, and saved to a csv file so that it can be used for further evaluation in the CNN model. After feature extraction, baseline, word space and slant are the three handwriting features that have been used to compare as the output in both the models. In MLP, all the seven features are taken to compare and in CNN only the pre-processed images are given as input and finally the outputs of both CNN and MLP are concatenated. For comparing personality of each individual is matched with its corresponding handwriting feature. Primary personalities are predicted based on combined CNN and MLP model whereas secondary personalities are predicted based on the seven extracted features. In the last module, for better accurate result the author passed the images along with the extracted features to a combined model consisting of CNN and MLP. By observing probabilities of different output labels, he predicted the most dominant class to which the test handwriting sample belongs or the personality traits that person has. In this project, by performing a handwriting analysis on the input image the author detected personalities or traits. The system compared the input image and features with the combined model. The key feature of this project was to extract all the possible traits using combined model (CNN and MLP).

**PLANNING AND METHODOLOGY**

We have taken samples of 10 individuals. Each individual has written 20 sentences each in an A4 sheet and images of these handwritings were taken. Those images were cropped and saved as PNG images. The images are saved in specific directories allotted to them with their label names, such as Optimistic and Stable. Class-1 named as ‘stable class contains 888 images whereas Class-2 named as ‘stable optimistic’ contains 907 images of the scanned text of handwriting.

Graphologists are the handwriting analysts who identify the characteristics traits of an individual just by examining the handwritten samples of the individual. This is a manual process; the skilfulness of the Graphologist defines the accuracy. But this a time consuming and we proposed methodology that focuses on the development of a computer aided tool with minimum human intervention. It would be able to predict the characteristic traits of a person intelligently. The most common methods that help in identifying the personality traits of a person through handwriting analysis are Baseline, writing pressure, slant of word, spacing between letters, words and lines, width of margins, size of letters, strokes connecting the letters, starting strokes, ending-strokes, etc. This project would be mainly 2 parameters such as Baseline, Height of the T-bar.

**Zone:** The slopes of the letters are broken into three zones – the upper zone, the middle zone, and the lower zone.  The Personality can be identified by analysing following three zones.

* Upper Zone represents: Future, upper body, conscious spiritual, intellectual and cultural aspirations, mental perceptions, concepts, fantasy.
* Middle Zone represents: Present, middle body, realistic, practical and social expression of the ego, emotional expression.
* Lower Zone represents: Past, Lower body, memory, sensual perception, basic drives, unconscious drives and biological needs.

**Baseline:** Baselines are the imaginary lines on which we write, indicating the goals and objectives of the person. We can think of a baseline as a road that leads to your goals or destination. The baseline can be ascending, descending, straight. It is used to find the emotional outlay and nature of the writer.

* Ascending baseline represents the person is Optimistic, cheerfulness, Hopefulness, stay busy and active, excitability, choleric behavior.
* Descending baseline shows that the person is pessimistic, mental tiredness of a temporary nature and has digestive trouble.
* Straight baseline tells that the person has stable outward behavior, realism, straightness and disciplined.

**Slant:** The slant of the writing is defined by the direction of letter slope and is measured by the angle formed between the letter (down stroke) and the baseline. The slant shows the emotions, degree of sentimental control, emotional control of the writer. There are three types of slants:

* Vertical slant
* Rightward slant
* Leftward slant

Height of ‘t’ bar: The lower-case letter 't' is another important feature that reveals lot of accurate information about the writer. A lowercase 't' can be written in different ways. This paper focusses specifically on the height of the “t” bar. The T-bar can cut the stem of the letter 't' at various positions like the upper-portion, lower-portion, mid-portion of the stem or not cut the stem at all. Depending on where the writer crosses the T-bar, different characteristics traits of a writer can be identified.

Template matching all the lowercase ‘t’ letters will be cancelled out and compared with the predefined templates. Four templates of letter ‘t’ will be predefined each consisting of T-bar at very high height on the stem, at a height very low on the stem, at mid- level on the stem and out of the stem respectively. using Hamming distance, the given template will be matched with the predefined ones. Hamming distance will measure the minimum number of substitutions required to change one into the other. The predefined template which is achieved with minimum number of substitutions on the input template will be the matching template.

**Pressure**: The amount of the force applied at the time of writing is known as pressure of pen. It can be light, heavy or medium.

* **Heavy pressure:** heavy pressure indicates that the writer has lot of energy. The energy mentioned here is physical as well as mental energy. This type of person may keep fighting to the end until he gets what he wants, he may also complete his tasks in spite of the difficulties that he faces and he may like physical activities.
* **Light pressure:** light pressure indicates that the writer has got less energy compared to a writer writing with heavy pressure. This person may prefer activities that require less physical work and he may also be a little [sensitive](https://www.2knowmyself.com/emotional_sensitivity). This person may also be more [flexible](https://www.2knowmyself.com/miscellaneous/flexability) than one who writes with heavy pressure and make him more adaptable to changing situations. Light pressure is sometimes referred to as feminine writing.
* **Moderate pressure:** In real life we aren't always going to find extremes. This means that we won’t be able to classify the writing sample of people under the two major categories of light pressure and heavy pressure.

**Word Spacing:** The space between the ending of first word and starting of the second word is called as word spacing. It represents the distance that the writer would like to maintain between himself and other people. Spacing describes the closeness of writer with the other people and his intelligence. Wide spacing shows independence, discrimination, exclusiveness, good taste, pride, snobbery, has clear thought, ability to organize his work. Narrow spacing describes inability to be alone, poor taste, obtrusiveness, friendliness.

**Margin:** Margin is the amount of space that the writer leaves in the left side or right side or the top or at the bottom of the page. Margins can either be narrow or wide with each telling certain personality traits about the writer.

* **Narrow left margin:** A narrow left margin may tell that the writer is attached more to his past than to his future and that his past could be having an effect on his future decision making.
* **Wide Left margin:** A wide left margin shows that the writer may not be much attached to his past and that he could leave it behind and continue moving.
* **Wide right margin:** A wide right margin shows that the writer may be afraid to take a future step as he may still be making his mind about it.
* **Narrow right margin:** A narrow right margin tells that the writer may be willing to take a forward step and that he may not be experiencing uncertainties at that period of his life.
* **Balance:** If the person left margins of a balanced size, then this indicates that he is a balanced person when it comes to risk taking and being attached to the past.

**STEPS UNDERTAKEN TO SOLVE THIS PROBLEM**

As the concept of deep learning and machine learning fascinates any human out, they’re in this universe, but rather this also has some hardcoding methodologies and to put lot of effort and time in making the data ready and delicious to be served to the user. With this we would like to introduce some of the main steps involved in any machine learning or deep learning project, that has helps us achieve the desired results which we are hoping for, they are:

1. Data Extracting / Data Creation
2. Data Cleaning / Data preprocessing
3. Model Creation
4. Training
5. Testing
6. Evaluation

**Data Extraction / Data Gathering:**

We all know that how data plays a very important role in any business and how big the impact might be if it is used in an efficient way, similarly with any machine learning project / deep learning project, the data which we gather or extract plays a very huge role in getting the desired results which we have hoping for. The reason this part is very crucial is and most of the time is being spent of 1 and 2 steps of this process, which boils us down to one-point i.e., if we don’t have right data for the solution which we are working on, it is no use of how many hours we sit looking training our model. So, in order to get the idea of which data to be chosen, we have to understand more about the solution we are trying to achieve or the problem statement given by the user, only then we will be able to be in a path of exploring the data we need. Once we get an idea of what type of data we will be using, we can either go with downloading the existing data from the data repositories or we can even create our own dataset. The data repositories are nothing but a place where we can find plethora of datasets of different domains, fields, types etc. Some of the most popular data repositories that exists are, Kaggle, UCI machine learning repository, Data at WHO (IF we are working on any medical related issue to be solved), Dataset Gallery(IF we are working on any Climate issues), Data.gov, Open Sanctions Dataset and a lot more. But, in most of the cases if we are into research level project, we might not be getting the right data that we are looking for, so in these cases, we might have to change the system, go out and create our own dataset, this will take a lot of time along with your effort, but will yield better results if the data collected is of good quality and also make us feel better when we complete the project and look in the past of how much effort we have put to collect the dataset. Types of data which are widely used in the businesses around the world.

1. Images
2. Video
3. Comma Separated Values

**Data Pre-processing / Data Cleaning:**

Another important steps which involve in putting a millionth of effort is data cleaning, Data cleaning is a very crucial part in any machine learning or deep learning project, which involves in pre-processing the data to make it ready to send it the model, which will eventually help us to make some good prediction on the data. There are numerous ways to pre-process the data and it varies by type of data that we are using. When we are dealing with images, we have to apply image pre-processing techniques involving converting images to Gray scale or normalizing them, applying various filters on top of image to enhance the quality of the image, when we are dealing with tabular data, we have to fill in the null values if there are any, deleting rows, columns or even imputing the null values. Different types of data which can be seen in the dataset are:

1. **Missing Data:** Missing data is a type of data which is null and will be empty, in order to remove this type of error in the dataset, we have applied imputation techniques to fill in the null vlaues. The imputations technique’s must be relevant and the value which we are trying to imputer must be relevant and appropriate to the vlaues which are in the column.
2. **Noisy Data:** Noisy Data are nothing but he outliers which are caused by the human errors, the other name for noisy data is outliers, which are the extremely low or extremely high values which are there in the specific column and are not so related to the other values.
3. **Inconsistent data:** Inconsistent data is a type of data error which mostly occurs whenever there is any duplicate data created or when the number in format of the data is not consistent, mistake with name of the columns or values. Rectifying these errors are very important because it might lead to create a bias in the model, which can result in generating inaccurate analysis.

**Model Selection:**

Once we extract and prepare the data for modelling after getting rid of all the errors and inconsistencies present inside the data, we need to make a pipeline should allow the users to train the model and make appropriate prediction. Choosing a right model is a very crucial step where, depending upon the type of output we are expecting and the type of data we got, we have to choose a model which can deliver us both of the above requirements mentioned. So, when we deal with Imagery dataset, we might be tentative to use Convolutional neural networks, as CNN are a very good starting point when we are trying to classify images or even identify different objects in an image, as Neural networks are known for its immense capability of learning anything and everything. When we deal with Tabular data we might go with more efficient and low-end models such as Decision tree, Support Vector Machines and Logistic Regression, and these algorithms also differ with which type of dataset we are working on. In our case, we have used Convolutional Neural networks, since it can extract and learn about the minute patterns/detailing which we are trying to capture in the image, which are the curves of the alphabets which plays a very important role in identifying the personality of the person.

**Model Training:** Once the appropriate model has been selected, we can then train the dataset which we have collected and cleaned to make some predictions. The Model training time might differ based on the various factors such as, the amount of dataset which we are using for training, if we are using neural network as base model, the training time might soar up with added hidden layers and number of neurons, which also increase the computations which happened inside the neurons.

**Model Testing:** Once the model is trained, it should be passed on for testing to check how good it is performing on the test dataset, Test dataset is a reserved data which we use it for testing, this data is accessed only after the model is trained, the reason we use new data for testing the model is we will get to know the exact accuracy of the model of how well our model is performing. **Note:** We should not combine both the train and test datasets to train the model, if this is done, we might be in situation of believing that our model is performing very well, but in fact it doesn’t perform very well when we pass a new data which the model has never seen, this issue we call it as Overfitting where the model will try to learn everything in the data, but unable to make predictions when a new data is passed. So, in order to avoid this, we divide our dataset to train and test and use these datasets separately wherever they are required.

**Model Evaluation:** Model evaluation is s final Phase of the project where we evaluate our model on the basis of various evaluation metrics available such as Confusion matrix, Precision, Recall, Accuracy, AUC\_ROC curve and classification report. These evaluation metrics are used different based on the type of data that we are dealing with. Although they can used interchangeably with imagery dataset and tabular dataset, most common, all the above-mentioned metrics are widely used for tabular dataset, which will give us a better understanding of what we are doing and how the model is trained. Although, these metrics can be used to evaluate the performance of the model which have been trained on imagery dataset and video formats, Now, in order to evaluate the performance of the model we just can’t use the above-mentioned metrics, so we have to make some transformations like converting the probabilities of the output layer values to a proper label format and then pass the actual\_y\_pred values and predicted labels to generate any metric mentioned above.

**DATASET DESCRIPTION**

The dataset used in this project was created completely from scratch from data collection to data pre-processing no eternal data repositories have been referred or used. The data was collected from multiple individuals in our office space with each individual writing their own thoughts into an A4 size sheet paper. The written paper then was clicked through a High-Definition mobile camera using photography stand to keep the mobile still while clicking the images. The images are then loaded onto an image cropping website to crop individual sentences written by the participants. The reason for using a specific website to crop is to maintain the scalability of the dimensions when cropped without any external software the dimensions would mismatch and it would allow us to then again manually identify the images of different dimensions and change it accordingly, so in order to avoid this a specific dimension size of the image has been recorded and those dimensions have been used for the entire dataset. Each individual has written around 15 to 20 sentences in the paper without any external disturbances around them, so that there handwriting wouldn’t alter. Another most important thing about this dataset is while collecting or creating the dataset, a Pilot pen has been used and given to every individual who had participated in this data creation, so as to maintain that same colour and fluid with the pen. The reason for using only one type of pen is because, when used another fluid pen the pen pressure might get altered or the curvatures might be as smooth as the writer wanted and not having a proper pen can also alter the individual’s mood of writing where he/she would feel like not writing at that moment which can make their handwriting to be altered. So, it was very important to maintain the uniformity across all the individuals. The dataset was created based on one characteristic of a human personality i.e., **Baseline and ‘T’ Letter,** this characteristic is a very basic and most important thing any graphologist notices when he/she makes an assumption about the person. So, once the dataset is collected and pre-processed, it has been sent for labelling, where each image with individual sentence has been thoroughly checked and labelled according to its personality (It the sentence has got an Ascending Baseline, then it is classified as Optimistic person, If the Sentence has got a Straight Baseline, then it is Classified as Stable Person). Similarly, each image is been classified and stored Ina specific folder according to its Personality, i.e., all the Optimistic Images have been Stored in a specific Optimistic folder and similarly all the Stable Images have been stored in a Stable Folder. Then these folders have been pre-processed and augmented using image data generator function form TensorFlow-keras, unorder to increase the dataset size and bring diversity in the data. Since we had limited number of participants for this dataset i.e., only 10, we have thought of augmented it and increasing the dataset, there are also other augmentation techniques available but as of now, image data generator is the best suited for our case, as a future scope, we can even augment our dataset with Generative Adversarial Networks to get more accurate synthetic images. Each image in this dataset is of Dimension 859 x 149 pixels with RGB Format.

**ARTEFACT**

So, in order to solve this particular problem of increased workload type and effort of graphologist we have proposed an efficient or rather a better approach, which helps us to identify the personality of a person using handwriting just by passing an input image of the individual cropped to a specific size and just wait for like a minute to get the personality of that person. As this particular model has been trained on very limited amount of data as there were data related issues which involved in restricted access to individual’s handwriting, so we have created our own handwriting sample data set by motivating the volunteers to participate. Using which we have trained our model, since this is just in the development process as the days progress, we can increase the data set and train the model furthermore to get approximate predictions when an input is passed of any kind. As we know artificial intelligence has been drastically increasing from the past decade and helping a lot of businesses out there in the world in generating revenue and reputation, so as a matter of fact we have chosen artificial intelligence and incorporated deep learning to solve this problem using convolutional neural networks. The reason we have chosen convolutional neural network is as we are dealing with imagery data set there is no such algorithm which can perform better than convolutional neural network which has a special behavior of identify the patterns involved in the image and even going through minute details which can ultimately yield into learning more accurately and providing us better results unlike artificial neural networks which can also be inculcated but cannot yield better results when compared to convolutional neural networks. But, In the future perspective we are even enthusiastic to use transfer learning to get even better accuracy than the existing one. But as of now we have used convolutional neural networks to achieve the problem we have discussed in the starting.

**ALGORITHM DESCRIPTION:**

**Convolutional Neural Networks** have been in this world from past 30 years, but its true application shave been evolving from past decade, there are a lot of reasons for this, one such reason might be the fact of not having Adequate amount of money and another important reason might be not having Enough GPU machines to train the model on. Yes, earlier they might have used CNN and have waited form more than a day to get the results and due to that fact of having longer training times, they might have gone with other techniques, but h same task which has been done with our GPU can be completed withing minutes with today’s growth in GPU machine and the power it possesses. Before talking about Convolutional Neural Networks Let us get an understanding of how does a Neural Networks actually work, As the term itself giving us a hint of working with neurons, the neural networks comprise of neurons which are connected to every other neuron present in the next layer, this mechanism is similar to what brain does i.e., transmitting information from one neuron to other neuron when a signal is passe, over here the signal would be our activation fucntion. As the term suggest that the neuron is connected via network, the network over here is nothing but the layers which we add which are also called ad Dense Layers, which are responsible to extract features from the data and pass it to the next layer, using Activation fucntion. These dense layers can be varied depending upon the type of problem we are solving, for example, If we look an artificial neural network with a tabular data, the numebr of neurons in the starting layer should be equal to the number of columsn which we have in our dataset, and then the dense layers should be added with adequate number of neuron sit it, We can try out by increasing and decreasing the number of neuron sin the dense layers and in the base layers, to check the how the accuracy is been altered, this will give us an understanding of how does a neural network work and how the accuracy is been altered. But, the main root cause about this algorithm is with Images, we know that artificial neural networks works very well with tabular data and also with the images but given the fact that computations can soar up high when we are dealing with images comapred to tabular data and the results might also not be promising, so in order to tackle this issues, Convolutional Neural Networks have been developed as an advancement of Artificial Neural Networks which will try to capture almost everything and anything from the data given and similar to Artificial Neural network Convolutional Neural Networks can also work on tabular data and Images but as te time progressed, people have been using Convolutional Neural Networks for more and more Image processing works. So, how does a Convolutional Neural Networks perform better than Artificial Neural Networks. Convolutional Neural networks are built by different layers sucha as convolutional Layer, Pooling layer and Fully Connected Layer.

**Convolutional Layer:** Convolutional Layer is core building block in any CNN architecture, which is responsible to do bulk task involving in identifying features in the image and passing it to further layers. This convolutional layer outputs a feature map, which will be in the form of a matrix of numbers which then passed onto pooling layer and Activation layer and then to another convolutional layer if any and then finally the feature map is passed to the fully connected Artificial neural network which are connected with these layers and ultimately pass the information to the output layer. The reason for not connecting a Convolutional layer to the output layer is that since the feature map should eb flattened before passing it further and without flattening our feature map the model will not able to understand what we are trying to achieve, but where as in the case of having fully converted layer, the flow goes on and using appropriate activation fucntion and dense layers the information is passed to the output layer which will result a result of what we are hoping for. (Probability of a class, example: If we are working with 2 classes, we can either have 2 neurons in the output layer or 1 depending on our need, this will give us a probability score of [0.687, 0.96878], so using this data, we have to apply numpy argmax () to pick the index of highest t probability and that will be our output the model has predicted. The same follows with 3 classes, [0.754754, 0.98887685, 0.244124], the highest index probability vales are picked and it is classified as valid answer by the model). In Convolution Layer we have sub layers such as:

1. **Filter or kernel:** Filters and kernels are the most important elements in the convolutional layer which are responsible from extracting features from the images passed. Filters or kernel are usually a 3x3 matrix which basically crawls through the image and will try to identify the patterns and feature present in the image. Each time filter is applied onto the image dot product is calculated for both the matrices and the vlaues is stored in a feature map. This cycle goes until the entire area of image is covered by the filter, and after each filter’s dot product is calculated the filters shifts with a **stride** or usually 3x3 and it might vary based on user. The stride shifts 3x3 matrix sidewards in the image vector. Along with filter and stride, we also have another parameter i.e., padding, which is used to make the filter to be applied on the input matrix of an image fully, by default we have a zero padding, i.e., whenever the filter dosent match with the input vector of an image the elements which fall outside the input matrix are made zero, producing larger and equally sized output. Other than zero padding, we also have **Valid Padding**, **Same Padding** and **Full Padding**. So, the feature discussed above holds all the values of the dot product of the input image and the filter, this feature map can also be known as Activation map or Convolved feature.
2. **Pooling Layer:** Pooling layer is basically used to reduce the feature map size or to reduce the dimensions of the feature map, so as to reduce the number of parameters in the input. The only difference between filter and pooling layer is where the filter does a dot product by taking in the input image vector and the filter size and then saving that value in a feature map, whereas the pooling layer is applied on top of our feature map and aggregations are made to rescue the dimensions. The parameter involved in polling layer is just a matrix of usually 2x2, if we want to reduce the dimensions even further, we can even increase the size to 3x3 or 4x4. This pooling size is applied on top of the area of our feature map and crawls all the way until it goes through all the vector of the feature map of the size specified. There are 2 kinds of Pooling techniques available:
3. **Max Pooling:** As the filter specific in the pooling layer moves across the feature map, it will pick the pixel which has the highest value and passes it to the next feature map. Max pooling layer is often used when comapred to the average pooling.
4. **Average Pooling:** Similar to max pooling, instead of picking he highest values from the filter we just aggregate all the values present inside it and then pass those values to the feature map. This feature map is then passed on to the next convolutional layer and can be flattened to move to fully connected layer.

**Fully Connected Layer:** Fully connected layer is nothing but the artificial neural networks which takes the input from the flatten layer, the term flatten layer refers to converting our feature map to a single dimensional array. This flatten layer is then connected with the dense layers of desired number of neurons. The number of neurons to be kept in the base dense layer totally depends on the suer of how much time he has delivering the output and it is often referred to as a trial-and-error approach. Fully connected layers can have as many dense layers as possible and again it totally depends on time and how big the feature map is. But as a proven fact, the number of neurons after the base dense layer should be to the power of 2 as matrix multiplication is one of the central computations in deep learning SMID operation in the CPU happens in batch sizes, which are the powers of 2. So, using powers of 2 we can increase the efficiency and time constraint of the model training. As the model approaching to the output layer, the neuron in the dense layers should decrease, where it is quite opposite when we are looking with convolutional layers, where the filter size increase to extract more and more minute details from the images.

**Justification of Algorithm Choice:**

The main reason of involving convolutional neural networks that it is cool fact that convolutional neural networks work very well when dealt with images angels better accuracy comparative to any other algorithm the fact that we have gone with CNN is ease of use of manipulating the hyperparameters and estimating the approximate accuracy for our model. As we know that Convolutional neural networks are involved in identifying patterns and features in an image using the filter and convolutional operation, these filters helps us to extract the exact feature we are trying to capture and verify, We don’t know what happens keep down in a neuron and neural network and how does it learn to identify such complex patterns but it does and it involves in a lot of mathematical computations which cannot be described her and will be out of ones imagination to understand how does a neural network learn anything and everything with such good precision. Other than above things mentioned above There are several reasons to use convolutional neural networks (CNNs) for image classification. CNNs are able to automatically learn features from images, which means that they can learn to recognize patterns of pixels that are indicative of certain classes of images. This is a much more efficient approach than manually designing features, and it means that CNNs can be used for image classification tasks even if there is very little training data available. In addition, CNNs are very good at generalizing from training data, which means that they are likely to perform well on new, unseen images. Finally, CNNs are relatively fast to train, which makes them well suited to applications where real-time classification is required. An image or speech recognition application is mainly performed by a convolutional neural network (CNN or ConvNet), a type of neural network that is based on a convolutional layer that reduces the high dimensionality of images while maintaining the information they contain. As a result, CNNs are an excellent tool for solving this problem. For the computer, RGB notation is a summary of three different matrices. For that reason, CNNs are particularly suitable for this application. Each pixel in the image is described by a matrix that describes what colour it displays. For example, the red component of the image is defined in the first matrix, the green component in the second matrix, and the blue component in the third matrix. We can get three different matrices for an image whose size is 3 pixels by 3 pixels. To process an image, we need to enter each pixel into the network as input. So, hence now, the image pixels i.e., 200 \* 200 \* 3 = 120,000 input neurons to be created and then passed to the further layers. Over there the 200 x 200 represents the dimensions of the image and 3 represent the RGB channel of the image. Then, each matrix is 200 by 200 pixels in size, making a total of 200 \* 200 entries. Finally, there are three copies of this matrix—one for each of red, blue, and green. It is then the first hidden layer that presents the problem, because it will have 120,000 weights from the input layer for each of the neurons within that layer. As we increase the number of neurons in the Hidden Layer, the number of parameters would increase very rapidly. It becomes even more challenging when we are trying to process bigger images with more pixels and colour channels. As a result of such a large number of parameters, there is a high chance that the network will overfit itself. This means that it will make good predictions for the training data, but it will not be able to generalize well to new cases that are unknown to the model. The network would be unable to attend to individual image details in the event of a large number of parameters, as they would very likely become lost in the sheer mass of parameters. Nevertheless, these details, such as the nose or the ears, can be very decisive for determining whether or not an image contains a dog, if we want to classify it based on these features, for example, whether or not it contains a dog. The Convolutional Neural Network, in order to solve these problems, takes a different approach to learning, mimicking the way we perceive the world around us. It is our natural tendency to divide an image into many small sub-images and then analyse each one individually when we see it. In a Convolutional Neural Network, we can interpret and process an image by assembling those sub-images. How can this principle be implemented? The work occurs in the layer of the neural network referred to as the convolution layer. Our method for achieving this is to define a filter that determines how large the partial images should be, as well as a step length that determines how many pixels, we continue to calculate between each of the partial images, i.e., how close the partial images are to one another. The next step is to create a pooling layer in order to reduce the dimensionality of the image. As far as the computational aspect is concerned, it works much the same as when we are using convolution layers, the only difference is that we only take the average or maximum value from the result, depending on the situation. In addition, we also have a fully connected layer, just like we are used to from regular neural networks. It is now possible to use tightly meshed layers to recognize the connections and perform classification of the image now that the dimensions of the image have been greatly reduced. As a result, we are able to link the individual sub-images again in order to recognize the connections between them. Our understanding of what each layer essentially does has been greatly expanded, and we can now move on to examining the process by which an image becomes a classification. Our objective is to determine whether there is a dog in a 4x4x3 image. For this purpose, we try to recognize the presence of a dog within the image. As a first step, we want to reduce the dimensions of the 4x4x3 image, which is why we define a filter of 2x2 dimensions for each colour of the image. It is also important that we specify a step length of as you move the filter forward by exactly one pixel after each calculation step, you will not be able to reduce the dimension much. However, you will be able to preserve the image's details. The output of a Convolutional Layer will be a 3x3 matrix if we migrate a 4x4 matrix to a 2x2 matrix and advance a column or row in every step of the way. As you will see from the graphic, the individual values of the matrix are calculated by taking the scalar product of the two 2x2 matrices. In the (Max) Pooling Layer, the (Max) Pooling Layer attempts to reduce the dimensionality of the image by further reducing the dimensionality of the convolution matrix as well as taking into account the important characteristics of the image. This layer is supposed to generate a 2x2 matrix as an output, so to do this, we divide the input into all possible partial matrices that are 2x2, and then search for the highest value in each of these fields. The value for the output matrix will be calculated based on this value. Instead of using a max-pooling layer, if we were to use the average pooling layer, we would calculate the average of all four fields. It is also important for the pooling layer to filter out noise from the image, in other words, the elements of the image that are not contributing to classification. It is not important at first whether the dog is standing in front of a house or in front of a forest. A fully-connected layer now accomplishes the same thing as the original picture in the beginning: we had intended to do with the entire image at the beginning. For each entry of the 2x2 matrix, we create a neuron and connect it to every other neuron in the following layer, which is a smaller 2x2 matrix. As a result, we have significantly fewer dimensions to train, and thus require fewer resources to train. This layer then learns the best way to classify an image as either dog or non-dog. There is of course the option of setting the convolution layer and pooling layer multiple times in a row if we are dealing with images that are much bigger than the 5x5x3 example we used in this example. In this way, it is possible to reduce the dimensionality far enough to reduce the training effort in a way that makes sense. Image processing and speech processing is carried out by convolutional neural networks, which are based on the structure of the human visual cortex. There are three layers in convolutional neural networks: a convolution layer, a pooling layer, and the fully connected layer. It is possible to program convolutional neural networks using TensorFlow by just following a few steps, which will enable you to view the image separately for the first time. Depending on the particular use case, it is important to correctly arrange the convolutional layers and the max-pooling layers in order to achieve maximum performance.

**Difficulty of choices:**

The major challenge that we might face while designing the solution for this project is when in the early stages where, we have to collect the proper dataset from the individuals and we had to make sure the dataset is collected in a well-lit area so that while training and testing the lighting doesn’t involve in the feature extraction, where while training we might face an issue with the lighting conditions of the inputs given. So, in order to avoid this error with the model, we have avoided taking images in darker areas, in fact this will be our future scope of wherever the image is clicked and passed to the model, the model will try to make appropriate predictions, but as of now, we have taken images in a specific lighting condition which didn’t have big impact on the model. Apart from the lighting conditions, we have also faced a bigger issue while labelling the dataset, as everyone knows that when the images are captured, they won’t be in a condition which can be directly passed to the model for training, so, we have done some pre-processing which involves in a lot of time and effort to be put in, the pre-processing techniques which involved in cleaning the dataset are:

1. Manually adjusting the lighting conditions to enhance the image.
2. Cropping individual sentence from the images.
3. Labelling the dataset.

Although, all the images are being taken in a well-lit lighting condition, there needs to be some lighting adjustment done, as as to remove the unnecessary shadow from the image, which can also play a vital role in bringing bias into our model. Since the dataset is collected in A4 size sheet paper and the images are captured using mobile camera the dimensions of the paper will be around 1120 x 800, we cannot give the whole image as an input to the model, at least in this case, since the model training is very computationally expensive procedure that too when we deal with Deep Learning the dataset should be resized to some scale depending on the system capabilities of your system. Keeping this issue in mind, we have cropped our images to 849 x 149-dimension size to increase the dataset and also reduce the computation power on the machine. This has been done by using a specific website, where entire image has been uploaded overtime to crop each sentence from that image, once this procedure has been finished, we have saved each individuals images in their own specific folder, to maintain and to avoid data labelling issue. Another challenge faced while developing this project is while labelling the images, where each individuals cropped image has been carefully assessed and his/her personality has been identified (This personality is not meant to be accurate, it is just an approximate and for research purpose only, didn’t mean to harm anyone’s feelings) referred to more than 3 research papers on personality identification through hand writing along with the articles published by famous graphologists around the world to understand the right characteristics for the hand writing of the person. As this stage involves in putting a lot of time and effort for the process, just to make sure the dataset is very well labelled, because, if we had messed up in this stage the complete project would have gone in vein, because if we had labelled a image with ascending baseline as pessimistic/stable person, but actually, that is an optimistic person we would just load the same dataset to the model, and when the model is tainted and ready for evaluation, it will give the results just as opposite to the label which we have given. So as to avoid this huge error, we had to put a lot of effort in this section just to improve the quality of the data and get better results while evaluating. Apart from this issue, the model building a very hectic and time-consuming task which might take hours and even day to train the model and it all depends on the number of images we have. But as of now, since our project was in an early stage of this research, we have taken a sample of data or data is collected from sample of population which had not taken too many hours to train.

**Evaluation of the solution:**

As the solution for our problem i.e. Identifying the personality of the person using handwriting is done or at least for the early stages we can be confident that this model will give better results whenever a new data is passed as an input given there are some mandatory pre-processing techniques which should be done before passing the input directly to the model, such as resizing the image to specific format the model is trained and normalizing the image and make sure that the input we are giving has at least a sentence written on it. As the model is being trained and also tested on some random images, there might be a question of how can we possibly be 100% sure of this is the actual personality of the person. But, in reality we can be sure that nothing is evaluated to get a 100% accuracy, even if it did, we cannot entirely and truly believe in those predictions. So there is always a threshold which we have to believe of not trying to make the machine learn everything about us that might sometimes lead to an overfitting model, This situation of how much should we train our model varies from domain to domain i.e., in the case of medical fields we have to very accurate regarding the predictions and need to make the model get an accuracy of close to 100, as we are dealing with lives of living things, we need to put a lot of effort and time to make the model accurate because even the 0.1 % of error in this field might lead to loss of lives. There are a lot of evaluation metrics to be considered in the case of any machine learning or deep learning project but they vary depending on the type of data we are using. Unlike, the metrics which we get when we deal with tabular data, we cannot replicate the same with the imagery data, since the image labels will be in the form of probabilities by passing through the output layer, yes it might be different by various activation functions and number of classes we have, but most of the time the output layers gives the output in the form of probabilities of classes, so we had to convert those probabilities into actual labels and then store them in a list and use the predicted converted values and the actual values to generate appropriate metrics, the metrics which we can get through these results are:

**Confusion Matrix:** There is a confusion matrix that is used to assess the performance of a machine learning model. It consists of four different quadrants that each represent a different combination of predicted and actual values that is used to evaluate the performance of a machine learning model. There are four quadrants in a model: the first quadrant represents how many correct predictions were made, the second quadrant represents how many false positives there were, the third quadrant represented how many false negatives there were, and the fourth quadrant represented the number of true negatives. As a result of color coding the quadrants, different types of predictions are easier to differentiate visually. It is typically a green quadrant in which the first quadrant is positioned, a red quadrant in which the second quadrant is positioned, a blue quadrant in which the third quadrant is positioned, and a black quadrant in which the fourth quadrant is positioned. There are several different metrics that can be calculated using these quadrants, such as accuracy, precision, recall, and specificity. As far as accuracy and precision are concerned, accuracy is simply the percentage of predictions that were accurate. Precision is the percentage of positive predictions that were actually correct. In order to calculate the false positive rate and the false negative rate, the quadrants can also be used. The recall is the percentage of actual positives that were correctly predicted. The specificity is the percentage of actual negatives that were correctly predicted. In contrast, the false negative rate is the percentage of positives that are incorrectly predicted as negatives. The false positive rate, on the other hand, is the percentage of positives that are incorrectly predicted as negatives. Any machine learning model can be evaluated using the quadrants, but they are particularly useful for evaluating models used for classification tasks, which are a particular type of machine learning problem. A confusion matrix has four parametres such as False Positives, False Negatives, True Positives, and True Negatives are its four parameters. which once more was deduced into different formulas, such as TPR and FPR, which aids us in obtaining more performance analysis. The ideal confusion matrix is shown below. The more TP and Tn there are, the better the model is; however, depending on the project we are working on, we might also be concerned with FN and FP, so lowering their numbers could be a crucial step in the evaluation process.

* **True positives:** Both Predicted and Actual value is True.
* **True negatives:** Both Predicted and Actual Value is False.
* **False positives:** In this case, the actual value is False, but the model has predicted True.
* **False negatives: In this case, the actual value is True, but the model has predicted False.**
* Using The TP, TN, FP, FN we can derive some formulas such as:
  1. **Accuracy** = (TP + TN) / Total number of classes
  2. **Recall** = TP / (TP + FN)
  3. **Precision** = TP / (TP + FP)

**Recall:** Recall is a metric for how well a model can identify positive examples. Recall is calculated by dividing the total of true positives and false negatives by the number of true positives. In contrast to low recall, which indicates that many positive examples are mistakenly classified as negative, a high recall indicates that the majority of positive examples are correctly identified. Precision is a measurement of a model's capacity to correctly identify negative examples, and it is frequently used in conjunction with recall. These two measurements taken together can give you a good idea of how well a model is doing. The goal should always be to maximize the recall as well as the precision of a survey. While there are trade-offs between the two measures, it is often necessary to compromise in order to find a balance between the two measures. If, for instance, identifying all of the positive examples is more important to you than identifying some of the incorrect predictions, then a model with a high recall but low precision may be more effective. For applications where accuracy is more important than recall, a model with a high precision but a low recall may be more useful, even if some positive examples are not detected, than a model with a high precision but a low recall.

**Precision:** An accurate prediction made by a machine learning model, based on a new piece of data that has not been used in training, is called precision in machine learning. A model that is precise has a low error rate. It means that it is capable of predicting the labels of new data with a high degree of accuracy. Precision is a model metric that is closely associated with accuracy. In fact, the terms are often interchangeable. It is important to note that there is a subtle difference between the two. In other words, accuracy measures how often a model makes the right prediction, based on all of the data. Precision measures how often a model makes the right prediction based on all of the data. This is the relationship between accuracy and precision. It is a way of measuring the likelihood that a model will predict correctly, based on new data, on a regular basis. The accuracy and precision of a machine learning model are both important when assessing its performance. However, precision is typically more important than accuracy, especially when the cost of false positives is higher than the cost of false negatives in some applications. A false positive (predicting a patient has a disease when they do not) is more harmful than a false negative (predicting a patient does not have a disease when they do) in a medical diagnosis application. In this regard, we would prefer a model that is more precise, even if it is less accurate, since we would want it to be more precise. A model's accuracy generally increases with the increase in the number of training examples. The precision of a model, however, is not necessarily guaranteed to increase with the increase in the number of training examples. This is because the number of false positives will also rise as the number of training examples increases. In general, false negatives will decrease as a result of the reduction of false negatives, but not at the same rate. This means that as accuracy increases, the precision of a model may decrease as well. Taking this into account when assessing how well a machine learning model performs is something that should be kept in mind.

**Classification report:** It is a tool used to assess a classification model's predictive ability and to determine the classes for which it is most and least accurate. The table in this report contains a number of metrics, including the following: -Precision: The precision ratio, which indicates how frequently the classifier correctly predicted a positive class, is calculated by dividing the number of true positives by the total number of positives. This metric tells us how often our classifier correctly predicts the positive class based on how many samples actually belong to the positive class. It is a ratio of true positives compared to all samples that actually belong to the positive class. A F1-score measures the ratio of precision and recall. As a rule, it is calculated as the harmonic average of precision and recall. Support: The number of samples belonging to each class is the support. There is no doubt that in order to effectively compare different classification models and to understand their performance, a classification report is a valuable tool.

**CONCLUSION**

This project has proposed an efficient method of identifying or estimating the personality of a person using his/her handwriting using Deep learning methodologies involving in convolutional neural networks. The project mainly consisted of 2 classes each representing a specific personality of a person, the personalities have been chosen based on their handwriting style such as, if the person is having his/her handwriting in a style where the sentence is slightly dipping downwards as they progress writing the sentence then this type of style has been classified as the Optimistic Personality and if the person handwriting is being stable and straight from the start to end, then this type of personality has been classified as the Stable Personality. The above-mentioned labels have been carefully chosen and labeled by thoroughly checking and inspecting the handwriting sample given by the volunteers. This project specifically deals with the specific characteristic among many different characteristics available such as Baseline, Margin, and Slant of the word present in graphology to identify or characterize the person based on his/her handwriting.