



## **THESIS REPORT**

### **Personality Prediction Through CV Analysis**

**Submitted By**

<b>17-33673-1</b>	<b>Hanif, Md</b>
<b>17-33088-1</b>	<b>Fahim,Imtiaz Ahmed</b>
<b>17-35970-3</b>	<b>Ali,Shahrin binte</b>
<b>17-33108-1</b>	<b>Uddin,Khaja Kutub</b>

**Department of Computer Science**

**Faculty of Science & Technology (FST)**

**American International University-Bangladesh (AIUB)**

**November 2020**

---

## Declaration

---

We declare that this thesis is our original work and has not been submitted in any form for another degree or diploma at any university or other institute of tertiary education. Information derived from the published and unpublished work of others has been acknowledged in the text and a list of references is given.



---

Department of Computer Science



---

Department of Computer Science



---

Department of Computer Science



---

Department of Computer Science

---

## Approval

---

The thesis titled “Personality prediction through CV analysis” has been submitted to the following respected members of the board of examiners of the department of computer science in partial fulfilment of the requirements for the degree of Bachelor of Science in Computer Science on (12-11-2020) and has been accepted as satisfactory.

---

**Sajib Hasan**

Assistant Professor & Supervisor  
Department of Computer Science  
American International University-Bangladesh

---

**ASMA FARIHA**

Lecturer & External  
Department of Computer Science  
American International University-Bangladesh

---

**Dr Md. Mahbub Chowdhury Mishu**

Assistant Professor & Head  
Department of Computer Science  
American International University-Bangladesh

---

**Professor Dr. Tafazzal Hossain**

Dean  
Faculty of Science & Technology  
American International University-Bangladesh

---

**Dr. Carmen Z. Lamagna**

Vice Chancellor

American International University-Bangladesh

---

## Acknowledgement

---

First of all, we would like to thank the Almighty, for His grace in accomplishing our thesis timely. We would like to express our gratitude to the **Faculty of Science & Technology** to keep thesis credit in the curriculum of the graduation program and give us a scope of tasting the flavor of research and the field of work with our interest.

We are also thankful to **Sajib Hasan**, Assistant Professor of Computer Science, at American International University – Bangladesh for supervising us in this research. We are also grateful to him from the core of our heart for his kind support, guidance, supervision, instructions and for advising us and helping us through hands and pens and motivating us to do our thesis research smoothly.

---

## Abstract

---

At present personality of an employee plays a major role in the development of an organization. One of the ways to judge a candidate is by analyzing the Curriculum Vitae (CV). Usually, recruiters manually filter candidate's CV as per requirements. The usual process of candidate selection is time consuming and sometimes incorrect candidates are chosen by mistake. In this paper our goal has been building a system that can ease the process by analyzing CV's image data and numerical data with the help of machine learning algorithm so that the system can be a helping hand for employee recruitment process. This paper focuses on predicting personality by analyzing CVs of the candidates applying some machine learning tool. To predict personality, we tried to find out the features needed, a way to get the features and tested if it is possible to predict personality. At first, we tried to classify candidate personality among five big personality traits from CV images of the candidates. Then we acquired image data and numerical data from CV template. Following that, through some processes such as FER2013 Dataset, CNN model training, HOG, Haar cascades, Data Preparation and Train Model we stored the extracted features in key-value pairs. From there respectively through image acquisition, greyscale conversion, face detection process, facial expression detection, age prediction, colorfulness, and object detection we got features as output. Lastly determined trait by calculating the matching percentage of a particular trait. We also checked the validity of our result by analyzing a dataset using unsupervised learning techniques. The result shows we were able to predict personality from CV image. From our result it is evident that from CV image we can classify an individual among big five personality traits. As in this study, for the first time CV image has been used to predict and classify candidate's personality among big five personality trait, this study can be useful in job recruitment process in future.

---

# Table of Contents

---

<b>Chapter 1: Introduction</b>	<b>8</b>
1.1 Background of the Study	8
1.2 Statement of the Problem	8
1.3 Significance of the Study	8-9
1.4 Goals & Objective	10
1.5 Chapter Overview	10-11
<b>Chapter 2: Literature Review</b>	<b>12</b>
2.1 Personal trait	12
2.2 Face Recognition	12
2.3 Facial Expression Detection	12
2.4 Classification Algorithm	13
2.5 Related Work	14-15
<b>Chapter 3: Methodology</b>	<b>16</b>
3.1 Personality Trait Classification from CV	19
3.2 Preparing CV Template including important features	19-20
3.3 Data Acquisition	20
3.4 Feature Extraction	20
3.4.1 Dataset	20
3.4.2 CNN	20
3.4.3 HOG	21
3.4.4 HAAR CASCADES	21
<b>Chapter 4: Results &amp; Discussion</b>	<b>22</b>
4.1 Data Preparation	22
4.1.1 Data Type	22

4.1.2 Training CNN Model	23
4.1.3 Feature Storage	23
4.1.4 Image Acquisition	23-24
4.1.5 GreyScale Conversion	25
4.1.6 Face Detection	25
4.1.7 Facial Expression Detection	26
4.1.8 Age Prediction	26
4.1.9 Colorfulness	27
4.1.10 Object Detection	27
4.2 Trait Determination	28-30
4.3 Discussion	30
<b>Chapter 5: Conclusion &amp; Future Work</b>	<b>31</b>
5.1 Conclusion	31
5.2 Future Work	31

---

## List of Figures

---

Fig 3.4.4: Haarcascade Classifier Frontal face detection	22
Fig 4.1.1: CNN model Architecture	24



---

# Chapter 1 Introduction

---

## 1.1 Background of the study

This thesis focuses on the findings of a thorough study to establish the factors that have led to the prediction of personality through CV analysis. To achieve the goal of predicting personality through CV, the relevant study was performed in the field of Artificial intelligence, Machine learning & image processing. The factors that were considered were how an emotion detection algorithm is used to extract information from image [2], how color preference could be used as a clue to predict personality [1], how facial expression recognition using image processor could be a factor in personality prediction from CVs [3], how machine learning algorithm could be helpful in personality evaluation & CV analysis [5], how personality trait can be detected using profile picture [18], if personality could be predicted from email messages [19], how applicant's information provided in resume can be used in predicting applicant's mental ability & personality dimension [20]

## 1.2 Statement of the Problem

Selecting the right candidates from the huge pool of candidates. A huge number of candidates apply for a particular job. To choose the best candidate among all of them becomes crucial responsibility for the company. This process of CV analysis ensures fair selection of candidates. Makes the recruitment process faster. Ease of selecting candidates [5]

Research Question: -Is it possible to predict personality through CV analysis?  
-Identifying Important features to predict personality?  
-How is it possible to extract features from a CV?

## 1.3 Significance of the study

Personality, implies the way a human being thinks, behaves and acts to identify the characteristics of a human being. Predicting a personality involves what qualities or traits a

person possesses making a prediction of that. Curriculum Vitae or commonly referred as CV analysis is an important part of job recruitment.

CV's for job recruitment is analyzed with the purpose of finding that particular candidate who has required skills and the ability to perform his/her specific task accurately yet quickly [21]. Thorough CV analysis is necessary to make sure recruiters don't miss good candidates or put forward wrong candidates at the same time it is also important to process all applications quickly to make sure a rapid response for clients and move on to the interview and placement stage [21].

Personality is the most important factor which reflects an individual, which keeps on varying [5]. The personality of an individual can vary with respect to time. Human personality has played a vital role in an individual's life as well as in the development of an organization. One of the ways to judge human personality is by using standard questionnaires or by analyzing the Curriculum Vitae (CV) [5]. Traditionally, recruiters manually shortlist/filters a candidate's CV as per their requirements. In this paper, a system has been presented that automates the eligibility check and proper evaluation of candidates in a recruitment process. The system analyzes professional eligibility based on the uploaded CV. As far as employment is considered, selecting the right candidate for the recruitment process from a vast pool of candidates, has been a fundamental issue. Conducting personality and various technical eligibility evaluation tests, interviews, and group discussions have been traditional techniques. Generally, in this lengthy process the required qualification of the candidates is sometimes unnoticed by the recruiters. These traditional practices are very much time-consuming, and may result in unfair choices of candidate [21]. As compared to traditional recruitment process, if a selection process is conducted that involves predicting an individual's personality by analyzing the CV, then a fair selection of the candidate is possible. Through image analysis and information included in CV, we tried to predict the personality of the candidates as correctly as possible whose are most eligible for a particular job position.

## **1.4 Goals & objective:**

Personality prediction is done for various purposes. Mostly it is done widely for identifying a right candidate for jobs. The personality prediction empowers a more compelling approach to short run-down submitted hopeful CVs from an extensive number of candidates giving a predictable & reasonable CV positioning arrangement, which can be legitimately supported [5].

In our paper, we propose personality evaluation and CV analysis using machine learning algorithm to predict candidate's personality. The goal is to:

To find out the features needed to predict personality.

To find out how to get the features. To find out if it is possible to predict personality.

To empower a more compelling approach to short rundown submitted hopeful CV's from a substantial number of candidates.

To assist the human asset office with selecting right contender for specific occupation profile, which thusly give huge workforce to the association.

## **1.5 Chapter Overview**

Personality showcases various individual differences within people. Personality includes all types of patterns of conception and emotions that make us to do and say things in specific ways. In short our personality defines us as people how we are. That is why to judge someone by personality trait becomes more important when it comes to job handover to a candidate. Through CVs, interviews recruiter choses the right candidates which is quite time consuming and tiresome. To make sure that the most deserving & qualified candidates get the particular job position, we tried to approach

a fast, efficient and less time-consuming recruitment process by only analyzing the CVs of the candidates. In this system we have used some machine learning tools and algorithm to analyze the numerical data and image data extracted from CVs. After processing those data, we determined in which class of personality trait the extracted features belong to among the big five personalities, which are Extravert, Conscientious, Agreeable, Openness, Neurotic. In which personality trait an individual belongs to depends on the number of percentage matches most with a trait determined from processed data [20].

In this paper, we proposed an as much close to accurate result as possible to predict personality of the candidates. As predicting the right candidate depends on the personality trait an individual belongs to, determining the trait for an individual candidate was the most challenging task. The paper is structured as follows: Section II presents literature review. The methods that were used to predict personality in this paper are provided in Section III. The results and discussion for the classification of candidate's personality based on the proposed method are provided in Section IV. Finally, Section V concludes the paper with conclusion & future work.

This paper will explore the efficient and easy approach to choose right candidate in job recruitment procedure. It will describe the methodology, feature selection and trait determination methods. The results will be presented and discussed based on the several procedure using machine learning tools and image analysis, followed by conclusions and future work.

---

## Chapter 2 Literature Review

---

### 2.1 Personality Trait:

An individual's personality is the reflection of mental characteristics that makes them different from others. Usually, personality is expressed through our nature or emotional tone. However, personality also reflects our values, beliefs, and expectations. There are many potential factors that are involved in shaping a personality. Psychologists have mostly focused on big five personality traits that are widely accepted defined as Openness, Conscientiousness, Extraversion, Agreeableness & Neuroticism. Psychologists over last few decades have been emphasizing on hereditary factors being more important specially for basic personality traits. The factors that plays important role in shaping an individual's personality are biological and physiological factors, Environment factors including physical, social, family, cultural and school environment. However, the acquisition of values, beliefs, and expectations seem to play more important role in a person's life in terms of experiencing unique experiences, especially during childhood [6].

### 2.2 Face Recognition:

The human face is a unique item and has a serious level of inconstancy in its appearance, which makes face detection a troublesome issue in computer vision that includes discovering faces in photographs. It is a trifling issue for people to comprehend and has been settled sensibly well by traditional feature-based strategies, for example, the cascade classifier. All the more as of late profound learning strategies have accomplished cutting edge results on standard benchmark face detection datasets. One model is the Multitask Cascade Convolutional Neural Network, or MTCNN for short.

### **2.3 Facial Expression Detection:**

The main thing which depicts one's personality of the person is his face. The face gives data about age, gender, and appeal. This likewise identifies with how the client introduces himself before his crowd. The most noticeable facial feelings are outrage, appall, dread, happiness, trouble, and shock. The six feelings can be sorted into positive (happiness and shock) and negative (Sad, Happy, Neutral). Distinguishing and perceiving a face from the picture is a very intriguing theme. A little change in facial articulation is effectively perceived by individuals [7]. This is generally in light of the fact that particular pieces of the human mind can effectively perceive facial feelings [8]. Distinguishing proof of the client's face from pictures by PC innovation is called face identification. It likewise discovers the number of appearances is present in the picture. Discovery of the face starts with eyes location, nose discovery, and so on Hence, outward appearances of people depict their personality

### **2.4 Classification Algorithms:**

CNN:

CNN is a Deep Neural Network initially intended for picture analysis. CNN consistently contains two fundamental operations, to be specific Convolution and Pooling. The Convolution activity utilizing various channels can remove highlights (include map) from the informational collection, through which their comparing spatial data can be safeguarded. The pooling activity, likewise called subsampling, is utilized to decrease the dimensionality of highlight maps from the convolution activity. Max Pooling and normal pooling are the most widely recognized pooling tasks utilized in the CNN. Because of the complicity of CNN, relu is the regular decision for the initiation capacity to move inclination in preparing by backpropagation.

Viola – Jones Algorithm:

Viola-Jones Object Detection framework can rapidly and precisely identify objects in pictures and functions admirably with the human face [9]. Regardless of its age, the structure is as yet a main

part in face discovery close by numerous individuals of its CNNs partners. The Viola-Jones Object Detection Framework joins the ideas of Haar-like Features, Integral Images, the AdaBoost Algorithm, and the Cascade Classifier to make a framework for object location that is quick and precise.

## **2.5 Related Work**

In an article [10] of “Using Recruiter Assessments of Applicants' Resume Content to Predict Applicant Mental Ability and Big Five Personality Dimensions” the author predicted the applicant mental ability and big five personality traits from the point of view of recruiter assessments of applicants' resume. The authors focused on an applicant screening function and proposed five hypotheses. Hypothesis 1: Recruiters will reliably ( $> .60$ ) assess the extent to which specific biographical topics (e.g., volunteering for community service) are present on applicant resumes. Hypothesis 2: Academic achievement information as reported on job applicants' resumes (e.g., being on the Dean's list, scholastic awards received) will be positively related with applicants' general mental ability. Hypothesis 3: Applicants reporting GPA on their resumes will have higher mental ability and conscientiousness scores than applicants not reporting GPA. Hypothesis 4: Applicants' academic achievement information (e.g., received scholastic awards) as reported on job applicants' resumes will be positively related to applicants' conscientiousness. Hypothesis 5: Applicants' resume information involving social interaction (e.g., membership in social fraternity/sorority, elected offices held) will be positively related to applicants' extraversion.

In another article [11] entitled “Understanding Email Writers: Personality Prediction from Email Messages” authors attempt to infer the personality of email writer based on the contents of email. They framed personality prediction based on the well-known Big Five Personality model and trained predictor based on extracted email features. Their study results that personality prediction is feasible. They extracted Bag-of-words Features, Meta Features, word statistics, writing styles from an email to predict personality. They imagined three models such as Joint Model, Sequential Model, and Survival Model to generate features to predict personality. They presented the

extracted email features and their predictive power. They then explored three learning strategies showing that a method with label-independence assumption works best in their case.

A group of researcher Goa College of Engineering, Goa, India studied the correlation between the social media profile pictures of a user with the Big-Five personality traits in 2018 [18]. Facial features have been used by them. Each user with one of the personality traits posts a specific type of profile pictures have been concluded by them. In this paper the author mentioned that personality trait prediction uses the “Big Five” model which is used of five broad traits: Openness, Conscientious, Extroversion, Agreeableness, and Neuroticism. They founded that Extrovert and agreeable users mostly post vivid and positive mood expressions. Neuroticism and openness users post pictures holding one or two people or self-images and mostly express a bad mood. Conscientiousness users post what exactly is required for profile picture that is a self-face picture and have a happy mood.

In 2019 another bunch of analysts of Thakur polytechnic, Maharashtra, India works on personality prediction using CV analysis [21]. In this project an organization-oriented recruitment system that would help the human resource department in short listing the right candidate for a specific job profile have been implemented by them. Their objective of this research is to develop a system to deliver a more effective way of short-listing the candidates and to detect online attitude and personality test. They tried to extract candidates’ personality traits from their social presence using Curriculum Vitae analysis. They have two major modules- the Admin login and the User login. The recruiting company uses the admin login and the applicants applying for the jobs uses the user login.

Next another team of Usha Mittal Institute of Technology, Mumbai, India worked on personality evaluation and CV analysis using Machine Learning Algorithm in 2019 [5]. They have also used Big Five Personality Model to predict the personality of the candidate which includes Openness, Conscientiousness, Extraversion, Agreeableness and Neuroticism. They have used the person Automated Personality Classification. It is used to classify the person from a huge amount of people. The machine learning algorithm (TF-IDF) have been used by them to find out the important keywords in a CV. In here a machine learning approach has been used in analysis of data through content and collaborative filtering.

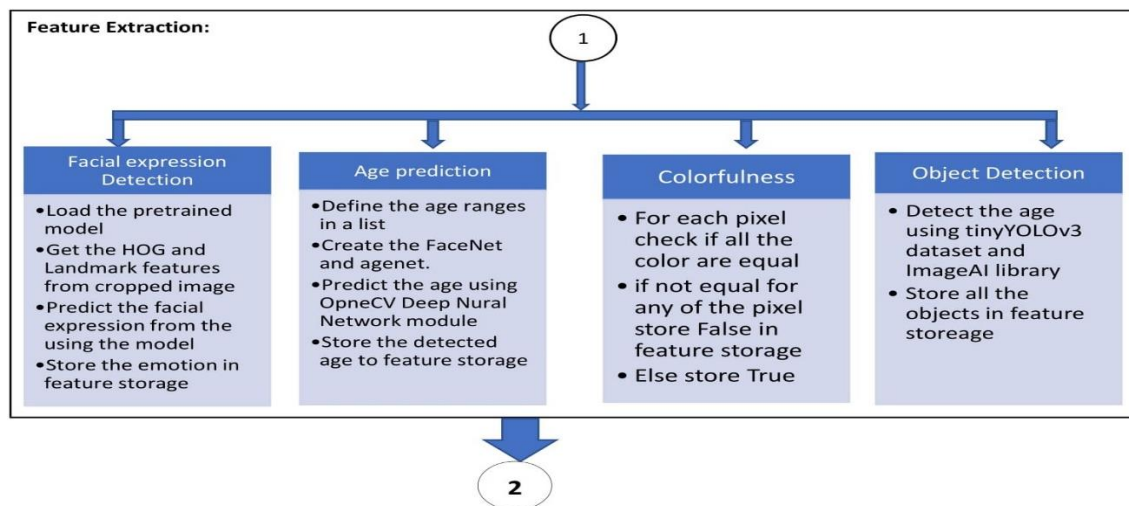
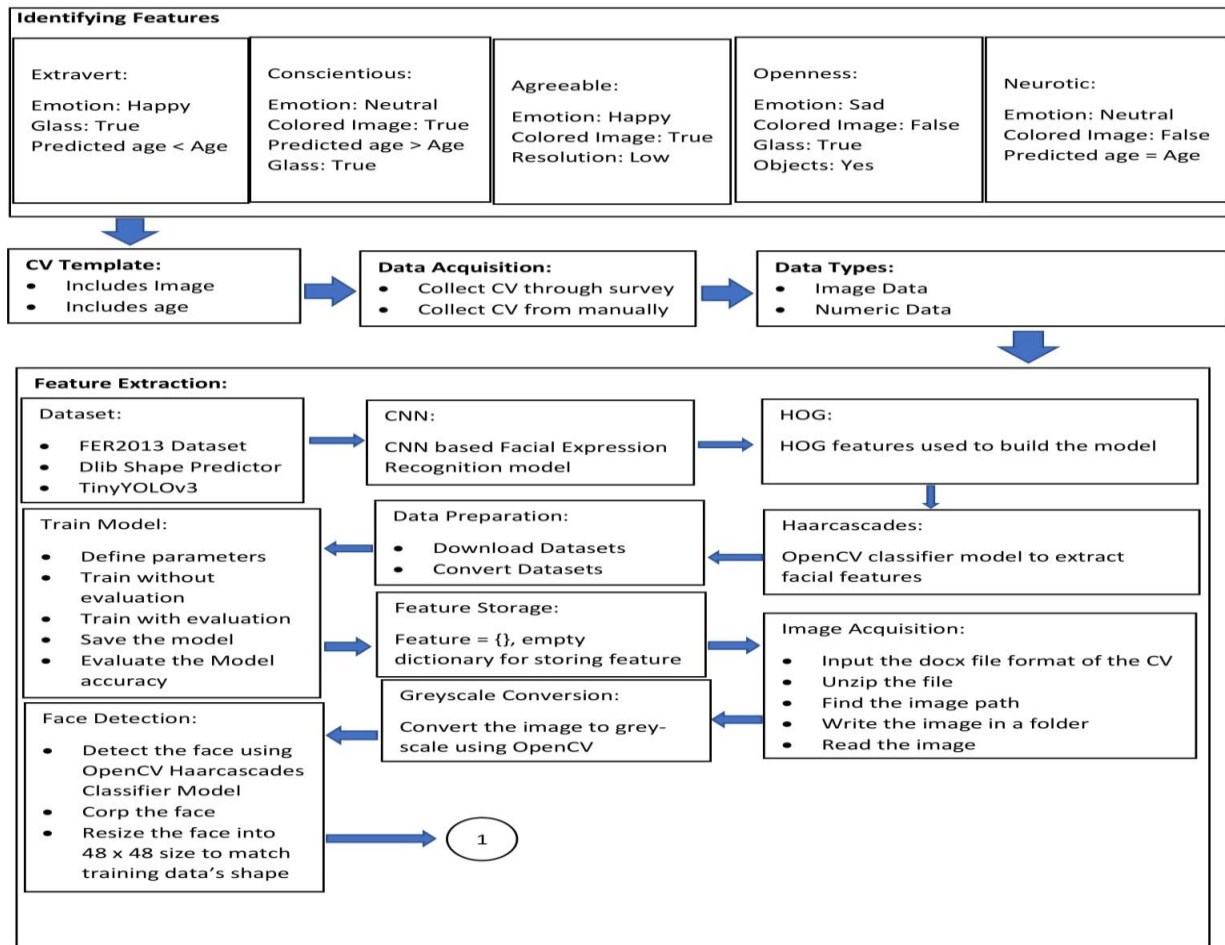


---

## Chapter 3 Methodology

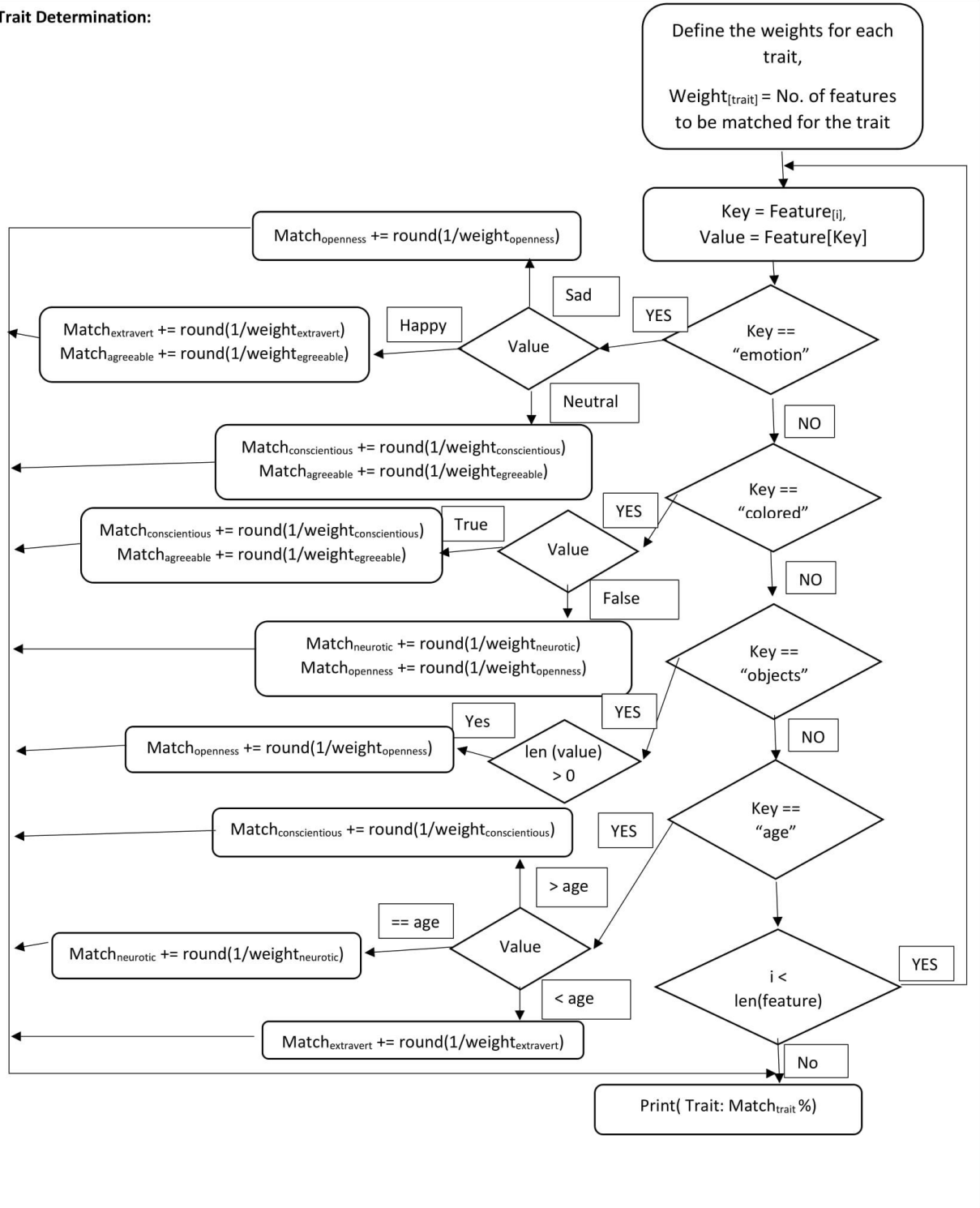
---

To find out the features needed to predict personality, to find out how to get the features and to test if it is possible to predict personality, at first we tried to classify candidate's personality among five big personality traits from CV images of the candidates. Then we acquired image data and numerical data from CV template. After that, according to our Image processing features we have used FER2013 Dataset for recognizing the facial expressions such as Sad, Neutral, Happy, Angry and also applied the dataset and Face Landmark model to train our CNN model in order to extract features. Following that HOG features were applied to build model & by applying Haarcascades algorithm OpenCV model was built to extract features. Then through Data Preparation and Train Model we stored the extracted features in key-value pairs. From there respectively through image acquisition, grey scale conversion and face detection process. Following the face detection, facial expression detection, age prediction, colorfulness and object detection processes are respectively applied on images, we get features as output. Our final step is trait determination where we put weight on each trait. After that, we initialized a matching for all five traits as 0%, we checked for each key-value in Feature Storage, if it matches a feature of a trait, we calculate the matching of a particular trait and finally got the matching percentage. There is another approach that we followed to check our machine performance. We collected a data set from kaggle.com which was taken in time of 2016 to 2018. An organization called International Personality Item Pool (IPIP) was conducting this survey on online. About 1015341 people participate in this survey. The following items were presented on one page and each was rated on a five points scale using radio buttons. The order of questions on page was EXT1, AGR1, CSN1, EST1, OPN1, EXT2, etc. The scale was labeled 1=Disagree, 3=Neutral, 5=Agree. We tried to perform some data analysis on this dataset. As the dataset has no specific label to classify. So we tried to classify the dataset by unsupervised learning techniques. The technique that we have performed in here is K-Mean clustering. After performing the technique and other analysis we clustered the data into 5 group. The Participant was given 50 question and told to scale the answer from 1 to 5. At the same time the participant image was also checked for classifying the personality.



2

# Trait Determination:



### **3.1 Personality Trait Classification from CV**

To predict the big five personality traits [14] from image attached to an applicant's CV or resume, the important features to be extracted have been determined from an article in Skillroads.com [12] and a journal paper [13] by Bhatti, Muneer, Lali and Din. The features are as follows:

- 1.Extravert: Extravert people tends to keep a smiling face in the image which makes them look happy. They tend to not wearing reading glasses and they look younger in their image then their actual age.
- 2.Conscientious: People with this personality tends to keep their image colorful and do not prefer grey images. The person looks older in his/her image then the actual age. They tend to wear glasses. They keep a neutral facial expression in their image.
- 3.Agreeable: This kind of people keeps a smiling face in image and the image are colorful and bright like extrovert people and the image have low resolution.
- 4.Openness: This kind of people tends to pose with some object, they tend to wear glasses. They like to keep their image colorless and they look sad in the photo.
5. Neurotic: This kind of people have negative emotions. They tend to keep their image colorless. They look neutral in their image. They look the same age in the picture as their actual age

### **3.2 Preparing CV Template including important features**

To use the image processing technique to determine an applicant's personality the CV must have to have an image of the applicant, and must have to include the age in the CV. Using that image the personality can be determined. For this We prepared a template how the CV format should be. There should be age and photo in the header of the CV like the following:

<b>[Name], [age]</b>	Photo
[Contact Details]	

### 3.3 Data Acquisition

To prepare our CV Dataset we have collected CV through survey and from our known people. To our known people we have provided our preferred template to give us their CV in our preferred template. Then we have collected their complete CV in our format. In survey we have collected the age, updated CV of the person with a photo of the person.

### 3.4 Feature Extraction

To extract the features according to our Image processing features we have used FER2013 Dataset for recognizing the facial expression such as Sad, Neutral, Happy, Angry and used the dataset and Face Landmark model to train our CNN model.

#### 3.4.1 Dataset

We have used the fer2013 dataset for our facial recognition. It consists of 48\*48 pixel grayscale images of faces. The faces have been automatically enlisted so that the face is pretty much focused and occupies about a similar measure of space in each image. We have used the tiny yolo v3 for object recognition and classification. This model was prepared with the Coco data set and can identify up to 80 classes.

#### 3.4.2 CNN

The convolutional neural network (CNN) is a class of **deep learning neural networks**. CNNs represent an enormous breakthrough in image recognition. CNNs are enlivened by biological processes. There are three layers in CNN. Input layer, output layer and hidden layers. [15]

### 3.4.3 HOG

Histogram of Oriented Gradients (HOG) has been broadly utilized in computer vision as feature descriptors for recognizing objects in scenes. This is a similar method like edge orientation histograms, shape context and scale-invariant feature transform descriptors. But the Hog descriptor has a couple of advantages than other different descriptors. [16]

### 3.4.4 HAAR CASCADES

Haar Cascade is a machine learning object detection algorithm. It is used to identify any object from any image or video. Here cascade function is trained from a lot of positive and negative images. Then it is used to detect any object in other images. This algorithm needs a lot of positive image with faces and a negative image without faces for train the classifier. Then we must extract features from it. We need to apply every feature on all the training images. For each feature, it finds the best threshold which will classify the faces to positive or negative. [17]

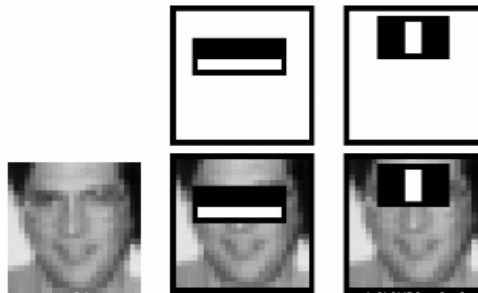


Fig 3.4.4: Haarcascade Classifier Frontal face detection

---

## Chapter 4 Results & Discussion

---

After building CNN model, it was trained with dataset where accuracy for the model was checked and thus we got extracted features that were stored in key value pairs. Then by applying image acquisition, Greyscale conversion, face detection, facial expression detection, age prediction, colorfulness and object detection we kept getting results after each process's appliance. After completing all these processes, we were able to extract the features. Using those features determined traits

### 4.1 Data Preparation

#### 4.1.1 Data Type

For determining personality from the image, we have to collect the age and the image from a CV. So, there are two types of data and they are Image and numeric data We have downloaded publicly available dataset FER2013, Dlib Shape Predictor mode land tinyYOLOv3 dataset. The FER2013 dataset contained training data for facial expression recognition. The tinyYOLOv3 dataset is a pre-trained model for detecting objects. After downloading datasets, we have converted the FER2013 and Dlib Shape Predictor Datasets to CNN based model. We defined the 'Convolution layer' for CNN model, where Convolution, padding, striding and RELU activation operation are being performed. Our CNN model architecture is given in Figure 4.1.1

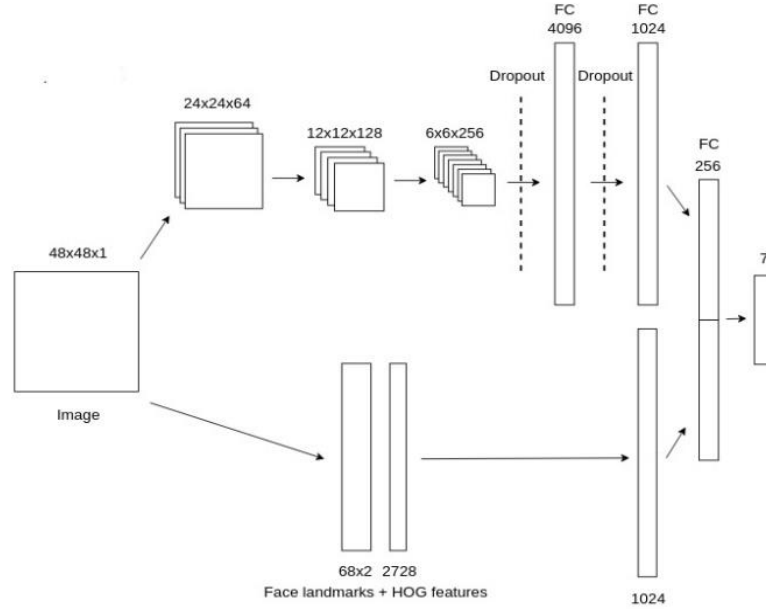


Fig 4.1.1: CNN model Architecture

#### 4.1.2 Training CNN Model

After preparing the model we have trained the model with FER2013 dataset. We first defined that ‘Happy’, ‘Neutral’, and ‘Sad’ facial expressions are needed to be determine. Then we have trained the model without evaluating its accuracy. Then we have trained the model again while evaluating accuracy. At the completion of the training we have saved the trained model. The model was 73% accurate.

#### 4.1.3 Feature Storage

In feature storage we store the extracted features in key-value pairs. At first we initialize the Feature Storage as an empty dictionary. Then after extracting a feature we store the feature in the feature storage.

Feature = { }

#### 4.1.4 Image Acquisition

For facial recognition we have took input the docx file format of the CV. This docx file is in a Zip file. So we have unzipped the file and find the image path. The file path



is 'word/media/'. The image file is jpeg or jpg file format. Then we have saved the image in a folder. After that we have read the image for extracting features.

```
docfile = "resume_012.docx"
z = zipfile.ZipFile(docfile)
all_files = z.namelist()
images = filter(lambda x: x.startswith('word/media/') and (x.endswith('.jpg') or
x.endswith('.jpeg')) , all_files)
imgName=''
for i in images:
    imgName=str(i).split('/')[-1]
    imgPath=str(i)
image1 = z.open(imgPath).read()
f = open(imgName,'wb')
f.write(image1)
z.extract(imgPath, r'CV')
imageName = 'CV/'+imgPath

image = cv.imread(imageName)
```



CV

Extracted image

#### 4.1.5 Greyscale Conversion

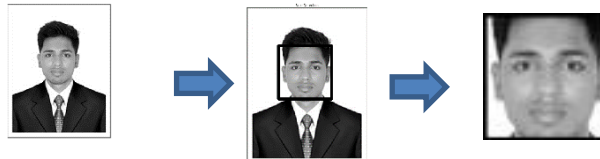
After extracting the image from the CV we have converted the image to greyscale image with OpenCV to pixel values of the image in a one dimensional array.

```
gray_image = cv2.cvtColor(image, cv.COLOR_BGR2GRAY)
```



#### 4.1.6 Face Detection

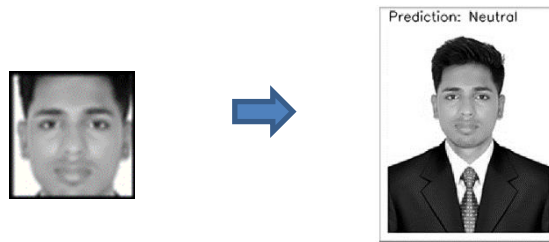
After converting the image to gray image we have detected the face using OpenCV haarcascade\_frontal\_face\_alt.xml classifier model. The XML file containing serialized Haar cascade detector of faces (Viola-Jones algorithm) in the OpenCV library.



```
face_cascade=cv.CascadeClassifier('haarcascades/haarcascade_frontalface_alt.xml')
faces = face_cascade.detectMultiScale(gray_image, 1.25, 6)
for f in faces:
    x, y, w, h = [ v for v in f ]
    cv.rectangle(gray_image, (x,y), (x+w, y+h), (0,0,0), 3)
    face_crop = gray_image[y:y+h, x:x+w]
```

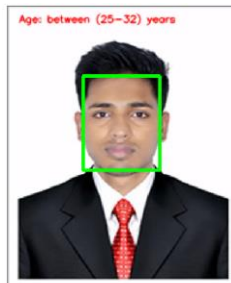
#### 4.1.7 Facial Expression Detection

To detect the facial expression, we first load our pre-trained model. Then we pass our landmark model to Dlib Shape Predictor which is a tool that takes in an image region containing some object and outputs a set of point locations that define the pose of the object. It identifies the locations of important facial landmarks such as the corners of the mouth and eyes, tip of the nose, and so forth. Then we identify hog features from the cropped image. After that, using the HOG features, facial landmarks and the CNN model we identify the facial expression. After finding the predicted emotion we store it in the Feature Storage as 'emotion' as key and predicted emotion as value.



#### 4.1.8 Age Prediction

For predicting age, we create ageNet through OpenCV Deep Neural Network using OpenCV pre-trained age detection model. Then we create a bold instance of the cropped face to set input for our ageNet. Then we run the forward pass from OpenCV library on the ageNet which gives us the predicted age. Then we store the result in Feature storage as 'age' as key and predicted age as value.



#### 4.1.9 Colorfulness

To check if the image is a colored image or gray image we have checked for every pixel value in the image if all red, green and blue are equal, cause a gray image will have only two color white and gray, we know for white and gray color red, green, blue will be equal.

So, if all the pixel satisfies the condition, we store 'colored' as a key and True as value in Feature storage. Else, we store False as value in feature storage.

#### 4.1.10 Object Detection

To check if the person posed with any object in the photo, we have detected objects. To detect object, we have used pre-trained tinyYOLOv3 model and ImageAI library.

```
detector = ObjectDetection()
detector.setModelTypeAsTinyYOLOv3()
detector.setModelPath('yolo-tiny.h5')
detector.loadModel();
detection = detector.detectObjectsFromImage(imageName,r'outputObjects.jpg')
objects=[]
for eachItem in detection:
    objects.append(eachItem["name"])
```

After getting the list of detected objects we stored the list as value and 'Objects' as key in Feature Storage.

At the end of all operations if feature extraction we get feature storage as the output.

Feature = {'emotion': 'Neutral', 'colored': True, 'Objects': ['tie'], 'age': '25-32'}

## 4.2 Trait Determination

After Feature extraction is complete, we get the Features in a dictionary as output. To determine the trait, we set the weight for each trait. The weight for a trait is number of features to be matched for the trait.

$$\text{Weight}_{\text{extravert}} = 3,$$

$$\text{Weight}_{\text{conscientious}} = 4,$$

$$\text{Weight}_{\text{agreeable}} = 3,$$

$$\text{Weight}_{\text{openness}} = 4,$$

$$\text{Weight}_{\text{neurotic}} = 3$$

After that, we initialize a matching of all five traits as 0%, we check for each key-value in Feature Storage, if it matches a feature of a trait, we calculate the matching of a particular trait by following equation:

$$\text{Match}_{\text{trait}} += (1/\text{Weight}_{\text{trait}}) * 100 \%$$

After all the key-value pair is checked in Feature storage, we get the matchings of all five traits in Percentage.

```

match={"extravert":0,"openness":0,"agreeable":0,"conscientious":0,"neurotic":
0}
weight={"extravert":3,"openness":4,"agreeable":3,"conscientious":4,"neurotic"
:3}
for key in features:
    value=features[key]
    if key=="emotion":
        if value=="Happy":
            match["extravert"]+=round((1/weight["extravert"])*100)
            match["agreeable"]+=round((1/weight["agreeable"])*100)
        elif value=="Sad":
            match["openness"]+=round((1/weight["openness"])*100)
        elif value=="Neutral":
            match["conscientious"]+=round((1/weight["conscientious"])*100)
            match["neurotic"]+=round((1/weight["neurotic"])*100)
    elif key=="colored":
        if value:
            match["conscientious"]+=round((1/weight["conscientious"])*100)
            match["agreeable"]+=round((1/weight["agreeable"])*100)
        else:
            match["neurotic"]+=round((1/weight["neurotic"])*100)
            match["openness"]+=round((1/weight["openness"])*100)
    elif key=="Objects":
        if len(value)>0:
            match["openness"]+=round((1/weight["openness"])*100)
    elif key=="age":
        lower=int(value[:2])
        upper=int(value[-2:])
        if age<lower:
            match["conscientious"]+=round((1/weight["conscientious"])*100)
        elif age>upper:
            match["extravert"]+=round((1/weight["extravert"])*100)
        else:
            match["neurotic"]+=round((1/weight["neurotic"])*100)
for personality in match:
    print(f'{personality.upper()}: {match[personality]}%')

```

From Age Detection, facial expression detection, object detection, image colorfulness checking we have got the features as output. Using the features, we have computed the matching of each trait of the person from image. Our final result is as follows:

EXTRAVERT: 0%
OPENNESS: 25%
AGREEABLE: 33%
CONSCIENTIOUS: 75%
NEUROTIC: 33%

#### **4.3 Discussion:**

From our result it is evident that from CV image we can classify an individual among big five personality traits. In our result we got most percentage for conscientious which is 75%. The person is 33% neurotic and agreeable equally, got 25% openness and 0% extravert. According to our results we reached to a decision that the image of this candidate reflects his personality as conscientious. Extravert users mostly use happy mood expressions and, they use glass. The extravert users predicted age tends to be less than their original age. Agreeable users also use happy mood expressions and colored image, but they do not use glass. Conscientious users use neutral emotion expression with colored image, and they use glass. The conscientious users predicted age tends to be greater than their original age. Neurotic and openness users do not use colored image. Neurotic user mostly expresses a neutral emotion. But openness users mostly express a sad emotion, and they use glass. Openness users mostly have some object inside their image. The neurotic users predicted age tends to be the same as their original age. To check our machine performance, we performed some data analysis on a dataset. We classified the dataset by unsupervised learning techniques. After performing the technique and other analysis we clustered the data into 5 group. At the same time the participant image was also checked for classifying the personality. Our main goal was to cross check our machine performance with dataset result. The result of image processing machine was able to give 75% correct result comparing to the dataset result.

---

## **Chapter 5 Conclusion & Future work**

---

### **5.1 Conclusion**

In this paper, we investigate automatic personality prediction from CV of a candidate with Big-Five personality traits. We used facial features for this. We inferred that each candidate with one of the personality traits attach a specific type of picture in their CV.

We didn't find any study as personality prediction from CV using machine learning algorithm or personality prediction from resume image. So, this study is the first study in this topic. After classifying candidates among big five personal traits by analyzing their images, we checked the validity of our result by analyzing data of a dataset using unsupervised learning techniques. After performing the technique and other analysis, we found that our prediction was 75% correct comparing to the dataset result.

We could have gotten more accurate results if we could detect the resolution and eyeglass from the image. So far from our study, we were able to achieve big five personality trait matching from resume image. The study can be improved much more in future

### **5.2 Future Work**

Although there is still room for improvement in the prediction of personality traits from CV. We had some limitations here such as, we could not detect the glasses of the users and could not define the resolution of their images. We hope to work on these next times. Moreover, we could not predict directly from the CV input. Rather we have used CV templates here. We may try to implement these in our work later.



---

## References

---

- [1] Faber B.,” COLOR PREFERENCE AS A CLUE TO PERSONALITY”, Art Psychotherapy, Vol. 1 pp. 13-16, Pergamon Press, 1973.
- [2] Moon Hwan K., j. Young Hoon , and Bae Park J.,” Emotion Detection Algorithm Using Frontal Face Image”, ICCAS2005, KINTEX, Gyeonggi-Do, Korea , july 2015.
- [3] Hsi-Chieh L., W. Chia-Ying and L. Tzu-Miao, “Facial Expression Recognition Using Image Processing Techniques And Neural Network”, *Advances in Intelligence System & Application*, SIST 21, pp. 259-267, copyright Springer-Verlag Berlin Heidelberg 2013.
- [4] A. Mercy Rani, R. Durgadevi, “Image Processing Techniques To Recognize Facial Expression”, *International Journal Engineering And Advanced Technology(IJEAT)*, ISSN: 2249-8958, Volume-6 Issue-6, August 2017.
- [5] Jayashre R., B. Sudhir, Y. Pooja and P. Nirmiti, “ Personality Evaluation and CV Analysis using Machine Learning Algorithm”, *International Journal of Computer Science and Engineering(JCSE)*, E-ISSN: 2347-2693, Vol.-7, Issue-5, May 2019.
- [6] Rosmadi Hj Mohd D., “PERSONALITY”, *Journal of Open University Malaysia*, September 2015.
- [7] Miriam Redi, Daniele Quercia, Lindsay T. Graham, and Samuel D. Gosling ,“Like Partyng? Your Face Says It All. Predicting the Ambiance of Places with Profile Pictures”, Proceedings of the Ninth International AAI Conference on Web and Social Media, 2015.
- [8] Goldman, A.I., and Sripada, C.S. “Simulationist models of face-based emotion recognition” *Cognition* 94(3):193-213.

- [9] P. Viola and M. Jones, "Rapid object detection using a boosted cascade of simple features", *Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. CVPR 2001*, 2001. Available: 10.1109/cvpr.2001.990517 [Accessed 11 November 2020].
- [10] Cole, M., Feild, H. and Giles, W., 2003. Using Recruiter Assessments of Applicants' Resume Content to Predict Applicant Mental Ability and Big Five Personality Dimensions. *International Journal of Selection and Assessment*, 11(1), pp.78-88.
- [11] Shen, J., Brdiczka, O. and Liu, J., 2013. Understanding Email Writers: Personality Prediction from Email Messages. *User Modeling, Adaptation, and Personalization*, pp.318-330.
- [12] Skillroads.com. 2020. *What Does CV Design Tells About Your Personality? Skillroads.Com - AI Resume Career Builder*. [online] Available at: <<https://skillroads.com/blog/what-does-cv-design-tells-about-your-personality>> [Accessed 7 November 2020].
- [13] Bhatti, S., Muneer, A., Lali, M., Gull, M. and Din, S., 2017. Personality analysis of the USA public using Twitter profile pictures. *2017 International Conference on Information and Communication Technologies (ICICT)*.
- [14] Roccas, S., Sagiv, L., Schwartz, S. and Knafo, A., 2002. The Big Five Personality Factors and Personal Values. *Personality and Social Psychology Bulletin*, 28(6), pp.789-801.
- [15] Y. Li, J. Zeng, S. Shan and X. Chen, "Occlusion Aware Facial Expression Recognition Using CNN with Attention Mechanism," in *IEEE Transactions on Image Processing*, vol. 28, no. 5, pp. 2439-2450, May 2019.

- [16] T. J. Alhindi, S. Kalra, K. H. Ng, A. Afrin and H. R. Tizhoosh, "Comparing LBP, HOG and Deep Features for Classification of Histopathology Images," *2018 International Joint Conference on Neural Networks (IJCNN)*, Rio de Janeiro, 2018, pp. 1-7.
- [17]. L. Cuimei, Q. Zhiliang, J. Nan and W. Jianhua, "Human face detection algorithm via Haar cascade classifier combined with three additional classifiers," *2017 13th IEEE International Conference on Electronic Measurement & Instruments (ICEMI)*, Yangzhou, 2017, pp. 483-487.
- [18] N. Angeeta, Kavita A., "Personality Prediction Using Profile Picture", *International Journal of Engineering Research in Computer Science and Engineering(IJERCSE)*, ISSN(ONLINE) 2394-2320, Vol 5, Issue 4, April 2018.
- [19] S. Jianqiang, Oliver B. and Juan L., "Understanding Email Writers: Personality Prediction from Email Messages", Palo Alto Research Center, 3333 Coyote Hill Road, Palo Alto, CA 94304, USA
- [20] S. Michael, Hubert S. and William F., "Using Recruiter Assessments of Applicant's Resume Content to Predict Applicant Mental Ability and Big Five Personality Dimensions", *International Journal of Selection and Assessments*, Vol 11, Number 1, March 2003
- [21] R. Allan, Kaushik S., Kaushik A., Keval J. and Shalmali J., "Personality Prediction System Through CV Analysis", *International Research Journal of Engineering and Technology(IRJET)*, e-ISSN: 2395-0056, p-ISSN: 2395-0072, Vol 6, Issue 2, February 2019

