FACIAL RECOGNITION FOR HUMAN DISPOSITION IDENTIFICATION

ANBANANTHAN PILLAI A/L MUNANDAY

BACHELOR OF ELECTRICAL ENGINEERING (ELECTRONICS) WITH HONOURS

UNIVERSITI MALAYSIA PAHANG

UNIVERSITI MALAYSIA PAHANG

DECLARATION OF THESIS AND COPYRIGHT

Author's Full Name	:	ANBANANTHAN PILLAI A/L MUNANDAY
Date of Birth	:	28/07/1997
Title	:	FACIAL RECOGNITION FOR HUMAN DISPOSITION IDENTIFICATION
Academic Session	:	20/21

I declare that this thesis is classified as:

☐	CONFIDENTIAL	(Contains confidential information under the Official Secret Act 1997)*
☐	RESTRICTED	(Contains restricted information as specified by the organization where research was done)*
☒	OPEN ACCESS	I agree that my thesis to be published as online open access (Full Text)

I acknowledge that Universiti Malaysia Pahang reserves the following rights:

1. The Thesis is the Property of Universiti Malaysia Pahang
2. The Library of Universiti Malaysia Pahang has the right to make copies of the thesis for the purpose of research only.
3. The Library has the right to make copies of the thesis for academic exchange.

Certified by:

ANBANANTHAN PILLAI	
(Student's Signature)	(Supervisor's Signature)
970728-04-5509	Associate Professor Ir Ts Dr Fahmi Bin Samsuri
New IC/Passport Number	Name of Supervisor
Date: 11/2/2022	Date: 13-02-2022

NOTE : * If the thesis is CONFIDENTIAL or RESTRICTED, please attach a thesis declaration letter.

THESIS DECLARATION LETTER (OPTIONAL)

Librarian,
Perpustakaan Universiti Malaysia Pahang,
Universiti Malaysia Pahang,
Lebuhraya Tun Razak,
26300, Gambang, Kuantan.

Dear Sir,

CLASSIFICATION OF THESIS AS RESTRICTED

Please be informed that the following thesis is classified as RESTRICTED for a period of three (3) years from the date of this letter. The reasons for this classification are as listed below.

Author's Name

Thesis Title

Reasons (i)

(ii)

(iii)

Thank you.

Yours faithfully,

____

(Supervisor's Signature)

Date:

Stamp:

Note: This letter should be written by the supervisor, addressed to the Librarian, *Perpustakaan Universiti Malaysia Pahang* with its copy attached to the thesis.

SUPERVISOR'S DECLARATION

I/We* hereby declare that I/We* have checked this thesis/project* and in my/our* opinion, this thesis/project* is adequate in terms of scope and quality for the award of the Bachelor of Electrical Engineering (Electronics) with Honours.

____

(Supervisor's Signature)

Full Name :
Position :
Date :

ASSOC. PROF. IR. TS. DR. FAHMI BIN SAMSURI
ASSOCIATE PROFESSOR
DEPARTMENT OF ELECTRICAL ENGINEERING
COLLEGE OF ENGINEERING
UNIVERSITI MALAYSIA PAHANG
LEBUHRAYA TUN RAZAK
26300 GAMBANG, KUANTAN, PAHANG
TEL : +609-549 2338 FAX : +09-424 5055

13-02-2022

____

(Co-supervisor's Signature)

Full Name :
Position :
Date :

STUDENT'S DECLARATION

I hereby declare that the work in this thesis is based on my original work except for quotations and citations which have been duly acknowledged. I also declare that it has not been previously or concurrently submitted for any other degree at Universiti Malaysia Pahang or any other institutions.

ANBANANTHAN PILLAI

____

(Student's Signature)

Full Name : ANBANANTHAN PILLAI A/L MUNANDAY

ID Number : EA18181

Date : 11 February 2022

FACIAL RECOGNITION FOR HUMAN DISPOSITION IDENTIFICATION

ANBANANTHAN PILLAI A/L MUNANDAY

Thesis submitted in fulfillment of the requirements
for the award of the Bachelor of Electrical
Engineering (Electronics) with Honours

College of Engineering
UNIVERSITI MALAYSIA PAHANG

FEBRUARY 2022

ACKNOWLEDGEMENTS

First and foremost, I would like to convey my huge thanks to my supervisor, Prof. Dr. Ir Fahmi bin Samsuri for the patient guidance, encouragement and advice he has provided throughout our time as his students. I have been extremely lucky to have a supervisor who cared so much with my work, and who responded to my doubtful questions and queries so promptly. With the confusion in this project, my supervisor taught me on simplifying the deep learning and understanding with programming language used in my project. We would also like to thank all the members of staff at University Malaysia Pahang who helped me succeed in my project. Also, I would like to thank Dr. Ikhwan for giving me advices on how to conduct best PSM presentation.

Moreover, I must express my gratitude to our family and fellow friends, for their continued support, encouragement as well as their patience because of experienced my ups and down of my project. Friends from other faculties such as FKM and FKP, they have been provided many great suggestions in software used for image processing and they also gave me advice since they are undertaking final year project along with me.

Not forgetting, the postgraduate students from University Malaysia Pahang and University Malaya, who provided many ideas for my research, also deserve thanks for helping me keep things in perspective. Furthermore, I would like to thank the university seniors who contribute a lot to my projects in terms of cost or ideas and also gave me advice on what are the challenges arise when it comes to final year project and suggested ways to overcome it easily. These people contribute a lot in this project. A huge thanks to these people.

In addition, I would like to thank the FYP coordinator named Madam Nur Huda binti Mohd Ramlan had been guiding me with keeping track on PSM progress and timeline. Also, huge thanks to my FYP 1 panels named Dr. Ahmad Afif Bin Mohd Faudzi and Dr. Asrul bin Adam, FYP 2 panels named Dr. Norizam bin Sulaiman, Dr. Norazlianie binti Sazali and Dr. Ikhwan Hafiz bin Muhammad for helping me sort out the minor problems and provide positive response towards my projects.

Finally, I would like to thank the College of Engineering, not only for providing the opportunity to learn the deep learning, which is very new to my extent of knowledge, but it is pretty easy to handle and learn faster. Also, I had opportunity to meet so many interesting and brilliant people. These FYP helped me to get prepared for handling real projects or contracts in working environment. Thus, final year project helped me to feel the pressure as an engineer and get prepared to be an engineer in future.

ABSTRAK

Dalam projek ini, terdapat perbincangan yang lebih lanjut mengenai model pembelajaran yang mendalam dan gunakan kaedah dengan betul yang boleh membantu pemprosesan imej. Terdapat banyak model pembelajaran mendalam dan model yang sesuai dipilih supaya boleh memenuhi keperluan operasi sistem seperti kelajuan dan ketepatan keputusan. Evolusi metodologi dilaksanakan dalam perancangan sistem dengan menggunakan teknik pemprosesan imej termasuk teknik dapatan imej, penambahbaikan imej (atau dikenali sebagai tahap pra-pemprosesan) dan pengeluaran ciri-ciri imej. Pada peringkat awal, berlaku beberapa peringkat pra-pemprosesan untuk penambahbaikan imej. Kemudiannya, muka akan didapatkan. Akhirnya, wajah akan diklasifikasikan ke dalam salah satu kelas yang berbeza menggunakan ‘CNN’ modal. Sistem ini Berjaya mencapai objektif yang disasarkan berdasarkan keputusan ujian yang menunjukkan 92.86%. Untuk projek ini, ia bertujuan untuk mendapatkan pengesanan yang tepat mengenai ungkapan manusia melalui aplikasi dan mengekstrak emosi / kelas dalam bentuk peratusan.

ABSTRACT

Human disposition identification and recognition has become one of the popular topics under OpenCV based on deep learning. The importance of this project is to recognize facial expressions. Here, the discussion will be done about the deep learning models and use it properly that can assist the image processing. There are many deep learning models and the suitable model for this project chose according to the ability to meet the system operation requirements such as speed and accuracy. Evolutionary methodology was implemented in this system design by using several image processing techniques include image acquisition, image enhancement (or known as pre-processing stages) and feature extraction. The system first applies some pre-processing stages to enhance the input image and reduce the noise. The face boundary will then be detected. The region of interest such as mouth and eyes will be determined, from which, features will be extracted. Finally, the face will be classified into classes using the CNN model based on the features extracted. The method was applied and tested on a dataset of faces (FER-2013) and the success rate obtained was 92.86%. For this project, it is targeted to get the accurate detection of human dispositions through the application and extract the emotions/classes in percentage.

TABLE OF CONTENT

LIST OF TABLES

LIST OF FIGURES

LIST OF SYMBOLS

*	Multiplication Symbol
/	Division Symbol
“”	Index to implement text in Python
=	Equal Symbol

LIST OF ABBREVIATIONS

OpenCV	Open Computer Vision
FAPU	Facial Animation Parameter Units
FDP	Facial Definition Parameter
HHI	Human-Human Interaction
IDLE	Interactive Development and Learning Environment
CNN	Convolutional Neural Network
UMP	University Malaysia Pahang
HRI	Human Robot Interference
AU	Action Units
FACS	Facial Action Coding System
PCA	Principal Component Analysis
PNN	Probabilistic Neural Networks
AAM	Active Appearance Model
ASM	Active Shape Model
DNN	Deep Neural Network
AI	Artificial Intelligence
CNTK	Cognitive ToolKit
FER-2013	Facial Emotion Recognition-2013
ROI	Region of Interest

LIST OF APPENDICES

CHAPTER 1

INTRODUCTION

1.1 Background Research

This chapter will give a summary of the problem and extract the topic's aims and objectives. The field of computer science known as facial recognition deals with methods and procedures for estimating a person's mood based on their human dispositions. Several technological advances in the fields of Machine Learning and Artificial Intelligence have simplified the process of identifying dispositions. The next computer communication channel is expected to be disposition. The majority of research in this topic focuses on recognising human moods or dispositions from input (human face) and real-time applications. Most of the research has been on face recognition and matching, but there is no Convolutional Neural Networks (CNN) have been utilised to infuse human dispositions from real-time applications.

Deep Learning is a type of machine learning that models data to achieve a certain goal. Deep learning in neural networks offers a wide range of applications in the disciplines of image identification, classification, decision making, and pattern recognition. Other deep learning approaches, such as multimodal deep learning, are used for feature selection and image recognition.

According to Mehrabian [1,] spoken words constitute for around 7% of what a listener understands in human-human interaction (HHI), with the other 93 percent consisting of nonverbal communication behaviour such as body language and tone. Based on other research, the most expressive methods for humans to transmit their sentiments are human facial dispositions and body movements [2, 3]. According to research [4], people are more likely to see computers as human-like when they comprehend and demonstrate acceptable nonverbal conversational behaviour. As a result, if computers can comprehend and recognise their human counterpart's nonverbal behaviour and affective state, human-computer interaction (HCI) will become more natural.

Since it allows individuals to transmit their moods, feelings, and dispositions without using words, facial recognition is one of the most significant parts of human communication. According to Darwin, facial recognition for human disposition identification has biological underpinnings and play a critical role in the survival of species, including humans. Ekman identified seven main human dispositions: anger, sad, fear, disgust, surprise, neutral, and happy, as shown in Figure 1.1, and believes that these emotions are universal. Human disposition is determined by the activation/dilation of one or more of the forty-three facial muscles. Several research have recently focused on the extraction and identification of facial traits in order to determine human temperament. It is largely used to foster seamless interaction between computers and their users, with the purpose of allowing computers to detect and respond to the user's mood based on their identified dispositions.

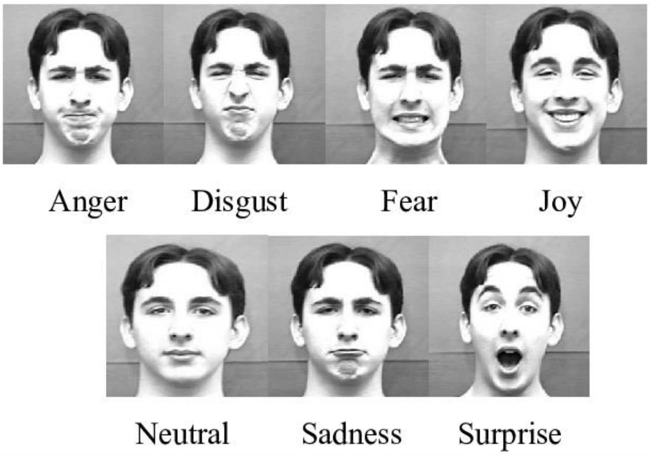


Figure 1.1 Seven basic human dispositions

1.2 Problem Statement

Individuals experience seven primary dispositions: fear, sadness, surprise, disgust, happy, anger and neutral. The activation of distinct groups of face muscles expresses our facial dispositions. These seemingly little but complex signals in our expressions might disclose a lot about our mental state. The effects of content and services on the audience/users may be measured using facial recognition, which is a simple and low-cost technology. As an example, this technology might be used by retailers to gauge customer interest. More information on a patient's mental state during treatment might aid healthcare providers in providing better care. Entertainment producers can measure audience participation throughout events in order to present desired content on a consistent basis.

Darwin's studies were the foundation for Ekman and Friesen's work. They claimed that the relationship between facial muscle patterns and specific emotions (happiness, sadness, anger, fear, surprise, disgust) may reveal the universals of human dispositions [6]. They predicted that cultural variations would show up in some of the signals that learn to activate emotions, the rules for managing facial behaviour in certain social circumstances, and many of the emotional arousal consequences [7].

A person's ability to recognise facial based on their human dispositions is influenced by a variety of circumstances. Social aspects that impact one's perception of another's dispositional state include deception and display norms. Several projects on this topic have already been performed, and our goal is to not only build an autonomous Facial Recognition System for Human Disposition Identification, but also to increase its accuracy in comparison to other systems.

Figure 1.2 Normalization of face illumination

Last but not least, certain common challenges that hamper the development of a human disposition identification system based on facial recognition still exist in the real-time environment, such as extracted features being sensitive to changes in illumination and noise reduction, as shown in Figure 1.2. As a result, even little changes in illumination and noise can affect the accuracy rate of recognition. Another factor that affects the performance of such systems is the large amount of data involved.

1.3 Objectives

By considering to the problem as stated in previous chapter, the main objectives are highlighted as below:

1. To investigate and study on applications development for facial recognition and disposition using Deep Learning algorithm that allows to recognize and differentiate multiple human facial dispositions.

2. To identify and perform analyses that will determine the optimum architecture and the suitable image processing techniques for facial disposition recognition.

3. To verify the performance of the apps being used as an automated monitoring system for human disposition identification based on facial expressions, thus enabling for Intelligent Human-Computer Interaction.

1.4 Scope

For my research, I will be choosing a model for my study that can detect human dispositions with high accuracy while fast in detection and efficient. The identification system should also be automated, so that it can run for long periods of time and recognise human dispositions automatically through PC webcam.

The study's main purpose is to use deep learning techniques to use facial recognition to assess human dispositions identification. The Convolution Neural Network (CNN) model is used by Python's Integrated Development and Learning Environment (IDLE) to perform high-accuracy disposition identification in real-time applications using OpenCV. The system was initially tested in the Jupyter Notebook. It was selected above other deep learning systems not just for its accuracy, but also for its speed, which allows it to process more frames per second. Furthermore, a brief peek at each frame indicates how rapidly everything operates.

This system might then be used for educational purposes, such as obtaining feedback on how UMP students behave during online or face-to-face sessions. This project also seeks to test a camera that can automatically recognise faces and dispositions of human faces, as well as implement the output in percentages and collect sample data. This is to see how well it is in detecting and correctly predicting human dispositions based on facial recognition.

CHAPTER 2

LITERATURE REVIEW

2.1 Human Dispositions

This research is about human dispositions based on expression, which has been recognized by facial recognition for over a decade and a vital topic in computer vision and deep learning. This system is applicable to wide range of applications, including healthcare, education, criminal investigation, and Human Robot Interface (HRI). Various approaches for face dispositions have been presented throughout the years. Many methods have been proposed for developing an application that can detect effectively. In many encounters, computer programmes might improve communication by adjusting replies based on the emotional state of human users. The eyes are often referred to as "windows to the soul", which leads to the apparent conclusion that the entire face, not just the eyes, may reveal the individual's "hidden" feelings. The face of the human is most complex and flexible to all the animals. The human face is a sophisticated and adaptable tool that may be used for many purposes, including conveying one's facial words. This helps people to communicate better by making their actions more predictable and understandable to others. When person speaks attitude toward the information being given can be revealed lively through facial expression [5].

As face can indicate emotion and pain, regulate social behaviour, and reveal brain function, research in psychology has indicated that at least six facial expressions are universally associated with distinct human dispositions. Figure 2.1 shows six principle dispositions are: happiness, sadness, surprise, fear, anger, and disgust. Several other emotions and many combinations of emotions have been studied but remain unconfirmed as universally distinguishable [8].

According to psychological study, faces may express emotion and suffering, influence social behaviour, and reveal brain activity. Studies shows that, minimum six human dispositions are frequently linked with well-defined facial expressions. Figure 2.1 depicts the six basic emotions of happiness, sadness, surprise, fear, anger, and disgust. A number of other emotions have been investigated, although their universality has yet to be determined [8].

Figure 2.1 Six principal emotions

2.2 Case Study on Existing System

Darwin's study pioneered the topic of facial recognition, which has been actively explored in psychology during the previous twenty years [9]. In this part, the studies will be addressed.

2.2.1 Vision-based Facial Expression Recognition

In the recent decade, facial recognition for human disposition identification analysis has generated interest in artificial intelligence and machine vision, with the objective of building systems that can interpret and utilise this nonverbal communication way of human. The majority of technology for analysing human facial dispositions automatically classified into two categories:

I. Facial expression recognition systems that can discriminate between archetypal facial expressions and emotions (happy and angry)

II. Facial actions recognition systems (eyebrow raise and frown)

System that can recognise a restricted set of prototypic emotional expressions from static images or visual sequences, such as pleasure, surprise, rage, sadness, fear, and disgust, have undergone extensive investigation. The work of Ekman [10] [11], who claimed that fundamental emotions have prototypic expressions inspired this focus on emotion-specific expressions.

2.2.2 Systems that Recognize Prototypic Facial Expressions

Analysis of face may be done automatically in two different ways: from static images or from frames of video. People are shown images of facial expression and the link between the expression components and the observers' judgements is analysed in studies of facial expression recognition from static pictures. An expressive face and neutral face were employed in these assessments, both of which are static representations of facial expressions. Facial expression recognition from the sequence of photo based on recognising features of facial and analysing the quantity of facial movement to categorise 5-7 kinds of prototypic face emotions.

A variety of possibilities have been examined. Among them are methods for relating face images to physical models of musculature and facial skin [12, 13, 14], shapes measurements and facial features and their spatial arrangements [15], holistic spatial analysis of the pattern using principal component analysis (PCA) techniques [16,17], and methodologies of relating images of face to physical models of skin and musculature [12, 14]. All these systems work in the same way: they take information from photos or videos, then feed it into a classification system, which generates one of the already selected categories of emotions. The retrieved properties and the classifiers used to discriminate between the various emotions are the fundamental differences.

2.2.3 System that Recognizes Facial Action

Six universal face expressions are insufficient to explain all facial expressions, according to study [18]. Although prototypic emotions such as fear, happiness and surprise are common in everyday life, they only explain a portion of face expression.

One or two facial traits are employed to communicate emotion, such as obliquely lowering the lips corner in sadness and lips tightening in the anger [14]. Other feelings like befuddlement, boredom, and impatience have no classic expressions. To capture the paralinguistic communication and subtlety human emotions, automated detection of fine-grained changes in face expression is necessary. As a result, algorithms for recognising facial expressions based on vision have been created for personality tasks. In general, as illustrated in Figure 2.2, Paul Ekman's Facial Action Coding System (FACS) [15] favours methods that try to recognise action units (AUs).

2.2.3.1 Previous Work on Recognizing Facial Action

Optical flow over the full facial measurements or face feature measurement were utilised to directly identify action units in various earlier studies. Despite the fact that several AUs had optical flow patterns that matched them but did not attempt to identify them. The use of an animation style coding system influenced by FACS [20, 21] was presented as a method for analysing expressions into fundamental movements. Then it was observed that local principal component analysis (PCA) outperformed whole face Eigen faces for expression detection [13].

The Facial Action Coding System (FACS) is the most extensively used approach for measuring face movement in behavioural research. FACS has been successfully used to distinguish between simulated and real pain, telling the truth vs lying, and suicidal versus non-suicidal patients [22]. Ekman and Friesen devised the Facial Action Coding System (FACS) in 1978 to use AUs to characterise facial expressions. This approach was created to identify changes in human facial expressions. This method discusses how various portions of the face are positioned to convey distinct emotions. These includes the chin, cheeks, lips, root of nose, eyes, forehead, and the eyebrows. There are 44 AUs in all, as seen in Figure 2.2.

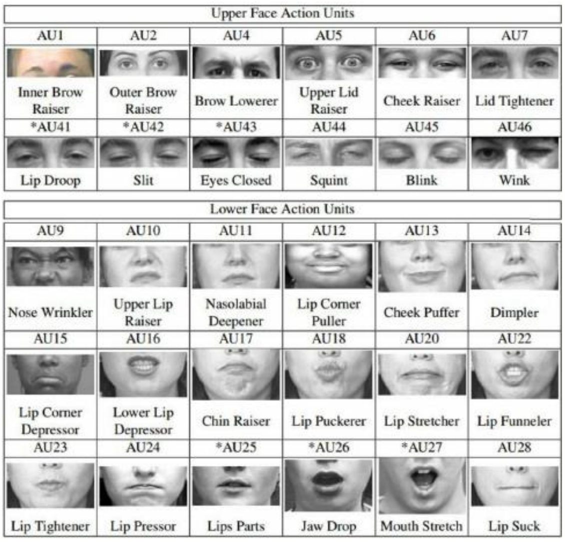


Figure 2.2 Sample AUs and appearance changes as they describe

Table 2.1 FACS Action Units (AU)

Action Unit	**Description**
Brows	
AU 1+2	The brows are elevated on the inside and outside.
AU 1+4	The middle of the brow is brought together and lifted.
AU 4	Brows are brought together and lowered.
Eyes	
AU 5	The upper eyelids are elevated, causing the eyes to expand.
AU 6	The lower eye and infra-orbital furrows have been elevated and deepened, and the eye opening has been narrowed.
AU 7	Lower eyelids are drooping, narrowing the eye aperture.
Mouth	
AU 27	The jaw is expanded, and the mouth is stretched open.
AU 26	The jaw is open, and the lips are relaxed.
AU 25	Lips are relaxed, and the jaw is not dropped when parted.
AU 12	The corners of the lips are dragged up and backward.
AU 12+25	AU 12 with gaping mouth
AU 20+25	Lips are parted in the middle and pulled back laterally, with the upper lip slightly lifted or pulled down.

AU 15+17	Lip corners are pulled down and stretched laterally (AU 15), and chin boss is raised, which pushes up the lower lip (AU 17).
AU 17+23+24	AU 17 tightens, narrows, and presses the lips together (AU23+24).
AU 9+17±25	With AU 17, the infra-orbital triangle and the top lip's centre are dragged upwards (AU 9). In 25% of cases, AU 9+17 was found with AU 25.

To identify the MPEG-4 general's Facial Animation Parameter Units, a face recognition device uses feature monitoring of Facial Definition Parameter (FDP) factors, which are likewise described inside the MPEG-4 framework (FAPUs) [22].

Both CMU/Pittsburgh and UCSD are of the maximum crucial studies organizations focusing on automated FACS recognition as a tool for behavioural analysis [18]. They developed an automatic AU (Action gadgets) evaluation machine that uses facial traits to realize sixteen CMU/Pittsburgh movement devices and any mixture of them. Face traits including the eyes, brows, lips, and cheeks are defined the use of multistate templates. Based at the residences of those multistate templates, a Neural Network-based totally classifier recognises the movement units. The amount of human pre-processing is reduced when automatic facial detection is used. However, because the method necessitates manually turning in the templates within the first body of the series, it is not definitely automatic. The UCSD researchers used a hybrid technique that included holistic spatial analysis, explicit measurement of attributes (local feature analysis) such as wrinkles, and motion flow field estimation to identify 6 upper face actions but no AUs occurring in combinations, but this system achieved 96.4% for upper face AUs and 96.7% for lower face AUs recognition rates [10].

The technology has a 91 percent accuracy rate. However, only results from image sequences that had been manually pre-processed were provided [9]. Based on profile contour fiducial points in a profile-view video, the facial gesture recognition system analyses tiny changes in face emotions. A profile contour and 10 profile contour fiducial points are produced to follow the profile face. Using a rule-based strategy to recognise 20 unique AUs that occur alone or in combination, an 85% identification rate is achieved. Pantic has presented a self-adaptive facial expression analyser that divides recorded facial muscle activity into a variety of quantifiable and user-defined interpretation categories.

The identification of upper action units is the focus of a fully automated system that analyses face activity without requiring human input. An infrared sensitive camera with infrared LEDs is used to recognise the pupils. The pupil locations are utilised to locate and normalise the eyes and eyebrow areas in each frame, which are then analysed with PCA (Principal Component Analysis) to get facial feature form parameters. These characteristics are fed into Support Vector Machine-based classifiers, which recognise upper facial action units in all of their potential combinations. For all conceivable AU combinations, the system achieves a recognition accuracy of 62.5 percent. When the person is wearing glasses, however, this technique fails. Because the gadget relies on infrared LEDs, direct sunshine may cause it to malfunction. It must be expanded to identify lower facial action units [11].

Temporal segmentation of face actions in spontaneous facial behaviour captured in actual-international contexts is a vital, hard, and basically omitted discipline of facial image processing. The exponential individual of viable facial motion combos, non-frontal placement, slight to important out-of-plane head movement, a huge sort of temporal scales in facial gestures, and the exponential nature of viable facial movement combinations are all examples. This approach was examined the usage of recorded facial behaviour at some point of face-to-face contacts. With guide FACS (Facial Action Coding System) annotation, the method acquired moderate convergent validity. Furthermore, whilst used to pre-method video for manual FACS annotation, the method extensively boosts productiveness even as additionally pleasant they want for ground-fact statistics for face picture analysis [14].

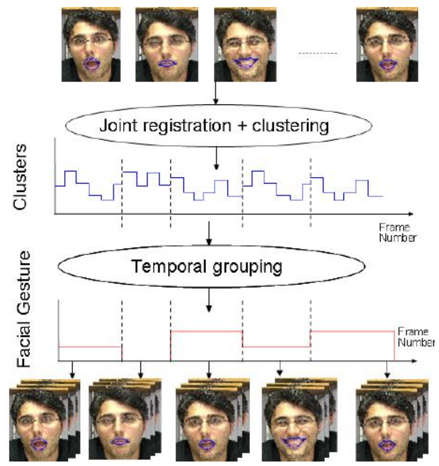


Figure 2.3 Temporal segmentation of facial gestures.

Figure 2.3 shows a processing method for recognizing human emotional behaviour from a speaker's communication using signals representative of verbal and/or nonverbal communication as part of an emotion recognition system. The signal attributes are derived by the processing machine from the input indicators [14]. The processing machine is also designed to enforce at least one intermediary mapping between the signal traits and one or extra components of an emotional ontology for you to produce an emotion recognition judgement. To seize human emotional behaviour, the emotional ontology employs a gradient version.

Changes in one or two discrete face characteristics represent most human emotions. Action Units are used to code these changes (AUs). The authors create a lip form extraction and lip movement tracking device in each static and dynamic facial snap shots the usage of a unique two-step lively contours model. A knowledge-primarily based approach is used to estimate the mouth's authentic position. In the first step [18], lively contour hooks onto stronger higher lip margins the usage of these excessive thresholds. The outlines are deflated with the use of Canny edge detector and balloon energy. Inside the anticipated mouth vicinity, an oval-shaped first lively contour is evaluated.

To control motion of points, four energy terms have been used. At first step active contour locks onto upper lip. Then using lower threshold image gradient snake inflates and locks onto lower lip edges. Then using lower threshold image gradient as well as balloon energy for inflation, snake inflates and locks onto weaker lower lip edges. Extracted lip feature points are used to extract some [17] geometric features to form a feature vector which is used to classify lip images into AUs as shown in Figure 2.4, using Probabilistic Neural Networks (PNN). Experimental results show robust edge detection and reasonable classification where an average AUs recognition rate is 85.98% in image sequences and 77.44% in static images.

The mobility of control points is governed by four energy factors. The active contour is applied to the top lip in the first stage. The snake inflates and hooks onto the bottom lip borders using a lower threshold picture gradient. Using a lower threshold image gradient and balloon energy, the snake then extends and grasps onto weaker edges of lower lip [21]. Extracted lip feature function points are utilised to extract positive geometric attributes to build a feature vector, which is then used to classify lip images into AUs with the use of Probabilistic Neural Networks (PNN), as illustrated in Figure 2.4. The average AUs recognition rate in picture sequences is 85.98 percent, whereas it's 77.44 percent in static photos, indicating strong edge detection and classification.

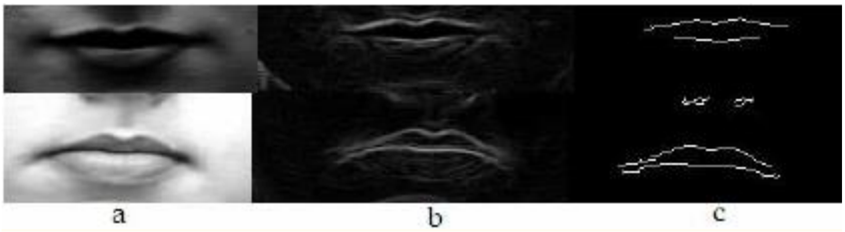


Figure 2.4 a) Original lips b) magnitude of gradient c) canny edge detector

After research of FACS-classified facial muscle movements in the creation of diverse emotions at the Figure 2.2, the landmarks for the active appearance model (AAM) form parameter, as shown in Figure 2.5, were chosen. The goal of AAM was to provide a framework for recognising emotions in facial expressions. The six well-known feelings, as well as a impartial expression, have been labelled in such a way that expressions might be recorded [21]. An AAM was developed using training data and tested on a separate dataset. Test face images have been classified as one of the six emotion-based expressions or a neutral expression using the AAM parameters as classification features. The approach worked effectively in categorising these different face expressions based on still images, as shown in Figure 2.6.

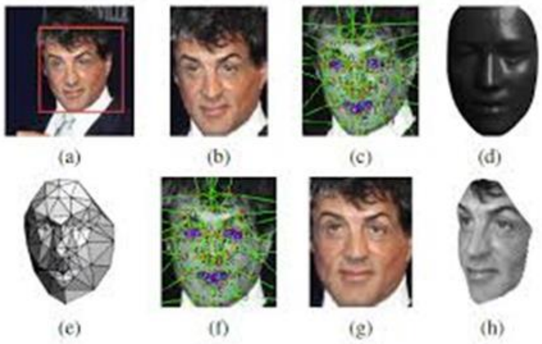


Figure 2.5 Points used in training AAM

Figure 2.6 Viewing AAM generation using training data.

Fear 90 %, Joy 93.3 %, Surprise 79.7 %, Anger 63.9 %, Disgust 93.3 %, Sadness 63.9 %, and Neutral 93.3 % as the perfect. Using merely a Euclidean distance measure, AAM parameters had substantial success and delivered results that were comparable to other strategies.

2.3 Review of Various Techniques

We live in a visually exciting environment with a wide range of shapes, patterns, colours, and textures, as well as movement and tranquillity. Human eyesight is akin to that of a computer when it comes to understanding the visual information included in still images, graphics, and video or moving images in our sensory world. Therefore, it's critical to comprehend the procedures for storing, processing, transferring, recognising, and finally interpreting visual circumstances.

2.3.1 Principal Component Analysis

Principal component analysis, often known as the karhunen-loeve [24] transformation, is a typical data reduction approach used in statistical pattern identification and signal processing. Because patterns typically contain redundant information, mapping them to feature vectors can remove it while keeping the bulk of the pattern's intrinsic information value. These extracted features are crucial in recognising input patterns. A one-dimensional vector of dimension N^2 may be thought of as a two-dimensional face portrait of size N x N.

Each of these vectors has a length of N^2 and represents a linear combination of the original facial images. These vectors are known as "Eigen faces" because they are the eigenvectors of the covariance matrix corresponding to the original face images. Significant eigenvectors of the covariance matrix are formed once the covariance matrix has been calculated. The number of Eigen-vectors used is determined by the application and precision required by the system, and it is clear that as the number of Eigen-vectors increases, the method's accuracy improves but its computational complexity increases.

The method of recognising patterns in data and displaying the data in such a way as to emphasise similarities and contrasts is shown in Figure 2.7. PCA is used to extract features from the input image in order to recognise facial expressions utilising eigen faces. To begin, they construct a training dataset with which to compare the results [24]. Once the face image has been pre-processed, it is compared to the training dataset, which has previously been calculated, but they separated the training set into six fundamental groups based on universal expression (Happy, Surprise, Disgust, Sad, Angry, Fear).

Figure 2.7 Identifying patterns in data

2.3.2 Active Appearance Model

When using the Active Appearance Model, we first gather a large number of face images in various shapes to use as a training set. Then, as shown in Figure 2.8, we use a collection of points to annotate face form, such that the coordinates of these landmarks may be used to indicate face shape.

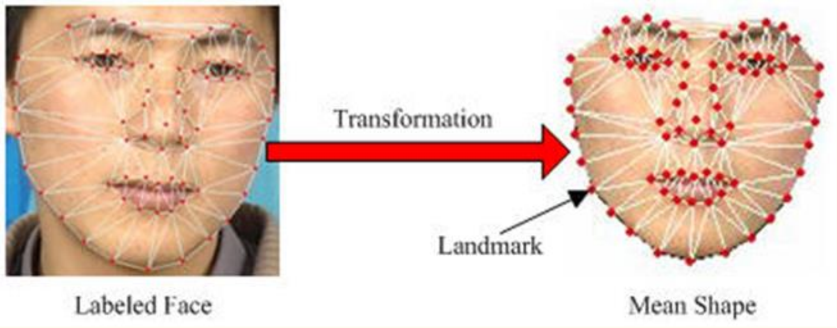


Figure 2.8 Construction of mean face

The mean shape of all the faces may be determined through a number of transformations, such as Principal Component Analysis, as shown in Figure 2.8, to develop a shape model for face alignment. We estimate the model's beginning position given a new face image, compute the proposed moves, and finally produce a decent face alignment result.

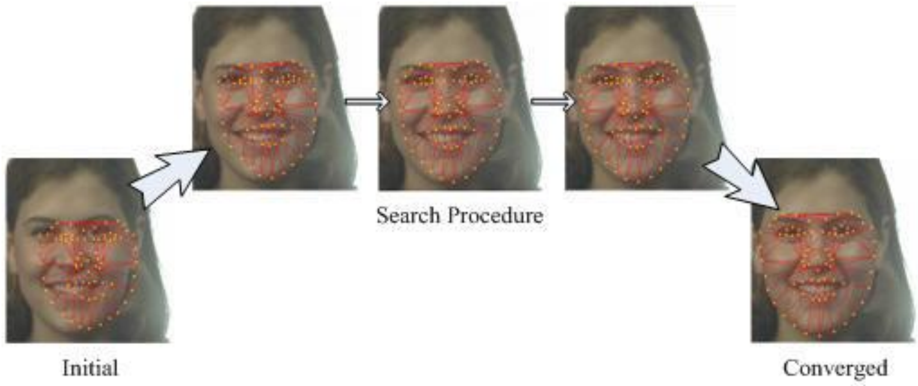


Figure 2.9 Example of face alignment

The alignment for shape and texture modelling and feature extraction is shown in Figure 2.9. It's a popular choice for computer vision applications. By integrating a model of form variation with a model of texture variation, AAM creates statistical appearance models. As a result, the AAM builds a training face picture sequence form and texture combination model. "Textures" are the target image's pixel intensities [25].

2.3.3 Facial Action Coding System (FACS)

The Facial Action Coding Technique (FACS), shown in Figure 2.2, is a system for quantifying facial expression established by Paul Ekman and Wallace Friesen in 1976 [22]. FACS is based on a study of the links between muscle contraction and changes in the appearance of the face. Upper and Lower Face Action Units can be separated on the face. Changes in the face induced by a single or a group of muscles are referred to as Action Units. Changes in face expression are represented by 46 AUs, while eye gaze and head orientation are represented by 12 AUs.

2.3.3.1 Image Processing

We recognise face regions and track facial landmarks established on the contours of brows, eyes, nose, and lips in films at the initial step of the automated FACS system [17]. The Viola-Jones face detector detects an approximate area of a face in each frame of the video. The detector has been shown to be reliable in the face of high inter-subject variability and lighting changes. Using a deformable face model, we seek the precise placement of facial landmarks inside the face area. For a variety of reasons, the Active Shape Model (ASM) is one of the many types of deformable face models. ASM is the simplest and fastest approach among the deformable models, and it matches our purpose to track a huge number of frames in multiple videos.

Figure 2.10 Sample tracking result.

On each face, yellow dots represent the positions of 159 landmarks, while red lines denote distinct facial components, as shown in Figure 2.10.

2.3.3.2 Feature Extraction

Geometric and texture detection are the two forms of detection. Previously, face recognition algorithms were just concerned with geometry or texture data, rather than both. Both characteristics are included since they contribute to the overall picture [23]. For example, geometric modifications can be used to identify certain AUs: Despite the lack of evident textural alterations, the Inner/Outer Brow Raiser (AU1/2) induces eyebrow displacements (increased horizontal creases in the forehead). When Inner Brow Raiser (AU1) and Brow Lowered (AU4) are present at the same time, the geometric displacement of the eyebrows is less evident. In this situation, texture changes (increased vertical creases between eyebrows) provide additional evidence of AU4 presence.

We use similarity transformations to fit the face to the template in order to obtain geometric information from a test face, decreasing within-subject head posture variances and inter-subject geometry discrepancies. Then, utilising edge length differences between each face and a neutral face of the same subject, we build a 436-dimensional vector of geometric attributes, as shown in Figure 2.11, emphasising changes associated with facial emotions while suppressing changes not associated with facial emotions. This figure depicts the method for acquiring geometric features.

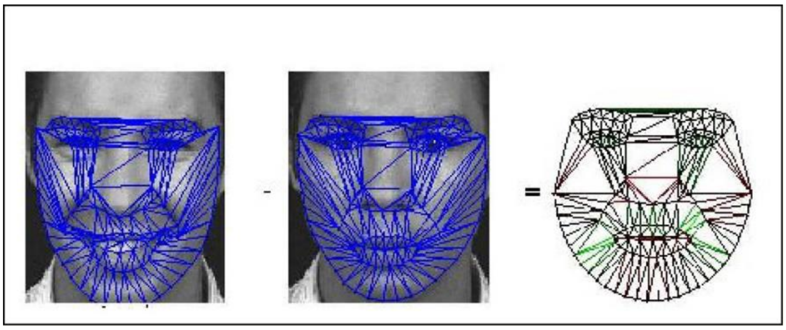


Figure 2.11 Geometric features for Action Unit Classification

We calculate a Gabor wavelet response as Figure 2.12, which has been frequently utilised for face analysis, to extract texture information. We use nine different spatial frequencies in pixel units, ranging from 1/2 to 1/32, and eight different orientations ranging from 0 to 180 degrees with a 22.5-degree step. Each face image is aligned to a template face using landmarks and shrunk or resized to about 100–120 pixels before applying the filters.

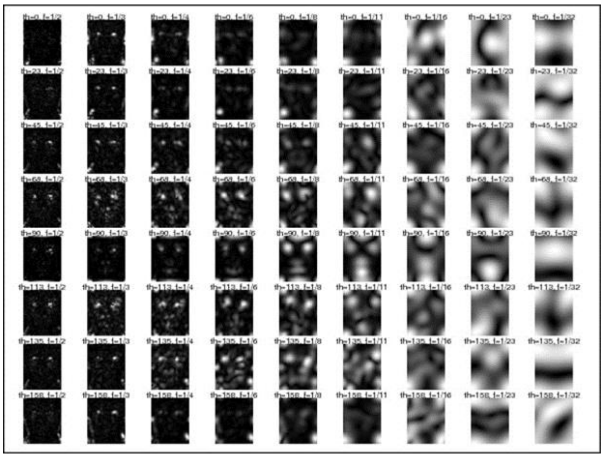


Figure 2.12 Gabor features of multiple spatial frequencies and orientations from a face image

2.3.4 Haar Classifier

Haar classifier-based method is chosen for face detection owing to its high detection accuracy and real time performance. Consists of black and white connected rectangles in which the value of the feature is the difference of sum of pixel values in black and white regions. The computational speed of the feature calculation is increased with the use of Integral image [26].

Since excellent detection accuracy and real-time performance, the Haar classifier-based technique is used for face detection. The value of the feature, which is made up of black and white linked rectangles, is determined by the difference between the sum of pixel values in black and white regions. The usage of an Integral image [26] improves the efficiency of feature calculation.

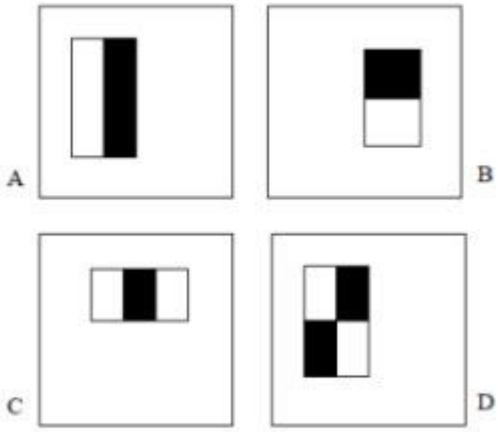


Figure 2.13 Rectangle features of black and white

Rectangular features are depicted in Figure 2.13 in respect to the detection window that surrounds them. When you subtract the total number of pixels in the white rectangles from the total number of pixels in the black rectangles, you get the total number of pixels in the black rectangles. Two-rectangle characteristics are shown in (A) and (B). Figures (C) and (D) shows a three-rectangle feature and a four-rectangle feature respectively.

It calculates the proper threshold for each attribute in order to categorise the faces as positive or negative. Choose the features with the lowest error rate, which means they properly categorise both facial and non-facial images. The weights of misclassified photos are increased with each categorization, starting with an equal weight for each image. The same method is followed after that, the error rates have been recalculated at new values and new weights as well. The method is repeated until the desired accuracy or error rate is obtained, or until the required number of features is determined.

2.4 Software Approach

2.4.1 Introduction to Jupyter Notebook

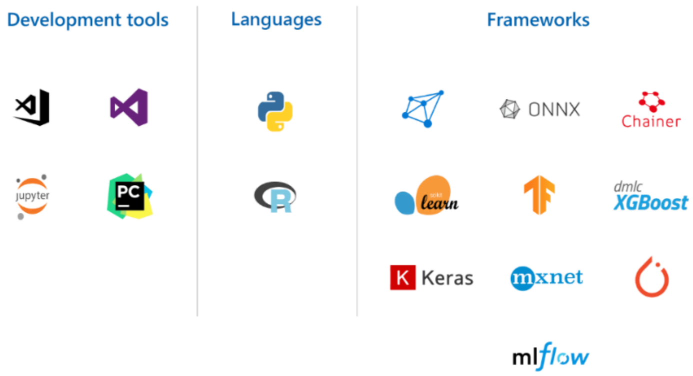


Figure 2.14 Example of Deep Learning Frameworks

Jupyter Notebook where it is an algorithm that assist fast real-time web application that contain both computer code such as Python and rich text elements such as paragraph, equations, figures and an essential tool for data analysis and visualization. Jupyter Notebook caught my attention where as it was most popular at this time. Jupyter Notebook only applies a single neural network to a whole targeted image. According to the Jupyter Notebook, the images will be divided into regions with different weighted which will be then predict the face and create the bounding boxes.

Each region of the image has its own probability, but Jupyter Notebook look once at the facial emotion and process the output. Therefore, Jupyter Notebook is useful to process more facial expressions which it makes extremely fast.

2.4.2 Deep Learning Framework

There are many open-source libraries that are available to develop a deep neural network system. But the most popular among these libraries are Pytorch, Tensorflow and Keras. In order to build a most effective DNN is depends on the usage of libraries and keep updated with latest features of libraries. The reason of using these open-source libraries is that they have pre-installed executable functions that can develop deep neural network easily without having to write the whole coding from the scratch. In short term, deep learning framework ease the whole process of creating a detection system using deep learning.

In that case, we will be focusing what are the functions for each of the framework and the most compatible one will be chosen based on the comparison and understanding. First and foremost, let's take a closer look on Tensorflow 2.0. It is a most popular open-source library developed and highly maintained by Google. The reason why tensorflow is still conquering the AI world is that it makes many improvements on removing inconsistent API and better integration with the Python runtime with new Eager execution. A little drawback is that tensorflow have complex coding to understand and it require a well experiencing on programming.

Next, Keras is known as DNN python library that mainly runs on top of the Tensorflow, Theano or Cognitive ToolKit (CNTK). In short term, Keras uses Tensorflow on the backend and it functions just to refer important libraries from Tensorflow. Hereby, Keras is easy to use and perform faster operation in python coding.

At last, Pytorch 1.6 is a deep learning platform that was built based on Torch and developed by Facebook. This platform uses simple language (English language) as library codes so it makes the platform easier to understand and gives quick start for beginners on deep learning. One of the attractions is that Pytorch is well known among the people due to the group that worked behind Facebook which gain most popularity among people. Thus, figure below summarize the comparisons between the deep learning platform in short term.

	Keras	TensorFlow	PyTorch
Level of API	high-level API[1]	Both high & low level APIs	Lower-level API[2]
Speed	Slow	High	High
Architecture	Simple, more readable and concise	Not very easy to use	Complex[3]
Debugging	No need to debug	Difficult to debugging	Good debugging capabilities
Dataset Compatibility	Slow & Small	Fast speed & large	Fast speed & large datasets
Popularity Rank	1	2	3
Uniqueness	Multiple back-end support	Object Detection Functionality	Flexibility & Short Training Duration
Created By	Not a library on its own	Created by Google	Created by Facebook[4]
Ease of use	User-friendly	Incomprehensive API	Integrated with Python language
Computational graphs used	Static graphs	Static graphs	Dynamic computation graphs[5]

Figure 2.15 Comparison between deep learning

By studying the comparison between different open-source libraries, I can examine that each of the frameworks has similar feature where all of them support python language, have their own pretrained models and also all can work either in GPU or CPU mode Every deep learning framework has its own advantages and disadvantages. But when it comes to my project, I would like to choose easier and understandable technique to implement the project properly.

In a nutshell, I would choose Pytorch as suitable framework to implement my deep neural network and work with it in python environment. The main reason I chose his over other libraries because Tensorflow need deep experience in coding and need to understand everything from scratch and Pytorch libraries are easy to understand and can be implement faster to test out the neural network, I consider Pytorch because of tutorials provided in the internet to build easy neural network faster and get good understanding at the same time.

CHAPTER 3

METHODOLOGY

3.1 Introduction

This chapter will discuss more about the method on understanding the project flow and thus will give clarity to the viewers on what are the processes involve in successful human disposition identification through facial recognition. In this chapter, we will cover the whole process on how the facial recognition and human disposition identification will be applied in the evolutionary methodology.

3.2 Project Flow

The whole process for facial recognition for human disposition shown in Figure 3.1 as a block diagram. This diagram shows the system modules that will be used to produce the output from the system's input. The system's input will be a human face from webcam, and the output will be in percentage based on the predictions. The techniques and process of image acquisition, image enhancement and feature extraction will be discussed in further detail.

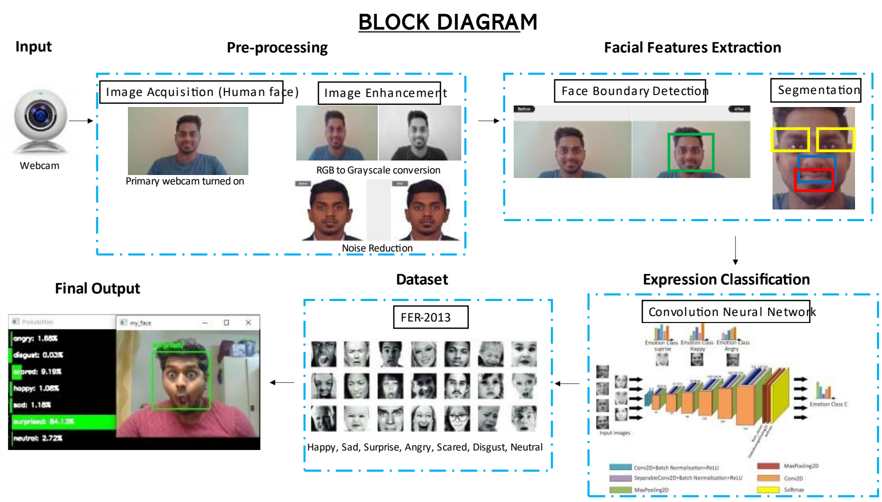


Figure 3.1 Block diagram

Based on the block diagram as shown Figure 3.1, firstly, the human face will be captured from the webcam and the image will be turned into grayscale thus easy to reduce noise or unwanted distortion, so that can enhance the image quality. Then, feature extraction process extracts an image's region of interest for image processing. Here facial boundary detection and segmentation done by Haar Cascade Classifier to recognise human facial organs such as nose, eyes and mouth. Next, The CNN used for filters which are sliding windows in our images that are responsible for detection the features such as color, shapes, edges of images and will compare with the FER-2013 database and classifies the dispositions. Dataset consists of thousands of sample pictures for neutral, happy, sad, anger, disgust, fear and surprise. So, the CNN architecture will test the input with the dataset and gives the accurate output of expressions in percentage in terms of confidence level.

3.3 Flowchart for the Project

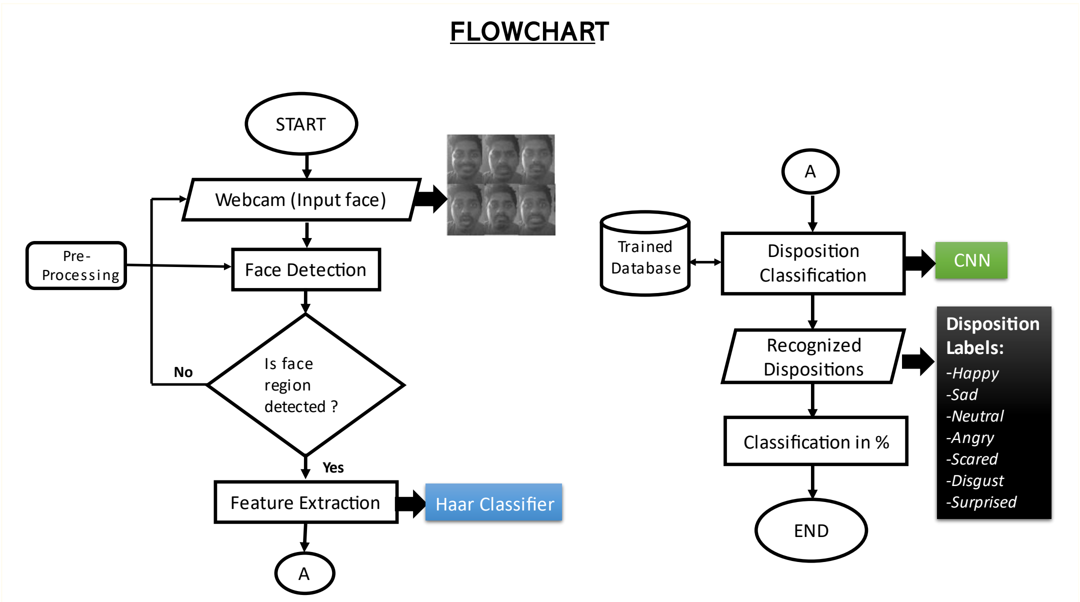


Figure 3.2 Flowchart for the system

Based on the flowchart at Figure 3.2 three major stages which is (i) Face Detection, (ii) Feature Extraction, and (iii) Expression Classification.

During the initial stage, this system requires an input which is a human face captured from the webcam. Once the face has been detected in pre-processing stage, the facial components such as nose, eyes, brows, and mouth will be detected from the region. When move to the second stage (feature extraction), region of interest will be extracted from different parts of the face using the model of Haar Cascade Classifier. The extracted features given as input to the classifier and that significantly helps the classifier to recognising the facial dispositions. In the last stage, a classifier needs to be trained before been used to generate labels for the dispositions using the training data. And final output will given in percentage value as a confidence on predictions.

3.4 Pre-Processing

Here, we will improve the performance of human facial disposition recognition since it addresses noises. The process were image acquisition and enhancement.

3.4.1 Image Acquisition

The acquisition of images is the first step in the development of system. The human faces obtain from a computer camera that was placed between 50cm to 200cm away from our faces. Figure 3.3 shows the images that were used as sample data. The HP True-vision HD camera was utilised to test this system. The human face image input was done in a variety, but Figure 3.3 only depicts one of the test scenarios.

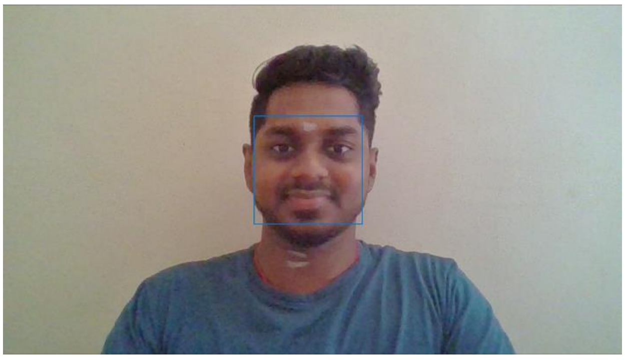


Figure 3.3 Face image

First and foremost, we need to classify the type of inputs to the detection system which are the real time application through webcam. Testing the raw images is easy where we just need to feed the image to the frames in order to run the detection and recognition for each N frames. As a result, OpenCV comes pre-installed with the

Video-Capture module, which allows you to extract images from a camera. The 'camera= cv2.VideoCapture(0)' function as shown Figure 3.4 below; the value 0 in VideoCapture informs the programme to use the laptop's primary webcam.

```
import cv2
# starting video streaming
cv2.namedWindow('my_face')
camera = cv2.VideoCapture(0)
time.sleep(2)
```

Figure 3.4 Method to use webcam and OpenCV

3.4.2 Image Enhancement

In this phase, three sub processes applied, at first the image was resized to get smaller image with smaller pixels. Then, the image is converted to grayscale from RGB as shown Figure 3.5. After that, noise removal filtering was implemented to reduce or remove noise from the image.

Figure 3.5 RGB to Grayscale conversion

3.4.2.1 Noise Reduction

Due to out of focus, there is some undesired blur and noise. Filters are used to remove these unwanted blurs.

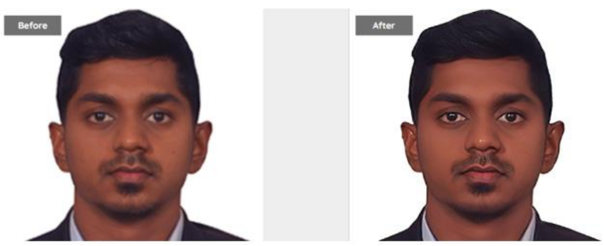


Figure 3.6 Before and after noise reduction

The original image which was processed in the system is shown in Figure 3.7.

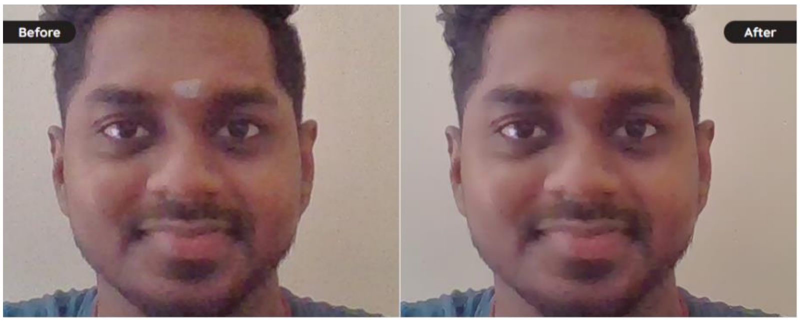


Figure 3.7 Result after noise removal

Median filter, Wiener Filter and Gaussian filters are applied to reduce the noises from the image and the filtered outputs are analysed using the image quality metrics Peak Signal to Noise Ratio (PSNR) and Root Mean Square Error (RMSE). The input image is pre-processed by these filters are shown in Figure 3.14.

To minimise noise in the picture, the Median, Wiener, and Gaussian filters are used, and the filtered outputs are analysed using the image quality metrics Peak Signal to Noise Ratio (PSNR) and Root Mean Square Error (RMSE). These filters, as shown in Figure 3.14, pre-process the input image.

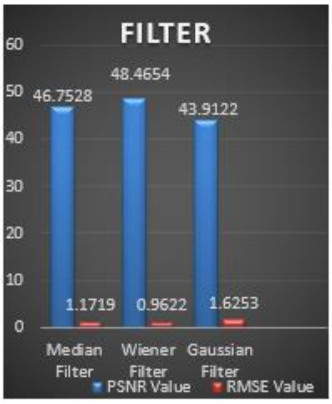


Figure 3.8 Comparison chart

The graph shows PSNR and RMSE values for various filter outputs. Based on the study, Wiener Filter beats the other two filters in terms of PSNR and RMSE noise.

3.5 Facial Feature Extraction

At this phase, the algorithm initially recognizes a face, after which it segments and extracts the recognised image or human face. Following the acquisition of a noise-free picture, the procedure moves on to feature extraction. The goal of feature extraction is to extract information from an image's region of interest for image processing. The image must go through many steps under feature extraction, including facial boundary detection and segmentation. Feature extraction is the process of converting pixel data into a higher-level representation of the face's form, motion, colour, texture, and spatial arrangement.

3.5.1 Face Boundary Detection

Among the ways for detecting the facial boundary, the Successive Mean Quantization Transform method (SMQT) is utilised. Other approaches include the Fourier Transform method, region growth method, level set method, skin region detection method, and Successive Mean Quantization Transform method (SMQT). This technique uses a mechanism that splits down information into its structural components automatically. These characteristics will be utilised to extract lighting insensitive elements from certain areas of an image. On a noise-free image, Figure 3.9 shows the results of face boundary detection, as well as the region of interest, which includes the eyes and mouth, and the frame created around the identified face.

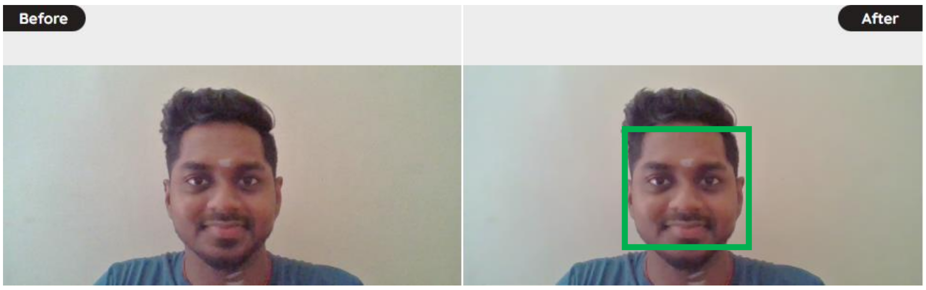


Figure 3.9 Boundary detection from frames

This method is implemented because it detects accurately on any human face at any lighting condition. Moreover, it also detects picture that contains many faces accurately. For persons who wear spectacles and have moustaches, this approach also recognises human emotions quite precisely.

Based on the code Figure 3.10, the picture or frame that was converted to grayscale with OpenCV's cvtColor() function and the cv2.COLOR BGR2GRAY attribute to the function that transforms the BGR colour to Gray is stored in the variable grey. The detectMultiScale function recognises faces in the image using the haar cascade classifier and returns a list with the coordinates of the rectangle around the face, which is kept in the variable faces.

```
frame = imutils.resize(frame,width=300)
gray = cv2.cvtColor(frame, cv2.COLOR_BGR2GRAY)
faces = face_detection.detectMultiScale(gray,scaleFactor=1.1,minNeighbors=5,minSize=(30,30),flags=cv2.CASCADE_SCALE_IMAGE)
```

Figure 3.10 Method for boundary detection

3.5.2 Segmentation

For developing the contour of human organs such as the nose, mouth, eyes, and cheek expansion, edge detection is one of the most often utilised image analysis techniques. This means that if the borders of a picture can be clearly specified, all of the parts can be identified and basic qualities like area, perimeter, and shape may be measured. Edge detection is a segmentation technique for locating regions of interest within an image.

On this system, Haar Cascade Classifier as Figure 3.11, utilised which provide high accuracy for face detection due to suitable Haar features. Face, eyes, mouth detection using Haar Classifier algorithm and passing them to the model as shown Figure 3.13. This method will allow objects to be detected in various sizes. Haar classifier will identify a set of features which are most contributing for the face detection problem in training phase itself. Therefore, it is suitable for face detection in training phase as it may indicates to high detection accuracy since the computation complexity (number of resources required to run) is small. In addition, the presence of spectacles, such as beard, hair and makeup have a considerable effect in the facial appearance as well.

```python
# parameters for loading data and images
detection_model_path = 'haarcascade_files/haarcascade_frontalface_default.xml'
emotion_model_path = 'models/_mini_XCEPTION.102-0.66.hdf5'
```

Figure 3.11 Haarcascade classifier

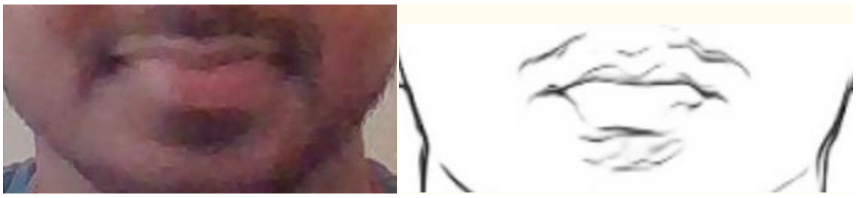


Figure 3.12 Edge detection

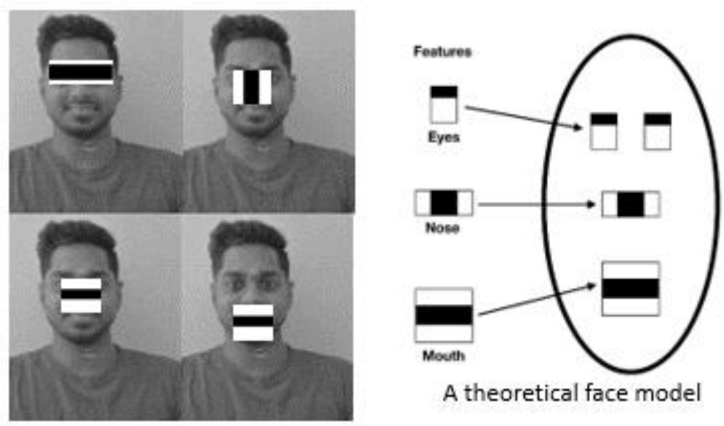


Figure 3.13 Haar features

```python
    # the ROI for classification via the CNN
roi = gray[fY:fY + fH, fX:fX + fW]
roi = cv2.resize(roi, (64, 64))
roi = roi.astype("float") / 255.0
roi = img_to_array(roi)
roi = np.expand_dims(roi, axis=0)
```

```python
cv2.rectangle(frameClone, (fX, fY), (fX + fW, fY + fH),
              (0, 255, 0), 2)
```

Figure 3.14 Method for region of interest

In one loop iteration, the cv2.rectangle function produces a rectangle using the (x,y,w,h) of one face. In the current iteration, the roi gray variable saves only that face from the webcam as an array of values in variable grey, which has the rectangle surrounding it, and then resizes the array to fit the rectangle (64,64). When resizing an image, need a mechanism to calculate the new image's pixel values from the old.

The pixel values in the roi gray array are then divided by 255.0 to get values between 0 and 1, which is kept in the roi variable and helps the model forecast to predict the results. The region on inteterest (ROI) is transformed to an array, which is then used in the model to provide a prediction.

3.6 Classification (Convolution Neural Network)

This system employs a Convolution Neural Network (CNN) as Figure 3.15, a deep learning model for processing information with a grid sample, including images, this is stimulated by way of the corporation of animal visual cortex and is designed to research spatial hierarchies of functions, from low to excessive-stage patterns, routinely and adaptively. As a result, this approach is less susceptible to noise than others, and it is more likely to detect actual weak edges. This also aids in the filling in of gaps in the margins that have been noticed. The default threshold is computed heuristically based on the provided data in each circumstance. We may load the model and classifier as indicated in Figure below to detect the human face. This study employed the 'haarcascade frontalface default' classifier to recognise a person's facial dispositions in real-time.

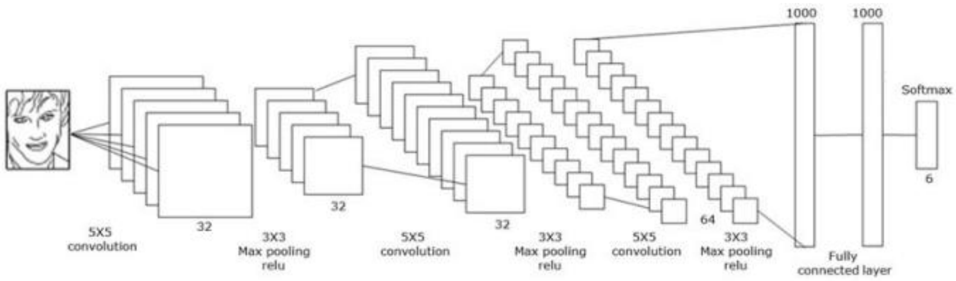


Figure 3.15 The proposed CNN model for basic facial expression recognition

It consists of 3 convolution layers, 3 max pooling layers, and 2 fully connected layers. We used 5X5 filters for each convolution layer. And we used 3x3 filters for max pooling layer to reduce the input image size to half. It has 2 fully connected layers which has 1000 nodes. And the activation function for each convolution layer is ReLU (Rectified Linear Unit) for preventing gradient vanishing problem. For regularizing the network, we used dropout method. The top layer is softmax layer which has 6 nodes. Each node corresponds to 6 emotions such as such as happiness, sadness, anger, disgust, surprise, neutral.

On this system CNN used to recognise human disposition identification, however, the recognition rate may be lower because there are a variety of individual differences in facial expressions. Therefore, additional model ‘haarcascade frontalface default’ classifier also known as haar cascade used to recognise a person's facial expressions in real-time.

```
# parameters for loading data and images
detection_model_path = 'haarcascade_files/haarcascade_frontalface_default.xml'
emotion_model_path = 'models/_mini_XCEPTION.102-0.66.hdf5'
```

Figure 3.16 Proposed CNN model

The feature points are then processed to form the inputs for the neural network. The neural network has been taught to differentiate between the emotions neutral, happy, sad, angry, disgust, surprise, and scared. In the last stage of the human disposition system, the classifier categorises expressions such as happy, sad, surprise, angry, fear, disgust, and neutral.

```
for (i, (emotion, prob)) in enumerate(zip(EMOTIONS, preds))
            # construct the label text
            text = "{}: {:.2f}%".format(emotion, prob * 100)
            print (label)
            if label=='happy':
                print ('inside')
            elif label=='sad':
                print ('inside')
            elif label=='scared':
                print ('inside')
            elif label=='surprised':
                print ('inside')
            elif label=='angry':
                print ('inside')
```

Figure 3.17 The proposed dispositions

3.7 FER-2013 Database

This dataset generated by compiling the results of each disposition from the google images as well as synonyms for those feelings. To train our CNN architecture, FER-2013 dataset utilised which consists of grayscale image measuring 48x48 pixel for each image. The total FER-2013 dataset is 35,887 consisting of seven different types of dispositions. The number of sample images as Figure 3.18 for each disposition is represented in Table.

Figure 3.18 Sample images for each dispositions

Table 3.1 Number of images for each dispositions

Disposition	Number of images
Angry	4593
Disgust	547
Scared	5121
Happy	8989
Sad	6077
Surprise	4002
Neutral	6198

3.8 Output Based on Percentage

If the person is smiling, the probability used to identify emotions with the percentage will signal 'Happy' near the frame within a second. If the person is doing nothing on expression it will show neutral and final output based on the percentage shown in Figure 3.19. The same approaches may be used to deal with other emotions.

Following the completion of the executable programme that generates output libraries based on the CNN model, the system must be tested to ensure that it meets the system's needs and the supervisor's expectations.

```
w = int(prob * 300)
cv2.rectangle(canvas, (7, (i * 35) + 5),
(w, (i * 35) + 35), (0, 255, 0), -1)
cv2.putText(canvas, text, (10, (i * 35) + 23),
cv2.FONT_HERSHEY_SIMPLEX, 0.45,
(255, 255, 255), 2)
cv2.putText(frameClone, label, (fX, fY - 10),
cv2.FONT_HERSHEY_SIMPLEX, 0.45, (0, 255, 0), 2)
cv2.rectangle(frameClone, (fX, fY), (fX + fW, fY + fH),
              (0, 255, 0), 2)
```

```
cv2.imshow('my_face', frameClone)
cv2.imshow("Probabilities", canvas)
if cv2.waitKey(1) & 0xFF == ord('q'):
    break
```

Figure 3.19 Final output based on Percentage

3.9 Applications

Here we will go through how to utilise Pyinstaller to learn how to make an application for facial recognition for human disposition identification. It will read the python script and analyse the code to see whether the script needs any additional modules or libraries to execute.

3.9.1 Pyinstaller

When the installer is run, the system instals all necessary dependencies without the need of intervention. Therefore, CNN-based segmentation does not need Python or any modules.

3.9.2 Requirement

3.9.2.1 Windows

PyInstaller can generate graphical windowed programmes that don't require a command window and operates on Windows 8 or later. PyInstaller necessitates the installation of two Python modules on a Windows machine. It requires the Windows Python extensions PyWin32 or pypiwin32. If you use pip to install PyInstaller and PyWin32 isn't already installed, pypiwin32 is installed as well. The pefile package is also required by PyInstaller.

3.9.3 Installation Process

3.9.3.1 Add Python to Windows path

Downloading a recent version of Python and then clicking the option to 'Add Python to PATH' at the beginning of the installation is a simple approach to add Python to the path.

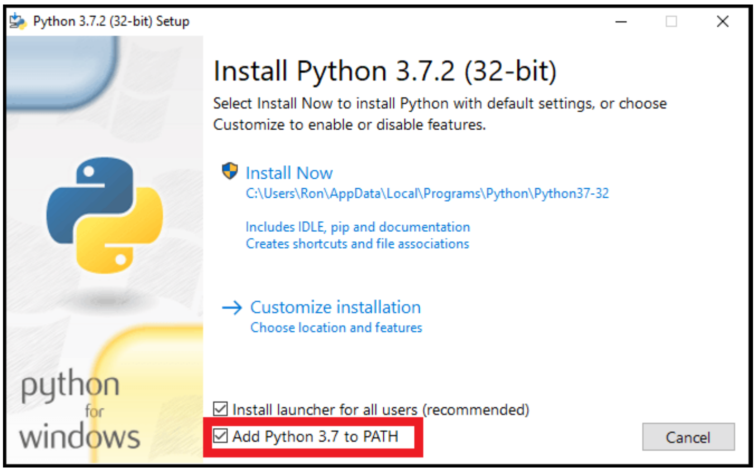


Figure 3.20 Install Python to PC

3.9.3.2 Creating Pyinstaller Packages

Open the Windows command prompt (cmd), select the directory of code and type '**pip install pyinstaller**' to install the packages which will generate the executable file.

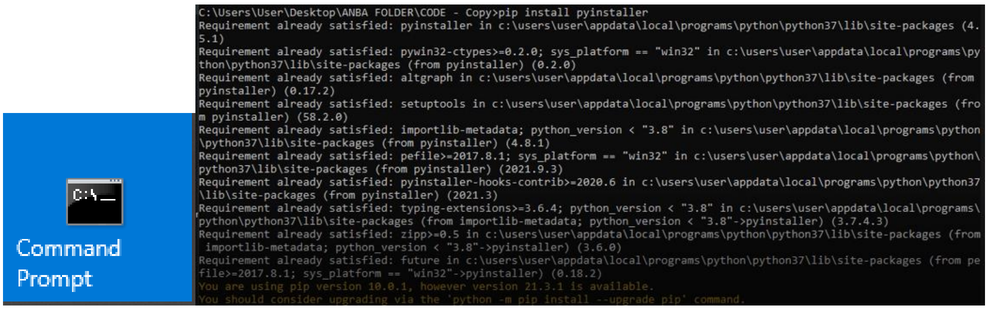


Figure 3.21 1st command

3.9.3.3 Adding Data Files to the Executable

Execute **'pyinstaller face emotion1.py'** once you've done installing Pyinstaller, as shown in Figure, and Pyinstaller will analyse the code script and write face emotion1.spec to the same folder as the script. The spec file instructs PyInstaller to process the python script. The spec file, as shown in Figure, is Python code that builds the app by executing the contents of the spec file.

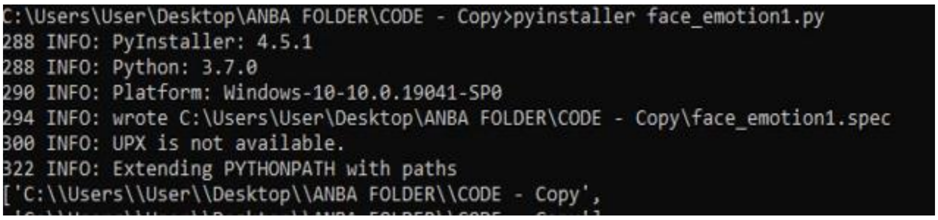


Figure 3.22 2nd command

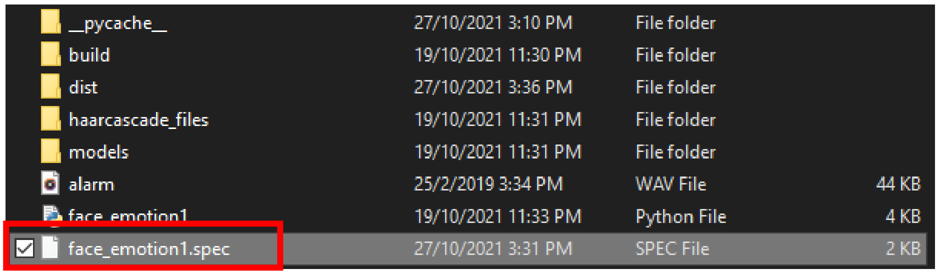


Figure 3.23 Copy file

Then, to install the Google API Python client library for Google's discovery-based APIs, execute **'pip install google-api-python-client'**. Pyinstaller builds a distribution directory to the path 'C:Users\User\AppData\Local\Programs\Python\Python37\Lib\site-packages' after a successful installation. Determine where the.exe file should be packed with these dynamic libraries.

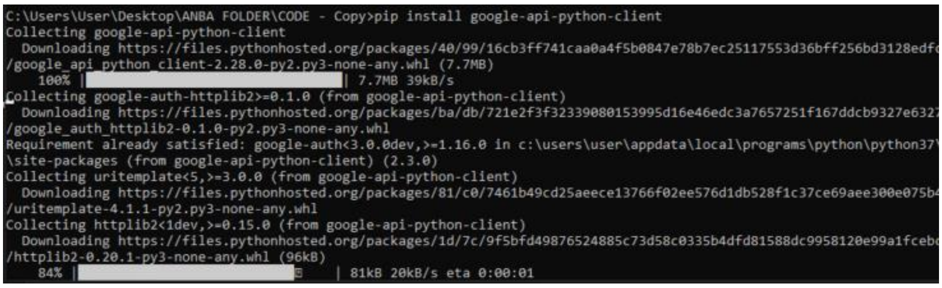


Figure 3.24 3rd command

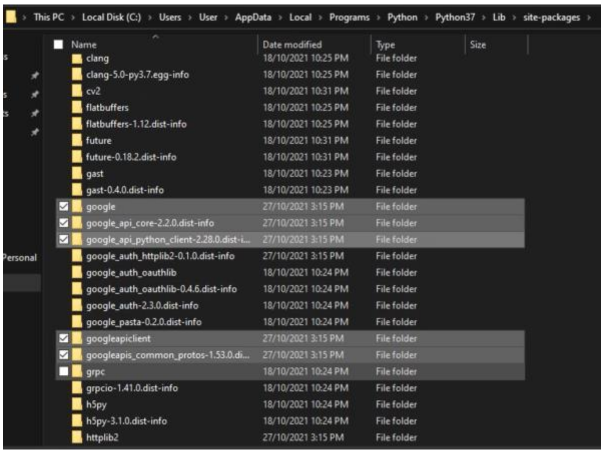


Figure 3.25 Copy API files

In addition, a DIST subdirectory will exist in the project directory. This, in turn, provides a packaged app directory with an .exe file to launch, as well as all the libraries and other required files. Figure 3.26 shows how to copy the specified addition folders from the path 'C:UsersUserAppDataLocalProgramsPythonPython37Libsite-packages' into a DIST file. Apart from that, the script creates a BUILD subdirectory in the same directory as the script and writes certain log and working files to it. Finally, copy the 'haarcascade files' and ‘models’ subfolders to the DIST folder.

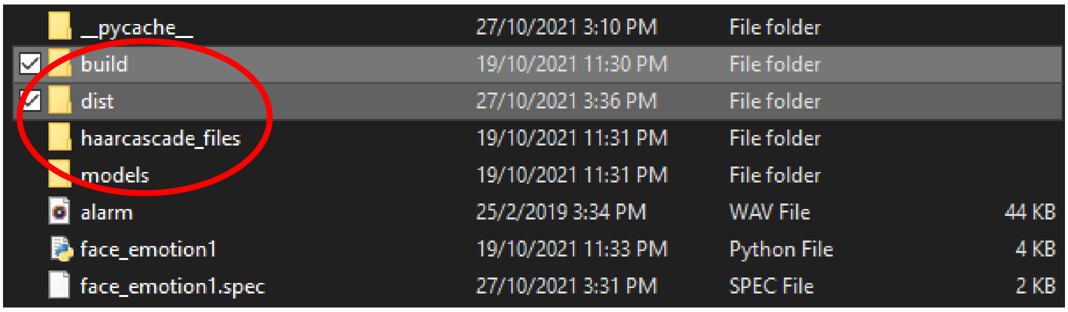


Figure 3.26 Copy all required files and bundled

3.9.3.4 Initiate Pyinstaller module

After bundling the files into one unique file, PyInstaller attempts to produce an executable from the programme, bundled with all of its dependencies, in this project 'DIST.' To start the system, navigate to the directory containing the bundled executable and launch the.exe file, or use the command prompt to type **'cd face emotion',** as shown in Figure.

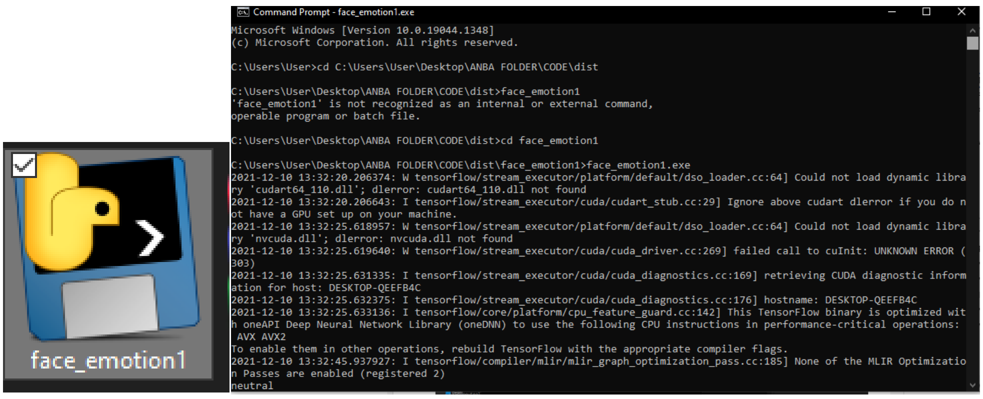


Figure 3.27 Execute file

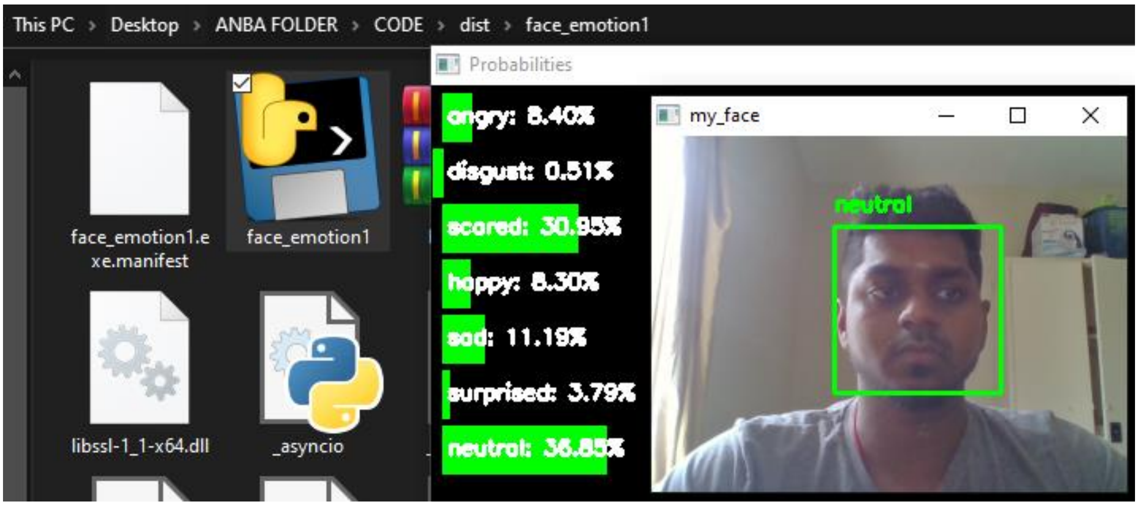


Figure 3.28 Application for human disposition identification

3.10 Software and Hardware Requirements

3.10.1 Software Requirement

Table 3.2 Software requirement describes software requirements needed in system development.

Table 3.2 Software requirement

Item	**Name**	**Purpose**
Operating System	Windows 10 Home Single Language	As the operating system.
Software	Python 3.7.0, Pytorch	As a platform in running and developing the system.
	Microsoft Word 2020	To document the system
	Pyinstaller 3.7.0	To create a single executable file (.exe) for an application.
	HP True-vision Camera	To capture images in real-time.

3.10.2 Hardware Requirement

Table 3.3 Hardware requirement describes hardware requirements needed in system development.

Table 3.3 Hardware requirement

Item	Minimum Requirement
Processor	Intel® Core™2 Duo Processor P7350 (2.0Ghz, 1066Mhz FSB, 3MB L2 cache)
RAM	8GB DDR4
CD-ROM	HL-DT-ST-DVDRAM GSA-T20N
Hard Disk	250 GB HDD + 120 GB SSD
Laptop Camera	2 Mega Pixel

CHAPTER 4

RESULTS AND DISCUSSION

4.1 Introduction to Results

In this chapter, we will do the analyses based on the result obtain from the stages of project flow and based on applications. As a testing phase of this project, I have tested with myself, and 7 images of human faces were used as sample data for each disposition as shown in Appendices B. All the human faces samples had gone through the phases of this system that has produced output, and then we will analyse the accuracy of human disposition identification at the final output in percentage.

4.2 Data Analysis on Application Setup

After created a desktop application through Pyinstaller from the designed Python code script, we able to get into this system by clicking the icon as shown Figure 4.1. Once the system start-up, it will automatically enable the PC camera to capture and identify the facial disposition of user's and output can be seen upside of the frame as well as the probability bar gives the detection accuracy in percentage. This application can recognize 6 basic human facial dispositions such as angry, happy, scared, sadness, surprised and neutral.

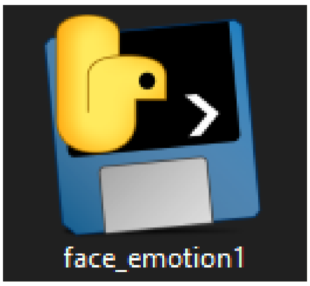


Figure 4.1 System application

4.3 Result and Analysis on Facial Recognition

The initial and most critical step in the processing pipeline was face detection. We had to detect the face first as Figure 4.2, which made determining the region of interest and extracting characteristics from it much easier. These operations will be carried out automatically by this system.

After the frame has been drawn to the identified face, the algorithm will quickly identify the human disposition near the frame, and we may calculate the accuracy of detection in percentage.

Figure 4.2 Result on accuracy for high resolution input

Thus, it is clearly saying that we need high resolution webcam to be fed into the detection system so that human disposition identification able to predict the expressions with clearer view and gives high accuracy for detection. Even though, the face detected in blurry this system may be fails to predict the disposition since the image is not clearer, even little illumination and noises.

4.4 Data Analysis on Human Disposition Identification

Here we will discuss each human disposition identification and how the detection happens for human facial organs including the nose, eyes, mouth, and cheeks which are mostly employed to recognise dispositions. The six basic emotions which are recognized from the system and easily interpreted through specific facial expressions which are happy, sad, neutral, scared, anger, and surprise.

4.4.1 Happy

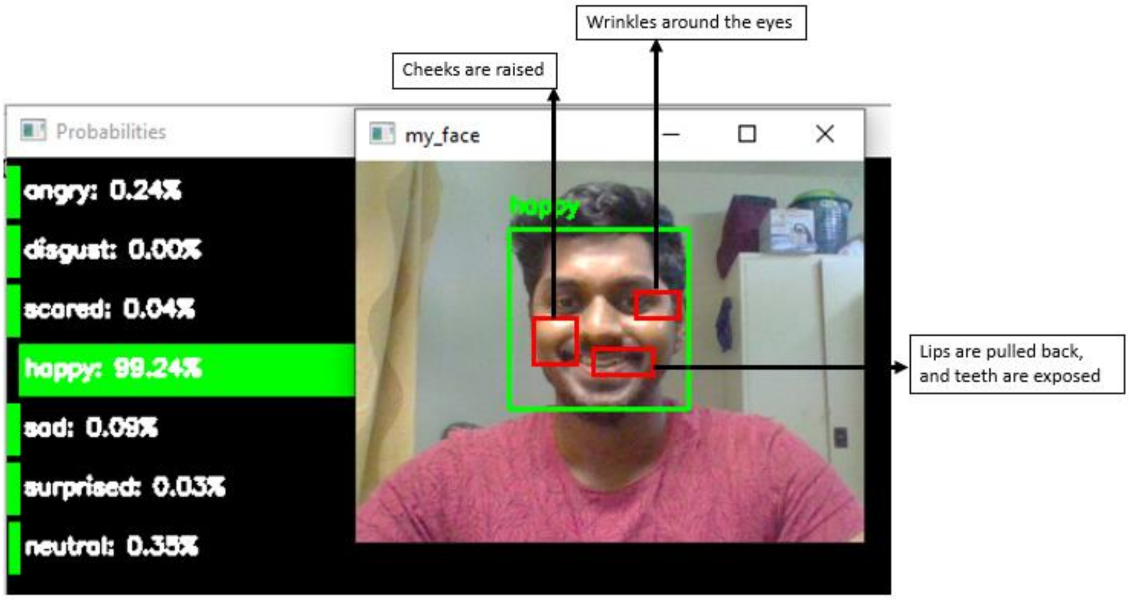


Figure 4.3 Happy disposition

According to the diagram above, when the user shows a cheerful smile or smiling it causes the muscles around the eyes to tighten, "crow's feet" wrinkles around the eyes to appear, cheeks to rise, and the corners of their lips to shift up at a diagonal as widening their mouth. Based on the Figure, it shows the confidence of predictions about 99.24%.

4.4.2 Sad

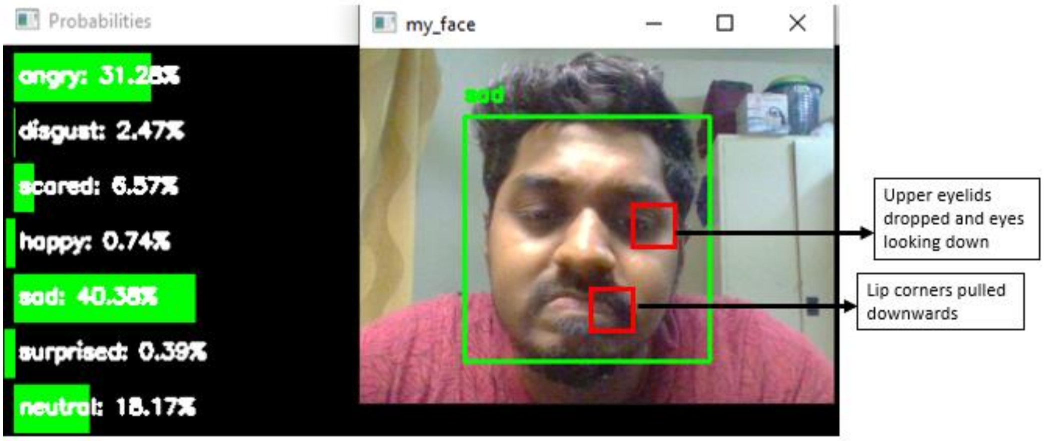


Figure 4.4 Sad disposition

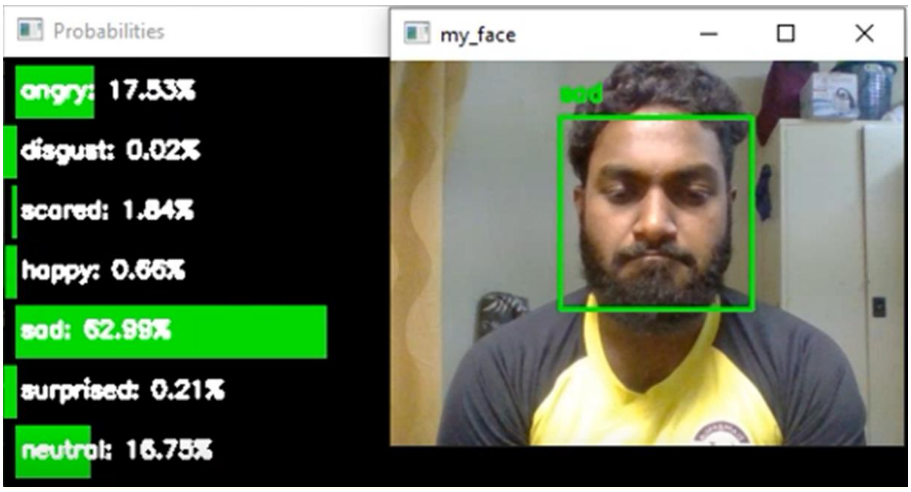


Figure 4.5 Sad disposition with moustache

When the user shows a sad face as shown Figure 4.4, it causes their eyebrows to fall and pull closer together, the inner corners of their eyebrows to be tilted up, the corners of their mouth to be dragged downwards, and their lips to be drawn in tightly or pouting outwards with the confidence of predictions 40.38%.

Sadness is a tough emotion to depict because when person mouth directs downwards then the system gives results based on the identified shape of the mouth. The challenging part here is when a person with moustache and gives sad expression as shown Figure 4.5, it detects about 62.99% of sad result. This is because as the moustache directs downwards and mouth gives expression of sad where it detects well with high accuracy.

4.4.3 Neutral

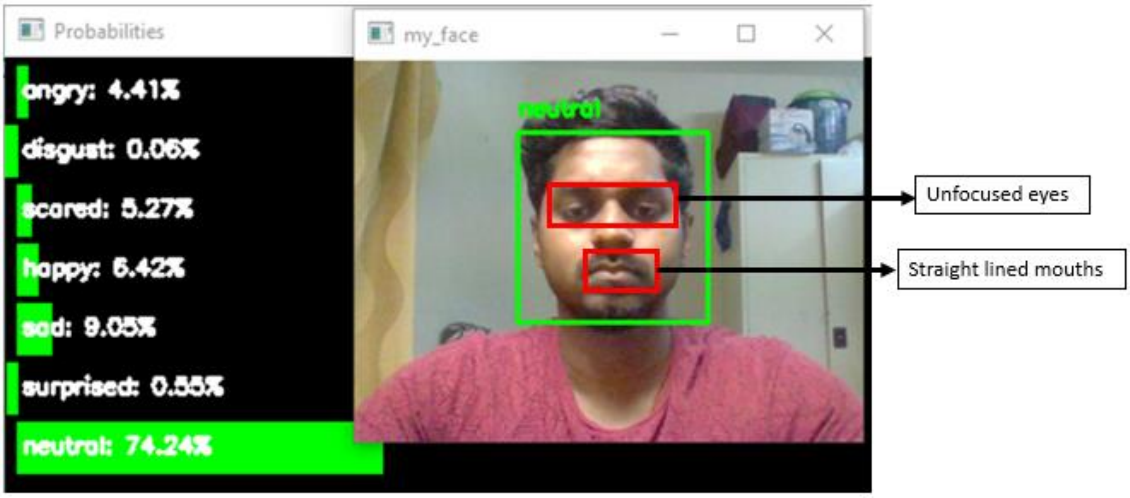


Figure 4.6 Neutral disposition

A neutral face is a blank expression that implies a lack of perceptible emotion. This face is defined by straight-lined mouths, unfocused eyes, and slack cheeks with the accuracy of 74.24%. Though it communicates negativity to some, others see it as a reflection of calmness.

4.4.4 Angry

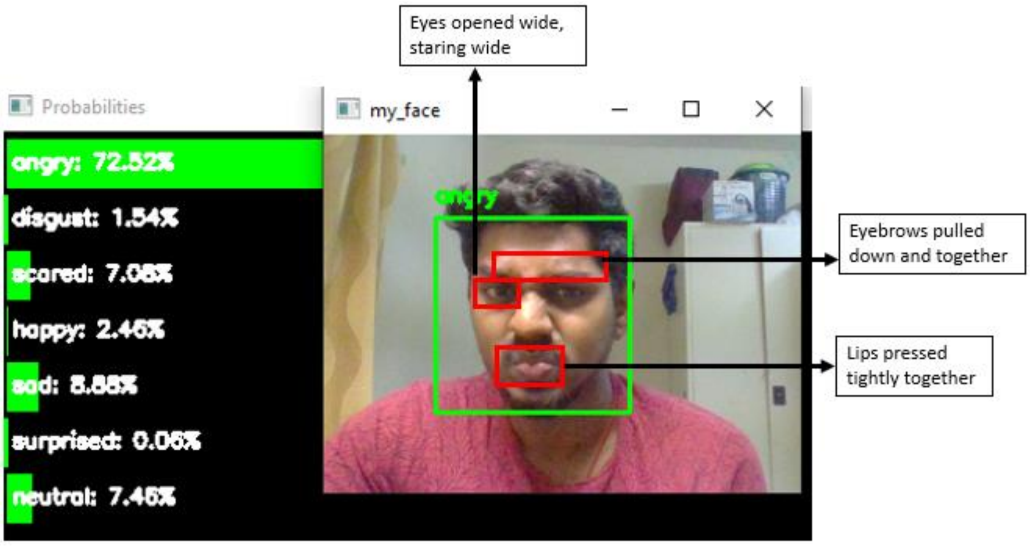


Figure 4.7 Angry disposition

As shown the Figure 4.7, brows of a person are lowered, drawn closer together, and vertical creases emerge between their brows when the user giving an angry disposition. The corners of the mouth would point downwards, the eyes would squint or lift, the lips would tighten or curl inwards, and the eyelids would squint or raise. When the system detects those characteristics, the output will be shows ‘Angry’ with the 72.52% confidence of prediction.

4.4.5 Scared

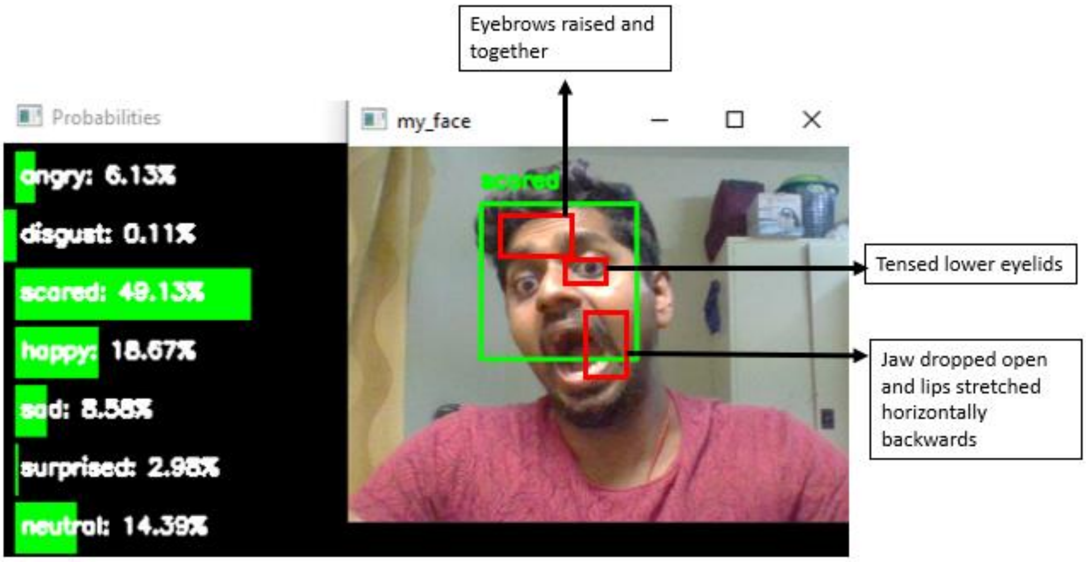


Figure 4.8 Scared disposition

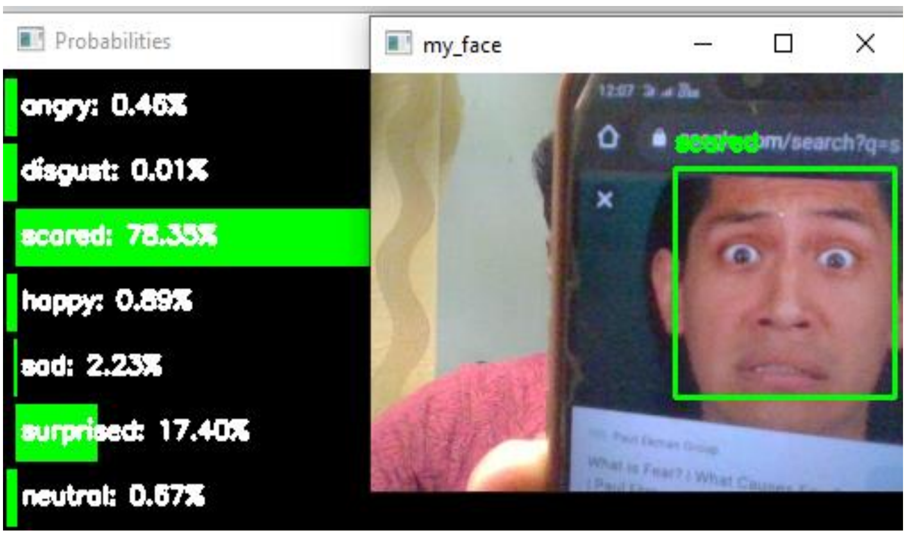


Figure 4.9 Another example for scared disposition

Scared is frequently confused with surprise on the face. While both dispositions have raised eyebrows, the scared expressions are straighter and more horizontal, while the surprise expressions are raised and curved. Scared causes the upper eyelid to lift higher than surprise, revealing more sclera (white of the eye). Finally, in scared, the lips are contracted and strained, while in surprise, they are more open and looser.

The facial disposition of scared shows that vertical wrinkles may emerge between their brows, user mouth would be stretched and drawn back, their upper eyelids would be drawn up, and their lower eyelids would be tense and dragged up as well, and their eyebrows would be pushed up and together for the scared disposition. By comparing both Figure 4.8 and Figure 4.9, the output on confidence of predictions are different because it depends on the facial structure of human due to God's creature and the way people expressing the emotions since it based on physical expressions.

4.4.6 Surprise

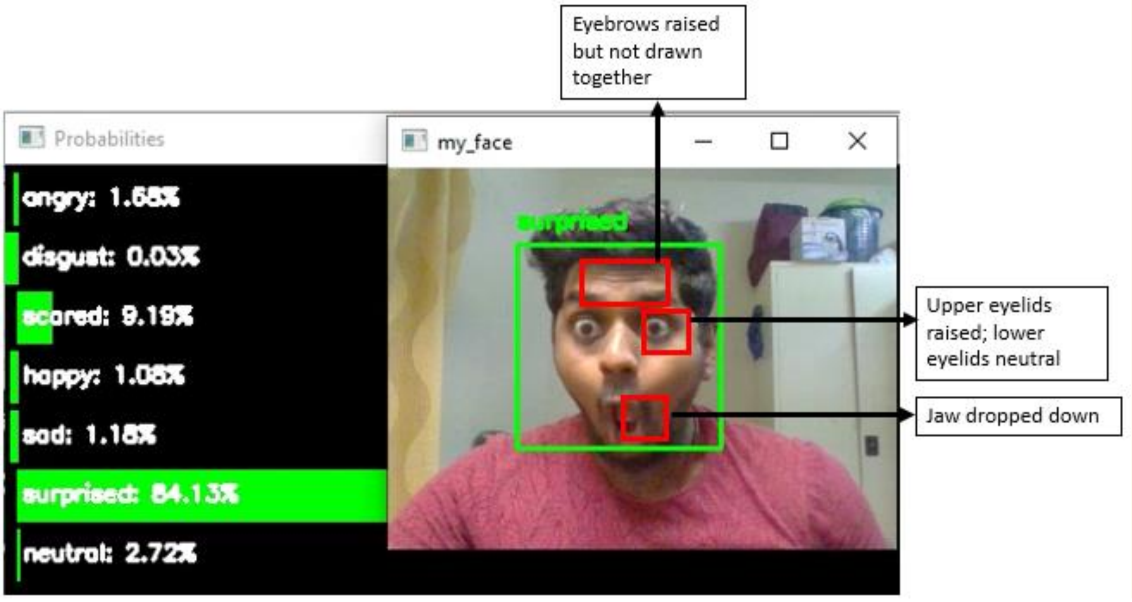


Figure 4.10 Surprised disposition

Scared and surprise are two of the most perplexing facial emotions since they share the same basic elements: brows, eyes, and lips. However, the system I built, extracts those features and is able to distinguish between them, providing correct results based on the recognized facial dispositions.

The brows are elevated in surprise, yet they curve more than they do in scared. When displaying surprise, the top eyelids and mouth are also more relaxed. Surprise is a transient emotion that causes eyebrows to lift, horizontal creases to develop on the forehead, and the jaw to sag with widening eyes. Based on the Figure 4.10, this system recognised the surprised disposition with the confidence of predictions 84.13%.

4.5 Data Analysis on Each Dispositions Classifications

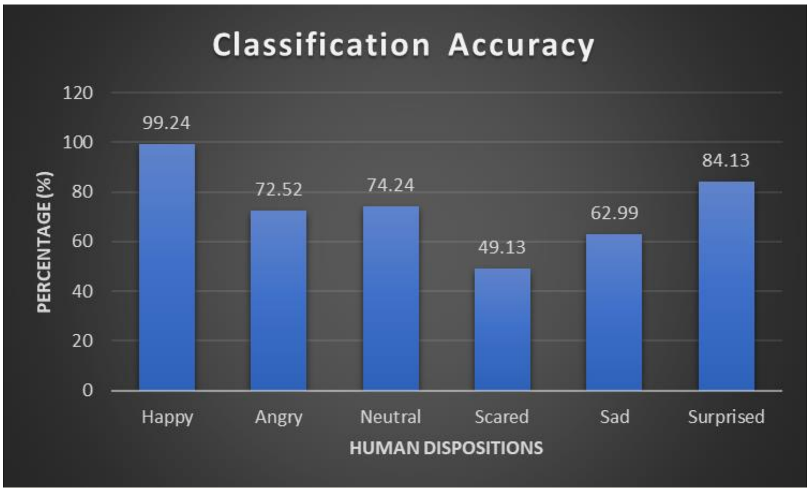


Figure 4.11 Classification accuracy

Based on the results, Happy disposition achieves highest which is 99.24%, Surprised 84.13%, Neutral 74.24%, Angry 72.52%, Sad 62.99% and Scared 49.13% in confidence of predictions. This is because most people will be smiling by pulling their lips backward and cheeks are raised so it automatically detects highest accuracy through the CNN classifications, plus consists of 8989 of sample images on dataset which provides better accuracy in detections based on the Table 3.1 Number of images for each dispositions. However, other facial dispositions it depends on the facial features expressions of a human being.

4.6 Data Analysis on False Positive Rate

Table 4.1 False positive rate

Human Dispositions	False Positive Result (%)
Happy	0.76
Angry	27.48
Neutral	25.76
Scared	50.87
Sad	37.01
Surprised	15.87

This table shows the summary of the false positive rate for each disposition. Based on the results some of the dispositions misclassified with others. This because when CNN train the input face with the dataset, noisy images on the sample pictures effects the accuracy in predictions, even little changes on illumination and noise. Moreover, human being cannot express same dispositions all the time, for an example some have sad face by puling downward the lips and some may be shows dull face, so this reduces the effectiveness on predictions. Moreover, highest number of images on dataset as shown Table 3.1 recognition rate will be higher thus reducing false positive results like in Happy Dispositions.

4.7 Data Analysis on Confusion matrix

Table 4.2 Confusion matrix for recognition system

Human Dispositions	Happy	Angry	Neutral	Scared	Sad	Surprise
Happy	**99.24%**	0	0	0.46%	0	0.30%
Angry	0	**72.52%**	0	0	16.98%	10.50%
Neutral	12.30%	0	**74.24%**	0	13.46%	0
Scared	15.70%	20.97%	0	**49.13%**	0	14.20%
Sad	12.90%	0	15.20%	8.91%	**62.99%**	0
Surprised	0	8.37%	0	7.50%	0	**84.13%**

Based on the table above, when happy faces analysed, 99.24% are properly recognised, whereas 0.76% are confused by the proposed methods. Among 0.76%, 0.3% and 0.46% corresponds with surprised and scared dispositions respectively. Since, both scared and surprised dispositions make the same facial gestures, such as opening their mouths wide and raising their upper eyelids, this causes rise in false rate.

When angry dispositions analysed, 72.52% are recognised confidently, however 27.48% are confused with 16.98% sad and 10.50% surprised. This due to of a person when they show sad dispositions by pulling down their brows and eyes as staring wide in surprised which shares same characteristics of angry dispositions.

Moreover, neutral human face has a false positive rate of 25.76%, where 12.30% shares with happy and 13.46% with sad. When human being shows the neutral dispositions with unfocused eyes and straight-line mouths, this effects the accuracy on predictions which shares same features as happy by pulling mouth backward without curve or bend and sad with lower eyes. However, this system able to distinguish between those dispositions and able to recognise neutral with 74.24% on confidence of predictions.

For the scared disposition, false positive results achieved 50.87%. Among that, 15.70% happy, 20.97% angry and 14.20% surprised dispositions shared same facial features expressions such as wrinkles on the forehead and mouth openly. For the scared disposition, it successfully achieves to recognise with the accuracy of 49.13%

When sad faces are tested, 62.99% correctly recognised by this system, however 37.01% confused with the proposed method. Among 37.01%, 12.90% matches with happy, 15.20% with neutral and 8.91% matches with scared. By comparing these dispositions with sad, this system recognised based on their physical expressions and can be affected when user not able to show appropriate dispositions.

Lastly, surprised is a micro disposition, which shares same characteristics as angry and scared by dropping the jaw down and eyes widely staring. This leads to rises in false positive results as 15.87%, however this system able to recognise surprised disposition with the accuracy of 84.13%.

4.8 Data Analysis on Testing Results

Due to the pandemic, testing results cannot be done by University Malaysia Pahang (UMP) students, however testing phase was finished by 7 google images with various skin colour, various ages and various types of faces analysed for testing purposes. Table 4.3 illustrates the tested results for each type of dispositions.

$$Rate\ of\ success = \frac{Successful\ recognition}{Total\ tested\ images} \times 100$$

Table 4.3 Testing results

Human Dispositions	Successful Recognition	Rate of Success (%)
Happy	7/7	100.00
Angry	5/7	71.43
Neutral	7/7	100.00
Scared	7/7	100.00
Sad	6/7	85.71
Surprised	7/7	100.00
TOTAL	**39/42**	**92.86**

From the result shown in Table 4.3, 7 images of human faces were shown a positive result out of total based on the ranges shown in Table 4.3. This means that the Facial Recognition for Human Disposition Identification recognises with a success rate of 92.86%. The result might be due to the fact that each person's mouth is different in size as a result of God's creation. It might also cause by some samples who cannot make or control their emotions during the test especially when they need to make the face look in sad. Not everyone can act through their facial expressions is the main thing that affects my system. Moreover, it is based on physical expression.

4.9 Data Analysis on Execution Time

Table 4.4 Execution time results

Number of reboots	Time taken for recognition (s)	
	CPU only	CPU+GPU
1	45.0	7.0
2	47.0	6.0
3	51.0	3.0
4	46.0	3.0
5	41.0	3.0

Table 4.5 Specification of computing platforms

System	Specification
CPU	Intel ® Core ™ i5-2500 CPU@3.30Ghz
GPU	NVIDIA GForce GTX 1060 3GB • Build in Memory: 3GB • NVIDIA CUDA Cores:1152 • Memory: 3GB GDDR5 • Memory Speed: 8Gbps
Software	OS: Windows 10 Pro (64bits) • Tensorflow Version: tensorflow-gpu-1.0.1 • CUDA version: cuda_8.0.61 • CuDNN Version: cuDNN v5.1

This analysis to evaluate the performance using GPU-enabled multi-core parallel computing systems and deep learning computation. While rebooting the system several times, we noticed multi-fold speed increases with GPU multi-core computing systems. When comparing the performance of CPU-only and CPU-enabled and GPU-enabled, GPU excels CPU-only, as GPU is considered the brain of deep learning. It is a single-chip processor that performs complex graphical and mathematical calculations while freeing up CPU resources for other tasks.

In terms of computation speed, based on the performance improvement with GPU-enabled platforms for deep learning computation. We used the multilayer perceptron to create the MNIST data classifications in order to test the performance on real-world platforms. We also utilised the TensorFlow software library and GPU-enabled systems (with 1152 cores) to discover out how much faster GPU-enabled computations are than CPU-only computations.

As a future research study, we will consider various algorithms and applications of deep learning to measure the achievement on improvements with CPU and GPU enabled computing platforms.

4.10 Analysis on Performance Comparison

In this section, we will discuss about the performance comparison in terms of accuracy and time execution of the project with already available systems on the market.

4.10.1 Accuracy Comparison

4.10.1.1 Inception Layer versus Convolution Neural Network (CNN)

There are several factors that can be compared with the previous researchers, one of it is the accuracy of the system. There numerous of researchers had used different kind of method to achieve a good number of accuracy rate. The author Mollahosseibi on 2016 [27] had shown a new method of adding inception layers in the Human Disposition Identification using Facial Recognition system. The inception layers used in the system are MultiPIE, MMI, DISFA, FERA, SFEW, and CK+. The researchers also investigate the performance of the system when the new method is added to this system. There are about six expression was observed by the researchers which are anger, scared, happiness, neutral, sadness and surprised. The accuracy of the system is as shown in table below.

Table 4.6 Accuracy of the facial detection by using Inception layer

Human Dispositions	Accuracy (%)
Anger	55.0
Scared	47.0
Happiness	86.6
Neutral	75.0
Sadness	56.1
Surprised	89.3

The accuracy of previous researchers is not that accurate, so to overcome this problem, a new system was created by using Convolution Neural Network (CNN). The input (human face) will be extracted from the OpenCV and the image enhancement process occurs, where the RGB will be converted to grayscale. After the image enhancement process, the noise reduction has to be done to reduce the noise from the input (image of the face). And lastly segmentation have to be done to the image before the image is analysed by the Convolution Neural Network (CNN). Convolution Neural Network (CNN) will classify the expression given by the image (input). These processes would be carried out automatically once the input (human face) recognized from the webcam. By using all this method, the accuracy of the system increases rapidly as shown in Table 4.7 Accuracy of current system.

Table 4.7 Accuracy of current system

Human Dispositions	Accuracy (%)
Anger	72.52
Scared	49.13
Happiness	99.24
Neutral	74.24
Sadness	62.99
Surprised	84.13

Figure 4.12 shows the comparison between the previous system that was created by author Mollahosseibi (blue) through Inception Layer and the current system that was develop by using CNN (red).

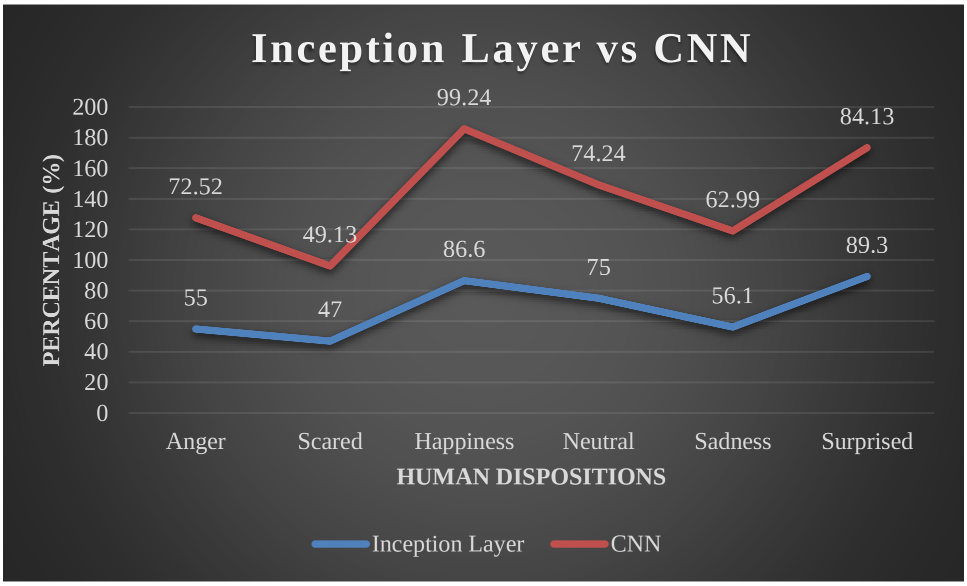


Figure 4.12 Graph comparison of Inception layer vs CNN

4.10.2 Time Execution Comparison

4.10.2.1 TREE, KNN, NB, SVM, DA, RF, MLP versus CNN

Time execution is the time taken to complete the testing stage, which is the average time taken to classify an input (human face) whether the facial disposition are Anger, Scared, Happiness, Neutral, Sadness and Surprised. Adel Alti in 2020 [14] had investigated the execution time for a system for the Facial Recognition for Human Disposition Identification by using CK+ dataset. Figure 4.13 shown the comparison on the execution time between CK+ and FER-2013 (current) dataset. There about eight classification that had been done, such as TREE, KNN, NB, SVM, DA, RF, MLP and using CNN FER-2013 dataset. From the comparison below we can conclude that by using FER-2103 dataset the average execution time can be reduced to 0.002 seconds and it is the shortest time among other systems. The longest time take to execute the classification is when the system is using RF and MLP from CK+ dataset, which is 0.7 seconds and 0.2 seconds respectively.

Figure 4.13 Comparison in execution time of FER and CK+ dataset.

4.10.3 System Comparison

4.10.3.1 Closed Eyes versus Open Eyes

Simona Maria Banu and et al in 2012 [27] have develop a new system for the facial recognition. In this paper, the authors have listed out the techniques used for their system, such as hear function, edge detection and Bezier curves. They also used feed-forward neural network to analysis the input data (face of the humans). But the system developed by this team does have some disadvantage, where when the human closes their eyes during the detection process, the system unable to detect the human dispositions.

After the detection of the face, the system will determine the class of the face such as angry, scared, happy, neutral, sad and surprised. Some of the parameter analysed by the authors in their research are the distance of the eye opening for the upper and lower eyelids, corner distance of the interior corner of the eyes and eyebrow, the width of the mouth opening and lastly the height difference between the distance of the corner of the mouth and the corresponding external eye corner. The image is obtained by OpenCV HaarCascade Classifier, this able to detect the position of the eyes and the connecting line between the eyes. By using the connecting line between the eyes, the input image (face) is rotated. By using the OpenCV their system detected the eye, mouth and the nose as shown in Figure 4.14. Algorithm are used to extract the skin region of the face. The skin region is replaced with the pixel value with the lowest contrast as shown in Figure 4.15. After the skin was extracted from the image, a mask approach was used to locate the eyes. Lastly the extraction for the eyes and mouth was done by using Bezier curves. Lastly the neural network analyser was used to determine the facial expression from the six-type listed previously. The neural analyser uses K-mean classification to determine the correct facial expression. Neural network analyser normally will show 3 parameter which is train, validation and test as shown in Figure 4.16. Validation is the best choice of the result for features selection.

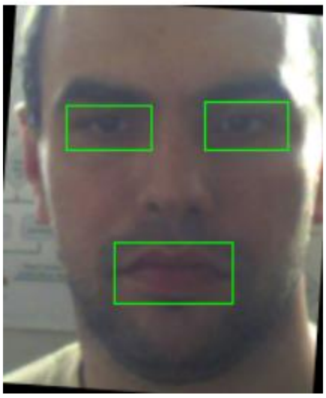


Figure 4.14 Eyes and mouth extraction

Figure 4.15 Lowest contrast images

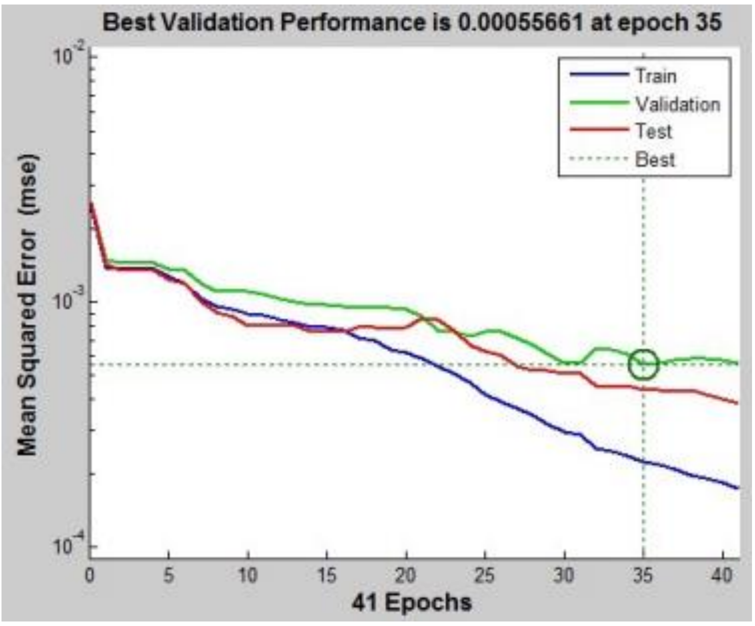
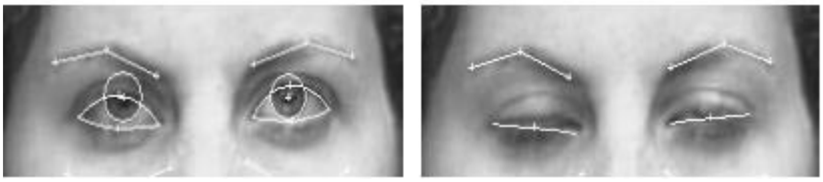


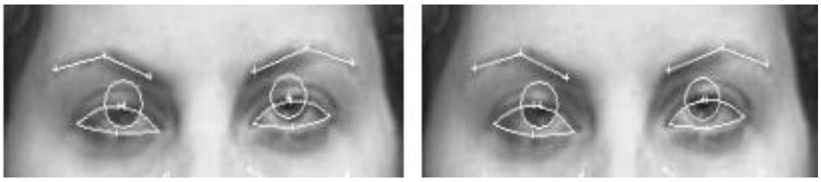
Figure 4.16 Validation test

Moreover, previous systems Simona Maria Banu and et al in 2012 [27] are unable to detect the closed eyes, but in this new system when the human closes their eyes the system still able to detect the location of the eye by using CNN classification through FER-2013 dataset and Haarcascade Classifier model. Besides that, the system also able to recover the parameter of the eye model such as the radius of the eyeball and the height of the eye opening. Figure 4.17 shows when the human blinked multiple time but the system able to detect the parameter of the eyes such as the iris of the eyes. Even when the iris is partially visible the system able to detect the parameter of the eyes very well.

(a) (b)



(c) (d)



(e) (f)

Figure 4.17 Detection of eyes when human blinked multiple times

This new system is very needed by all the application as human keep blinks few times in every one second. By having this modification this can reduce the delay in execution time and it won't disturb the function of the system.

4.10.4 Conclusion

By adding some modification in this new system, it makes the system to function more advance compare with the current available system at the market. Besides that, the new system also able to fulfil the need of human and applications in this advancing technology.

4.11 Benefits of this Project to Users

This project uses a webcam to capture the human face and recognize a person's expressions, where we can use by lecturers at University Malaysia Pahang to capture the student's human disposition during the online or face-to-face classes. The staffs can capture their emotions in real-time to understand their expressions during the class time. For example, if the staff or lecturers offer a difficult task, they can use this technology to discern whether the students are angry or sad. If students maintain a sad countenance all of the time, instructors might minimise their tension by assigning fewer assignments or tasks.

Furthermore, when conducting an interview with a candidate, this initiative aids the interviewer by allowing them to recognise their emotions during the interview session. For example, if the candidate's answer is given with a fearful look, the interviewer can give them some time to calm down and lessen anxiousness so that they can do well in the interview. Even for the employer, capturing their face dispositions while assigning responsibilities might be beneficial. For example, if an employee is given a lot of work to them and they are often depressed, they should take some time off or lower their duties to the bare minimum. Employees will perform better as a result of this, corporate production will grow, and they will be more loyal and like their work.

This project can be used in crime investigation by the facial expression of criminals or accused. When the accuse giving fear expression during the investigation for the committed crime, then the police can recognize by using this system through identify their facial dispositions

The facial dispositions of offenders or accused people can be employed in this research to help solve crimes. When a suspect expresses anxiety or fear during an inquiry for a crime he committed, the police can recognise him by utilising this technique to recognise the facial dispositions.

Most significantly, this research may be utilised to assist deaf and dumb people in recognising facial emotions. When a deaf or dumb person is sad or angry, for example, the engaging person can quickly recognise their emotion and find a method to make them happy.

Finally, the entrepreneur gains from this project since it identifies the customer's expression when consuming their products or eating foods. When the term Happy is detected, it offers the entrepreneurs a sense of fulfilment and confidence in their enterprises.

4.12 System Advantages

This section will address the system's advantages which may spark an improvement suggestion for the future.

4.12.1 Advantages

The following are some of the benefits of this system:

i. It can recognise human faces reliably and quickly in any lighting situation.

ii. This system offers a superior alternative to show someone your feelings without oral communication.

iii. Recognition that is completely automated

4.13 Assumptions

There were three assumptions to be consider during the development of this Facial Recognition of Human Disposition Identification system.

i. The camera placed in the range of 50cm to 200cm far from human face.

ii. The region of interest or the cropped mouth to be accurate on this system.

iii. Human face must be direct to camera under 90degrees to obtain better results.

CHAPTER 5

CONCLUSION

5.1 Conclusion

As to summarize the initial process that have done for facial recognition for human disposition identification, the pathway to complete the objectives is not easy and the project will be continuing to improve because of uprising technology advancement on deep learning. Facial has been recognized and the predictions of human disposition were able to show through the percentage and also in real-time application. The project taught me on how to implement a detection system using convolutional neural network. Python based on OpenCV has some advantages and disadvantages on detecting and predicting the expressions but overall, the system able to detect the human dispositions more accurately. In a nutshell, it is an honour to learn more about deep learning and get hands on working with Python based on OpenCV.

Although the proposed technique has been successfully implemented on a desktop application for real-time facial expression detection with good results, the work might be improved further by using other characteristics such as pre-trained facial expression deep learning features.

5.2 Future Recommendations

There are number of improvisations still can be done to upscale the accuracy of facial recognition of human disposition identifications. Due to time constraint, we could not test out the system to fully capable of accomplishing higher accuracy. Below are few recommendations that I would like to suggest improving the project.

Human disposition identification requires a larger dataset for training and a larger sample size for testing in the future. Making the CNN model deeper and broader can enhance it, but it will take longer time to process. It's also more difficult to change the learning rate and drop out with a more complicated CNN model. For instance, if the facial detector has a 5% error rate, the human disposition recognizer will have an unmanageable 5% error rate. Eyes and brows detectors can be added to the face detector to improve it. The outcome of the haarcascade classifier may be double-checked by detecting eyes in real-time. If both eyes and one mouth detected from the real-time application, this can reduce the facial emotion detection error rate and increase the maximum possible accuracy for the facial expression recognizer. The human disposition recognizer needs to process faster to satisfy the need of human computer interaction project. This system needs to be more compatible for different platform and different human computer interaction system.

Next, to detect more than one's expression in real-time processed. This could be done by segment the multiples detected images into single image before processing with the region of interest process. This could increase the efficiency of Real-time human disposition identification whereby it used in webcam.

Data cleaning, such as eliminating noisy photographs from the FER2013 and using supplemental data to improve model performance, may be included in future work. Autistic youngsters can benefit from better detection of human emotions, blind people facial expressions can be read, robots can communicate more effectively with humans, and drivers can stay safe by monitoring their attention while driving. Applications emotional intelligence and consumer experience may both benefit from the use of facial expression recognition.

In short, the system has achieved to detect and show results of facial expression of a person which is optimum to complete our project objectives. But more objectives can be invented to improve the project to be more productive.

5.3 Impact to the Society

This project has a positive impact on our daily lives. This project can also be utilised for educational purposes, such as receiving feedback on how UMP students behave during online or face-to-face classes. In addition, face expression analysis is a useful way to move beyond the traditional survey technique. It's a means of recognising what the user is going through while also receiving feedback. When feedback is given in this way, it is truly non-intrusive in terms of the user experience. Due to ongoing modification, this project is one of the most sophisticated technologies that has yet to reach the market.

Even, there are still detection devices are produced for private uses in many industries.

This project will be one of the steps to lead the world into future. Moreover, I personally believe that deep learning projects will innovate the youngstres to join the artificial intelligence (AI) field and make it as a interesting job for them since it is new and can be mastered. I trust this is a great opportunity for electrical & electronic engineering students to widen their knowledge on Python programming and deep learning. Sooner, I hope deep learning will be a mandatory subject for all engineering students. As for conclusion, deep learning is new and interesting to learn as it has high capability of innovate the human life to more technological and greater.

REFERENCES

[1] A.Mehrabian, (1972). *Nonverbal Communication*, Chicago: Aldine-Atherton, Illinois.

[2] P. Ekman and W. V. Friesen, (1975). *Unmasking the face: a guide recognizing emotions from facial clues*, New Jersey: Imprint Englewood Cliffs, Prentice-Hall.

[3] M.D. Meijer (1989), *The contribution of general features of body movement the attributions of emotions*, vol. 13, pp. 247-268, Englewood Cliffs, N. J.Prentice- Hall.

[4] J.Casell, (2000) A framework for gesture generation and interpretation, In R. Cipolla and A. Pentland (eds.), *Computer vision in human machine interaction*, Cambridge University Press

[5] Charles Darwin, (1898) *The expression of the emotions in man and animals*, D. Appleton & Co., New York.

[6] Ekman, P., (1969), Darwin and cross cultutal studies of facial expression, In P.Ekman (Ed.), *Darwin and facial expression: A century of research in review*. Academic, Press, New York.

[7] Ekman, P., & Friesen, W, (1969). *The repertoire of nonverbal behaviour: Categories, origins, usage, and coding*. Semiotica 1, 49-98

[8] Hattice.G, Massimo.P and Tony J (2007). *Face and Body Gesture Recognition for a Visual-Based Multimodal Analyzer*. Australia: University of Technology

[9] Yasser Yacoob and Larry S.Davis (1994). *Recognizing Human Facial Expression*. Maryland: University of Maryland

[10] Ekman, P. & Friesen, W., V. (1978). *The facial action coding system: a technique for measurement of facial movement*. Psychologists Press, San Francisco, CA.

[11] Ekman, P. (1982*). Emotions in the human faces. studies in emotion and social interaction*. second edition, Cambridge University Press.

[12] Lanitis, A., C., Taylor A., & Cootes, T., F. (1997), *Automatic interpretation and coding of face images using flexible models. Transactions of Pattern Analysis and Machine Learning*. Vol. 19, No. 7, p. 743-756

[13] Padgett, C & Cottrell, G. (1997). Representing face images for emotion classification. In M. Mozer, M. Jordan, and T. Petsche, editors, *Advances Neural Information Processing Systems*, Vol 9, Cambridge, MA. MIT Press.

[14] Joseph, Hager, C., & Ekman, P.(1995) *Essential behavioural science of the face and gesture that computer scientists need to know.*

[15] Cowie, R., Douglas-Cowie, E., Tsapatsoulis, N., Votsis., G, Kollias, Fellenz., W, & Taylor, J., G. (January 2001). *Emotion recognition in human computer interaction*. IEEE Signal Processing Magazine, Vol 18(1). Pp33-80

[16] Tian, Y. Kanade, T., & Cohn., J., F. (2001). *Recognizing action units for facial expression analysis. Pattern Analysis and Machine Intelligence*, Vol 23. (2)

[17] Donato, G., Bartlett, M., Hager, J., Ekman, P., & Sejnowski, T. (1999) *Classifying facial actions*. IEEE Pattern Analysis and Machine Intelligence, Vol. 21(10). pp 974-989.

[18] Bartlett, M., A., Hager, J., F, Ekman., P. & Sejnowski, T. (1999). *Measuring facial expressions by computer image analysis.* Psychophysiology, 36(2); p 253-263

[19] Mase, K. (1991). *Recognition of facial expressions for optical flow*. IEICE Transactions, Special Issue on Computer Vision and its Applications, E 74(10)

[20] Essa, I & Pentland, A. (1997). *Coding analysis, interpretation and recognition of facial expressions*. Pattern Analysis and Machine Intelligence, Vol. 701). pp 757-763

[21] Yacoob, Y. & Davis, L. (1994). *Computing spatio-temporal representations of human faces*. In Proceedings of the Computer Vision and Pattern Recognition Conference. pp 70-75.

[22] Simon. L, A. Bilal Ashraf and Jeffrey. F. *Investigating Spontaneous Facial Action Recognition through AAM Representations of the Faces*. Carnegic Mellon University, USA (unpublished)

[23] Chao-Fa Chuang, Frank Y.Shih (2006) *Recognizing Facial Action Units Independent Component Analysis and Support Vector Machine*. New Jersey Institute of Technology, USA.

[24] Principal Components Analysis. (n.d.). https://www.stat.cmu.edu/~cshalizi/uADA/12/lectures/ch18.pdf

[25] A. M. Baumberg. Learning Deformable Models for Tracking Human Motion. PhD thesis, University of Leeds, 1995.

[26] M.A. Turk, A.P. Pentland, "Face Recognition Using Eigenfaces", Proceedings of the IEEE *Conference on Computer Vision and Pattern Recognition* (CVPR'91), 3-6 June 1991, Maui, Hawaii, USA, pp. 586-591.

[27] li Mollahosseini, Behzad Hassani, Michelle J. Salvador,Hojjat Abdollahi, David Chan, and Mohammad H. Mahoor. *'Facial Expression Recognition from World Wild Web'* Retrieved May 2016, from University of Denver, Denver, C

APPENDICES

Appendix A: Python Coding

face_emotion1.py - C:\Users\User\Desktop\ANBA FOLDER\CODE\face_emotion1.py (3.7.0)

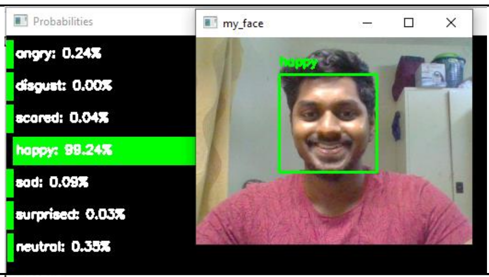
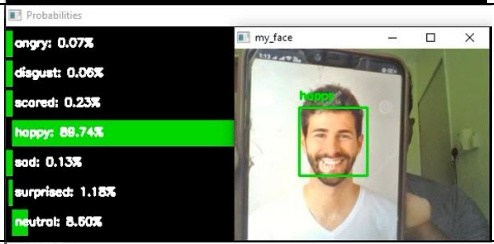
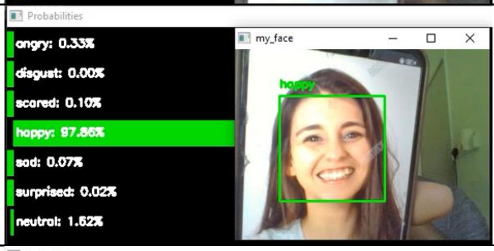
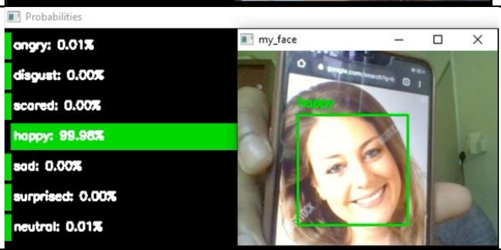
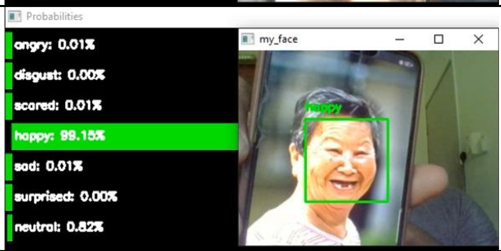
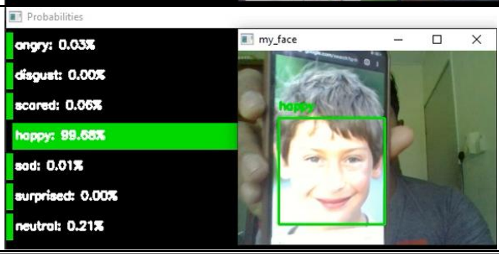
File Edit Format Run Options Window Help

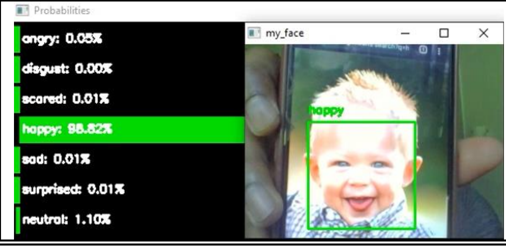
```python
from keras.preprocessing.image import img_to_array
import imutils
import cv2
from keras.models import load_model
import numpy as np
import time

# parameters for loading data and images
detection_model_path = 'haarcascade_files/haarcascade_frontalface_default.xml'
emotion_model_path = 'models/_mini_XCEPTION.102-0.66.hdf5'

# hyper-parameters for bounding boxes shape
# loading models
face_detection = cv2.CascadeClassifier(detection_model_path)
emotion_classifier = load_model(emotion_model_path, compile=False)
EMOTIONS = ["angry" ,"disgust","scared", "happy", "sad", "surprised",
 "neutral"]

# starting video streaming
cv2.namedWindow('my_face')
camera = cv2.VideoCapture(0)
time.sleep(2)
while True:
    frame = camera.read()[1]
    #reading the frame
    frame = imutils.resize(frame,width=300)
    gray = cv2.cvtColor(frame, cv2.COLOR_BGR2GRAY)
    faces = face_detection.detectMultiScale(gray,scaleFactor=1.1,minNeighbors=5,minSize=(30,30),flags=cv2.CASCADE_SCALE_IMAGE)

    canvas = np.zeros((600, 700, 3), dtype="uint8")
    frameClone = frame.copy()
    if len(faces) > 0:
        faces = sorted(faces, reverse=True,
        key=lambda x: (x[2] - x[0]) * (x[3] - x[1]))[0]
        (fX, fY, fW, fH) = faces
                    # Extract the ROI of the face from the grayscale image, resize it to a fixed 28x28 pixels, and then prepare
            # the ROI for classification via the CNN
        roi = gray[fY:fY + fH, fX:fX + fW]
        roi = cv2.resize(roi, (64, 64))
        roi = roi.astype("float") / 255.0
        roi = img_to_array(roi)
```

```
        roi = np.expand_dims(roi, axis=0)


        preds = emotion_classifier.predict(roi)[0]
        emotion_probability = np.max(preds)
        label = EMOTIONS[preds.argmax()]
    else:
        continue

    for (i, (emotion, prob)) in enumerate(zip(EMOTIONS, preds)):
                # construct the label text
                text = "{}: {:.2f}%".format(emotion, prob * 100)
                print (label)
                if label=='happy':
                    print ('inside')
                elif label=='sad':
                    print ('inside')
                elif label=='scared':
                    print ('inside')
                elif label=='surprised':
                    print ('inside')
                elif label=='angry':
                    print ('inside')


                # draw the label + probability bar on the canvas
               # emoji_face = feelings_faces[np.argmax(preds)]


                w = int(prob * 300)
                cv2.rectangle(canvas, (7, (i * 35) + 5),
                (w, (i * 35) + 35), (0, 255, 0), -1)
                cv2.putText(canvas, text, (10, (i * 35) + 23),
                cv2.FONT_HERSHEY_SIMPLEX, 0.45,
                (255, 255, 255), 2)
                cv2.putText(frameClone, label, (fX, fY - 10),
                cv2.FONT_HERSHEY_SIMPLEX, 0.45, (0, 255, 0), 2)
                cv2.rectangle(frameClone, (fX, fY), (fX + fW, fY + fH),
                              (0, 255, 0), 2)
```

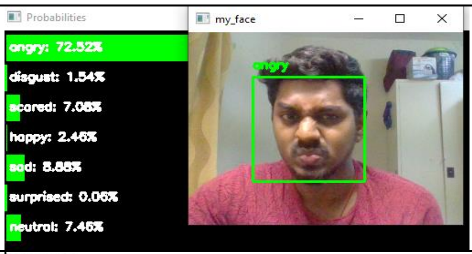
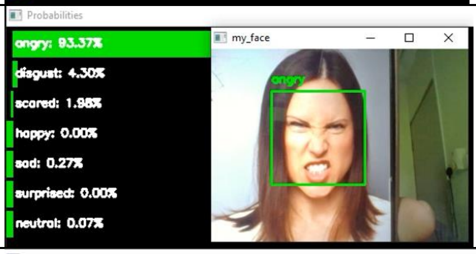
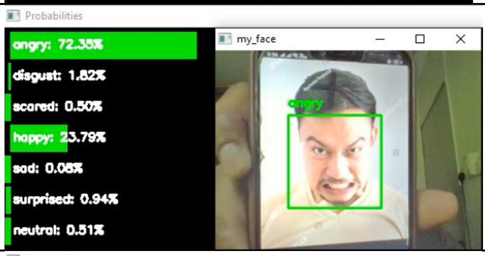
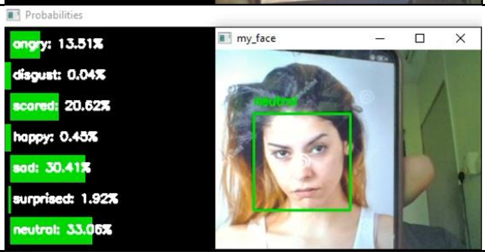
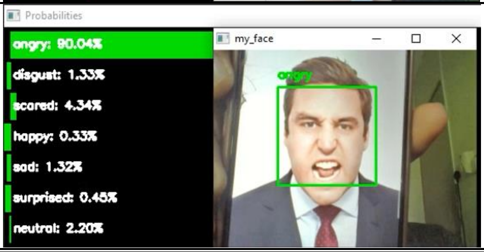
Appendix B: Testing Results

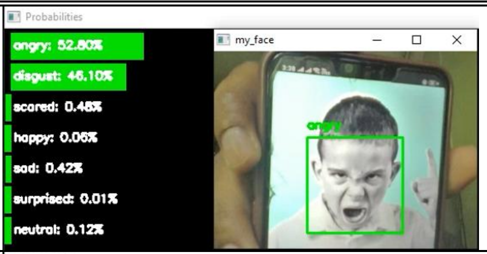
Happy Disposition

<table>
<tr><th>NO.</th><th>SAMPLE IMAGES</th><th>RESULTS</th><th>CONFIDENCE ON PREDICTIONS (%)</th></tr>
<tr><td>1</td><td></td><td>Happy</td><td>99.24</td></tr>
<tr><td>2</td><td></td><td>Happy</td><td>89.74</td></tr>
<tr><td>3</td><td></td><td>Happy</td><td>97.86</td></tr>
<tr><td>4</td><td></td><td>Happy</td><td>99.98</td></tr>
<tr><td>5</td><td></td><td>Happy</td><td>99.15</td></tr>
<tr><td>6</td><td></td><td>Happy</td><td>99.68</td></tr>
</table>

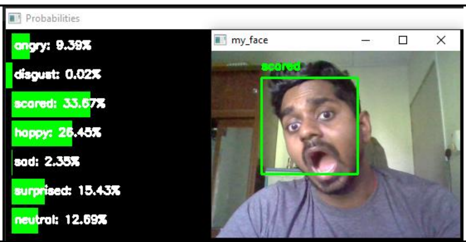
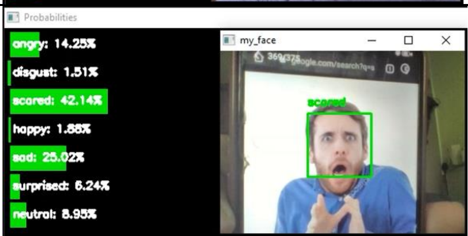
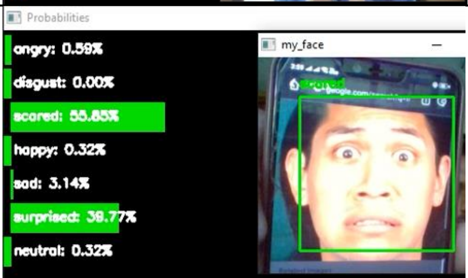
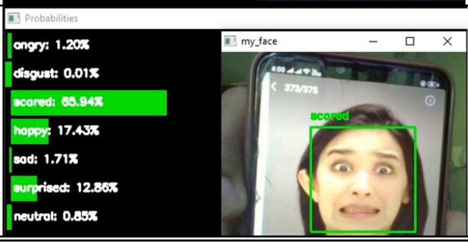
7		Happy	98.82

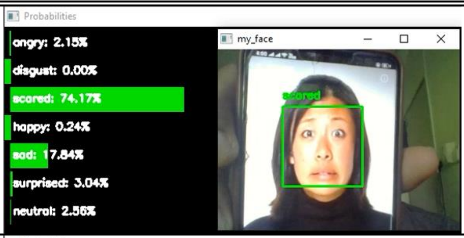
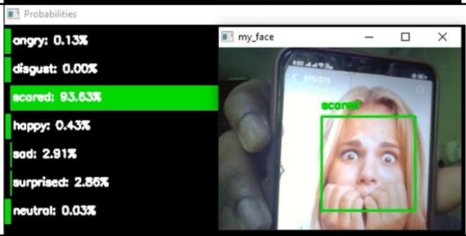
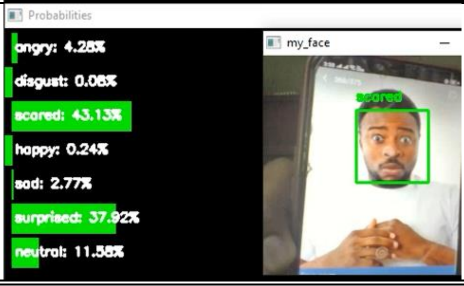
Angry Disposition

NO.	SAMPLE IMAGES	RESULTS	CONFIDENCE ON PREDICTIONS (%)
1		Angry	72.52
2		Angry	93.37
3		Angry	72.35
4		Neutral	Fail
5		Angry	90.04

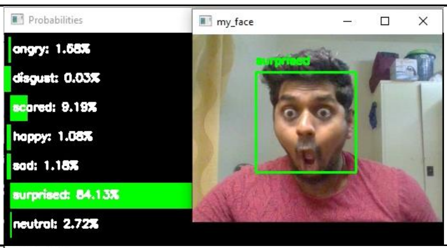
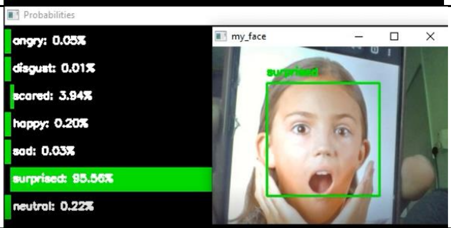
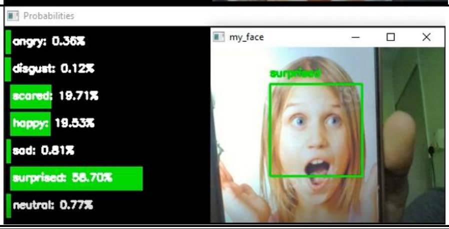
6		Angry	52.80
7		Disgust	Fail

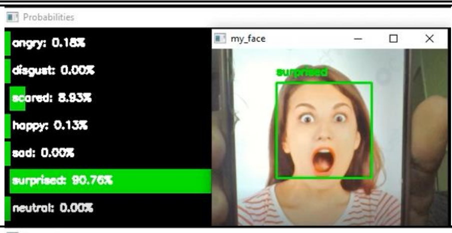
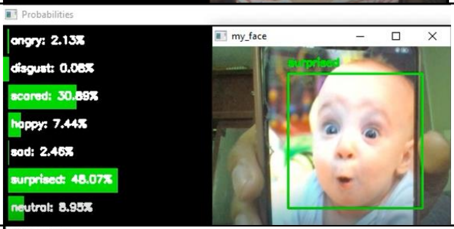
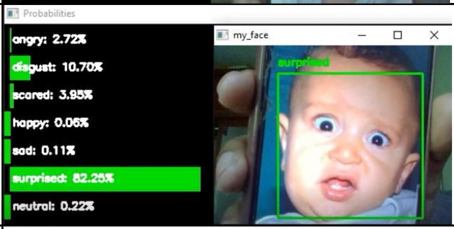
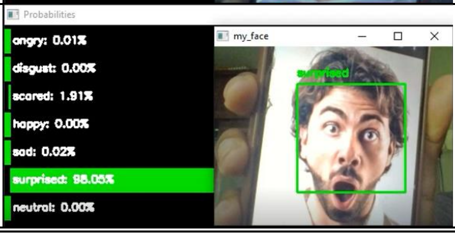
Scared Disposition

NO.	SAMPLE IMAGES	RESULTS	CONFIDENCE ON PREDICTIONS (%)
1		Scared	33.67
2		Scared	42.14
3		Scared	55.85
4		Scared	65.94

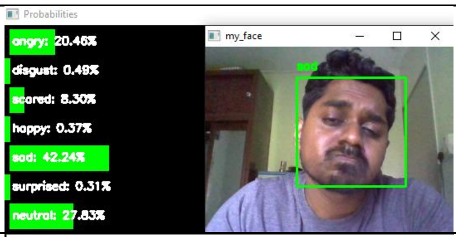
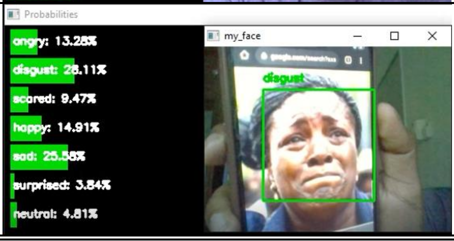
5		Scared	77.14
6		Scared	93.63
7		Scared	43.13

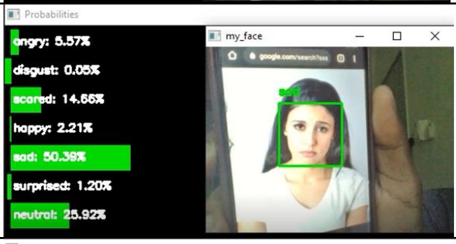
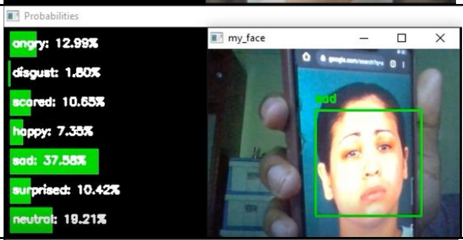
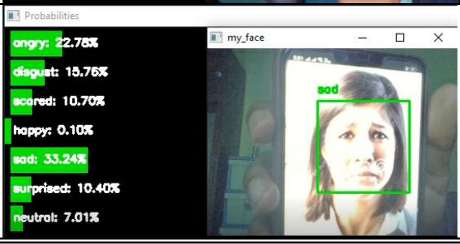
Surprised Disposition

NO.	SAMPLE IMAGES	RESULTS	CONFIDENCE ON PREDICTIONS (%)
1		Surprised	84.13
2		Surprised	95.56
3		Surprised	58.70

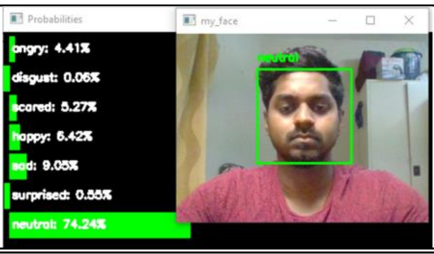
4		Surprised	90.76
5		Surprised	48.07
6		Surprised	82.25
7		Surprised	98.05

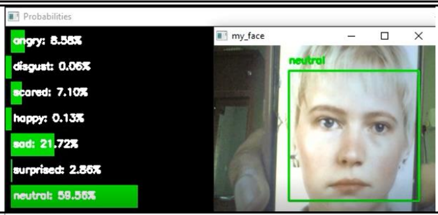
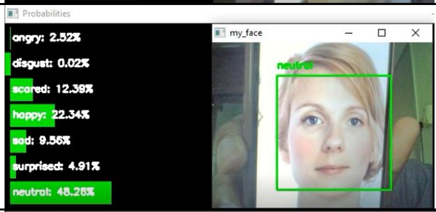
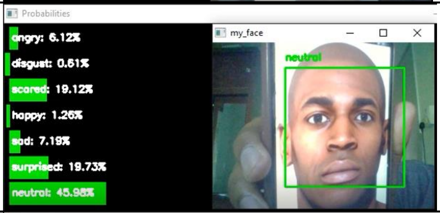
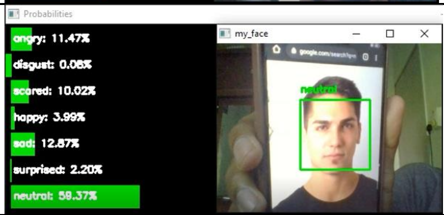
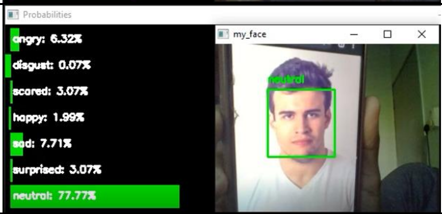
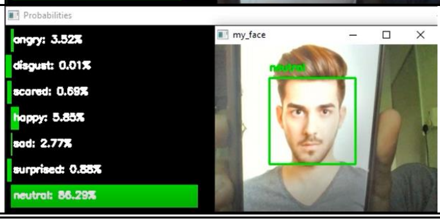
Sad Disposition

NO.	SAMPLE IMAGES	RESULTS	CONFIDENCE ON PREDICTIONS (%)
1		Sad	42.24
2		Disgust	Fail

3		Sad	64.10
4		Sad	50.39
5		Sad	56.37
6		Sad	37.58
7		Sad	33.24

Neutral Disposition

NO.	SAMPLE IMAGES	RESULTS	CONFIDENCE ON PREDICTIONS (%)
1		Neutral	74.24

<table>
<tr><td>2</td><td>
</td><td>Neutral</td><td>59.56</td></tr>
<tr><td>3</td><td>
</td><td>Neutral</td><td>48.26</td></tr>
<tr><td>4</td><td>
</td><td>Neutral</td><td>45.98</td></tr>
<tr><td>5</td><td>
</td><td>Neutral</td><td>59.37</td></tr>
<tr><td>6</td><td>
</td><td>Neutral</td><td>77.77</td></tr>
<tr><td>7</td><td>
</td><td>Neutral</td><td>86.29</td></tr>
</table>