HMR INSTITUTE OF TECHNOLOGY & MANAGEMENT

HAMIDPUR, DELHI 110036
Affiliated to
GURU GOBIND SINGH INDRAPRASTHA UNIVERSITY Sector - 16C Dwarka, Delhi - 110075, India 2017-21
MINOR PROJECT

Emotional Recognition using facial expressions and speech analysis

Submitted in partial fulfillment of the requirements for the award of the degree of

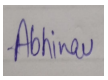
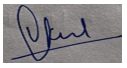
Bachelor of Technology
In
(Computer Science & Engineering)

Guide(s):
Ms. Teena Verma

Submitted by:
Abhinav k.Gupta 00296502718
Sahil Niranjan 02396502718
Vinay Kumar 03496502718
Yash Vashist 03696502718

DECLARATION

We, students of Btech (Abhinav, Sahil, Vinay, Yash) hereby declare that the Minor project entitled **"Emotional Recognition using facial expressions and speech"** which is submitted to Department of CSE, HMR Institute of Technology & Management, Hamidpur Delhi, affiliated to Guru Gobind Singh Indraprastha University, Dwarka(New Delhi) in partial fulfillment of requirement for the award of the degree of Bachelor of Technology in Information Technology, has not been previously formed the basis for the award of any degree, diploma or other similar title or recognition. The list of members involved in the project is listed below: -

S. No.	Student Name	Enrollment Number	Student Signature
1	Abhinav k.Gupta	00296502718	
2	Sahil Niranjan	02396502718	
3	Vinay Kumar	03496502718	
4	Yash Vashist	03696502718	

This is to certify that the above statement made by the candidate(s) is correct to the best of my knowledge.

Signature of Guide

Ms. Teena Verma

New Delhi

Date:

.

Designation & Department

Head of the Department	Dr. Mohd Izhar
Name of the Department	Computer Science & Engineering
HMRITM Hamidpur, New Delhi	

.

ACKNOWLEDGEMENT

It is a great pleasure for me to acknowledge the kind of help and guidance received to me during my project work. I want to express my sincere gratitude to our guide Mr Anupam Sir for giving us this opportunity to undergo the project and constantly motivating me to complete the project.

I would like to thank God for his blessings and I am extremely grateful to my parents for their love, prayers, caring and sacrifices for educating and preparing me for my future.

To my friends, you should know that your support and encouragement was worth more than I can express on paper.

.

ABSTRACT

Emotions play a vital role in people's daily lives. Understanding feelings and knowing how to respond to the feelings of others is important to involve effective communication with the community. Currently, emotional awareness is not only important in people’s daily lives, but also a hot topic in curriculum research, such as new strategies as the perception of emotions from the context of speech encourages us to feel what emotions are related to the content we are talking about.

The need and importance of emotional recognition has increased dramatically in many programs in recent years, such as video games, online communication, computer computing, and compatible computing. Emotional recognition can be made from Many sources including text, speech, hand, body language and facial expressions.

Currently, most sensory systems use only one of these sources. People's feelings change every second and one method used to process emotional recognition may not reflect emotions in the right way. This research recommends the desire to understand and explore people's feelings in many similar ways speech and face. In this case, various emotional states have been used.

The proposed framework can detect emotions in speech, facial expressions, and both to them. We have chosen to explore textual, sound and video inputs and develop an ensemble model

.

that gathers the information from all these sources and displays it in a clear and interpretable way.

By improving the emotion recognition accuracy, the proposed multi-sensory emotion recognition system can help to improve the naturalness of human computer interaction.

.

CONTENTS

LIST OF FIGURES

LIST OF TABLES

INTRODUCTION

1.1GENERAL

Computers currently play a very important role in many fields of people's lives. Traditional computer integration such as keyboard and mouse cannot satisfy the needs of the people. Therefore, to contribute to the co-operation of the environment between a computer and a person is much needed. To date, human communication through computers is much less natural than interconnected people. It is not yet possible to make face-to-face contact with a person complete computer communication.

To improve natural birth, computers should be able to support the way people communicate. An important need is that they should be able to find and express emotions. They should also have the ability to perceive and understand the feelings of one's fellow human beings so that there is a natural connection like communication between people.

Emotions are one of the most complex, both physical and mental. In the daily life of the community, knowing the emotional feelings of our partners is important and accurate. However, when it comes to computers, it is very difficult to get the feeling of colleagues.

In human communication, there are many things like speech, gestures, eye contact, and facial expression. Basic units for verbal communication in speech and basic units for verbal communication include facial expressions, body movements, gestures and eye contact.

Two popular methods utilized mostly in the literature for the automatic FER systems are based on geometry and appearance.

Facial Expression Recognition is usually performed in four-stages consisting of pre-processing, face detection, feature extraction, and expression classification. In this project we applied various deep learning methods (convolutional neural networks) to identify the key seven human emotions: anger, disgust, fear, happiness, sadness, surprise and neutrality

The basic units of verbal communication are words in speech and the basic units of the nonverbal communication include facial expression, body movements, and gestures as well as the eye contact. Sometimes, speech itself is sufficient for communication, take the phone call for example. Nonverbal communication units are also important, especially in face-to-face communication.

Moreover, the emotions reflected from facial expression are far more intense than words. So, to approach more natural human computer interaction, the computer should have the ability to recognize the emotion from speech, facial expression, or body gestures. In this dissertation, we proposed a multi-sensory emotion recognition framework to recognize emotion from speech and facial expression.

Experiments show that multi-sensory emotion recognition can help the computer understand emotions of human beings more accurately.

Hence, the proposed multi-sensory emotion recognition can help to make human computer interaction more natural.

Several Projects have already been done in this fields and our goal will not only be to develop an Automatic Facial Expression Recognition System but also improving the accuracy of this system compared to the other available systems

1.2 OVERVIEW OF THE PROJECT

Emotions and Emotion Recognition

What is emotion?

From the psychological viewpoint, emotion is often determined as a complex state of feeling which represents the physical and psychological changes that can affect human beings' thought and behavior. Emotions can be reflected from voices, speeches, hand and body gestures, as well as facial expressions that play an important role in our daily life. When a person smiles, it may indicate that the person is happy at that moment, while if one is frowning it may establish that the emotion of this person is sadness or anger. The number of emotions that human beings use in everyday life is hard to specify because of different situations and environments.

It is widely accepted by psychological theory that the emotions of humans can be classified into six general categories: happiness, anger, sadness, fear, disgust, and surprise. However, at different times humans express different moods about how they are feeling by showing

different facial actions and gestures during communications. The emotional expressions used are much more, and some of them are combinations of more than one.

With the rapid development of the computer and the Internet, understanding emotions and recognizing how to react to people's expressions are becoming important in video games, human computer interaction, cognitive computing, and affective computing. Thus, emotion recognition and its application gain more and more popularity in both scientific research field and industry field. Nevertheless, to observe or interpret human emotion is not an easy task for computers since emotion itself is complicated.

Currently, emotion recognition can be done via many ways including facial expression, voice, body gestures, as well as body movements, in which facial emotion recognition and speech emotion recognition get increasing interest in designing human machine interaction. Oral communication is a rich source of information for emotion recognition when people communicate with others. Moreover, sometimes it is not important what we have spoken, but how we spoke. The facial expression is the most visible form of emotional communication, but it is also most easily controlled by the speakers in different social environments when compared to speech and other kinds of expression.

Thus, emotion recognition by a single way, such as facial expression or speech, may not reflect the person's emotional state correctly.

To overcome the limitation of a single way emotion recognition, multimodal emotion recognition was proposed theoretically. Schuller, Lang, and Rigoll (2002) discussed multimodal emotion recognition by analyzing the user's speech signal and interaction on a touch screen or via mouse. In their work not only the common prosodic speech features (pitch and energy), but also the semantic and intention-based features (wording, degree of verbosity, temporal intention and word rate, as well as the history of user utterances) were used. The limitation of this work is that only a small video data set was used in the experiment. This dissertation focuses on multi-sensory emotion recognition.

Our goal is to design a framework for emotion recognition from speech via the microphone and facial expression via the built-in webcam or other camera device and then fuse the information of these two parts to improve the accuracy of emotion recognition.

There are plenty of challenges in the research of emotion recognition, such as what kind of emotions should be conceived to recognize and where to observe emotions since the information about emotion can be found in facial expressions, body gestures, speeches, or conversation.

Actually, human beings are the experts in emotions since we use emotions all the time in our daily lives. We can also name the emotion we expressed to others and tell the emotion states

the people we interacted with. However, it is hard to describe emotions by a computer automatically. Moreover, to classify emotions by some rules is even harder.

Thus, the models of emotions are significant. Psychologists have done plenty of detailed research on emotions, and they have proposed multiple models or theories for the description of emotions. Generally, emotions can be divided into three categories: the basic emotion model, the dimensional model, and the componential appraisal model (Grandjean, Sander, & Scherer, 2008).

According to the experiments of Ekman (1992), judging the static images with facial expressions of humans, there are six basic emotions that can be recognized universally. These emotions are happiness, sadness, surprise, fear, anger, and disgust. From his research higher level emotions can be combined from six basic emotions. The higher-level emotion recognition is not in the research scope of this dissertation, since most of recent emotion recognition research focuses on the basic emotion model such as the six basic emotions represented by Ekman (Ekman, 1992). Figure 1 is an example of the facial expressions of the six basic emotions.

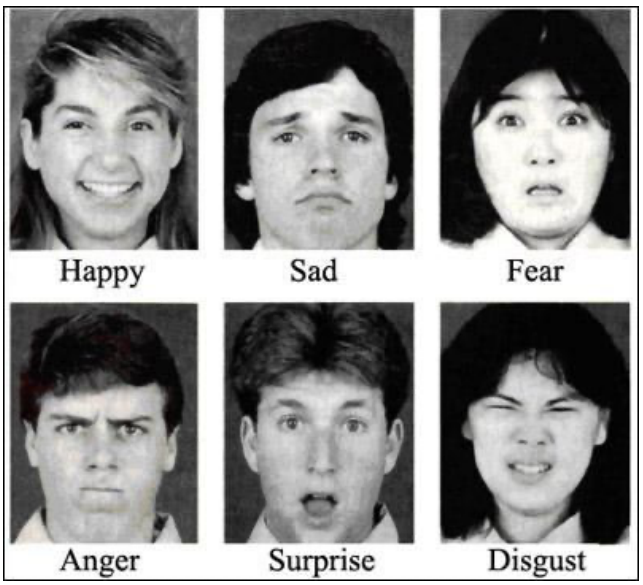


Fig 1: Six Basic facial expressions

EMOTIONS	DESCRIPTION
Happy	The eyebrows are relaxed. The mouth is open and the mouth corners upturned.
Sad	The inner eyebrows are bent upward. The eyes are slightly closed. The mouth is usually relaxed.
Fear	2 The eyebrows are raised and pulled together. The inner eyebrows are bent upward. The eyes are open and tense.

Anger	3 The inner eyebrows are pulled downward and together. The eyes are wide open. The lips are tightly closed or opened to expose the teeth.
Surprise	4 The eyebrows are raised. The upper eyelids and the eyes are wide open. The mouth is opened.
Disgust	5 The eyebrows and eyelids are relaxed. The upper lip is raised and curled, often asymmetrically.

Table 1: Depicting the features of emotions

Justification of study

Specific facial expressions that are recognized around the world and seem universal are not linked to emotions, but rather to certain conditions that also elicit emotions. Furthermore, it is often not enough to describe the emotions of daily life by the mixing of basic emotions. A real emotion may be between joy and surprise, shadowed by anticipation, and the intensity of all the components is also important. Ekman also reported that there is some confusion from the judgment study of the six basic emotions. For example, anger and disgust, as well as fear and surprise, are commonly confused. Surprise is also confused with the emotion of interest.

Emotions can also be represented in a dimensional framework where emotions can be mapped by two or three variables. According to this approach, the emotional states are not independent from each other. Variables for a two dimensional space are usually valence and arousal. For a third dimension energy or control is usually included. The valence dimension usually represents the positive or negative degree of the emotion, and the range is from uncomfortable feelings to comfortable feelings. The arousal dimension represents how excited the emotion is, and it ranges from low to high. While the energy or control dimension represents the degree of the energy or control over the emotion. These variables enable a more accurate description of emotions, since multiple aspects of emotions are used simultaneously in a continuous range of values. Basic emotions can still be represented in a

dimensional model as a point or an area. According to the differences of valence and arousal, emotion can be represented in the two dimensional space. For example, happiness can be elicited by a beautiful sunset or a smiling baby (the yellow rectangle) in the arousal and valence dimensions. There are some studies on using dimensional models for emotion recognition (Grimm, Mower, Kroschel, & Narayanan, 2006; Wöllmer, et al., 2008), and usually only two dimensional models (valence-arousal model) are used. The advantage of the dimensional model is that it is very intuitive to represent emotions on some continuous scale. However, the reduction of emotion differentiation to two or three dimensions will lose some information (Ekman, 1982; Tomkins, 1962). Moreover, some of the basic emotions proposed by Ekman, such as happiness or sadness, are easy to recognize in the dimensional models. However, some other emotions such as anger and disgust are hard to distinguish, and some of the emotions cannot even be described. Take Figure 3 for example, surprise is missing in the two-dimensional model.

1.3 Literature Survey

As per various literature surveys it is found that for implementing this project four basic steps are required to be performed.

i. Preprocessing

ii. Face registration

iii. Facial feature extraction

iv. Emotion classification

Description about all these processes are given below-

1. Preprocessing: Preprocessing is a common name for operations with images at the lowest level of abstraction both input and output are intensity images. Most preprocessing steps that are implemented are –

    a. Reduce the noise

    b. Convert the Image To Binary/Grayscale.

    c. Pixel Brightness Transformation.

    d. Geometric Transformation

2. Face Registration: Face Registration is a computer technology being used in a variety of applications that identifies human faces in digital images. In this face registration

step, faces are first located in the image using some set of landmark points called "face localization" or "face detection". These detected faces are then geometrically normalized to match some template image in a process called "face registration"

3. Facial Feature Extraction: Facial Features extraction is an important step in face recognition and is defined as the process of locating specific regions, points, landmarks, or curves/contours in a given 2-D image or a 3D range image. In this feature extraction step, a numerical feature vector is generated from the resulting registered image. Common features that can be extracted.
    - Lips
    - Eyes
    - Eyebrows
    - Nose tip
4. Emotion Classification: In the third step, of classification, the algorithm attempts to classify the given faces portraying one of the seven basic emotions.

Information about emotion is encoded in many aspects of speech such as what is said in a speech and the way it is said. The same sentence can be uttered in different ways to convey totally different emotions, while different people may prefer different words to describe exactly the same thing. Generally, there are two major types of messages (Johnstone &

Scherer, 2000) in speech: the explicit (linguistic) message and the implicit (paralinguistic) message.

Explicit messages are the contents of what was spoken, while the implicit messages are the way how the contents were spoken.

In speech emotion recognition, if one only considers the content, what was spoken, important information may be missed from the speaker's utterance, and even the content will be misunderstood. Furthermore, the linguistic content and emotion are language dependent, and a general rule is hard to conclude .

Currently, there are many approaches that have been done by researchers on detecting emotion from the implicit messages. However, the interpreting of paralinguistic features has not been fully discovered. Some ambiguities regarding how different acoustic features affect different emotional states still exist.

The advantages for recognizing emotion from audio information are easy to realize, low cost, and multiple available databases. The disadvantages are language dependent, lower recognition accuracy when compared to facial expression and ambiguity about the implicit messages.

Currently, emotion recognition from speech grows to an important research area. Several approaches to recognize emotions from speech have been proposed that are capable of detecting emotion from speech by acoustic or linguistic information.

So far, there are several off-line approaches for speech emotion recognition, and on-line speech emotion recognition has been achieved scarcely.

1.4 PROBLEM STATEMENT

As mentioned in the previous chapter, emotion recognition from speech is difficult because of several events and challenges. First of all, human beings' emotions are extremely varied, various and ambiguous. It is frequently very hard for humans themselves to recognize others' emotions correctly, since each individual has his/her own manifestations of different emotions. This makes picking out a person's emotion via a machine or a computer even harder.

The full scope and variety of emotions make the emotion recognition much harder. **It is inconceivable to count all the spectrum of emotions.** Some subtle emotions like pride, appreciation, etc. may not yet be given voice. Most emotional recognition systems can simply make out a little set of emotions, for example the six basic emotions.

Another question for emotion recognition from speech is **what data is the best.** Since there are different ways to show emotions, what data is chosen to best represent different emotions on the spot? There is a great amount of information that can be collected, and many characteristics can be drawn out. So far, researchers have not been able to zero down on the feature-sets which are better than others for speech emotion recognition.

Finally, the **lack of standardized data sets** makes the comparison of recognition results more difficult. Since researchers use different data sets and different features to perform emotion recognition, to compare the resolutions and achieve a consensus on the practice is hard. **How to collect the data for emotion recognition is also a major problem.**

The recognition rate is high in experiments while testing in the actual world the similar result is harder to achieve. The high quality of records held for training purposes have significantly better performance than what we can collect in our daily lives during the experiments

1.5 SCOPE OF STUDY

As we described previously, the research of emotion has a long history in psychology. In this section several excellent research works on emotion recognition from speech will be briefly summarized. In the final few decades, a number of researches on affective computing have been nominated. In history, the **first practical** pilot work on automatic emotion recognition was conducted in 1996 by Dellaert, Polzin, and Waibel. In their work, several statistical pattern recognition techniques to classify utterances according to the emotional content were proposed. They have collected over 1000 utterances from different speakers in four different, acted emotional states: happiness, sadness, fear, and anger. In their work, only the pitch features were taken out from the utterances for classification purposes. Although there are many achievements for emotion recognition from facial expression, there are still many difficulties and limitations in emotion recognition due to the complexity of emotion expression, especially when in a conversation. Another one is that the recognition results depend heavily on the quality of the image or video

In speech emotion recognition, feature sets are very important for training the classifier and doing classification. These features can be either local or global and linguistic or acoustic. They concluded that there are significant differences in the feature sets of different types of databases.

CHAPTER-2 SYSTEM ANALYSIS

INTRODUCTION

The chapter describes the system study, analysis, design strengths and weaknesses of the current system, Contest level diagrams, Entity Relationship Diagram, Architectural design.

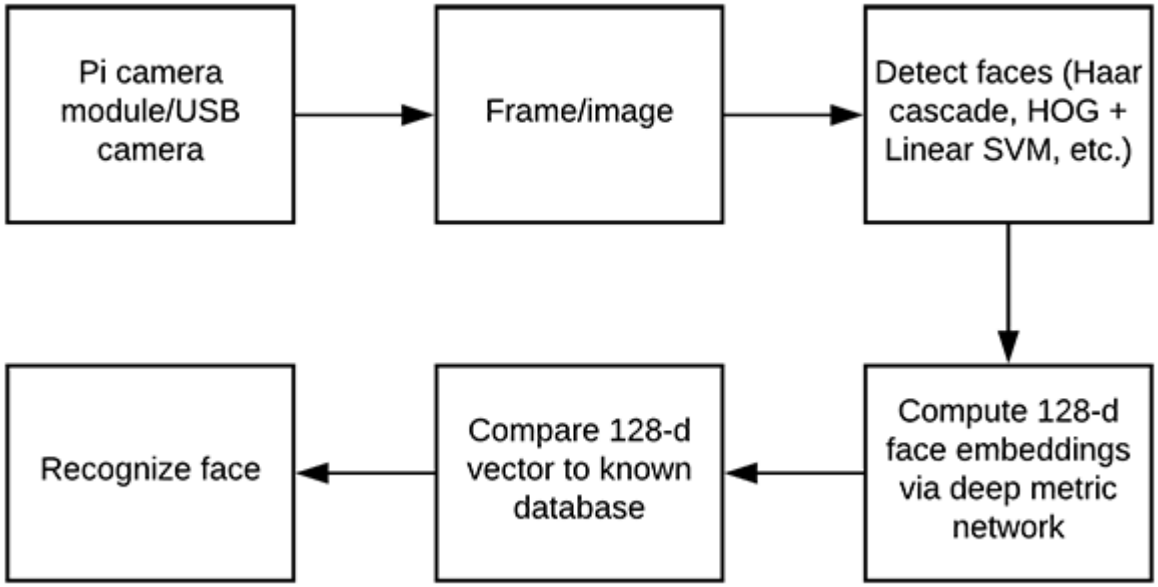


Fig 2.1 Depicts the process involved

2.2 Preliminary Investigation

System Study

Factors of Speech Emotion Recognition

From the related works on speech emotion recognition, we have noticed that for speech emotion recognition there are usually 3 main steps: signal processing, feature extraction and selection, as well as classification. Six steps involved are:

1. Voice Recording,
2. Audio Discretization (breaking of audio),
3. Log-mel-spectrogram (sound filtering),
4. Split spectrogram using equal size (filtered audio splits in small equal size),
5. Finally, in the classification step, models are trained with emotion labels and then used to predict the emotion of a new instance. In this section some of the basics of speech signal concepts will be introduced. Then the details of the three major steps will be described. Finally, the feature we used in this project and some widely used speech databases will Be given.

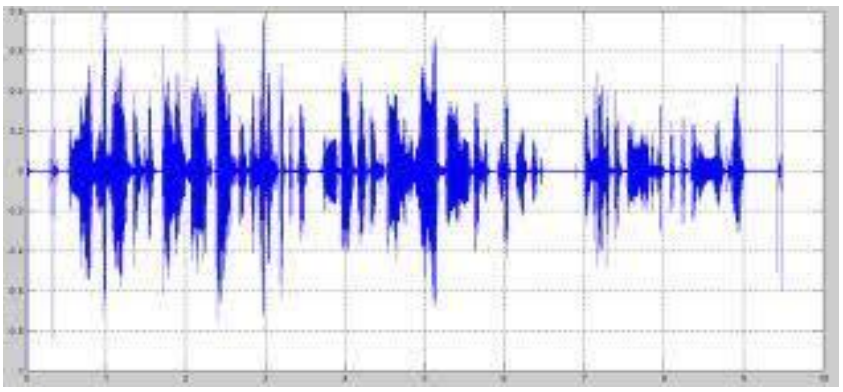


Fig 2.2: Analog signal of speech

Factors of Facial Emotion Recognition

Vision based emotion recognition is mainly focused on facial expressions because of the significance of face in emotional expression and perception. Many approaches have been done in this area. However, there are still many challenges in the facial emotion recognition area. As we know faces are non-rigid and have different colors and pose. Some of the facial features are not common and not suitable for pattern recognition. Lighting of the background and the illumination conditions can also change the overall recognition rate of facial expression. The above problems make the emotion recognition from facial expression more complicated. The input data for the recognition system can be video streams from a Web Camera or some recorded video clips, as well as some static facial images. After the 3 steps, the final decision of the facial emotion will be given as the output.

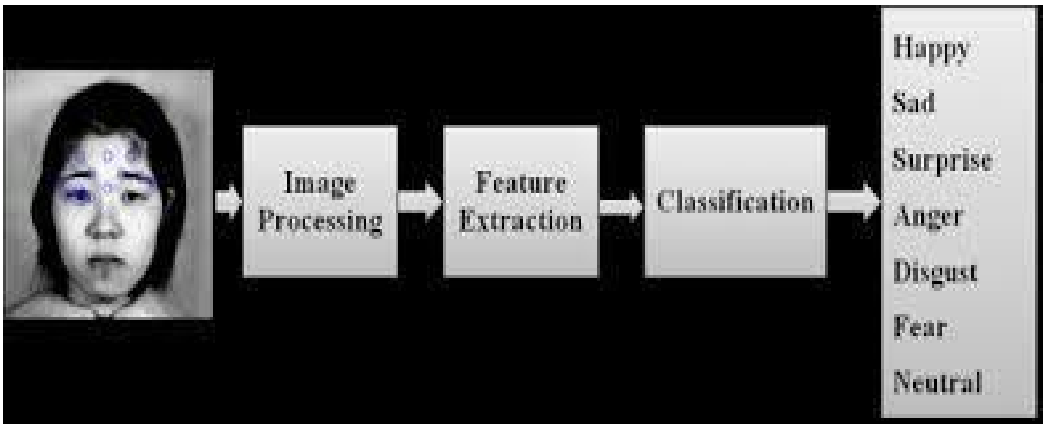


Fig 2.3: Facial recognition steps

Face Detection

Though there are several types of input to the facial emotion recognition system, images containing faces are still essential to intelligent vision-based human computer interaction. Face detection tries to identify all image regions which contain a face regardless of its 3D position, orientation, and lighting conditions from a given image. The challenge here is that faces are non-rigid and have a high degree of variability in size, shape, color, and texture. A variety of face detection techniques have been developed. In the Yang, Kriegman, and Ahuja (2002) survey, the authors classified the detection methods into four categories:

Knowledge-based methods, Feature invariant approaches, Template matching methods, and Appearance-based methods.

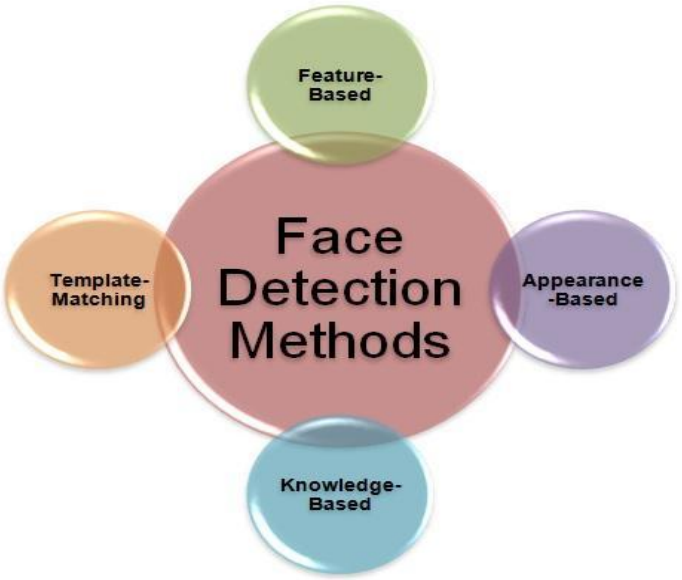


Fig 2.4: Face detection methods

2.3 Feasibility Study

After doing this project, study and analyze all the existing or required functionalities of the system, the next task is to do the feasibility study for the project. All the projects are feasible -given unlimited resources and infinite time.

Feasibility study includes consideration of all the possible ways to provide a solution to the given problem.

The proposed solution should satisfy all the user requirements and should be flexible enough so that future changes can be easily done based on the future upcoming requirements.

2.3.1 Technical Feasibility

This included the study of function, performance and constraints that may affect the ability to achieve an acceptable system. For this feasibility study, we studied complete functionality to be provided in the system, as described in the System Requirement Specification, and checked if everything was possible using different type of frontend and backend platforms

2.3.2. Economic Feasibility

This is a very important aspect to be considered while developing a project. We decided on the technology based on the minimum possible cost factor.

- All the hardware and software cost has to be borne by the organization.
- Overall, we have estimated that the benefits the organization is going to receive from the proposed system will surely overcome the initial costs and the later on running cost for the system.

2.3.3. Operational Feasibility

No doubt the proposed system is fully GUI based and is very user friendly and all inputs to be taken are all self-explanatory even to a layman. Besides, a proper training has been conducted to let know the essence of the system to the users so that they feel comfortable with new system .

2.4 Software Requirements

Software Minimum System Requirement: As the project is developed in python, we have used Anaconda for Python 3.6.5 and Jupyter Notebook.

Anaconda

It is a free and open-source distribution of the Python and R programming languages for data science and machine learning related applications (large-scale data processing, predictive analytics, scientific computing), that aims to simplify package management and deployment. Package versions are managed by the package management system conda. The Anaconda distribution is used by over 6 million users, and it includes more than 250 popular data science packages suitable for Windows, Linux, and MacOS.

Fig 2.5: Depicts Anaconda Logo

Features include:

- editor with syntax highlighting and introspection for code completion
- support for multiple Python consoles (including IPython)
- the ability to explore and edit variables from a GUI

Hardware Interfaces:

- Processor: Intel CORE i5 processor with minimum 2.9 GHz speed.
- RAM: Minimum 4 GB.
- Hard Disk: Minimum 500 GB

Software Interfaces:

- Microsoft Word 2003
- Database Storage: Microsoft Excel
- Operating System: Windows10

2.5 Data Flow Diagram

A data flow diagram (DFD) is a graphical representation of the "flow" of data through an information system, modelling its process aspects. A DFD is often used as a preliminary step to create an overview of the system without going into great detail, which can later be elaborated. DFDs can also be used for the visualization of data processing (structured design).

FOR FACIAL EMOTIONS

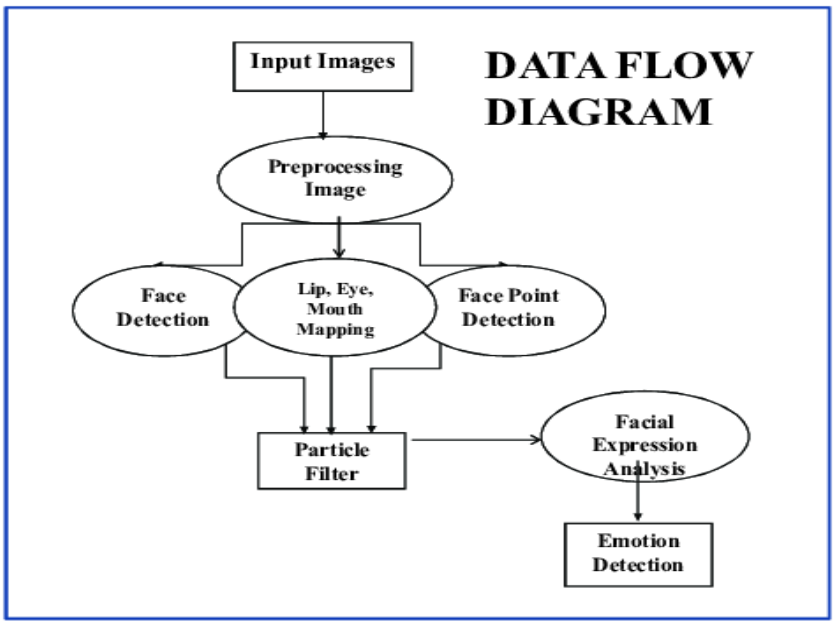


Fig 2.6: Data Flow diagram (Facial Expression)

FOR SPEECH

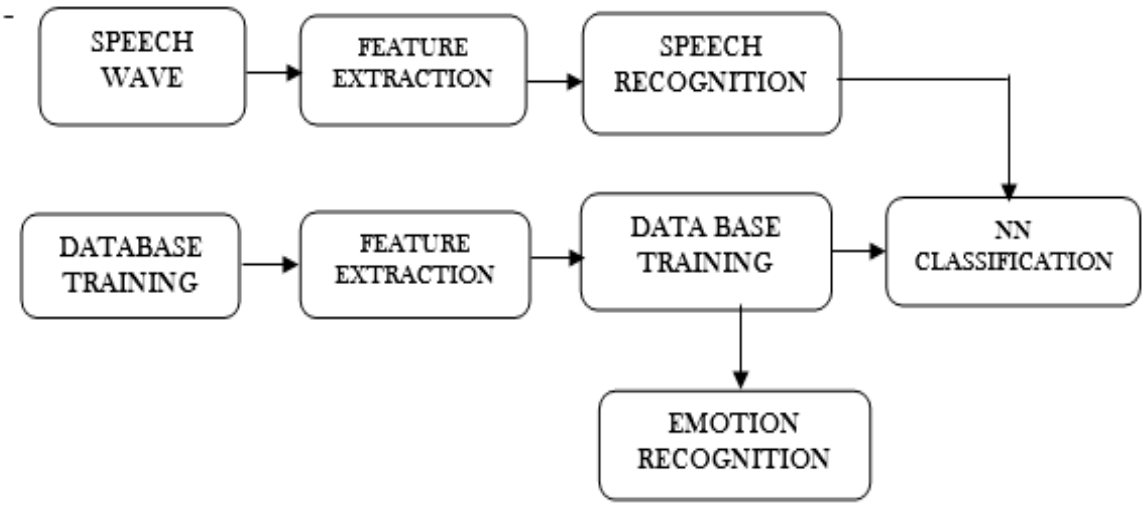


Fig 2.7 : Data flow Diagram (Speech)

CHAPTER-3

SYSTEM DESIGN

3.0 Methodology

Waterfall Model

3.1 Justification of Methodology

Every software developed is different and requires a suitable SDLC approach to be followed based on the internal and external factors.

Requirement Gathering and analysis: All possible requirements of the system to be developed are captured in this phase and documented in a requirement specification doc.

System Design: The requirement specifications from the first phase are studied in this phase and system design is prepared. System Design helps in specifying hardware and system requirements and also helps in defining overall system architecture.

Implementation: With inputs from system design, the system is first developed in small programs called units, which are integrated in the next phase. Each unit is developed and tested for its functionality which is referred to as Unit Testing.

Integration and Testing: All the units developed in the implementation phase are integrated into a system after testing of each unit. Post integration the entire system is tested for any faults and failures.

Deployment of system: Once the functional and nonfunctional testing is done, the product is deployed in the customer environment or released into the market.

Maintenance: There are some issues which come up in the client environment. To fix those issues patches are released. Also to enhance the product some better versions are released. Maintenance is done to deliver these changes in the customer environment. All these phases are cascaded to each other in which progress is seen as flowing steadily downwards (like a waterfall) through the phases. The next phase is started only after the defined set of goals are achieved for the previous phase and it is signed off, so the name "Waterfall Model". In this model phases do not overlap.)

The steps we followed while developing this project are-:

1. Analysis of the problem statement.
2. Gathering of the requirement specification
3. Analysation of the feasibility of the project.
4. Development of a general layout.
5. Going by the journals regarding the previous related works in this field.

6. Choosing the method for developing the algorithm.
7. Analyzing the various pros and cons.
8. Starting the development of the project
9. Developing an algorithm.
10. Analysation of algorithms and increase the accuracy.
11. Coding as per the developed algorithm in PYTHON.

We developed this project as per the iterative waterfall model.

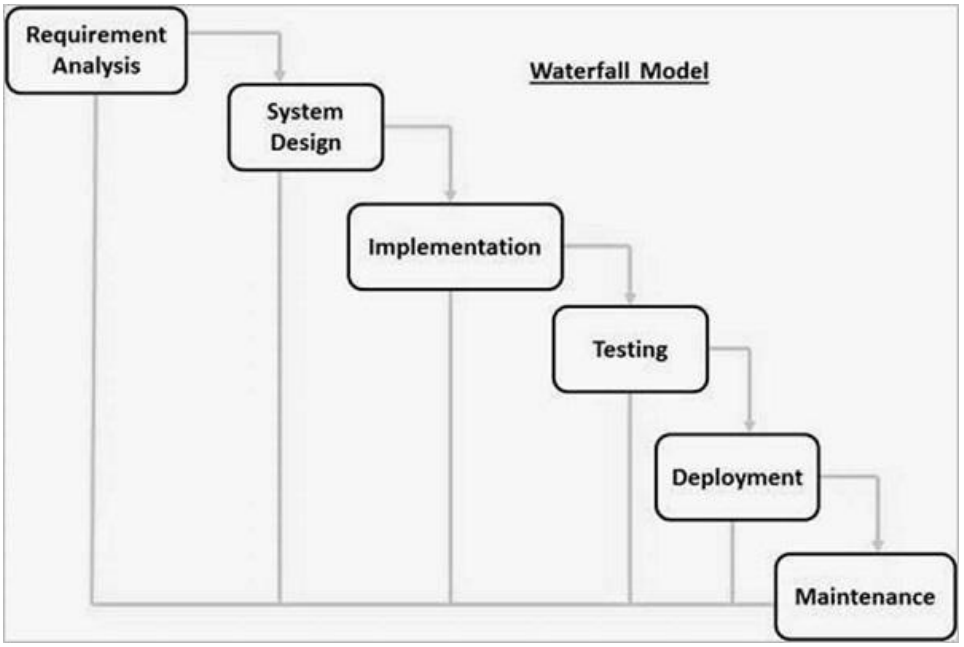


Fig 3.1 Depicts the waterfall model

3.2 User Interface Design

Comparison of Recognition Rate for Emo-DB

The following graph depicts the comparison of the recognition in our observed database. The emotional database taken in the project have following details:

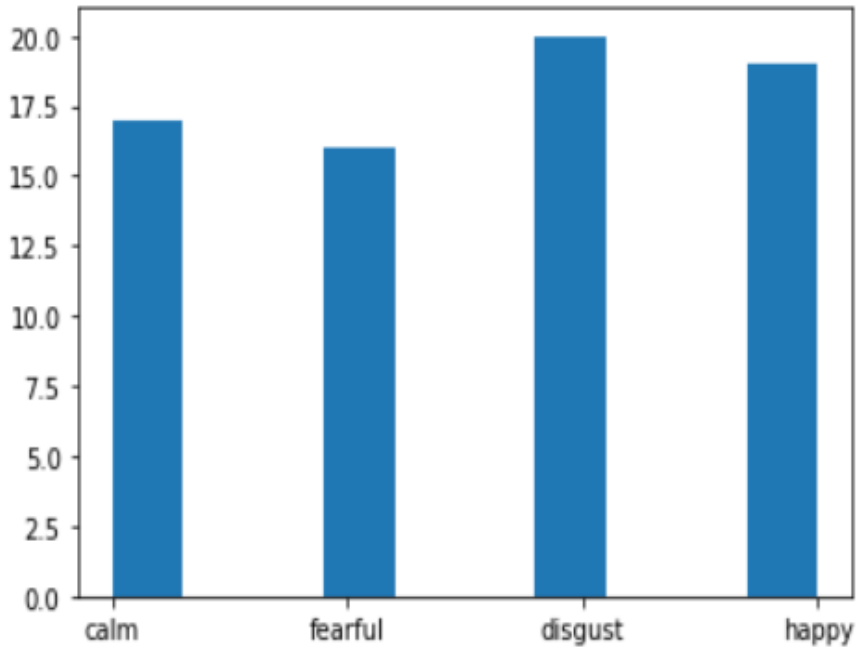


Fig 3.2 Depicts the recognition rate for our Database

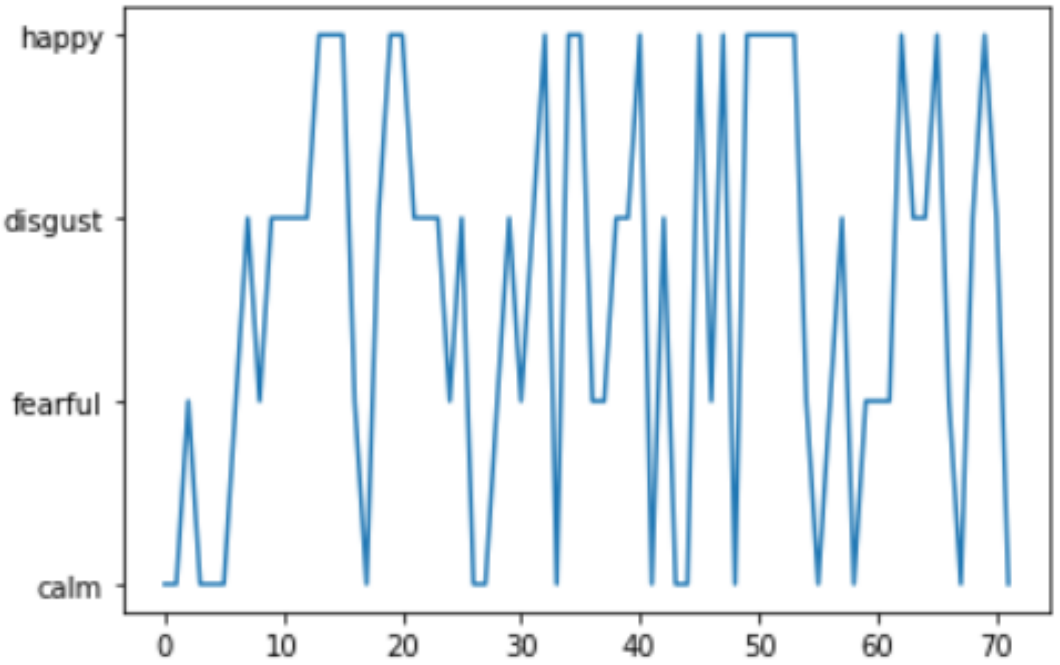


Fig 3.3 Depicts the recognition rate for our Database

This picture uses deepface technique to predict the expression of the boy

Deepface is a lightweight face recognition and facial attribute analysis (age, gender, emotion and race) framework for python. It is a hybrid face recognition framework wrapping state-of-the-art models: VGG-Face , Google FaceNet , OpenFace , Facebook DeepFace , DeepID , ArcFace and Dlib .

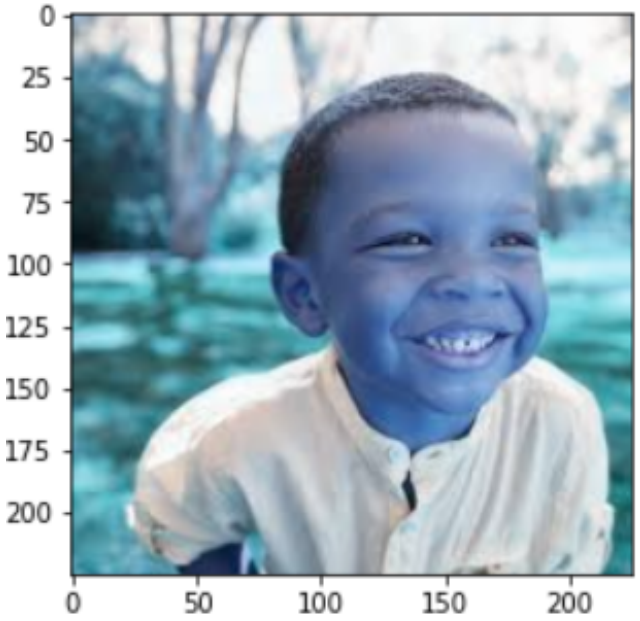


Fig 3.4 Boy's Face using deepface

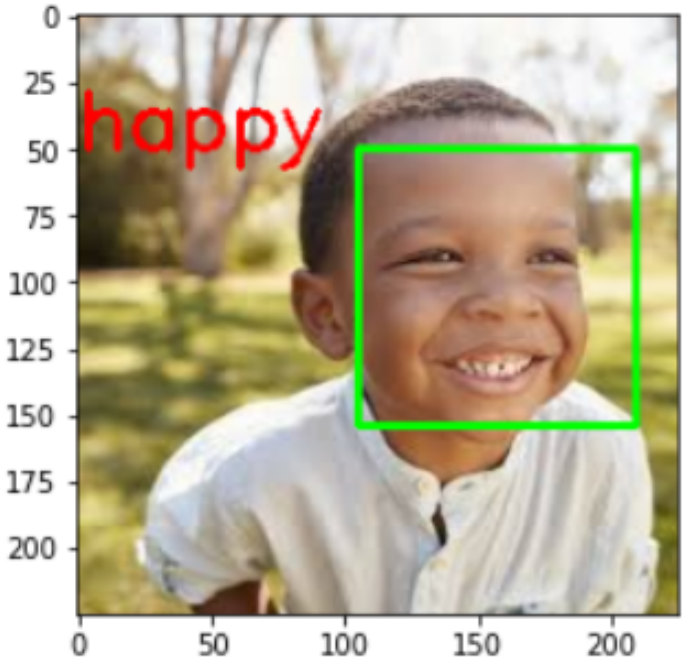


Fig 3.5 Happy Face using deepface

CHAPTER 4 -TESTING

ANALYSIS, DESIGN AND DEVELOPMENT

4.1 TYPES OF TESTING

Software testing determines the correctness, completeness, and quality of software being developed. Validation refers to the process of checking that the developed software meets the requirements specified by the user. The activities involved in the testing phase basically evaluate the capability of that system to meet its requirements. The main objective of software testing is to detect errors in the software. Errors occur if some part of the developed system is found to be incorrect, incomplete or inconsistent. Test techniques include, but are not limited to, the process of executing a program or application with the intent of finding software bugs (errors or other defects). It involves the execution of a software component or system to evaluate one or more properties of interest. In general, these properties indicate the extent to which the component or system under test: Meets the requirements that guided its design and development,

- Responds correctly to all kinds of inputs,4
- Performs its functions within an acceptable time,

- Is sufficiently usable,
- Can be installed and run in its intended environments, and
- Achieves the general result its stakeholders desire.

As the number of possible tests for even simple software components is practically infinite, all software testing uses some strategy to select tests that are feasible for the available time and resources. As a result, software testing typically (but not exclusively) attempts to execute a program or application with the intent of finding software bugs (errors or other defects). Software testing can provide objective, independent information about the quality of software and risk of its failure to users and/or sponsors. Software testing can be conducted as soon as executable software (even if partially complete) exists. The overall approach to software development often determines when and how testing is conducted. For example, in a phased process, most testing occurs after system requirements have been defined and then implemented in testable programs. In contrast, under an Agile approach, requirements, programming, and testing are often done concurrently.

Whitebox Testing Tests are performed to ensure that all internal operations of the software are performed according to the specifications of the client. This is called White box testing. White-box testing (also known as clear box testing, glass box testing, transparent box testing, and structural testing) is a method of testing software that tests internal structures or workings

of an application, as opposed to its functionality (i.e. black-box testing). In white-box testing an internal perspective of the system, as well as programming skills, are used to design test cases. White-box testing can be applied at the unit, integration and system levels of the software testing process. Although traditional testers tended to think of white-box testing as being done at the unit level, it is used for integration and system testing more frequently today. It can test paths within a unit, paths between units during integration, and between subsystems during a system–level test. Though this method of test design can uncover many errors or problems, it has the potential to miss unimplemented parts of the specification or missing requirements. The details entered by the administrator are saved and stored in the database, and testing is done to verify whether the control of each form or action is working in the exact way. Black box Testing Tests are performed to ensure that each function is working properly. This is referred to as Black box testing. Black-box testing is a method of software testing that examines the functionality of an application (e.g. what the software does) without peering into its internal structures or workings. This method of test can be applied to virtually every level of software testing: unit, integration, system and acceptance. It typically comprises most if not all higher level testing, but can also dominate unit testing as well. Test cases are built around specifications and requirements, i.e., what the application is supposed to do. Test cases are generally derived from external descriptions of the software, including specifications, requirements and design parameters. Although the tests used are

primarily functional in nature, non-functional tests may also be used. The test designer selects both valid and invalid inputs and determines the correct output without any knowledge of the test object's internal structure. Testing is conducted in the system so that the functions namely Login, Making, searching the nearest donor, getting routes to the nearest blood banks etc. are done properly.

TESTING STRATEGY

Condition Testing Test cases are derived to determine whether the logic conditions and decision statements are free from errors. Condition testing strategy is used to check if the operators used if the operators used are correct and to verify conditions such as if an error message is displayed if a non-registered user is signed in to the application, or a user is registered without providing his email, address or full name. Loop Testing This testing is used to check the variety of loops present in programming. The working of the loops such as while, for and do while are checked for its proper execution. The statements inside the loop body are executed line by line for every condition that satisfies the loop. Unit Testing This testing is performed to test the individual units in the system. Each module in the system is tested individually and executed line by line for accurate functioning of the system. The appointment part of the user module has been tested for its proper functioning, since it’s the core part of the application. Integration Testing The objective of integration testing is to take

all tested individual modules, integrate them, test them again and develop the system. The user module, the appointment module as well as the doctor & admin should be integrated together for the proper functioning of the whole system. Testing is conducted at this stage to check whether the requested action has completed successfully and changes are affected properly at both the doctor and admin module

Acceptance Testing

This testing is performed to ensure that the functional, behavioural, and performance requirements of the system are met. It may involve physical tests or performance tests. The acceptance test suite is run against the supplied input data or using an acceptance test script to direct the testers. Then the results obtained are compared with the expected results. If there is a correct match for every case, the test suite is said to pass. If not, the system may either be rejected or accepted on conditions previously agreed between the sponsor and the manufacturer. The objective is to provide confidence that the delivered system meets the business requirements of both sponsors and users. The acceptance phase may also act as the final quality gateway, where any quality defects not previously detected may be uncovered.

So far, the framework of the multi-sensory emotion recognition has been introduced, including the speech emotion recognizer, facial expression recognizer, and information fusion. In this chapter, many experiments have been done based on the proposed work, and the results were examined in detail. Choosing Base Model for Speech Emotion Recognizer In the speech emotion recognition approach, we have already built many different models based on different databases, feature sets, as well as the classification algorithm.

CHAPTER -5 IMPLEMENTATION

SYSTEM IMPLEMENTATION

System implementation is the final stage of software development life cycle. For the successful implementation and cooperation of new systems users must be selected, educated and trained. Unless the users are not trained, the system will become complex and it will feel as a burden for them. A product software implementation method is a systematically structured approach to effectively integrate software based service or components into the workflow of an organizational structure or an individual end-user. A product software implementation method is a blueprint to get users and/or organizations running with a specific software product.

SYSTEM DEVELOPMENT

OpenCV:

OpenCV (Open-Source Computer Vision Library) is an open-source computer vision and machine learning software library. OpenCV was built to provide a common infrastructure for computer vision applications and to accelerate the use of machine perception in the commercial products. Being a BSD-licensed product, OpenCV makes it easy for businesses

to utilize and modify the code. The library has more than 2500 optimized algorithms, which includes a comprehensive set of both classic and state-of-the-art computer vision and machine learning algorithms. These algorithms can be used to detect and recognize faces, identify objects, classify human actions in videos, track camera movements, track moving objects, extract 3D models of objects, produce 3D point clouds from stereo cameras, stitch images together to produce a high resolution image of an entire scene, find similar images from an image database, remove red eyes from images taken using flash, follow eye movements, recognize scenery and establish markers to overlay it with augmented reality, etc. OpenCV has more than 47 thousand users and an estimated number of downloads exceeding 14 million.

The library is used extensively in companies, research groups and by governmental bodies. It has C++, Python, Java and MATLAB interfaces and supports Windows, Linux, Android and Mac OS. OpenCV leans mostly towards real-time vision applications and takes advantage of MMX and SSE instructions when available. A full-featured CUDA and OpenCL interfaces are being actively developed right now. There are over 500 algorithms and about 10 times as many functions that compose or support those algorithms. OpenCV is written natively in C++ and has a templated interface that works seamlessly with STL containers.

OpenCV's application areas include:

- 2D and 3D feature toolkits
- Ego motion estimation
- Facial recognition system
- Gesture recognition
- Human–computer interaction (HCI)
- Mobile robotics
- Motion understanding
- Object identification
- Segmentation and recognition
- Stereopsis stereo vision: depth perception from 2 cameras
- Structure from motion (SFM)
- Motion tracking
- Augmented reality

To support some of the above areas, OpenCV includes a statistical machine learning library that contains:

- Boosting
- Decision tree learning
- Gradient boosting trees

- Expectation-maximization algorithm
- k-nearest neighbor algorithm
- Naive Bayes classifier
- Artificial neural networks
- Random forest
- Random forest
- Support vector machine (SVM)
- Deep neural networks (DNN)

Fig 5.1: various packages used

Numpy

NumPy is an acronym for "Numeric Python" or "Numerical Python". It is an open-source extension module for Python, which provides fast precompiled functions for mathematical

and numerical routines. Furthermore, NumPy enriches the programming language Python with powerful data structures for efficient computation of multi-dimensional arrays and matrices. The implementation is even aiming at huge matrices and arrays. Besides that, the module supplies a large library of high-level mathematical functions to operate on these matrices and arrays. It is the fundamental package for scientific computing with Python.

It contains various features including these important ones:

- A powerful N-dimensional array object
- Sophisticated (broadcasting) functions
- Tools for integrating C/C++ and Fortran code
- Useful linear algebra, Fourier Transform, and random number

Numpy Array:

A numpy array is a grid of values, all of the same type, and is indexed by a tuple of nonnegative integers. The number of dimensions is the rank of the array; the shape of an array is a tuple of integers giving the size of the array along each dimension.

Sci-Py:

SciPy (Scientific Python) is often mentioned in the same breath with NumPy. SciPy extends the capabilities of NumPy with further useful functions for minimization, regression,

Fourier-transformation and many others. NumPy is based on two earlier Python modules dealing with arrays. One of these is Numeric. Numeric is like NumPy a Python module for high-performance, numeric computing, but it is obsolete nowadays. Another predecessor of NumPy is Numarray, which is a complete rewrite of Numeric but is deprecated as well.

Keras:

Keras is a high-level neural networks API, written in Python and capable of running on top of TensorFlow, CNTK, or Theano. It was developed with a focus on enabling fast experimentation. Keras contains numerous implementations of commonly used neural network building blocks such as layers, objectives, activation functions, optimizers, and a host of tools to make working with image and text data easier. The code is hosted on GitHub, and community support forums include the GitHub issues page, and a Slack channel. Keras allows users to productize deep models on smartphones (iOS and Android), on the web, or on the Java Virtual Machine. It also allows use of distributed training of deep learning models on clusters of Graphics Processing Units (GPU).

TensorFlow:

TensorFlow is a Python library for fast numerical computing created and released by Google. It is a foundation library that can be used to create Deep Learning models directly or by using wrapper libraries that simplify the process built on top of TensorFlow

Guiding Principles:

1. User Friendliness: Keras is an API designed for human beings, not machines. It puts user experience front and center. Keras follows best practices for reducing cognitive load: it offers consistent & simple APIs, it minimizes the number of user actions required for common use cases, and it provides clear and actionable feedback upon user error.
2. Modularity: A model is understood as a sequence or a graph of standalone, fully-configurable modules that can be plugged together with as little restrictions as possible. In particular, neural layers, cost functions, optimizers, initialization schemes, activation functions, regularization schemes are all standalone modules that you can combine to create new models.
3. Easy Extensibility: New modules are simple to add (as new classes and functions), and existing modules provide ample examples. To be able to easily create new modules allows for total expressiveness, making Keras suitable for advanced research

4. Work with Python: No separate models configuration files in a declarative format. Models are described in Python code, which is compact, easier to debug, and allows for ease of extensibility

Fig 5.2 Depicts tensorflow and keras logos

DeepFace:

Deepface is a lightweight face recognition and facial attribute analysis (age, gender, emotion and race) framework for python.

Face Registration

Haar Features:

The Haar feature is similar to Karnals, which is generally used to detect edges. All human faces share some similar features, like the eye region is darker than the upper cheek region,

nose region is brighter than eye region. By these match-able features, their location and size will help us to detect a face.

Here are some Haar features, using those we can say if there is a face or not. The Haar feature signifies that black region is represented by +1 and white region is represented by -1 .

Cascading: Suppose, we have an input image of 640X480 resolution. Then we need to move a 24X24 window throughout the image and for each window 2500 features are to be evaluated. Taking all 2500 features in a linear way it checks whether there is any threshold or not and then decides if it is a face or not.

Haar Cascade Classifier in OpenCv :

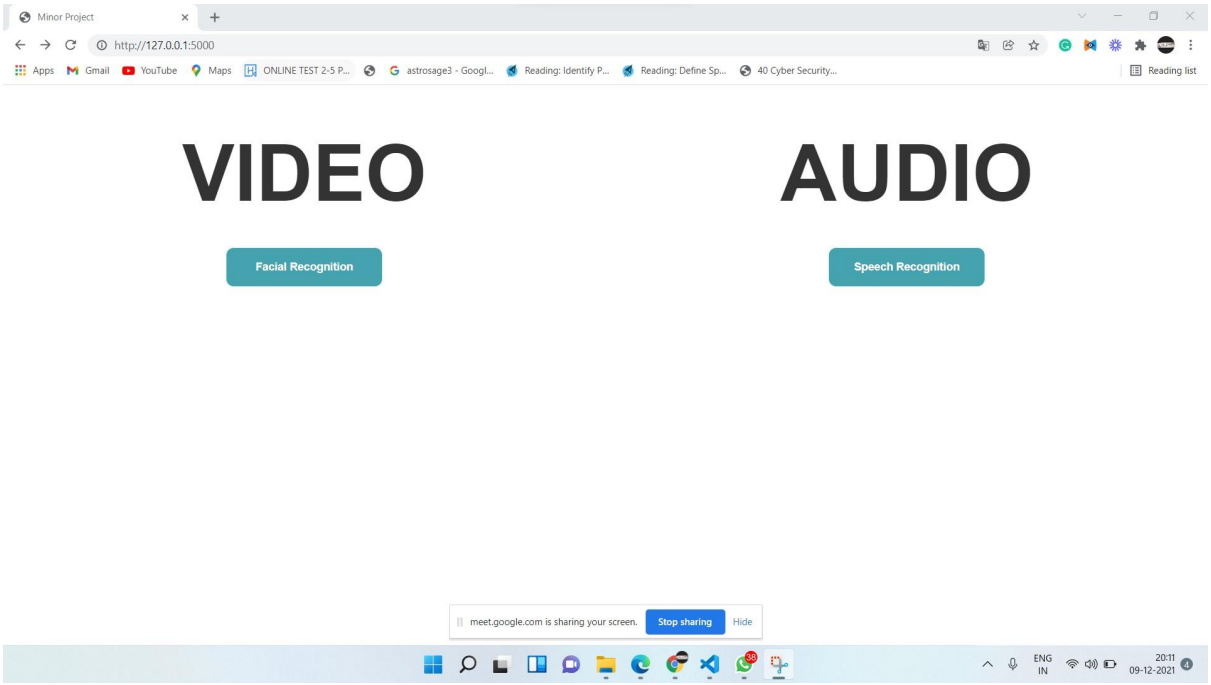
The algorithm needs a lot of positive images (images of faces) and negative images (images without faces) to train the classifier. Then we need to extract features from it. For this, haar features shown in the image below are used. They are just like our convolutional kernel. Each feature is a single value obtained by subtracting the sum of pixels under white rectangle from the sum of pixels under black rectangle. Now all possible sizes and locations of each kernel are used to calculate plenty of features. (Just imagine how much computation it needs? Even a 24x24 window results in over 160000 features). For each feature calculation, we need to find the sum of pixels under white and black rectangles. To solve this, they introduced the

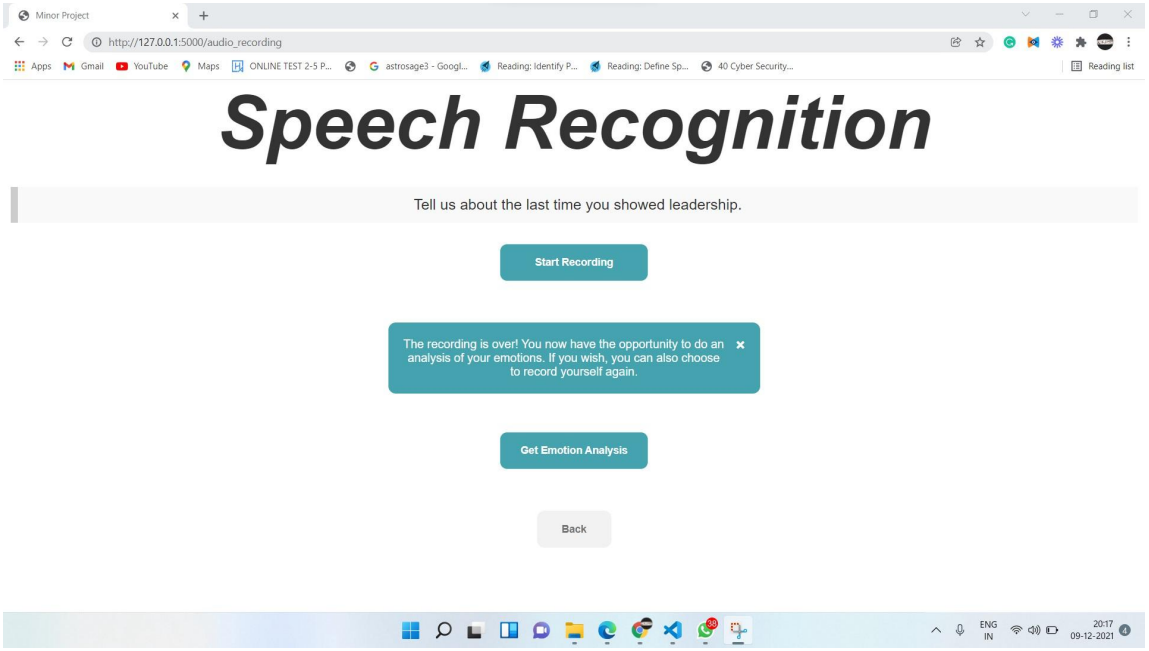
integral images. It simplifies calculation of sum of pixels, how large may be the number of pixels, to an operation involving just four pixels. Nice, isn‟t it? It makes things super-fast.

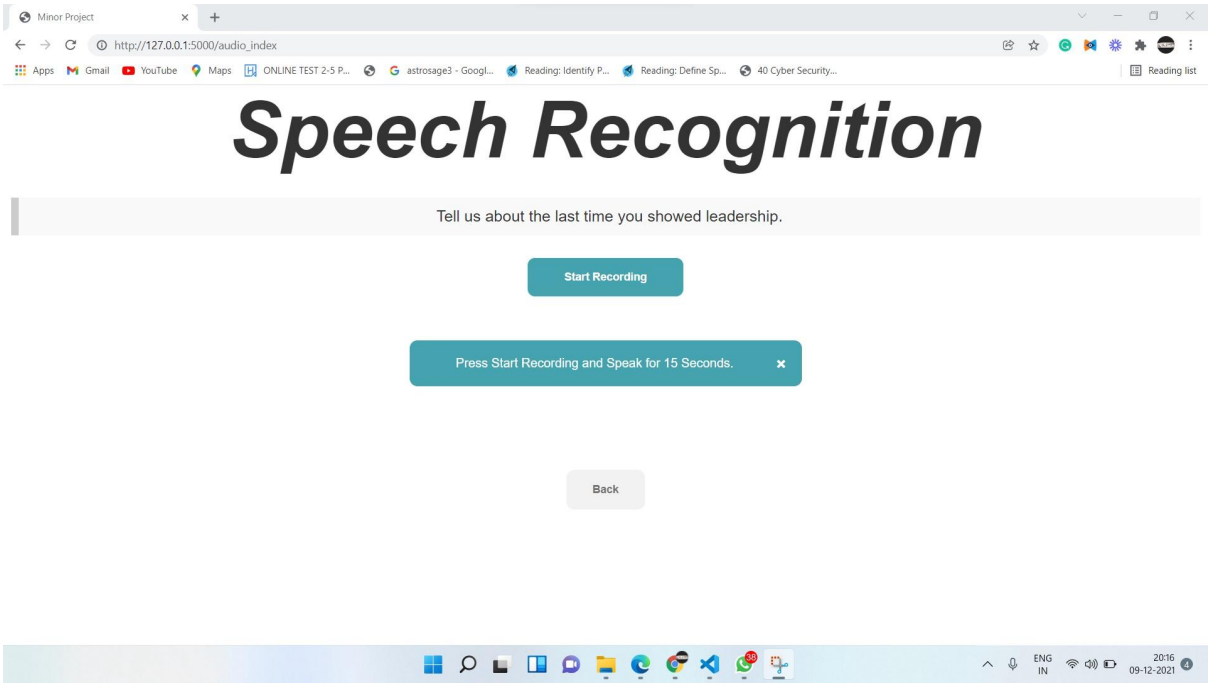
For this, we apply each and every feature on all the training images. For each feature, it finds the best threshold which will classify the faces to positive and negative. But obviously, there will be errors or misclassifications. We select the features with minimum error rate, which means they are the features that best classify the face and non-face images. (The process is not as simple as this. Each image is given an equal weight in the beginning. After each classification, weights of misclassified images are increased. Then again the same process is done. New error rates are calculated. Also new weights. The process is continued until required accuracy or error rate is achieved or required number of features are found).

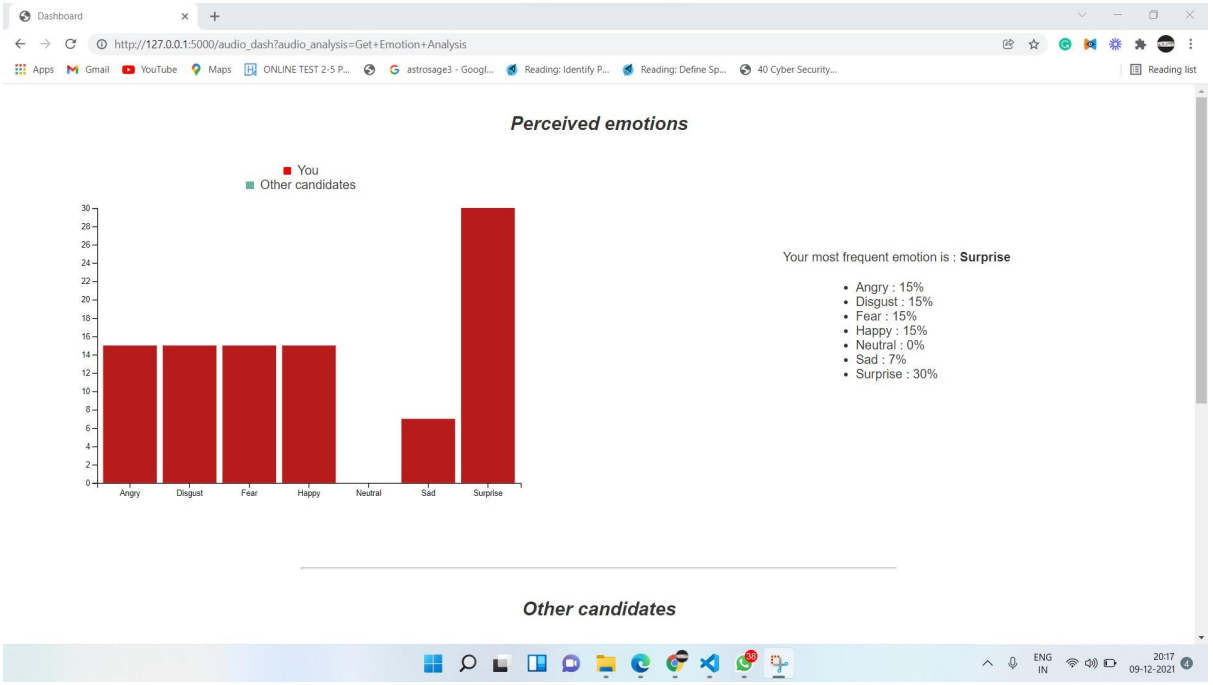
SOME SNAPSHOTS FROM OUR PROJECT

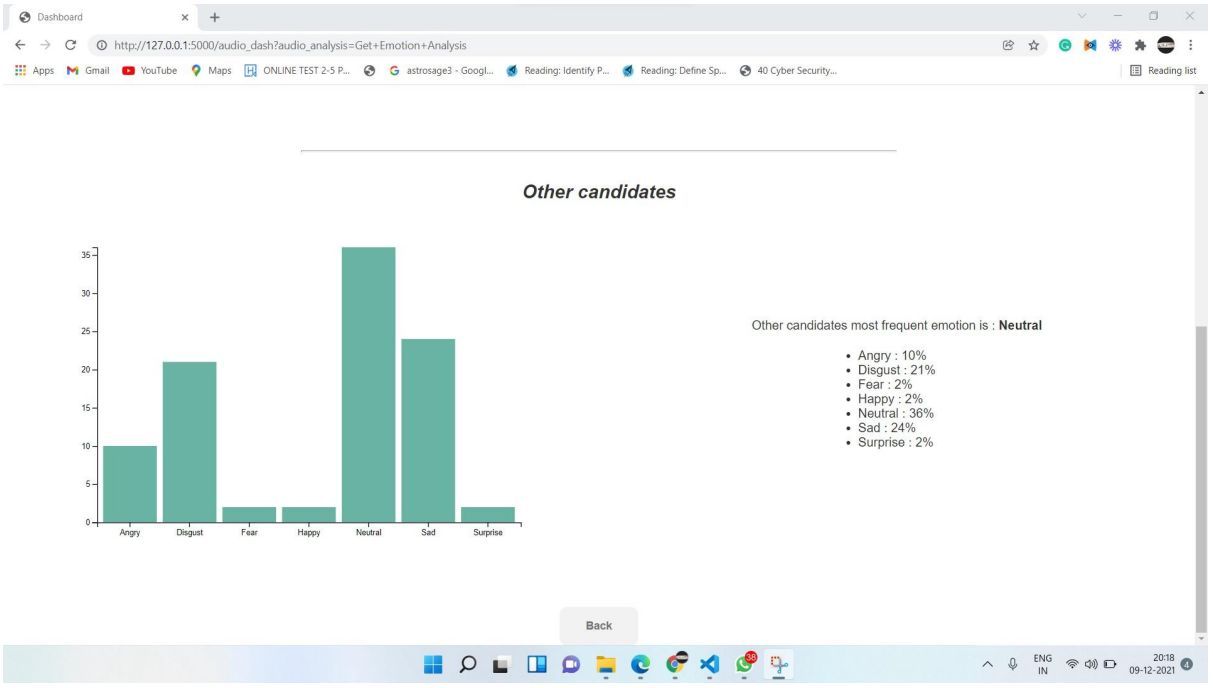
1. Audio



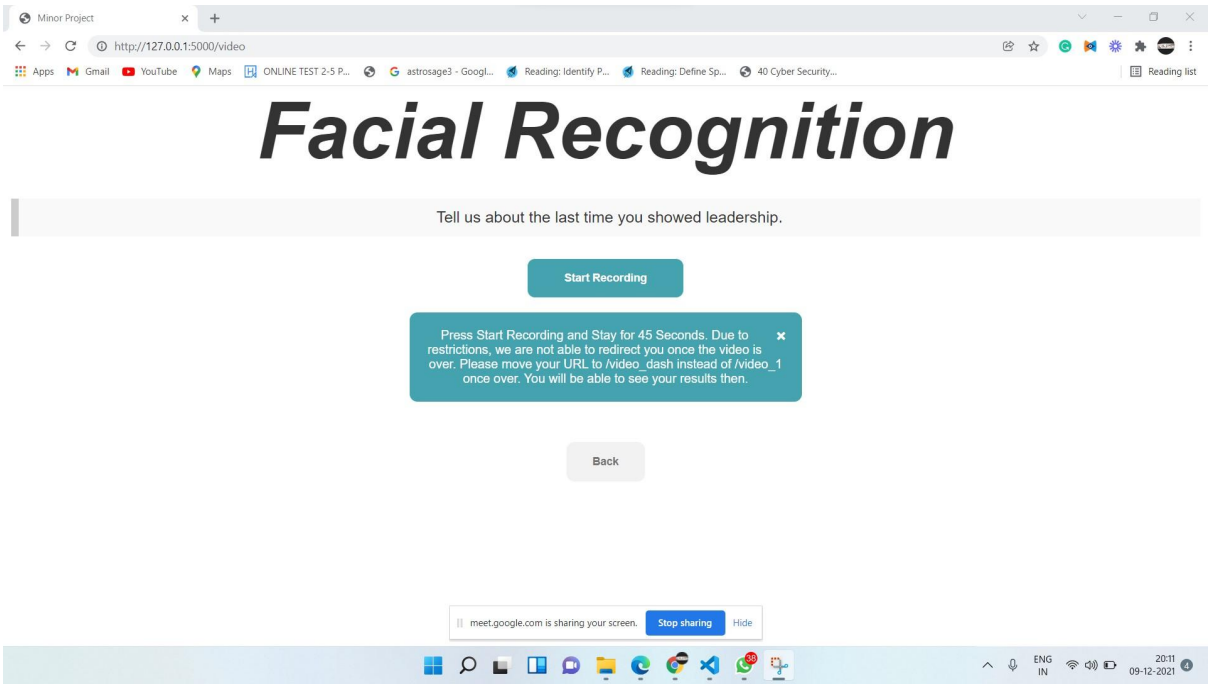

Minor Project
http://127.0.0.1:5000/audio_recording
Apps
Gmail
YouTube
Maps
ONLINE TEST 2-5 P...
astrosage3 - Googl...
Reading: Identify P...
Reading: Define Sp...
40 Cyber Security...
Reading list
Speech Recognition
Tell us about the last time you showed leadership.
Start Recording
The recording is over! You now have the opportunity to do an analysis of your emotions. If you wish, you can also choose to record yourself again.
Get Emotion Analysis
Back
ENG
IN
20:17
09-12-2021

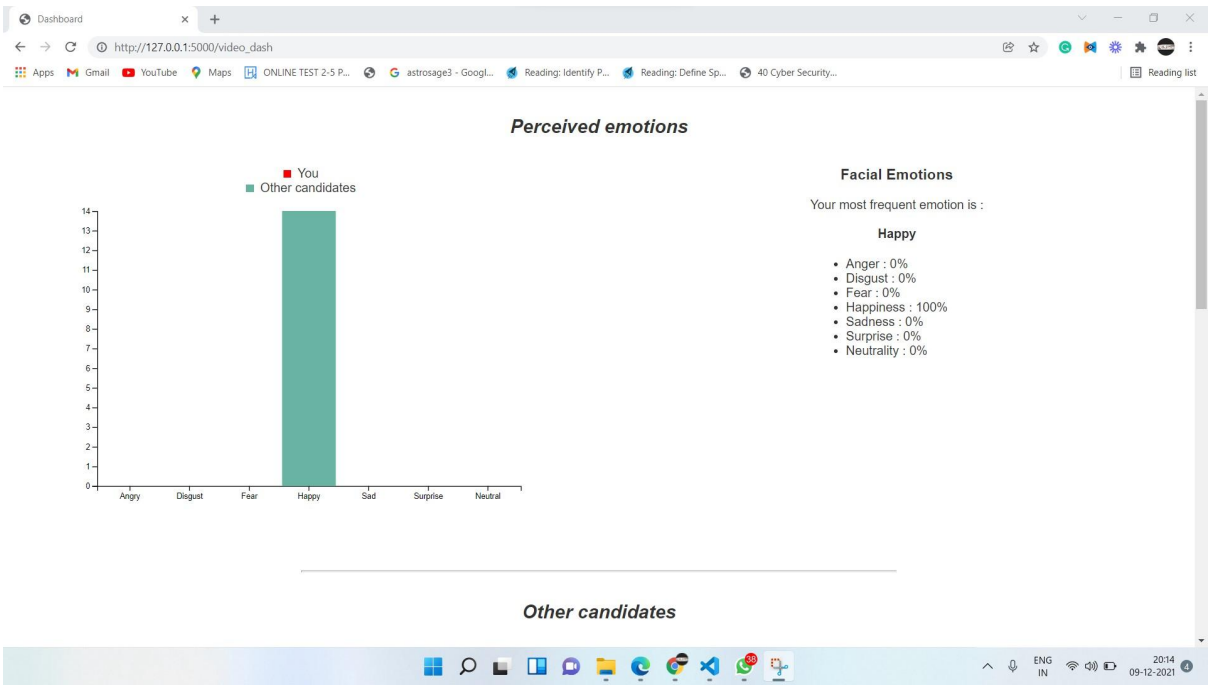

Minor Project
http://127.0.0.1:5000/audio_index
Apps
Gmail
YouTube
Maps
ONLINE TEST 2-5 P...
astrosage3 - Googl...
Reading: Identify P...
Reading: Define Sp...
40 Cyber Security...
Reading list
Speech Recognition
Tell us about the last time you showed leadership.
Start Recording
Press Start Recording and Speak for 15 Seconds.
Back
ENG
IN
20:16
09-12-2021

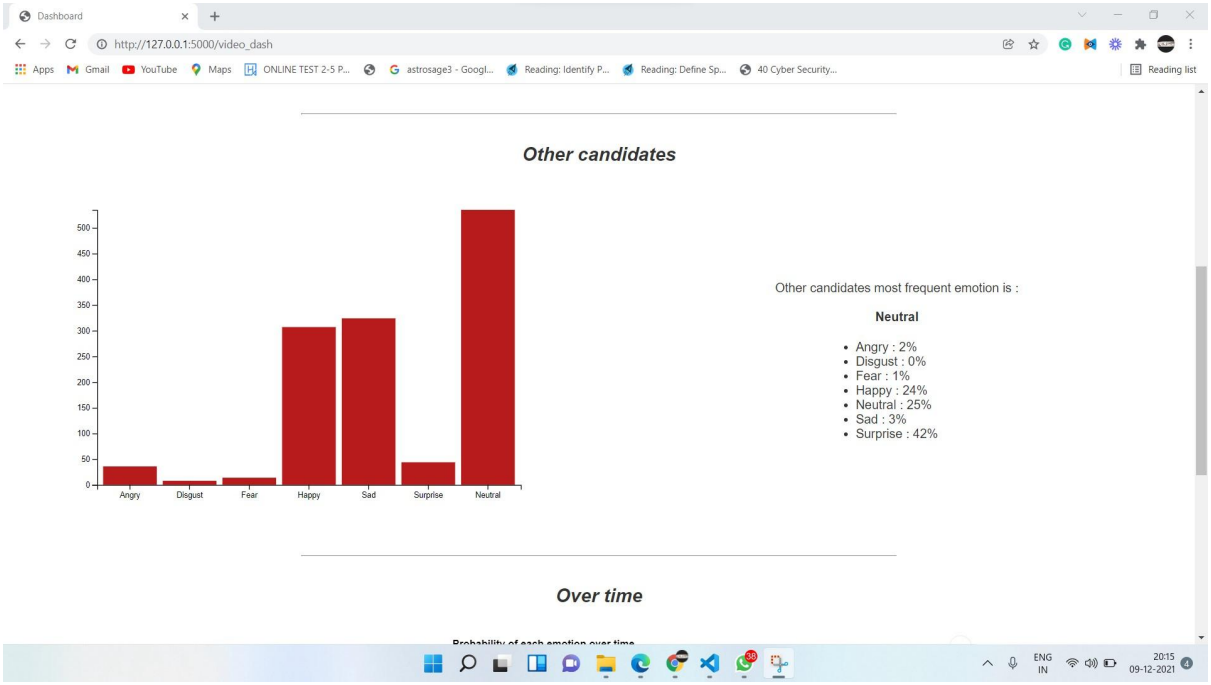

Dashboard
http://127.0.0.1:5000/audio_dash?audio_analysis=Get+Emotion+Analysis
Perceived emotions
You
Other candidates
Angry
Disgust
Fear
Happy
Neutral
Sad
Surprise
Your most frequent emotion is : Surprise
Angry : 15%
Disgust : 15%
Fear : 15%
Happy : 15%
Neutral : 0%
Sad : 7%
Surprise : 30%
Other candidates

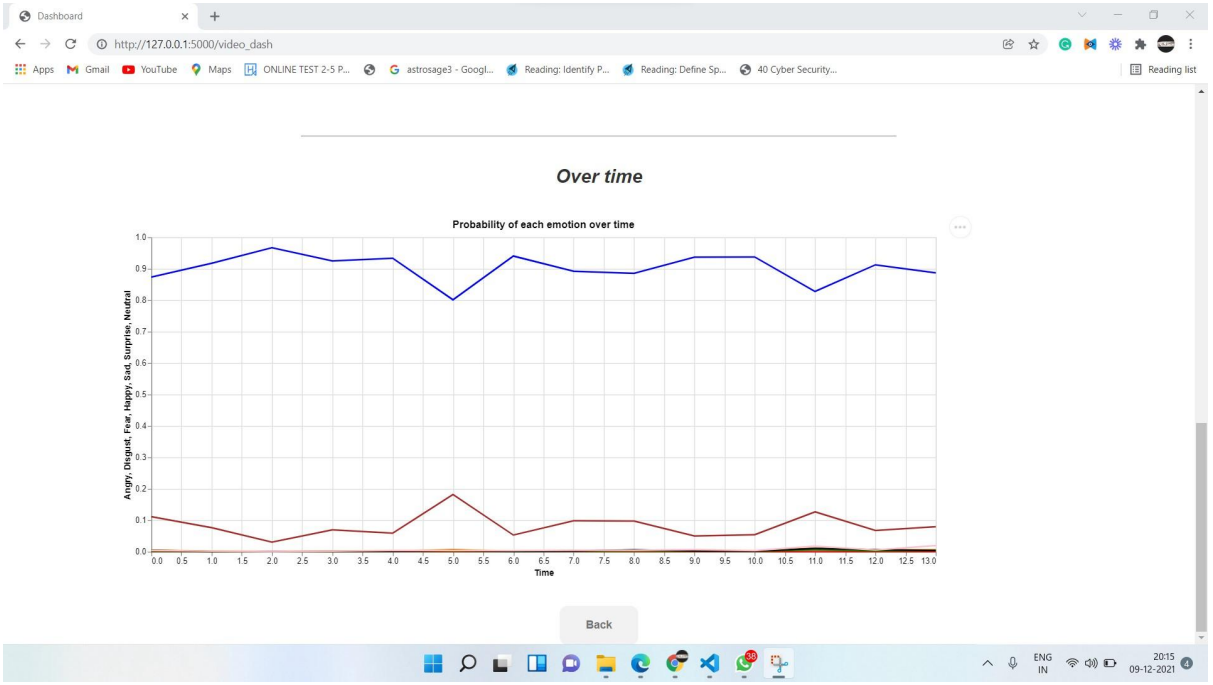

Dashboard
http://127.0.0.1:5000/audio_dash?audio_analysis=Get+Emotion+Analysis
Other candidates
Angry
Disgust
Fear
Happy
Neutral
Sad
Surprise
Other candidates most frequent emotion is : Neutral
Angry : 10%
Disgust : 21%
Fear : 2%
Happy : 2%
Neutral : 36%
Sad : 24%
Surprise : 2%
Back

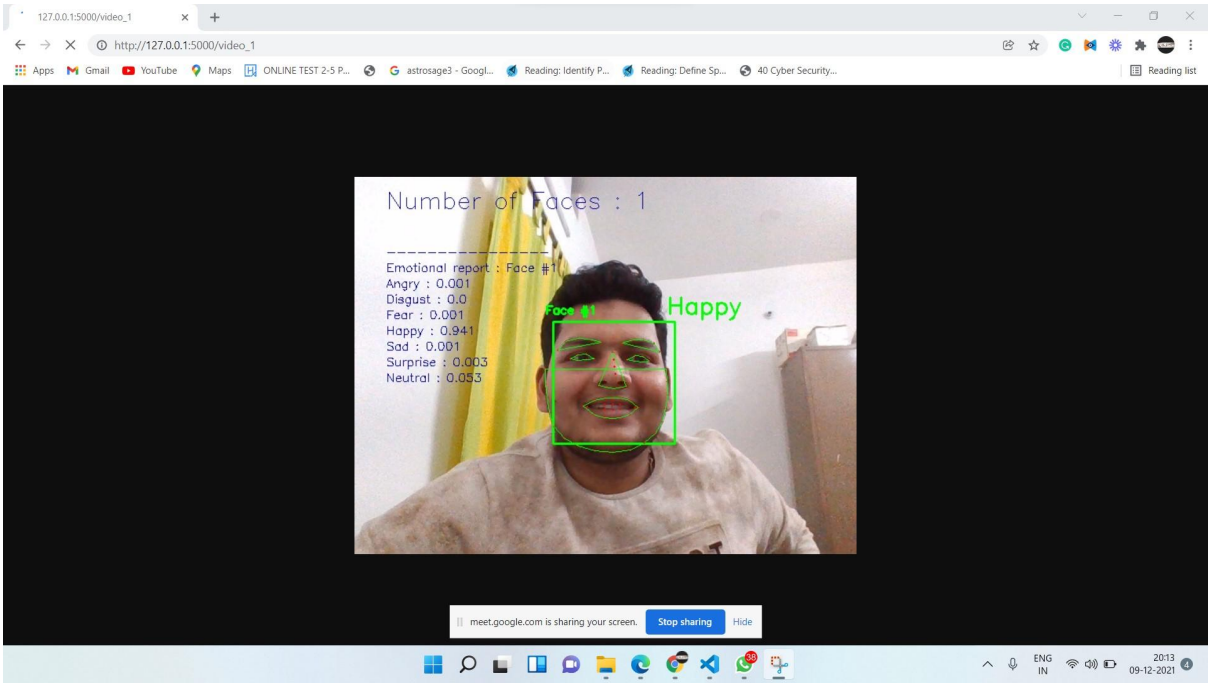
2- Facial


Minor Project
http://127.0.0.1:5000/video
Facial Recognition
Tell us about the last time you showed leadership.
Start Recording
Press Start Recording and Stay for 45 Seconds. Due to restrictions, we are not able to redirect you once the video is over. Please move your URL to /video_dash instead of /video_1 once over. You will be able to see your results then.
Back
meet.google.com is sharing your screen.
Stop sharing
Hide


Dashboard
http://127.0.0.1:5000/video_dash
Perceived emotions
You
Other candidates
Angry
Disgust
Fear
Happy
Sad
Surprise
Neutral
Facial Emotions
Your most frequent emotion is :
Happy
Anger : 0%
Disgust : 0%
Fear : 0%
Happiness : 100%
Sadness : 0%
Surprise : 0%
Neutrality : 0%
Other candidates


Dashboard
http://127.0.0.1:5000/video_dash
Other candidates
500
450
400
350
300
250
200
150
100
50
0
Angry
Disgust
Fear
Happy
Sad
Surprise
Neutral
Other candidates most frequent emotion is :
Neutral
Angry : 2%
Disgust : 0%
Fear : 1%
Happy : 24%
Neutral : 25%
Sad : 3%
Surprise : 42%
Over time


Dashboard
http://127.0.0.1:5000/video_dash
Over time
Probability of each emotion over time
Angry, Disgust, Fear, Happy, Sad, Surprise, Neutral
Time
Back


127.0.0.1:5000/video_1
http://127.0.0.1:5000/video_1
Number of Faces : 1
Emotional report : Face #1
Angry : 0.001
Disgust : 0.0
Fear : 0.001
Happy : 0.941
Sad : 0.001
Surprise : 0.003
Neutral : 0.053
Face #1
Happy
meet.google.com is sharing your screen.
Stop sharing
Hide

CONCLUSION AND FUTURE WORK

CONCLUSION

In this dissertation, a multi-sensory emotion recognition framework was proposed which mainly contains three parts: the speech emotion recognizer, the facial emotion recognizer, and information fusion. The proposed approach tries to distinguish between six basic emotions (anger, disgust, fear, happiness, sadness, and surprise) as well as the neutral state by integrating information from audio (speech) and video (facial expression) channels. The proposed framework can also do either speech emotion recognition or facial expression recognition individually.

The facial expression recognition system presented in this research work contributes a resilient face recognition model based on the mapping of behavioral characteristics with the physiological biometric characteristics. The physiological characteristics of the human face with relevance to various expressions such as happiness, sadness, fear, anger, surprise and disgust are associated with geometrical structures which are restored as base matching template for the recognition system. The behavioral aspect of this system relates the attitude behind different expressions as property base. The property bases are alienated as exposed and hidden categories in genetic algorithmic genes. The gene training set evaluates the

expressional uniqueness of individual faces and provides a resilient expressional recognition model in the field of biometric security. The design of a novel asymmetric cryptosystem based on biometrics having features like hierarchical group security eliminates the use of passwords and smart cards as opposed to earlier cryptosystems.. This research work promises a new direction of research in the field of asymmetric biometric cryptosystems which is highly desirable in order to get rid of passwords and smart cards completely. Experimental analysis and study show that the hierarchical security structures are effective in geometric shape identification for physiological traits.

One thing to be pointed out is that, though the result of facial emotion recognition is much higher than that of the speech, it can achieve an even higher result if the test instance is much closer to our requirement. Our requirement for the facial video is that at the beginning, the testing subject should have a neutral state for the initialization. However, most of the test instances do not satisfy this requirement, and we can only use the mean shape of the first frame (is the initialization threshold) of the test subject to initialize. Thus, a new audio-visual emotional database is needed to demonstrate the advantage of the proposed system in the future.

FUTURE SCOPE

Based on our research results in this dissertation, more opportunities for future research in this field can be done to extend this work. **First**, more emotional data samples can be collected from a large number of people, and these samples can be used in speech and facial emotion recognition for the training purpose or testing purpose and also to form an emotional database to convenient other researchers in this area. If such a database can be collected, it may become a milestone in the field of emotion recognition.

Second, it is possible to add feedback to the multi-sensory emotion recognition framework to improve the recognition accuracy by assigning a dynamic weight to the information fusion instead of using the fixed weight . Another possible extension is to integrate more modalities into the system such as body gesture, hand gesture, etc. Adding more modalities can provide more information about the emotional states of human beings, thus, may achieve better emotion recognition accuracy during human computer interaction. Finally, different acoustic features can be tested to improve the robustness of the base model in speech emotion recognition.

In this project we got an accuracy of almost 70% which is not bad at all compared to all the previous models. But we need to improve in specific areas like-

- number and configuration of convolutional layers
- number and configuration of dense layers
- dropout percentage in dense layers

But due to the lack of a highly configured system we could not go deeper into dense neural networks as the system gets very slow and we will try to improve in these areas in future. We would also like to train more databases into the system to make the model more and more accurate but again resources become a hindrance in the path and we also need to improve in several areas in future to resolve the errors and improve the accuracy. Having examined techniques to cope with expression variation, in future it may be investigated in more depth about the face classification problem and optimal fusion of color and depth information. Further study can be laid down in the direction of allele of gene matching to the geometric factors of the facial expressions. The genetic property evaluation framework for facial expression systems can be studied to suit the requirement of different security models such as criminal detection, governmental confidential security breaches etc.

REFERENCES

1. https://www.rcciit.org/students_projects/projects/it/2018/GR8.pdf
2. https://github.com/maelfabien/Multimodal-Emotion-Recognition/blob/master/README.md
3. https://towardsdatascience.com/building-a-speech-emotion-recognizer-using-python-4c1c7c89d713
4. https://github.com/maelfabien/Multimodal-Emotion-Recognition/blob/master/README.md#vii-research-paper
5. https://www.youtube.com/watch?v=Mubj_fqiAv8&list=PLeo1K3hjS3uu7CxAacxVndI4bE_o3BDtO
6. https://www.youtube.com/watch?v=fkgpvkqcoJc&t=1673s
7. https://pypi.org/project/deepface/