A Project Report on

ML Based Comprehensive Application to Enhance Soft Skills

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

Bachelor of Engineering

in

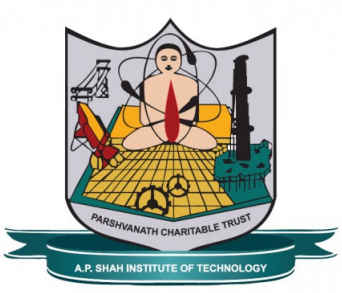
Information Technology

by

Sourav Joshi(19104068)
Akshay Bura(19104041)
Sakshi Deshpande(19104002)

Under the Guidance of

Mrs. Rujata Chaudhari



Department of Information Technology
NBA Accredited
A.P. Shah Institute of Technology
G.B.Road,Kasarvadavli, Thane(W), Mumbai-400615
UNIVERSITY OF MUMBAI

Academic Year 2022-2023

Approval Sheet

This Project Report entitled ***"ML Based Comprehensive Application to Enhance Soft Skills"*** Submitted by ***"Sourav Joshi"(19104068), "Akshay Bura"(19104041), "Sakshi Deshpande"(19104002)*** is approved for the partial fulfillment of the requirement for the award of the degree of ***Bachelor of Engineering*** in ***Information Technology*** from ***University of Mumbai***.

(Mrs. Rujata Chaudhari)
Guide

Dr. Kiran Deshpande
Head Department of Information Technology

Place: A. P. Shah Institute of Technology, Thane
Date:

CERTIFICATE

This is to certify that the project entitled ***"ML Based Comprehensive Application to Enhance Soft Skills"*** submitted by ***"Sourav Joshi" (19104068), "Akshay Bura" (19104041), "Sakshi Deshpande" (19104002)*** for the partial fulfillment of the requirement for award of a degree ***Bachelor of Engineering*** in ***Information Technology***, to the University of Mumbai, is a bonafide work carried out during academic year 2022-2023.

(Mrs. Rujata Chaudhari)
Guide

Dr. Kiran Deshpande
Head Department of Information Technology

Dr. Uttam D.Kolekar
Principal

External Examiner(s)

1.

2.

Place: A. P. Shah Institute of Technology, Thane
Date:

Acknowledgement

We have great pleasure in presenting the report on **ML Based Comprehensive Application to Enhance Soft Skills.** We take this opportunity to express our sincere thanks towards our guide **Mrs. Rujata Chaudhari** Department of Information Technology, APSIT thane for providing the technical guidelines and suggestions regarding line of work. We would like to express our gratitude towards her constant encouragement, support and guidance through the development of project.

We thank **Dr. Kiran B. Deshpande** Head of Department, Information Technology, APSIT for his encouragement during progress meeting and providing guidelines to write this report.

We thank **Prof. Sonal Jain** BE project co-ordinator, Department of IT, APSIT for being encouraging throughout the course and for guidance.

We also thank the entire staff of APSIT for their invaluable help rendered during the course of this work. We wish to express our deep gratitude towards all our colleagues of APSIT for their encouragement.

Sourav Joshi
19104068

Akshay Bura
19104041

Sakshi Deshpande
19104002

Declaration

We declare that this written submission represents our ideas in our own words and where others' ideas or words have been included, We have adequately cited and referenced the original sources. We also declare that We have adhered to all principles of academic honesty and integrity and have not misrepresented or fabricated or falsified any idea/data/fact/source in our submission. We understand that any violation of the above will be cause for disciplinary action by the Institute and can also evoke penal action from the sources which have thus not been properly cited or from whom proper permission has not been taken when needed.

_____

Sourav Joshi (19104068)

_____

Akshay Bura (19104041)

_____

Sakshi Deshpande (19104002)

Date:

Abstract

In today's world, soft skills are as important as technical skills, and students need to master them to succeed in the job market and in their personal lives. However, many students struggle with expressing themselves and communicating effectively in group settings, even if they possess a lot of knowledge. Inadequate communication skills may lead to missed opportunities and negatively impact one's success. To address this problem, we are developing a comprehensive application that will assist students in improving their soft skills and build confidence to speak up which has not been addressed by an application till now. The goal of this application is to help students to overcome their fear of speaking and to enhance their communication skills in a variety of settings. Social anxiety is often considered a lifelong condition that cannot be cured, but we believe that there are ways to help students build their confidence and overcome their anxiety. We are exploring state-of-the-art models that extract emotions and body postures to help students gain confidence. By analyzing sound and video inputs, we can measure the student's confidence level and assign them various activities such as logical game to enhance the soft skills and boost the confidence. Our application will provide a personalized learning experience that will enable students to identify their areas of improvement and focus on enhancing their skills in those areas. We aim that this application will be a valuable tool for students looking to improve their soft skills and build confidence in themselves, and empower them to succeed both in their personal and professional lives.

Contents

List of Figures

List of Tables

List of Abbreviations

ML:	Machine Learning
CNN:	Convolutional Neural Network
NLP:	Natural language processing
Opencv:	Open Source Computer Vision Library
LSTM:	Long Short-term Memory
ANN:	Artificial Neural Networks
SVM:	Support Vector Machine
MER:	Multi-model emotion recognition
ADAM:	Anxiety Detection And Management
HMM:	Hidden Markov Model
SER:	Speech emotion recognition
DNN:	Deep neural networks
GUI:	Graphical user interface
API:	Application programming interface
IQ:	Intelligence Quotient
EQ:	Emotional Quotient

Chapter 1

Introduction

These days, many students lack confidence and verbal ability in social environments like placement drives, which may affect their job interviews. It is not always possible to hire counselors in rural areas. Even if we have abundant sources and mediums to only detect lack of self-confidence, most of them don't help in building confidence.From past 2 years due to Covid-19 and online college verbal skills are lacking. To enhance these skills we will be creating an application which can be used free of cost in rural areas too. Predicting confidence of the personal is a challenging task for Machine learning. Getting accurate results may not be possible by Machine learning.

For this Project we are using the Ryerson Audio-Visual Database of Emotional Speech and Song (RAVDESS). The database consists of 24 professional actors (12 male,12 female), vocalizing lexically-matched statements in an accent of neutral North American. Speech includes happy, fearful, sad, surprise, calm, angry, and disgust expressions, and song contains happy, calm, sad, fearful, and angry emotions[3].The main idea of a Time Distributed Convolutional Neural Network is to apply a rolling window (fixed size and time-step) all along the log-mel-spectrogram[1].

Our main motive is to develop a model that will be able to provide a live sentiment analysis with a visual user interface. The convolutional neural network (CNN) [6] will give most accurate result and It will be easy to calculate the confidence of the person. For the video we are Implementing the XCeption model. The XCeption architecture is based on DepthWise Separable convolutions that allows us to train much fewer parameters, and hence it reduces training time on Collab's GPUs to less than 90 minutes. So we focus on deep learning Artificial Neural Network(ANN), Convolution Neural Network (CNN) and Long Short-Term Memory (LSTM)[2].In this project we are developing Convolution Neural Network (CNN).[1] The challenge lies in effectively interpreting and communicating the results, which can be complex and difficult to understand. To overcome this challenge, we're developing an innovative solution that incorporates advanced visualization techniques and explainable AI to facilitate clear and concise explanations of the CNN's output. We can plot class activation maps, which display the pixels that have been activated by the last convolution layer. We see how pixels are activated differently based on the tagged emotion. The expression of emotions can be attributed to specific facial features. For instance, happiness appears to be linked to the pixels around the eyes and mouth, while sadness or anger is commonly associated with the eyebrows. By examining these subtle variations in facial expressions, we can gain a deeper understanding of an individual's emotional state and respond accordingly.

1.1 Purpose

The purpose of creating an ML-based comprehensive application to enhance users' soft skills is to provide individuals with a tool that can help them improve their communication, leadership, teamwork, and other interpersonal skills. Soft skills are critical for success in the workplace and in life, and improving them can lead to better job performance, more effective communication, and stronger relationships with others.

The ML-based application can leverage advanced machine learning algorithms like natural language processing to analyze user communication patterns, identify areas for improvement, based on the communication style, tone, and word choice. This can help the application leading to more effective and efficient skill development.

The comprehensive nature of the application can also allow users to track their progress over time, and receive ongoing support and guidance. By providing users with a tool that is both personalized and comprehensive, the application can help individuals develop their soft skills more effectively and efficiently.

Overall, the purpose of creating an ML-based comprehensive application to enhance users' soft skills is to provide individuals with a powerful tool for personal and professional growth. By leveraging the latest advances in machine learning, the application can help users to improve their communication, leadership, teamwork, and other critical soft skills, leading to greater success and fulfillment in all areas of life.

1.2 Objectives

We intend to do this project implementation to meet following objectives

- To automate Xpression Club (Xpression Club is an initiative which is made by our organisation to boost the confidence level of students).
- To Improve logical thinking and verbal ability of the students using various games and sessions for guiding them.
- To predict the confidence level of students using ML Algorithm i.e. NLP and CNN and help them in improving it.
- To create an user friendly UI which contains various activities, mentoring sessions and games using Flutter which will be connected to cloud for storing the data.
- To perform feature extraction on the video using media pipe using which model will identify the posture of the user.

1.3 Problem Definition

- Problem Identified
  Nowadays students are facing lack of confidence while speaking to others. Most of the apps on market do not provide a solution to anxiety and verbal ability. The apps that do try are costly and their free versions don't have the features. They also contain lots of ads. Most of the apps use a Support Vector Machine (SVM) algorithm

to predict the confidence of the person. The SVM algorithm's accuracy is low, and this may lead to false predictions that can lower the self-esteem of the user. A user may need a personal mentor to guide them, track down their progress, and solve their problems. Also, students don't get enough mental attention, so they become anxious and don't talk with other people due to anxiety and low self-esteem. Apps in the market only do activities that are related to breathing and meditating, which solve problem temporary.They also charge a couple of dollars in the form of subscriptions.

- Solution Proposed
  Our Application provide a solution to anxiety and verbal ability. This app provides all the features free of cost. We will be using Convolutional Neural Network (CNN) algorithm to predict the confidence of the person. We are getting good accuracy using CNN algorithm. As user may need a personal mentor to guide them, track down their progress, and solve their problems, we will be having a separate credentials for mentor to access student details in our application. Also, students will get enough mental attention, so they can overcome anxiousness which will help them in boosting confidence and can speak with other people openly.

1.4 Scope

Many companies notice employers who got placed in their company are good in technical skills but lack in communication skills. So to overcome this problem our institute had started an activity named Xpression Club. Our project automates the process of Xpression Club and also it can be useful i rural areas where psychiatrist can not be reached.

Chapter 2

Literature Review

The purpose of literature review is to gain an understanding of the existing research on Confidence Enhancing and debates relevant to the area of study. The literature review helped in selecting appropriate algorithms and suitable feature extraction processes for getting efficient results.

According to a research paper by Gou Wei and et al [1], speech emotion recognition can be useful for determining the confidence and mood of the speaker. The researchers describe how emotions are analyzed and classified using audio features and signal processing techniques. With these methods, they were able to achieve an accuracy of around 76%. The model they used is based on a discrete emotion classification system that includes labels such as happy, sad, angry, disgusted, fear, and surprise. For this particular project, only the emotions of fear and surprise are needed. By accurately identifying these emotions in speech, the technology can provide valuable insights into the speaker's emotional state.

Research paper by Sabrina Begaj and et al [2], compares several classification algorithms to determine which algorithm provides most accurate results and was able to quickly predict labels. They used the Multimodal Emotion Recognition (MER) Dataset for their analysis. After comparing various combinations of algorithms and frame-level features, they found that the best accuracy was achieved using a Convolutional Neural Network (CNN), with an accuracy of around 50%. Based on these results, the authors decided to use a CNN for their emotion recognition model.

Paper entitled "BLSTM-Based Confidence Estimation for End-to-End Speech Recognition" [3], provides a research on confidence in speech and how this can be useful in different situations.For an instance author of this paper uses an example of an online meeting to calculate confidence metrics in speech of an employee. People can also use this to rehearse their speech and can receive live feedback based on their confidence level and also to improve those scores one can adapt their speech. These tools can also help the one with disabilities like autism perceive confidence, which might be crucial in situations like high-stakes.

Mannapperuma and et al [4], justifies that the main motive of building this application was to help the employees of Sri-Lanka's IT Industry in managing their anxiety. Anxiety is a huge burden for a large number of employees in workplaces around the globe, Which is also a challenge for employees in achieving their goals and also increasing their productivity. Due to lack Of applications related to mental health, Mental Illness is considered as a social stigma in Sri-Lanka. The author has created an application named ADAM (Anxiety Detection and Management), this application will help employees facing anxiety in realizing

their severity and also helps in following self managing activities and also encourages them in achieving professional goals. Appropriate anxiety-management activities were allocated using a Rulebased machine learning model for employees based on different parameters like age, gender and severity.

According to Kasthuri and et al [5], In this pandemic situation, everything has become online including Education. There's a lot of drawbacks in online education, one of the main drawbacks of this system is the interaction between students and teachers which has been quite low. Chatbots have become a popular means for students to study and clear their doubts. With the ability to provide support anytime and anywhere, chatbots offer great convenience to students. Students can text chat to ask a question, which will then be answered using natural language processing and deep learning technology.

"Real-time video emotion recognition based on reinforcement learning and domain knowledge" [6], provides information on how automation is important nowadays. Automating the process of the interview using machine learning is also a trending application which is used widely. Researchers find that using Natural Language Processing and deep learning technique systems can be automated and help in predicting accurate results. System predicts the result and gives candidate scores accordingly. System reduces the efforts of the HR and helps in selecting the suitable/applicable candidates. We can use this technique to automate the Xpression club by understanding the workflow of the system. Voice signal is processed using NLP and uses Hidden Markov Model (HMM) for the recognition of the voice.

According to Liu Wei and et al [7], Multimodal emotion systems are increasing rapidly in year due to automation of the interviews. System can predict the performance of the candidate by detecting emotions and the answers given by the candidate. By applying the algorithms on audio-video, the accuracy system can easily predict the analysis. To achieve real-time emotion recognition capabilities, many models that aim to extract the context from video are developed using the entire dialogues. However, this approach often leads to a lack of real-time ERC ability The proposed erldk model utilizes a combination of reinforcement learning , domain understanding, which distinguishes it from previous research in this field. By incorporating extracted domain knowledge, the ERLDK model can refine the results generated by the RL module. Additionally, the model's performance is evaluated on a separate dataset to confirm its efficiency.

According to Ryota Sato and et al[8], One of the most recent challenges in human-computer interaction is SER, or speech emotion recognition. A single emotion label is outputted each syllable as the estimation result in traditional SER classification algorithms. This is due to the fact that SER models are trained using traditional speech emotional databases, which only have one emotion label per utterance. But with human speech, it's common for numerous emotions to be communicated at once and in varying degrees of intensity. The presence of many emotions in a one syllable needs to be taken into consideration in order to achieve more natural SER than ever. As a result, we developed an emotional speech database that includes labels for various emotions and their varying degrees of intensity. The creation experiment was carried out by removing emotional spoken utterance segments from pre-existing video works.

According to Lotfian and et al [9],"Curriculum learning for speech emotion recognition from crowdsourced label" provides a solution to the challenge of establishing a natural order of difficulty in the training set for speech emotion recognition. The method uses the dis-

agreement between evaluators as a measure of difficulty for the classification task and defines the curriculum based on inter-evaluation agreement. The experimental results show that the proposed method leads to statistically significant improvements over baselines trained without a curriculum. The proposed method can be applied to different datasets and different DNN architectures, making it a useful tool in the field of speech emotion recognition.

Paper entitled "3-D convolutional recurrent neural networks with attention model for speech emotion recognition" [10], addresses the issue of emotionally irrelevant factors and achieves state-of-the-art performance in SER. The method uses personalized features calculated by deltas and delta-deltas and a three-dimensional attention-based convolutional recurrent neural network to learn discriminative features for SER. The attention mechanism in the network learns relevant feature representations for specific tasks, which can reduce the influence of emotionally irrelevant frames and silent frames. The experimental results demonstrate the effectiveness of the proposed method in reducing misclassification and achieving state-of-the-art performance in SER.

Paper entitled "Speech emotion recognition using deep learning techniques: A review." [11], provides an overview of deep learning techniques and discusses recent literature where these methods are utilized for speech-based emotion recognition. The review covers databases used, emotions extracted, contributions made toward speech emotion recognition, and limitations related to it. The reviewed literature demonstrates the effectiveness of deep learning techniques in improving the performance of SER and proposes new methods to handle the variability in speech signals. The limitations related to deep learning techniques in SER are also highlighted in the paper.

S.K Tiwari and et al [12], proposes a system that includes face mask detection and an automatic attendance system. The system aims to keep track of the users by using a novel proposed algorithm. The system's flow begins with collecting pre-integrated data and the system's newly generated data, followed by face mask recognition, face recognition, causing the datasheet and notifying the administrator. The advanced techniques used in this system include Graphic User Interference (GUI), a deep learning technique to extract targeted class features, Open CV to execute real-time operations, speech instructions, and the MobileNetV2 approach.

Chapter 3

Proposed System

The purpose of creating an ML-based comprehensive application to enhance users' soft skills is to provide individuals with a tool that can help them improve their communication, leadership, teamwork, and other interpersonal skills. Soft skills are critical for success in the workplace and in life, and improving them can lead to better job performance, more effective communication, and stronger relationships with others. Functions of the Proposed System are :

- User Authentication: The initial step is to register and login since students and mentor both will be the one to use the application.

- Attempt Quiz: Once the user has logged in, the next step is to attempt the quiz. This involves various questions based on the verbal and logical thinking. The user will be getting score based the correct answers and these scores are stored in cloud for further use.

- Upload Video: The user will be presented with a screen to upload a video, which will then be converted into base64 format. The base64-encoded video will be sent to a Google Cloud Function, which will process the video using advanced algorithms and machine learning techniques. This approach allows for seamless integration between the user interface and the backend processing, providing a smooth and efficient user experience.

- Video Analysis: After uploading a video's base64 format to the cloud, it can be converted to a standard video format for analysis. Landmark-based posture identification can then be performed using libraries such as MediaPipe.

- Speech analysis: The next speech will be processed using natural language processing (NLP) algorithms to extract the spoken text from the video. The text will then be analyzed for grammar and syntax using NLP techniques, and a score will be generated based on the accuracy and complexity of the language used in the speech. This score can provide valuable insights into the quality and effectiveness of the speech, as well as identify areas where the speaker may need to improve their language skills.

- Activites: Once the video has been processed and analyzed, the user will be presented with a range of interactive activities designed to improve their soft skills. These activities may include memory games, logical puzzles, and verbal games, which are specifically designed to help users enhance their cognitive abilities and communication skills

- Certificate: Upon successful completion of the various activities and consistently achieving good scores, the user will be awarded a certificate acknowledging their proficiency in the targeted soft skills.

The above mentioned functions have been implemented using following technologies :

- NLP - Natural Language Processing focuses on teaching computers to understand, interpret, and generate human language.NLP is used in a wide range of applications, including machine translation, sentiment analysis, chatbots, speech recognition, and more. We had used this algorithm for Speech Recognition.

- CNN - Convolutional Neural Network, is a type of artificial neural network designed to process and analyze data with a grid-like structure, such as images and videos. One of the key advantages of CNNs is their ability to automatically learn relevant features from raw data, reducing the need for manual feature engineering. This makes them well-suited for tasks such as image recognition, object detection, and natural language processing. We had used this algorithm for Emotion Recognition.

- Mediapipe - Mediapipe is an open-source cross-platform framework developed by Google for building real-time, data-driven applications and machine learning pipelines. It provides a rich set of easy-to-use, customizable, and pre-built components for processing multimedia data, such as images, video, and audio, and extracting valuable information, such as facial landmarks, hand and body poses, object detection, and semantic segmentation. We had used this framework for Posture Detection.

- Flutter - Flutter is a free and open-source mobile application development framework created by Google. It allows developers to build natively compiled apps for mobile, web, and desktop from a single codebase. Flutter uses the Dart programming language and provides a rich set of pre-built widgets and tools for building beautiful and high-performance user interfaces. It also has a "hot reload" feature that allows for fast development and iteration, making it a popular choice for building cross-platform apps. We had this framework for building a comprehensive application.

Chapter 4

Project Design

The project's key features, structure, criteria for success, and major deliverables are all planned out in this steps. The aim is to develop design in a way so that it can differ from existing system that can be used to achieve the desired project goals.

4.1 System Diagram

4.1.1 Use Case Diagram

Use-case diagrams describe the high-level functions and scope of a system. These diagrams also identify the interactions between the system and its actors. The use cases and actors in use-case diagrams describe what the system does and how the actors use it, but not how the system operates internally.

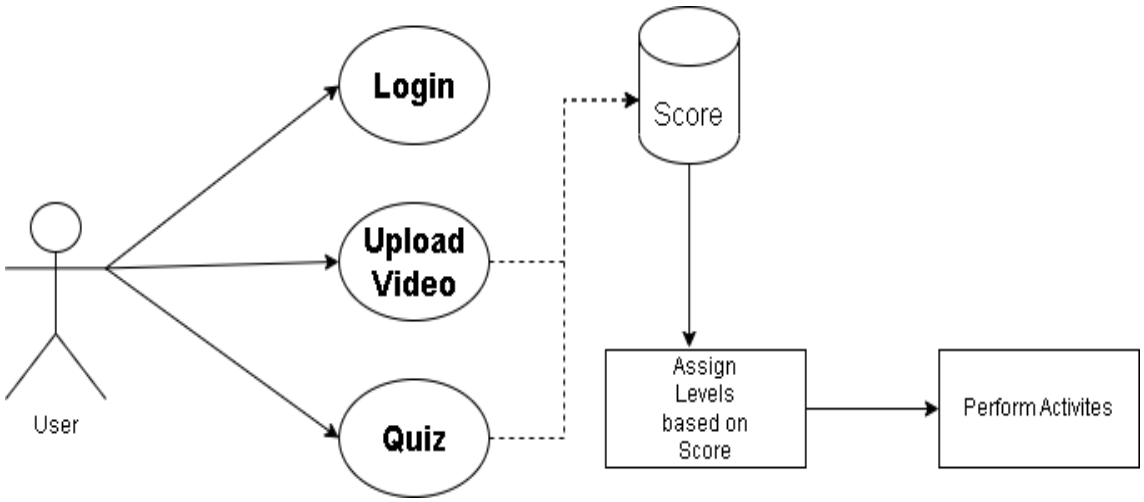


Figure 4.1: Use case Diagram

The use case diagram will show how the user will interact with the application. The user has three roles: to log in, to upload videos, and to attempt quizzes. The score will be stored in the backend, and based on the score, the user will perform activities.

4.1.2 Activity Diagram

An activity diagram is a type of UML diagram commonly used in IT to model the flow of activities and actions in a system or process. It provides a visual representation of the steps involved in performing a specific function, such as a use case scenario or business process. Activity diagrams are useful for analyzing complex systems and identifying potential issues or bottlenecks in the process. They can also be used to communicate the process flow to stakeholders and development teams, making it a valuable tool for system design and development.

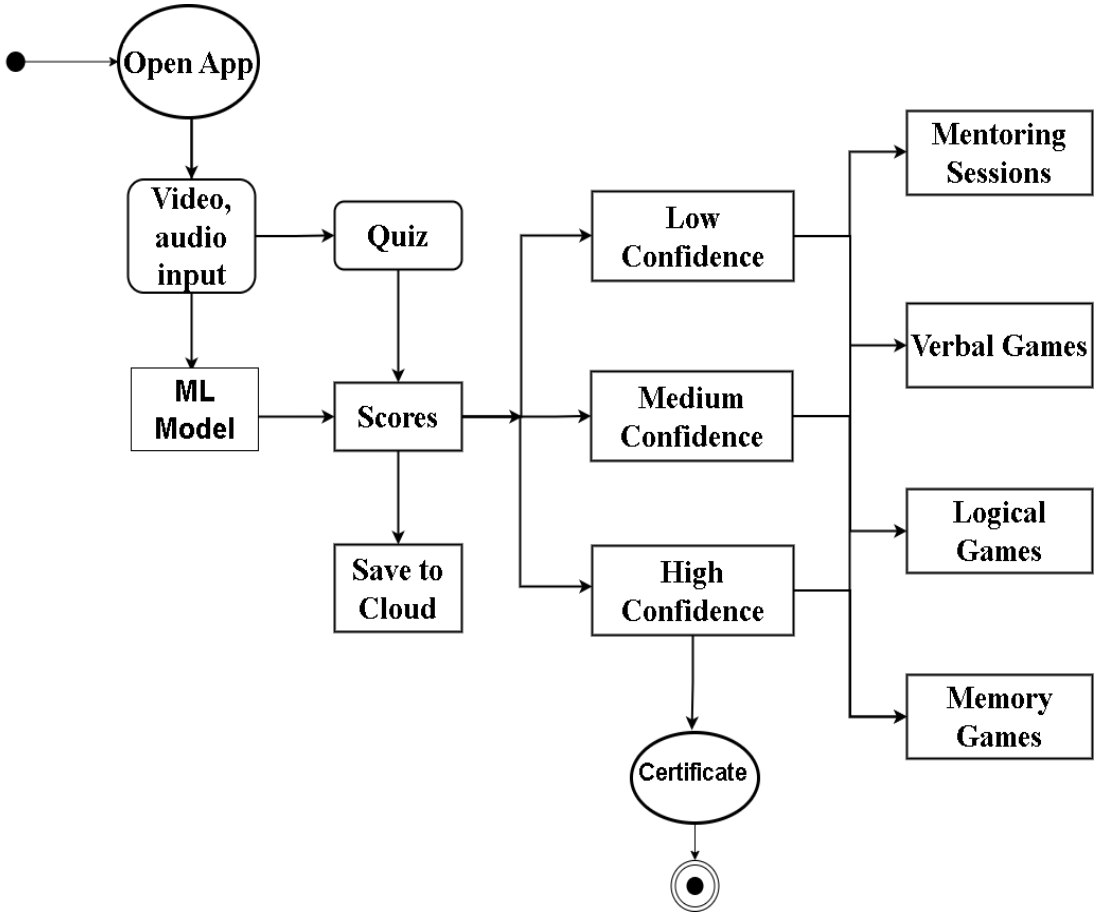


Figure 4.2: Activity Diagram

The proposed system is a complete system diagram that involves students/users downloading the application from play-store, signing up, uploading a video, and taking a quiz. Scores generated by a trained ML model will be stored in the cloud and used to assign users to three confidence levels: Low, Medium, and High. Each level will have various activities, including logical, verbal, and memory games, to enhance users' soft skills. Users will also receive guidance from a mentor to improve their communication skills. High Confidence users will receive a certificate upon completion.

4.1.3 Sequence Diagram

A sequence diagram is a type of UML diagram used to show the sequence of interactions between objects or components in a system or process. It helps developers and stakeholders visualize the flow of information and events between different parts of a system, which can aid in identifying potential problems and optimizing system performance.

The sequence diagram shows the entire sequence of flow of the data in the system.First we will gather all the required data and train and test them accordingly. For Posture we have used media-pipe and for the audio we have used NLP and CNN. When user adds the video, Model will separate audio and video. Based on Separated video it will identify the posture of the person using CNN and by using NLP on the audio model will tell the confidence level of the user. Using these two parameters model will give score which will be stored in a database.

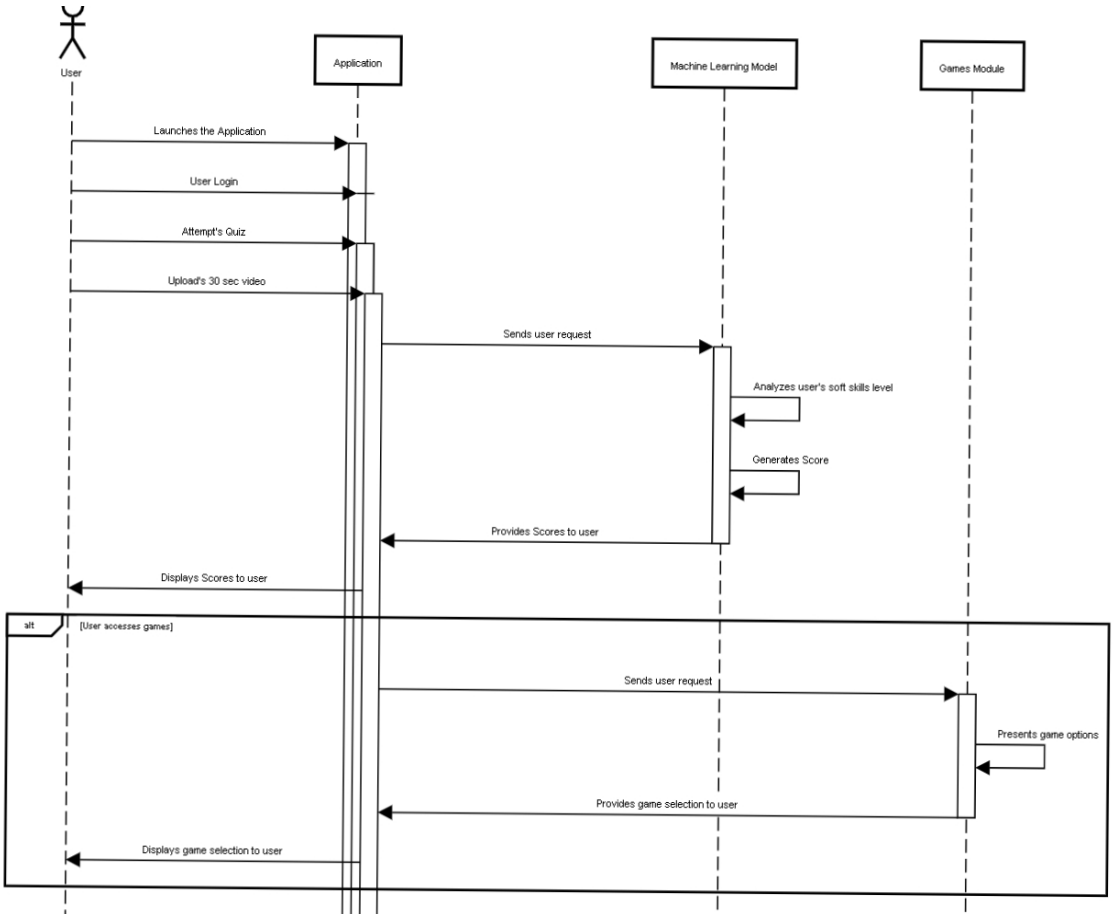


Figure 4.3: Sequence Diagram

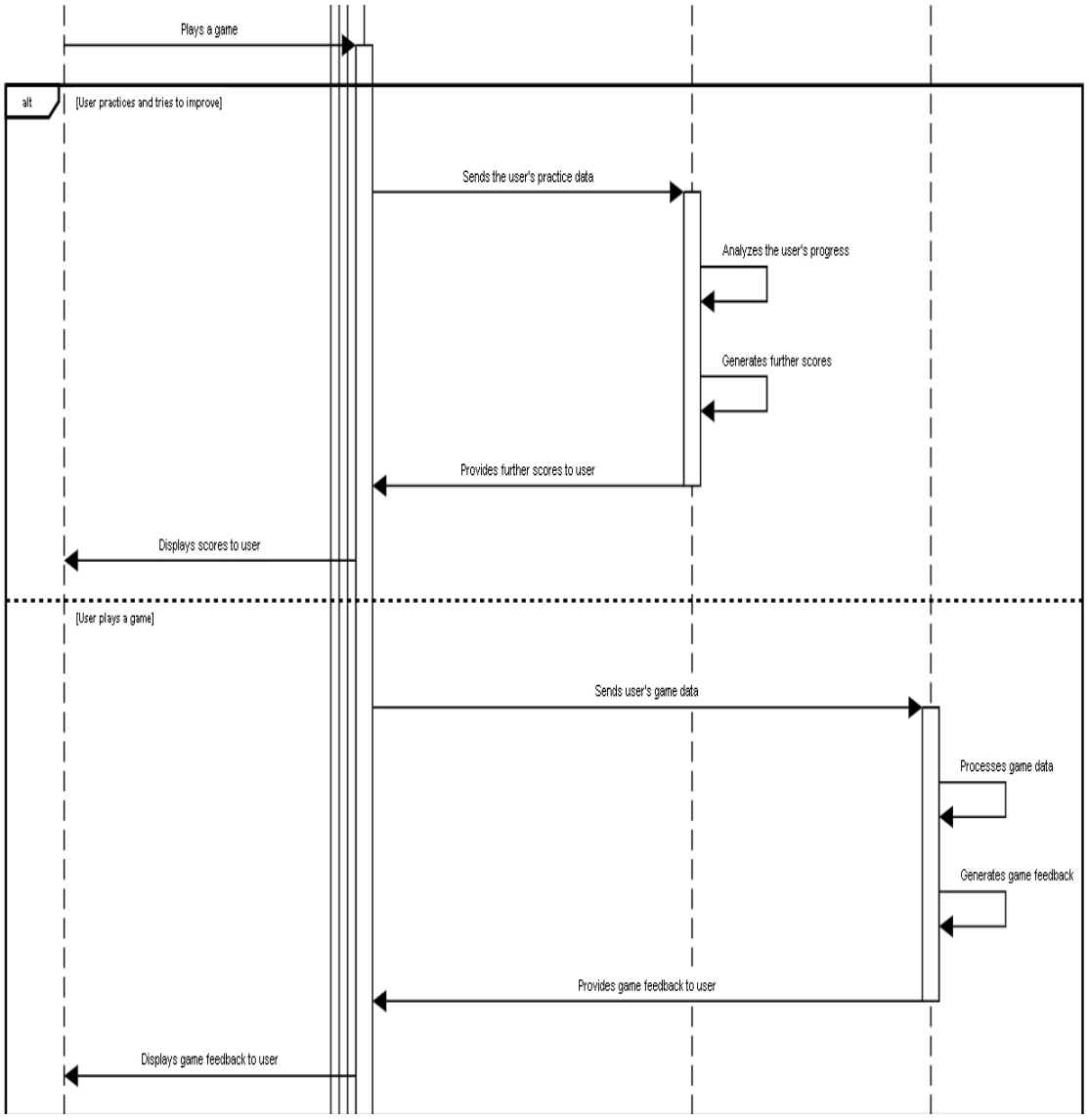


Figure 4.4: Sequence Diagram

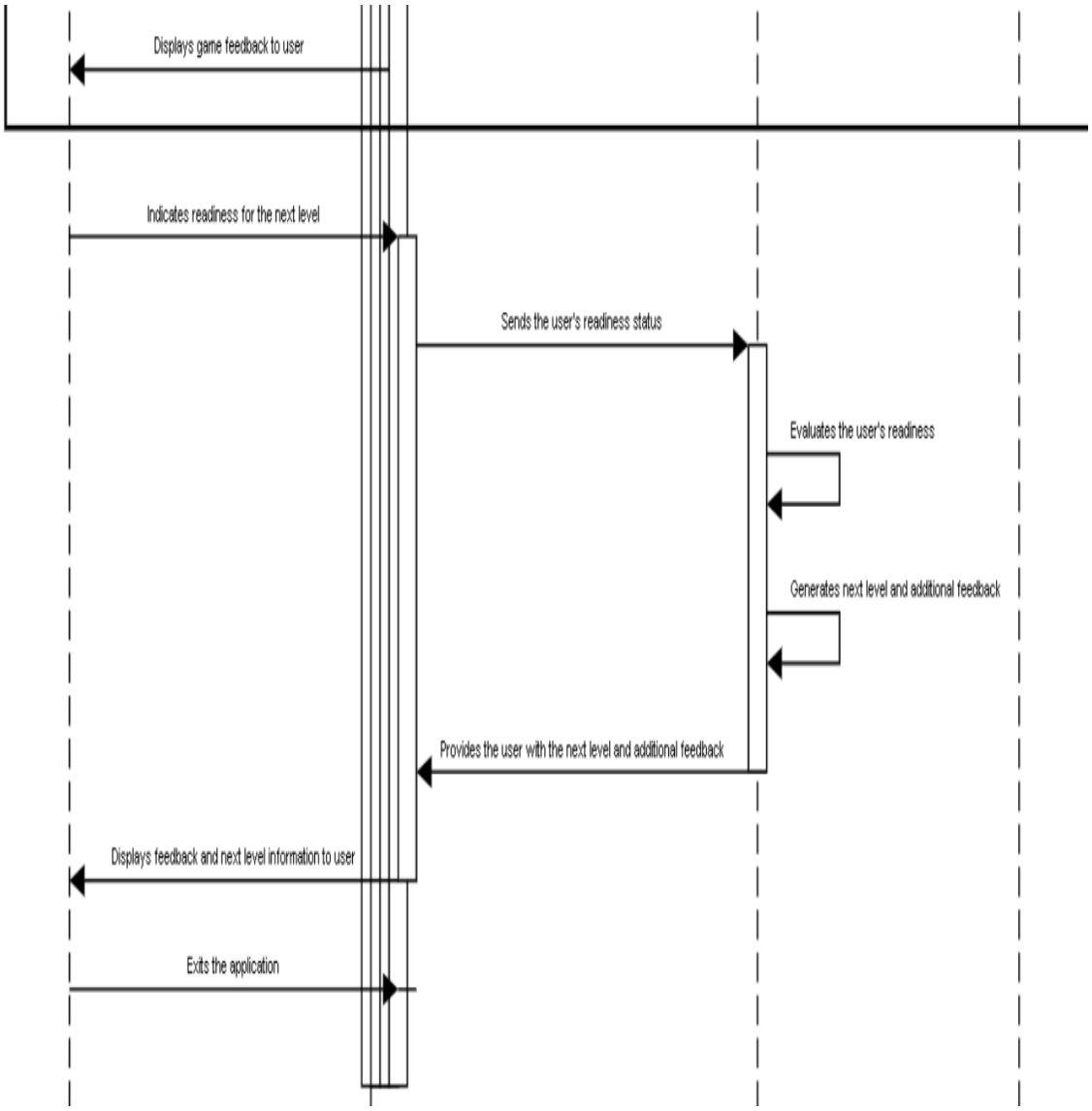


Figure 4.5: Sequence Diagram

Chapter 5

Project Implementation

Project implementation consists of visions and plans with which we are supposed to build end product. This includes the logical conclusion, after evaluating, deciding, visioning, planning and finding the other resources for the project. Technical implementation is one of the major aspects of executing a project.

- Code Snippet This code is for a machine learning model that uses several libraries, including cv2, mediapipe, numpy, speechrecognition, and flask. The code initializes the default values for the scores and sets the default parameters for mediapipe to analyze posture. These libraries and parameters work together to create a powerful tool for analyzing and processing data.

```
import cv2
import os
import mediapipe as mp
import numpy as np
import base64
from moviepy.video.io.VideoFileClip import VideoFileClip
import speech_recognition as sr
import requests
import flask from Flask,requrst

app=Flask(__name__)

speech_score = 0


mp_drawing = mp.solutions.drawing_utils
mp_pose = mp.solutions.pose

```

Figure 5.1: Importing Libraries

In the following code, a base64-encoded video is written to an output file named "output.mp4". The cv2 library is then initialized and fed the "output.mp4" file as input. This allows the cv2 library to process the video data and perform various operations on it. By using the base64 encoding and the cv2 library, the code can efficiently handle and manipulate video data.

```
filename = "output.mp4"

@app.route('/',methods=['POST'])
def confidence():
    video_base64=request.json['base64']
    with open('output.mp4', 'wb') as f:
        video_data = base64.b64decode(video_base64)
        f.write(video_data)
    cap = cv2.VideoCapture(filename)


    cv2.namedWindow("Sitting Posture")


    while cap.isOpened():
        success, image = cap.read()
        if not success:
            print("End of video.")
            break
```

Figure 5.2: Video to Base64 Converter

This code snippet initializes the positions of various landmarks on a human body, such as the left hip, right hip, left knee, and right knee. The code then compares the angles between these landmarks and calculates a score based on their relative positions. This allows the code to analyze the posture of a person and provide feedback on their alignment and balance.

```
        image = cv2.cvtColor(image, cv2.COLOR_BGR2RGB)

        with mp_pose.Pose(min_detection_confidence=0.5, min_tracking_confidence=0.5) as pose:
            results = pose.process(image)

            image = cv2.cvtColor(image, cv2.COLOR_RGB2BGR)
            mp_drawing.draw_landmarks(image, results.pose_landmarks, mp_pose.POSE_CONNECTIONS)

            left_hip = results.pose_landmarks.landmark[mp_pose.PoseLandmark.LEFT_HIP]
            right_hip = results.pose_landmarks.landmark[mp_pose.PoseLandmark.RIGHT_HIP]
            left_knee = results.pose_landmarks.landmark[mp_pose.PoseLandmark.LEFT_KNEE]
            right_knee = results.pose_landmarks.landmark[mp_pose.PoseLandmark.RIGHT_KNEE]
            left_ankle = results.pose_landmarks.landmark[mp_pose.PoseLandmark.LEFT_ANKLE]
            right_ankle = results.pose_landmarks.landmark[mp_pose.PoseLandmark.RIGHT_ANKLE]
            score = 0
            if left_hip.y < left_knee.y < left_ankle.y and right_hip.y < right_knee.y < right_
                score += 50
                finalscore.append(score)
            if abs(left_knee.x - left_hip.x) < abs(right_knee.x - right_hip.x):
```

Figure 5.3: Initialize the landmarks on body

This code uses the numpy library to calculate the mean value of an array of data. The resulting value is then rounded to the nearest integer. By using the numpy library and its powerful mathematical functions, the code can efficiently process large amounts of data and perform complex calculations.

```
                score += (left_knee.x - left_hip.x) * 50 // (right_knee.x - right_hip.x)
                finalscore.append(score)
            cv2.putText(image, f"Sitting Score: {score}", (50, 50), cv2.FONT_HERSHEY_SIMPLEX,

            cv2.imshow("Sitting Posture", image)

        if cv2.waitKey(1) == ord('q'):
            break

    avg = round(np.mean(finalscore),1)
    print(avg)
    cap.release()
    cv2.destroyAllWindows()


    video = VideoFileClip(filename)

    audio = video.audio
```

Figure 5.4: Calculate the Score

In this code, an audio file named ”output.mp3” is converted into a ”output.wav” file. This conversion is necessary in order to work with the speech recognition part of the code. Once the audio file has been converted, the code uses a transcription function to convert the audio data into text. This allows the code to analyze and process spoken language and extract meaningful information from it.

```
    audio.write_audiofile("output.wav")


    r = sr.Recognizer()

    audio_file = sr.AudioFile("output.wav")

    with audio_file as source:
        audio = r.record(source)
        transcript = r.recognize_google(audio)

    print(transcript)

    api_key = 'a6558ffdf5mshb040b6de6a19840p135346jsnba7122a2de16'

    text = transcript


    url = 'https://api.languagetool.org/v2/check'
```

Figure 5.5: Converting Video to Speech

In the next step, the code connects to an API that analyzes the user's grammar and provides feedback on any errors. Based on the accuracy of their grammar, the user is given a speech score out of 100. This score reflects their proficiency in using correct grammar and can help them identify areas for improvement. By connecting to the API and analyzing the user's speech, the code provides a valuable tool for improving language skills.

```
96    url = 'https://api.languagetool.org/v2/check'
97    data = {
98    'text': text,
99    'language': 'en-US'
100   }
101   headers = {
102   'Content-Type': 'application/x-www-form-urlencoded',
103   'Accept': 'application/json',
104   'Authorization': f'Bearer {api_key}'
105   }
106   response = requests.post(url, data=data, headers=headers)
107
108
109   json_data = response.json()
110   matches = json_data['matches']
111   if len(matches) > 0:
112       print(f'{len(matches)} grammar error(s) found:')
113       for match in matches:
```

Figure 5.6: Response from API

Chapter 6

Testing

In this process, we validate and verify that the application does what its supposed to do. The system or its components are tested to ensure the software satisfies all specified requirements.

6.1 Functional Testing

Functional testing is a type of software testing technique that focuses on verifying the specific functions or features of an application, such as ML-based comprehensive application. The goal of functional testing is to ensure that the application functions as expected and meets the requirements of the stakeholders.

This testing technique involves testing the functionalities of the ML based comprehensive application to ensure that it meets the requirements of the stakeholders. Functional testing can be done manually or through automated tools and includes testing the user interface, ML model, Quiz functionality, Back-end connection and other features of the application. Ultimately, the testing techniques employed will depend on the specific requirements and characteristics of the application. It is recommended to employ a combination of different testing techniques to ensure that the application is thoroughly tested and meets the requirements of all stakeholders involved.

We have taken video file as an input to test this model.

Sr.No	Video File	Expected Result	Actual Result	Pass or Fail
1	Output.mp4	Confident	Confident	Pass
2	Output.mp4	Less Confident	Less Confident	Pass
3	Output.mp4	Confident	Confident	Pass

Table 6.1: ML Model Testing

6.1.1 Unit Testing

Unit testing is the first level of testing, which is typically performed by the developers themselves. It helped us understand the desired output of each module, which we had broken down into separate units and in classifying the confidence of users on the basis of algorithm that we have used.

6.1.2 Integration Testing

Integration testing is a software testing technique that evaluates how different modules of an application work together when integrated. This testing method checks for defects or errors in the interfaces between different modules or systems and ensures that they work together as expected

Sr No.	Test Name	Expected Result	Actual Result	Pass or Fail
1	Empty MoodleID Field	Enter MoodleID	Moodle ID cannot be empty	Pass
2	Empty Password Field	Enter Password	Password cannot be empty	Pass
3	User already Exists	Enter MoodleID	Moodle ID is already registered	Pass

Table 6.2: Test Case: Signin

Sr No.	Test Name	Expected Result	Actual Result	Pass or Fail
1	Empty MoodleID Field	Enter MoodleID	Moodle ID cannot be empty	Pass
2	Empty Password Field	Enter Password	Password cannot be empty	Pass
3	Incorrect Password	Enter Password	Password Doesn't Match	Pass

Table 6.3: Test Case: Login

Sr No.	Test Name	Expected Result	Actual Result	Pass or Fail
1	Wrong Format File	Enter Video	Invalid File Format	Pass
2	Empty Filed	Enter Video	Please Enter MP4 File	Pass

Table 6.4: Test Case: Video Upload

Chapter 7

Result

Machine learning (ML) model has been created using a Convolutional Neural Network (CNN) algorithm, which is a type of deep learning method that has shown high performance in image recognition tasks. The ML model is designed to identify the emotions of a person based on their facial expressions. To create this model, several Python libraries were utilized, including OpenCV, which is an open-source computer vision library, and Keras, which is a high-level neural networks API. The provided image shows the successful prediction of the emotion "angry" in a woman's facial expression using this ML model. This means that the algorithm was able to correctly identify the facial features and patterns associated with the emotion of anger, and make an accurate prediction.

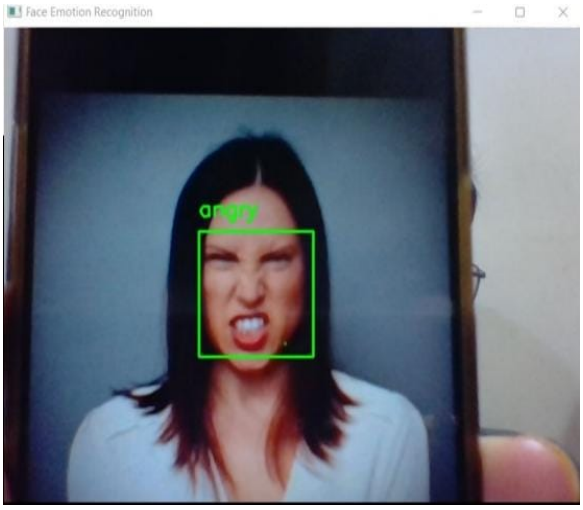


Figure 7.1: ML Model output for Emotion

A similar ML model is built using CNN (Convolutional Neural Network), media-pipe and NLP which has both posture as well as audio or speech recognition model and based on the correctness of the posture and the confidence in speech user will get scores. These scores are stored on cloud and will be combined with the scores generated by quiz. By calculating the average of these scores user will be classified into different levels i.e. Low Confidence, Medium Confidence and High Confidence. After successfully classifying, the user will be getting mentoring sessions and different activities to enhance their soft skills and boost their confidence. Below figures display's the score based on the posture of the user.

Figure 7.2: ML Model for bad Posture

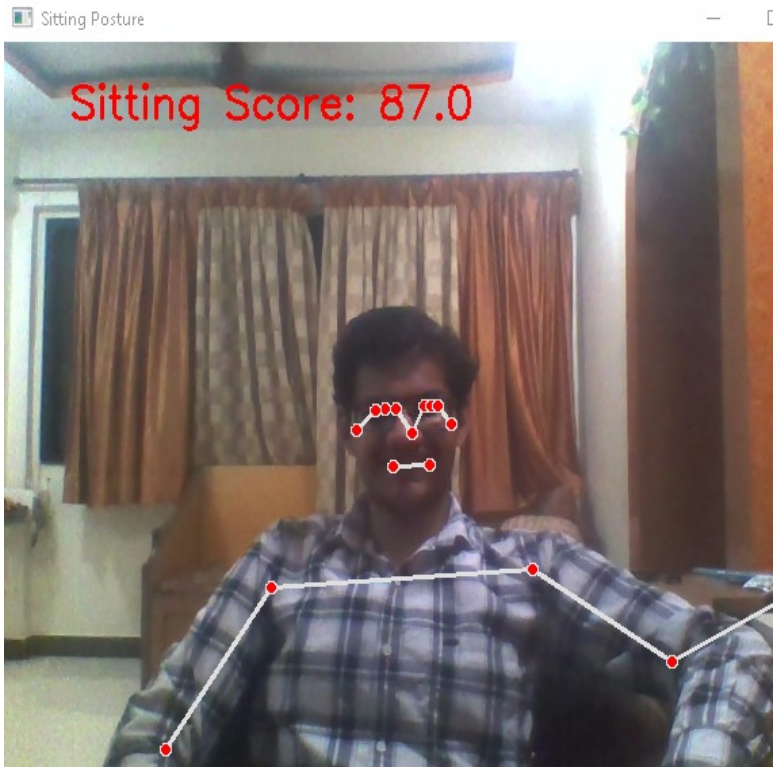


Figure 7.3: ML Model for good Posture

The text "Whoever ceases to be a student has never been a student" suggests a confident tone. NLP can be used to analyze the grammar of the sentence to confirm this. Specifically, the use of "whoever" to begin the sentence, as well as the use of parallel construction, both indicate confidence in the statement being made.

```
Person is confident
>>>
```

Figure 7.4: NLP speech model for high confidence

The text "tell me bad example of English" suggests a lower level of confidence, as indicated by the use of "tell me" rather than a more assertive statement. Additionally, the phrase "bad example of English" could be seen as a hesitancy or lack of certainty in the speaker's language abilities. NLP can be used to analyze the grammar and structure of the sentence, confirming the lower level of confidence in the speaker's statement.

```
Person is low confident
>>>
```

Figure 7.5: NLP speech model low confidence

The sign-in and sign-up page of an application built using the Flutter framework, a popular open-source mobile app development platform. These pages are typically the first points of interaction between the application and its users, and they are designed to allow students to enter their login credentials, such as username and password, to access the app's features.

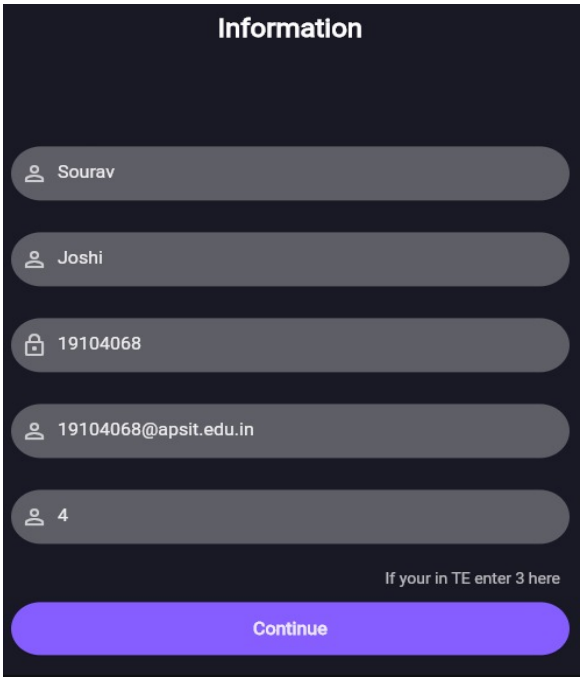


Figure 7.6: Registration Page

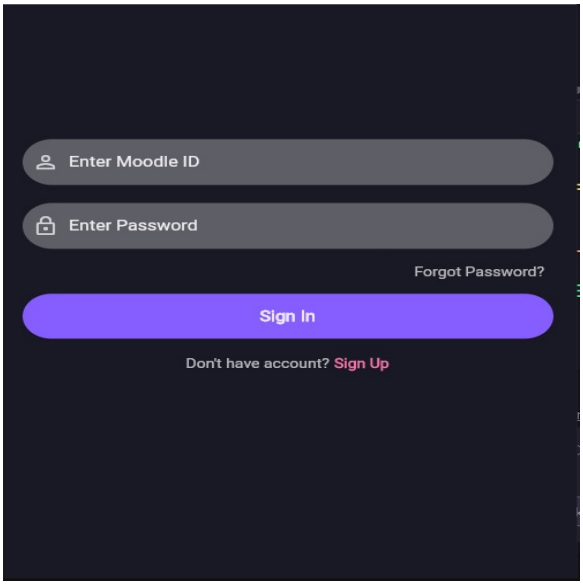


Figure 7.7: Login Page

After successful login students have various activities using which he/she can enhance their soft skills. The image below shows one such activity i.e. MCQ type questions which are asked to students, these questions will help in increasing EQ and IQ of students. Based on the answers the score is generated and stored in the database which will help in knowing the knowledge of the students and classify into different categories i.e. low, medium and high confidence level.

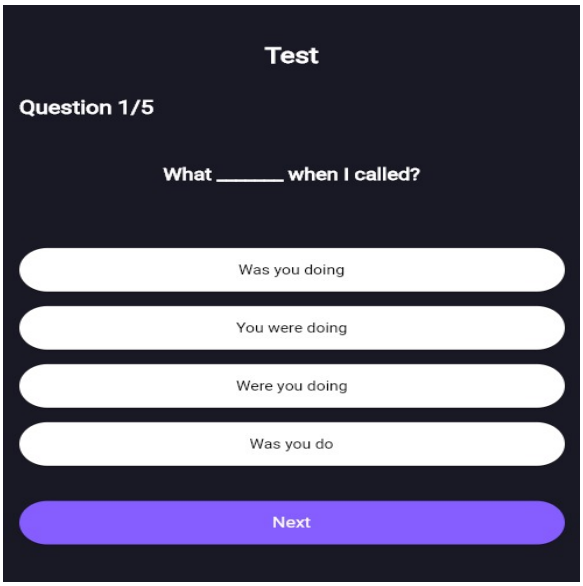


Figure 7.8: Quiz Page

After attempting the test user will need to upload a video and the instructions which need to be followed by user are mentioned on the page, screenshot of the same is provided above. Based on the video uploaded by the user, confidence score is generated and stored on Cloud. This score is generated based on the posture, emotion, tone of the speaker, grammar used in speech and how fluently user is speaking.

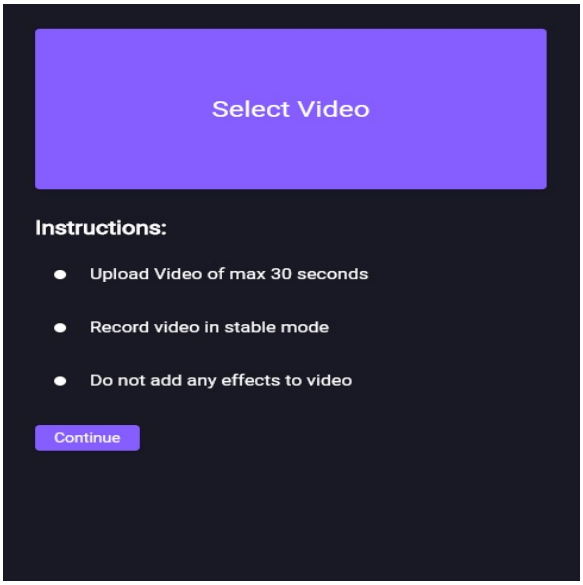


Figure 7.9: Video Uploading Page

User will Upload maximum 30 seconds video on platform after uploading it to application it will connect to cloud function and get output from cloud function and display it to user make it little bit bigger

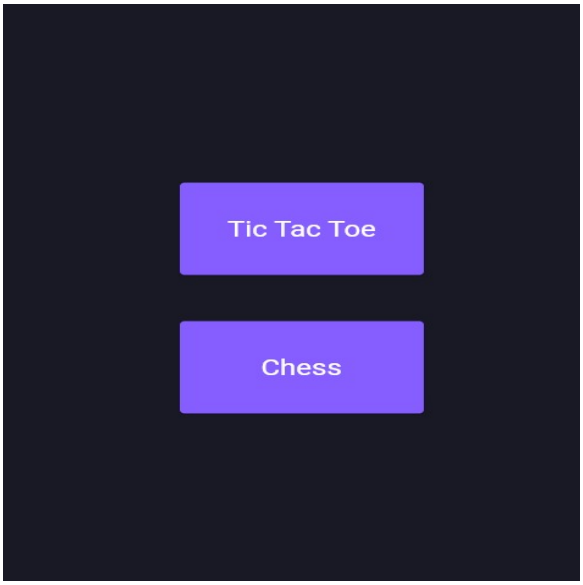


Figure 7.10: Activities

After receiving their score, users will be directed to a screen where they can play games with their friends in person. The platform offers two games to choose from: Tic Tac Toe and Chess. Both games are designed for two players and provide a fun and engaging way for users to interact with each other.

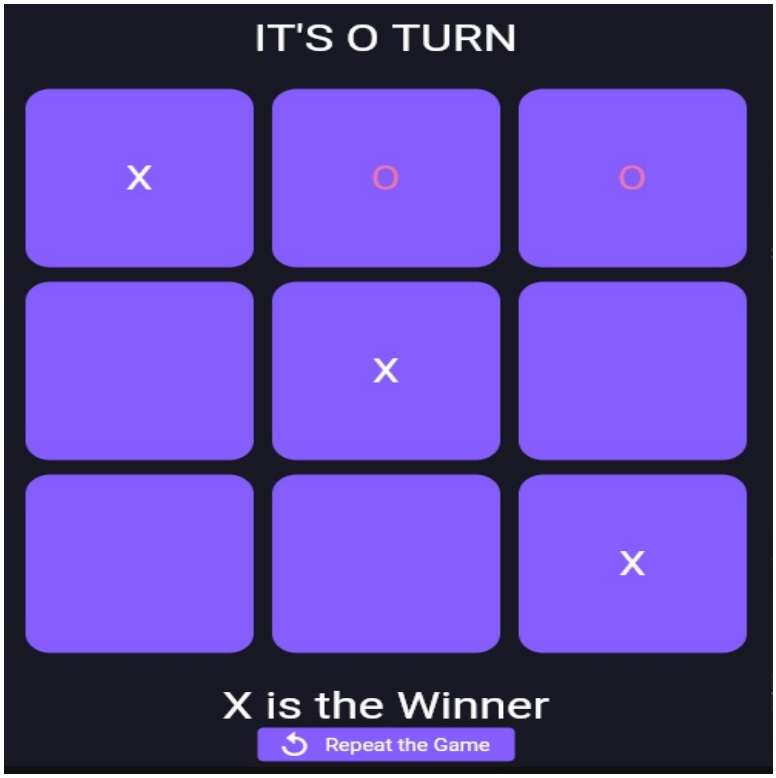


Figure 7.11: Tic Tac Toe game

Figure 7.12: Chess Game

Chapter 8

Conclusions

Our primary objective was to empower students to express themselves with confidence in front of a large audience. To achieve this goal, we had developed a cross-platform application that will enhance students' communication abilities and self-assurance. Our approach involves using both audio and visual inputs to create an ML model that can compile and analyze data from all sources to evaluate the student's confidence level. The model will have mentoring sessions for students in order to build their confidence. We believe that this project has the potential to be a valuable resource for colleges seeking to help their students improve their self-confidence. Based on different algorithms we have achieved higher accuracy using NLP for speech and CNN for emotion and posture.

Chapter 9

Future Scope

One potential avenue for further development is to expand the application's capabilities beyond just analyzing and improving confidence levels. For example, the application could be adapted to help students identify and address specific communication challenges, such as accent reduction, public speaking anxiety, or nonverbal communication issues. Additionally, the application could incorporate more sophisticated machine learning algorithms and natural language processing to provide personalized feedback and recommendations to each student based on their unique communication style and learning preferences. Another potential direction is to integrate the application with virtual or augmented reality technologies to provide students with realistic simulated environments in which to practice and develop their communication skills. With continued research and development, this application has the potential to become an indispensable tool for students seeking to enhance their communication abilities and build their self-assurance.

Bibliography

[1] Wei, Gou, Li Jian, and Sun Mo. "Multimodal (audio, facial and gesture) based emotion recognition challenge." 2020 15th IEEE International Conference on Automatic Face and Gesture Recognition (FG 2020). IEEE, 2020.

[2] Begaj, Sabrina, Ali Osman Topal, and Maaruf Ali. "Emotion Recognition Based on Facial Expressions Using Convolutional Neural Network (CNN)." 2020 International Conference on Computing, Networking, Telecommunications & Engineering Sciences Applications (CoNTESA). IEEE, 2020.

[3] A. Ogawa, N. Tawara, T. Kano and M. Delcroix, "BLSTM-Based Confidence Estimation for End-to-End Speech Recognition," ICASSP 2021 - 2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Toronto, ON, Canada, 2021, pp. 6383-6387, doi: 10.1109/ICASSP39728.2021.9414977.

[4] Mannapperuma, Dilara, and Abarnah Kirupananada. "ADAM-Anxiety Detection and Management: a Solution to Manage Anxiety at Workplaces and Improve Productivity." 2020 IEEE International Women in Engineering (WIE) Conference on Electrical and Computer Engineering (WIECON-ECE). IEEE, 2020.

[5] Kasthuri, E., and S. Balaji. "A chatbot for changing lifestyle in education." 2021 Third International Conference on Intelligent Communication Technologies and Virtual Mobile Networks (ICICV). IEEE, 2021.

[6] Zhang, Ke, et al. "Real-time video emotion recognition based on reinforcement learning and domain knowledge." IEEE Transactions on Circuits and Systems for Video Technology 32.3 (2021): 1034-1047.

[7] Liu, Wei, et al. "Comparing recognition performance and robustness of multimodal deep learning models for multimodal emotion recognition." IEEE Transactions on Cognitive and Developmental Systems 14.2 (2021): 715-729.

[8] Sato, Ryota, et al. "Creation and Analysis of Emotional Speech Database for Multiple Emotions Recognition." 2020 23rd Conference of the Oriental COCOSDA International

Committee for the Co-ordination and Standardisation of Speech Databases and Assessment Techniques (O-COCOSDA). IEEE, 2020.

[9] Lotfian, Reza, and Carlos Busso. "Curriculum learning for speech emotion recognition from crowdsourced labels." IEEE/ACM Transactions on Audio, Speech, and Language Processing 27.4 (2019): 815-826.

[10] Chen, Mingyi, et al. "3-D convolutional recurrent neural networks with attention model for speech emotion recognition." IEEE Signal Processing Letters 25.10 (2018): 1440-1444.

[11] Khalil, Ruhul Amin, et al. "Speech emotion recognition using deep learning techniques: A review." IEEE Access 7 (2019): 117327-117345.

[12] Tiwari, Shreesh Krishannarayan, et al. "A novel approach for attendance monitoring system with face mask detection." AIP Conference Proceedings. Vol. 2424. No. 1. AIP Publishing LLC, 2022.

[13] Mathur, Romir, et al. "Orthopantomogram (OPG) Image Analysis Using Bounding Box Algorithm." Computational Methods and Data Engineering: Proceedings of ICCMDE 2021. Singapore: Springer Nature Singapore, 2022. 55-65.

[14] Kolhe, Pranita, Kamlesh Kalbande, and Atul Deshmukh. "Internet of Thing and Machine Learning Approach for Agricultural Application: A Review." 2022 10th International Conference on Emerging Trends in Engineering and Technology-Signal and Information Processing (ICETET-SIP-22). IEEE, 2022.

Appendices

9.1 Appendix-I: Installation of Python Libraries

1. mkdir FinalProject
2. cd FinalProject
3. pip install virtualenv
4. virtualenv myenv
5. myenv/Scripts/activate
6. pip install opencv-python
7. pip install mediapipe
8. pip install numpy
9. pip install SpeechRecognition
10. pip install Flask
11. pip install moviepy
12. Download the flutter SDK From "https://docs.flutter.dev/get-started/install"
13. Extract the zip
14. Edit environment variables and add flutter path to it
15. check the location of flutter by where flutter dart
16. Run flutter doctor command
17. flutter create app
18. cd app
19. flutter analyze
20. flutter test
21. flutter run lib/main.dart

22. cd lib
23. mkdir pages
24. dart create pagenames
25. create firebase account on https://firebase.google.com/
26. flutter pub get cloudfirestore
27. flutter pub get required packages
28. create google cloud account on https://cloud.google.com/
29. go to concole
30. create cloud function
31. add code and dependencies
32. flutter pub get http
33. flutter run lib/main.dart

Publication

Paper entitled **"ML Based Comprehensive Application to Enhance Soft Skills"** is presented at **"2023 11th International Conference on Emerging Trends in Engineering & Technology - Signal and Information Processing (ICETET - SIP)"** by **"Sakshi Deshpande , Sourav Joshi, Akshay Bura , Rujata Chaudhari , Dr. Kiran Deshpande".**