

A Report on

# **ML based Comprehensive Application To Enhance Soft Skills**

Submitted in partial fulfillment of the requirements  
of the degree of

**Bachelor of Engineering**

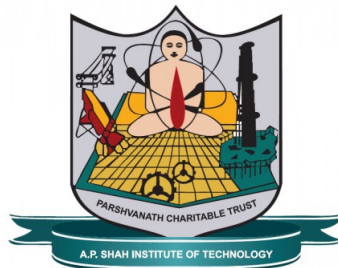
in

**Information Technology**

by

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

**Project Guide**  
**Ms. Rujata Chaudhari**



**Department of Information Technology**  
A.P. Shah Institute of Technology  
G.B.Road,Kasarvadavli, Thane(W), Mumbai-400615  
UNIVERSITY OF MUMBAI  
2022-2023

## CERTIFICATE

This is to certify that the project Synopsis entitled “*ML based Comprehensive Application To Enhance Soft Skills*” Submitted by “*Akshay Bura (19104041), Sourav Joshi (19104068), 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

(Ms.Rujata Chaudhari)  
Guide

(Dr. Kiran Deshpande)  
Head Of Department Of Information Technology

(Dr. Uttam D. Kolekar)  
Principal

External Examiner(s)

1.

2.

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

Date:

## Declaration

We declare that this written submission represents our idea 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.

---

(Akshay Bura 19104041)

---

(Sourav Joshi 19104068)

---

(Sakshi Deshpande 19104002)

Date:

# Abstract

Now a days most of the students are facing problem to talk in mob or having lack of confidence even though they knew many more things but cannot express it because of lack of confidence. To address this problem we are building an cross platform application through which students can improve their soft skills and build confidence to express themselves. Many people treat social anxiety like a disease which never going to cure and they will remain like this for the rest of their life ,but that is not true.In this project we are creating an application which will be helping in boosting the self confidence of students by exploring state of the art models to extract emotions and body postures which will help in building confidence. We have chosen to explore sound and video inputs and develop an ensemble model that gathers the information from all these sources and use them to analyse the confidence level of the student and assign them certain activities to gain confidence. This project can be used by colleges to help their students in enhancing or boosting their self-confidence.This application will be containing different activities i.e. games, quiz, mentoring sessions, etc. which will help in enhancing soft skills. We will also store the past details of the students so that he / she can track their growth in soft skills. Student as well as Mentor both can track the growth. Mentors can upload a video for the users so user will know how to communicate on stages and how to be a good listener and speaker.App will be automating Xpression Club and improve logical and verbal ability of the user.

# Introduction

These days, many students lack confidence and verbal ability in social environment like placement drives and that may affect their job interviews. It is not always possible to hire counselor in rural areas. There are abundant of sources and mediums to detect lack of self-confidence but 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 challenging task for Machine learning. Getting accurate result may not be possible by Machine learning.

For this Project we are using the Ryerson Audio-Visual Database of Emotional Speech and Song (RAVDESS). This database contains 7356 files (total size: 24.8 GB). The database contains 24 professional actors (12 female, 12 male), vocalizing two lexically-matched statements in a neutral North American accent. Speech includes calm, happy, sad, angry, fearful, surprise, and disgust expressions, and song contains calm, happy, sad, angry, and fearful 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]. Each of these windows will be the entry of a convolutional neural network, composed by four Local Feature Learning Blocks (LFLBs) and the output of each of these convolutional networks will be fed into a recurrent neural network composed.

We had selected this topic based on an Club in Our Institute named Xpression Club. This club is an initiative made by our institute which helps students in boosting their self confidence by performing various activities such as giving speech in front of whole class, debate, helping students in preparing how to speak in front of an interviewer or any professional person. Our project is an modified / automated version of this club. We will be having different models for different purposes few of them are emotion recognition, posture recognition, speech recognition, question generation, etc. using these models we will be creating our application.

Our aim is to develop a model 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 allow to train much fewer parameters, and therefore reduce training time on Colab'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] When it comes to applying CNNs in real life application, being able to explain the results is a great challenge. We can indeed plot class activation maps, which display the pixels that have been activated by the last convolution layer. We notice how the pixels are being activated differently depending on the emotion being labeled. The happiness seems to depend on the pixels linked to the eyes and mouth, whereas the sadness or the anger seem for example to be more related to the eyebrows.

## 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.

## Literature Review

The purpose of literature review is to gain an understanding of the existing research on Cry Analysis and debates relevant to area of study. The literature review helped in selecting appropriate algorithm and suitable feature extraction process for getting efficient results.

- According to Gou Wei and etl , Speech emotion recognition is automatically finding the emotions of the human voice. Emotion of Speech can be useful to identify confidence in the voice and also the mood of the person. From this research paper we have understood how emotions are analyzed and how to classify them .By extracting audio features and signal processing we can get around 76% (backslash percent) accuracy. The model is based on a discrete emotion classification system. It includes labels such as happy, sad, angry, disgusted, fear and surprise. For this project we only need fear and surprise.
- According to the Sabrina Begaj and etl , This paper compares several classification algorithms to find which one has an accurate result and also predict the label quickly. They use the Multimodal Emotion Recognition (MER) Dataset for this prediction. Then we compared Features as well as various combinations of the algorithm and frame-level features only got 50%(backslash percent) accuracy. So we stick with the Convolutional neural network(CNN).
- According to the DiDiChuXing and etl, Author had researched on confidence in speech and how this can be useful in different situations.For an instance author of this paper uses example of a 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.
- According to the Ali Osman Topal and etl , according to the author the main motive of 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 the Maaruf Ali and etl, 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. One of the most convenient ways of studying for students is Chatbot and it also clears student doubts any time anywhere without human support. This paper aims to approach a typical way to design a chatbot for MATLAB practical dataset. Students can ask a question in the chatbot in the form of text then, the question is processed with natural language processing and deep learning technology. Finally, the chatbot can answer the students with exact answers. For both students and teachers these chatbots are useful.
- According to the Li Jian and etl, Author explains the 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 finds that using Natural Language Processing and deep learning technique system can be automated and helps in predicting accurate results. System predicts the result and give candidate scores accordingly. System reduce 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 the Sun Mo and etl, Multimodal emotion system 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, accuracy system can easily predict the analysis. To fully extract the relative context from video clips, most of them build their models on the entire dialogues which make them lack of real-time ERC ability. Different from related researches, a novel multimodal emotion recognition model for conversational videos based on reinforcement learning and domain knowledge (ERLDK) is proposed in this paper. The extracted domain knowledge is used to revise the results from the RL module and is transformed into other dataset to examine the rationality.

## **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.

## Proposed System Architecture/Working

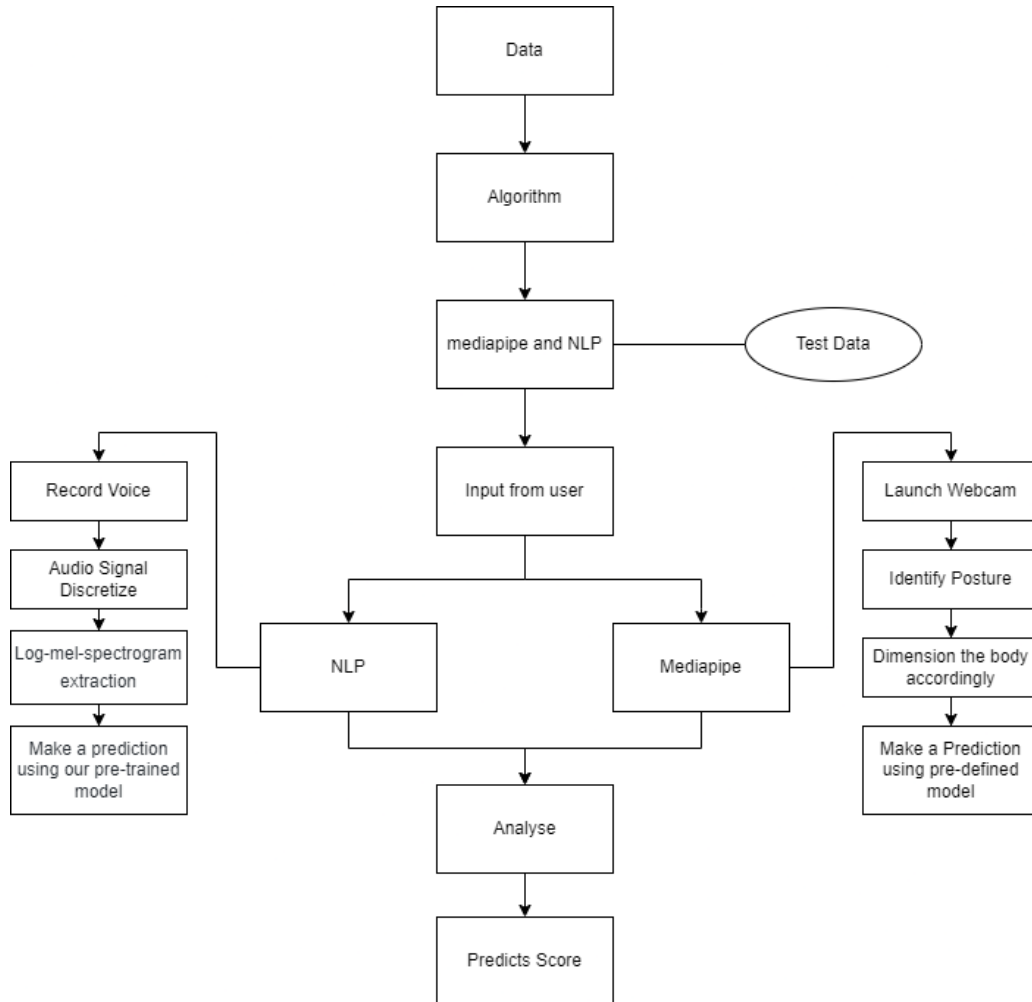


Figure 1: ML Model

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 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.

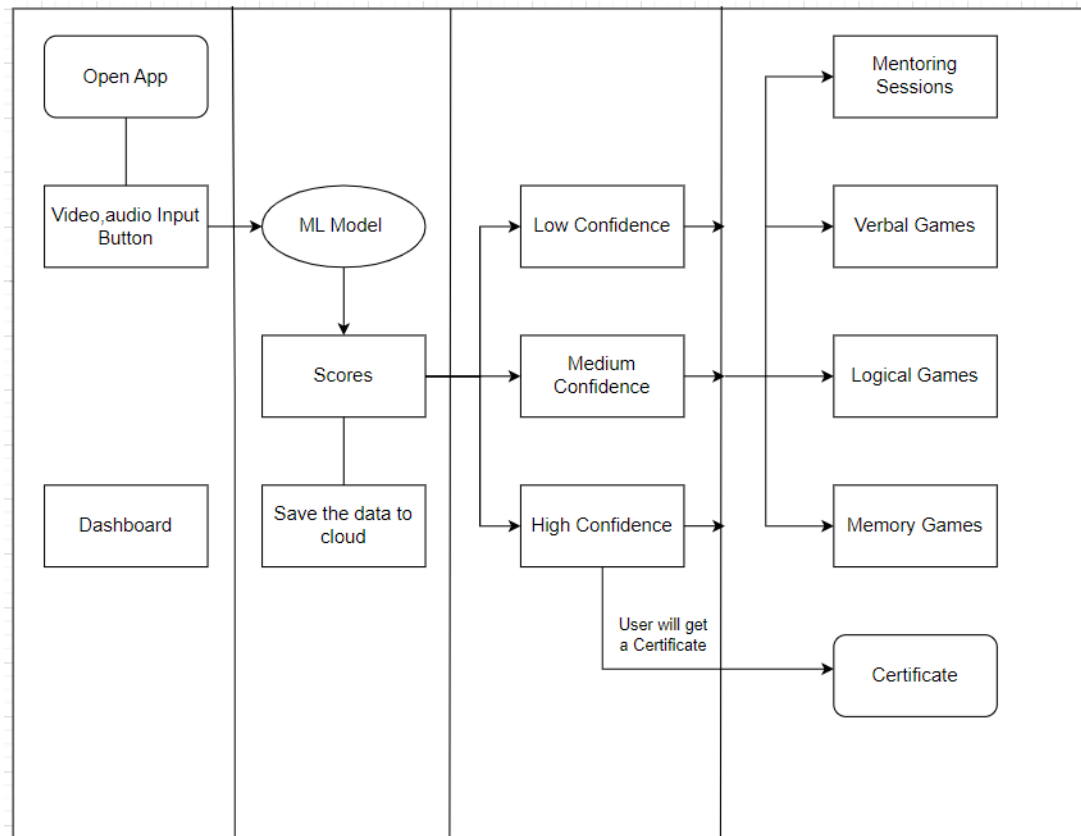


Figure 2: Design/Structure Of Application

Students / Users will have to download the app from play-store and open it. Then he/she has to upload a video. Based on the video user will get a score which is generated based on the ML model that we have trained which is explained in the previous section. These scores will be stored in the cloud for future use. On the basis of these scores, the user will be assigned to a specific level i.e. Low confidence, Medium confidence, and High confidence. These levels will have various activities such as logical, verbal, and memory games to improve soft skills of the user. High Confidence level student will get a certificate at the end.

## Summary

The work discussed in this paper focuses on enhancing students' self-confidence and assisting them in speaking their minds in front of a large audience. We are developing a cross-platform application so that students can hone their communication abilities and gain self-assurance. We will investigate cutting-edge models to extract emotions and body postures that can boost confidence. We have opted to investigate audio and visual inputs and create an ensemble model that compiles data from all of these sources, analyses the student's degree of confidence, and gives them specific tasks to complete in order to build confidence. Colleges can use this project to assist their students in improving or raising their self-confidence. This application will include various activities, such as games, quizzes, mentorship sessions, etc. that will aid in the improvement of soft skills. At the completion of this project, we will have automated Xpression Club, which was one of our key goals.

## References

- [1] Multimodal(Audio, Facial and Gesture) based Emotion Recognition challenge, 2020
- [2] Emotion Recognition Based on Facial Expressions Using Convolutional Neural Network (CNN)
- [3] pWAVE: A Novel Dataset for Emotional Confidence Detection , 2017
- [4] ADAM- Anxiety Detection and Management: a Solution to Manage Anxiety at Workplaces and Improve Productivity
- [5] A Chatbot for Changing Lifestyle in Education
- [6] Real-Time Video Emotion Recognition based on Reinforcement Learning and Domain Knowledge , 2021
- [7] Comparing Recognition Performance and Robustness of Multimodal Deep Learning Models for Multimodal Emotion Recognition