NAGARJUNA COLLEGE OF ENGINEERING AND TECHNOLOGY

(An Autonomous College under VTU, Belagavi)

A Phase II Report

on

"SPEECH EMOTION RECOGNITION USING MLP CLASSIFIER"

Submitted in partial fulfillment for the course

Project Phase II and Seminar (18CSP78)

Submitted by

Ms. Pallavi S	1NC18CS073
Ms. Sushma G S	1NC18CS112
Ms. Thrisha B V	1NC18CS116
Mr. Yakkali Mani Gopal	1NC18CS121

Under the guidance of

Mrs. Swathi S

Associate Professor

Dept. of CSE,NCET

DEPARTMENT OF COMPUTER SCIENCE & ENGNIEERING

NAGARJUNA COLLEGE OF ENGINEERING & TECHNOLOGY

(An Autonomous College under VTU, Accredited by NAAC with "A" Grade)
Mudugurki (V), Venkatagirikote (P), Devanahalli (T), Bengaluru-562164.

2021-22

DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING

CERTIFICATE

Certified that the project work entitled **"SPEECH EMOTION RECOGNITION USING MLP CLASSIFIER"** carried out by **Ms. PALLAVI S (1NC18CS073), Ms. SUSHMA G S (1NC18CS112) Ms. THRISHA B V (1NC18CS112), Mr. YAKKALI MANI GOPAL (1NC18CS121)** bonafide students of Nagarjuna College of Engineering and Technology, an autonomous institution under Visvesvaraya Technological University, Belagavi in partial fulfillment for the award of Master of Technology in **Computer Science & Engineering** during the academic year 2021-2022. It is certified that all corrections/suggestions indicated for internal assessment have been incorporated in the report deposited in the departmental library. The project work has been approved, as it satisfies the academic requirement in respect of project work prescribed for the said degree.

Name & Signature of the Guide	**Name & Signature of the HOD**	**Name & Signature of the Principal**
Mrs. Swathi. S	Dr. Dinesha H A	Dr. Ravishankar B V
Associate Professor	HOD,	Principal, NCET
Dept. of CSE	Dept. of CSE	

External Viva-Voce

Name of the Examiner   **Signature with date**

1. ....

2. ....

ACKNOWLEDGEMENT

ABSTRACT

Language is human's most important communication and speech is basic medium of communication. Emotion plays a crucial role in social interaction. Recognizing the emotion in a speech is important as well as challenging because here we are dealing with human machine interaction. Emotion varies from person to person were same person have different emotions all together has different way express it. When a person express his emotion each will be having different energy, pitch and tone variation are grouped together considering upon different subject. Therefore the speech emotion recognition is a future goal of computer vision. The aim of our project is to develop the smart emotion recognition speech based on the convolutional neural network. Which uses different modules for emotion recognition and the classifier are used to differentiate emotion such as happy sad angry surprise. The machine will convert the human speech signals into waveform and process its routine at last it will display the emotion. The data is speech sample and the characteristics are extracted from the speech sample using librosa package. We are using RAVDESS dataset which are used as an experimental dataset.

TABLE OF CONTENTS

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

Language is human's most important communication and speech is basic medium of communication. Emotion plays a crucial role in social interaction. Recognizing the emotion in a speech is important as well as challenging because here we are dealing with human machine interaction. Emotion varies from person to person were same person have different emotions all together has different way express it. When a person express his emotion each will be having different energy, pitch and tone variation are grouped together considering upon different subject. Therefore the speech emotion recognition is a future goal of computer vision. The aim of our project is to develop the smart emotion recognition speech based on the convolutional neural network. Which uses different modules for emotion recognition and the classifier are used to differentiate emotion such as happy sad angry surprise. The machine will convert the human speech signals into waveform and process its routine at last it will display the emotion. The data is speech sample and the characteristics are extracted from the speech sample using librosa package. We are using RAVDESS dataset which are used as an experimental dataset.

CHAPTER 2

LITERATURE SURVEY

Paper 1: Machine Learning Based Speech Emotions Recognition System

Authors: Dr. Yogesh Kumar, Dr. Manish Mahajan

Article in International Journal of Scientific & Technology Research · July 2019

The speech signal is one of the most natural and fastest methods of communication between humans. Many systems have been developed by various researchers to identify the emotions from the speech signal. In differentiating between various emotions particularly speech features are more useful and if not clear is the reason that makes emotion recognition from speaker's speech very difficult. There are a number of the dataset available for speech emotions, it's modelling, and types that helps in knowing the type of speech. After feature extraction, another important part is the classification of speech emotions so the paper has compared and reviewed the different classifiers that are used to differentiate emotions such as sadness, neutral, happiness, surprise, anger, etc. The research also shows the improvement in emotion recognition system by making automatic emotion recognition system adding a deep neural network. The analysis has also been performed using different ML techniques for Speech emotions recognition accuracy in different languages.

- In this paper , they have used KNN algorithm for detecting emotions in speech signals.

- Features like MFCC , LPCC are used for data preprocessing and feature extraction.

- Its accuracy is 60.7%

ADVANTAGES:

- KNN has less computation time.

- KNN is simple to interpret.

DISADVANTAGES:

- It has less accuracy.

- It's accuracy depends on data.

- KNN require high memory

Paper 2: Speech Emotion Recognition with deep learning

Authors: Hadhami Aouani , Yassine Ben Ayed

24th International Conference on Knowledge-Based and Intelligent Information & Engineering Systems 2020

This paper proposes an emotion recognition system based on speech signals in two-stage approach, namely feature extraction and classification engine. Firstly, two sets of feature are investigated which are: the first one , we extract an 42-dimensional vector of audio features including 39 coefficients of Mel Frequency Cepstral Coefficients (MFCC), Zero Crossing Rate(ZCR), Harmonic to Noise Rate (HNR) and Teager Energy Operator (TEO). And the second one, we propose the use of the method Auto-Encoder for the selection of pertinent parameters from the parameters previously extracted. Secondly, we use the Support Vector Machines (SVM) as a classifier method. Experiments are conducted on the Ryerson Multimedia Laboratory (RML). Speech is the main and direct means of transmitting information. It contains a wide variety of information, and it can express rich emotional information through the emotions it contains and visualize it in response to objects, scenes or events. The automatic recognition of emotions by analyzing the human voice and facial expressions has become the subject of numerous researches and studies in recent years. The fact that automatic emotion recognition systems can be used for different purposes in many areas has led to a significant increase in the number of studies on this subject.

- Features like 39 coefficients of Mel Frequency Cepstral Coefficients, Zero Crossing Rate, Harmonic to Noise Rate are used.
- In this paper , they have used SVM algorithm for training the machine.
- Its accuracy is 73.3%.

ADVANTAGES:
- It is memory efficient.
- It is more effective in high dimensional spaces.

DISADVANTAGES:
- It couldn't classify disgust emotion.
- Not suitable for large datasets.

CHAPTER 3

SYSTEM ANALYSIS

3.1 EXISTING SYSTEM

Speech Emotion Recognition is one of the booming research topics in the computer science world. Emotion is a medium by which one expresses how a person feels and one's state of mind. Predicting emotions is a tough task as every individual has a different tone and intonation of speech. Thus emotions are difficult to extract using current machine learning systems easily. Therefore, many researchers have used Deep Learning and Machine Learning techniques to extract the emotions of speech signals
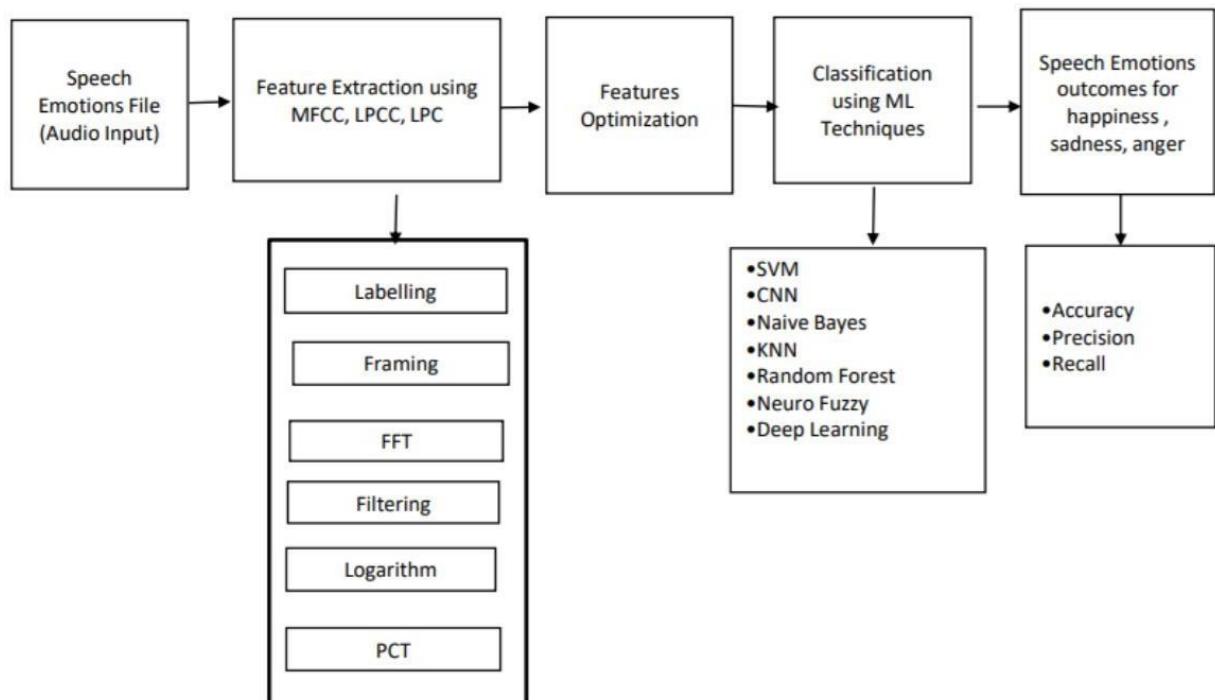

Fig.1 Framework for Existing Speech Emotions Recognition

3.2 PROPOSED SYSTEM

Speech Emotion Recognition is one of the booming research topics in the computer science world. Emotion is a medium by which one expresses how a person feels and one's state of mind. Emotions are difficult to extract using current machine learning systems easily. Therefore, many researchers have used Neural Network and Machine Learning techniques to extract the emotions of speech signals. Our proposed system consists of MLP classifier.
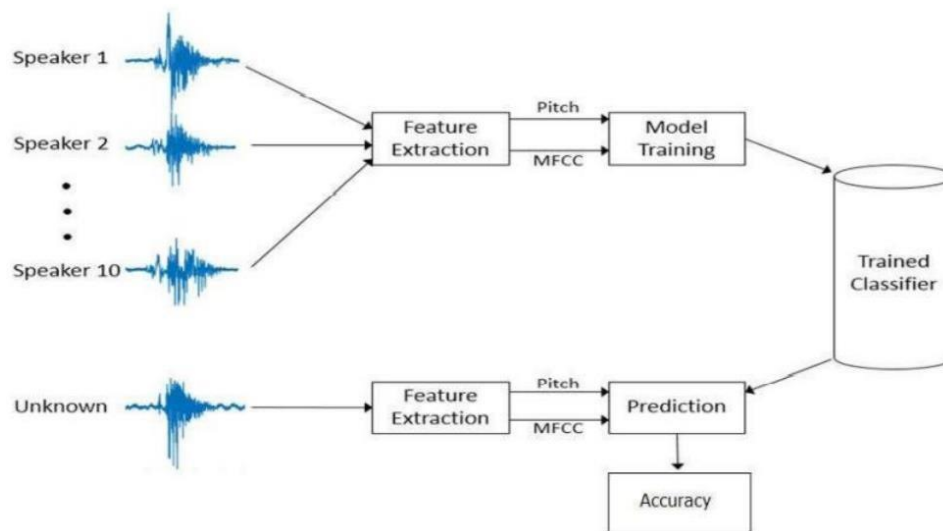
Fig.2 Framework for Speech Emotion Recognition

Dataset for this project we are using the RAVDESS dataset, Ryerson Audio-Visual Database of Emotional Speech and Song (RAVDESS) dataset[18]; in the RAVDESS dataset, there are 1440 Audio-speech files belong to 24 professional actors among them, 12 are male, and 12 are female, 60 trails for each of them totally 1440 Audio-speech files, and 1012 Audio-Song files belong to 24 professional actors among them 12 are male, and 12 are female, 44 trails for each of them totally 1012 Audio-song files. Where Audio- speech files contain sad, calm, angry, happy, fearful, disgust, and surprise expressions, and Audio-song files contain sad, calm, angry, fearful, and happy emotions

♦ **Feature extraction**

It is the function where we are extract the mfcc, chroma, and mel features from a sound file. This will take 4 parameters, the file name and three Boolean parameters. The three features are mfcc : mel frequency cepstral coefficient, represent the power spectrum of a sound. Chroma: pertains to the 12 pitch classes mel: mel spectrum frequency Sound file will be opened and it will be readied and result will be saved to array. For each of three, if it exists then a call will be made to the corresponding function from librosa. Mean value will be noted and result along with feature value and storing it in a file.

♦ **Training and Testing**

We are loading the data where it takes in the relative size of the test set as parameter. X and Y are empty lists, functions will checks whether the emotion are in the list of observed emotions. The feature will be send to X and emotions to Y. Now the testing and training function will be called. 75% of audio will be tested at the same time 25% of audio will trained. For classification we are using MLP Classifier.

```python
import numpy as np
x_train,x_test,y_train,y_test=load_data(test_size=0.25)
# print(x_train[1].shape,y_train[1])
```

Fig.3 Training and testing a speech

Finally the system will be ready and it will trained using fit/train model.

```python
#Train the model
model.fit(x_train,y_train)

MLPClassifier(alpha=0.01, batch_size=256, hidden_layer_sizes=(300,),
              learning_rate='adaptive', max_iter=500)
```

Fig.4 Training the model

CHAPTER 4

SYSTEM REQUIREMENT AND SPECIFICATIONS

4.1 Requirements Specification:

Requirement Specification provides a high secure storage to the web server efficiently. Software requirements deal with software and hardware resources that need to be installed on a serve which provides optimal functioning for the application. These software and hardware requirements need to be installed before the packages are installed. These are the most common set of requirements defined by any operation system. These software and hardware requirements provide a compatible support to the operation system in developing an application.

4.1.1 SOFTWARE REQUIREMENT:

The software requirements specify the use of all required software products like data management system. The required software product specifies the numbers and version. Each interface specifies the purpose of the interfacing software as related to this software product.

- Anaconda , Jupyter Notebook or Google Colab
- Kaggle for datasets.

4.1.2 HARDWARE REQUIREMENT:

The hardware requirement specifies each interface of the software elements and the hardware elements of the system. These hardware requirements include configuration characteristics.

Name of components	:	Specifications
CPU	:	Intel i5 processor with minimum of 7th generation.
RAM	:	Minimum of 8GB.
Operating System	:	Ubuntu or Microsoft Windows 10.

4.2 FUNCTIONAL REQUIREMENTS:

The functional requirement refers to the system needs in an exceedingly computer code engineering method. The main purpose of functional requirements within the requirement specification document is to define all the activities or operations that take place in the system. These are derived through interactions with the users of the system. Since the Requirements Specification is a comprehensive document & contains a lot of data, it has been broken down into different Chapters in this report.

The key goal of determinant "functional requirements" in an exceedingly product style and implementation is to capture the desired behavior of a software package in terms of practicality and also the technology implementation of the business processes.

4.3 NON-FUNCTIONAL REQUIREMENTS:

All the other requirements which do not form a part of the above specification are categorized as Non-Functional needs. A system perhaps needed to gift the user with a show of the quantity of records during info. If the quantity must be updated in real time, the system architects should make sure that the system is capable of change the displayed record count at intervals associate tolerably short interval of the quantity of records dynamic. Comfortable network information measure may additionally be a non-functional requirement of a system.

The following are the features:

➢ Accessibility

➢ Availability

➢ Backup

➢ Certification

➢ Compliance

➢ Configuration Management

➢ Documentation

➢ Disaster Recovery

➢ Efficiency (resource consumption for given load)

➢ Interoperability

4.4 PERFORMANCE REQUIREMENTS

Performance is measured in terms of the output provided by the application. Requirement specification plays an important part in the analysis of a system. Only when the requirement specifications are properly given, it is possible to design a system, which will fit into required environment. It rests largely with the users of the existing system to

give the requirement specifications because they are the people who finally use the system. This is because the requirements have to be known during the initial stages so that the system can be designed according to those requirements. It is very difficult to change the system once it has been designed and on the other hand designing a system, which does not cater to the requirements of the user, is of no use.

The requirement specification for any system can be broadly stated as given below:

- The system should be able to interface with the existing system

- The system should be accurate

- The system should be better than the existing system

The existing system is completely dependent on the user to perform all the duties.

4.5 Feasibility Study:

Preliminary investigation examines project feasibility; the likelihood the system will be useful to the organization. The main objective of the feasibility study is to test the Technical, Operational and Economical feasibility for adding new modules and debugging old running system. All systems are feasible if they are given unlimited resources and infinite time. There are aspects in the feasibility study portion of the preliminary investigation:

- Technical Feasibility
- Operation Feasibility
- Economic Feasibility

➢ Technical Feasibility

The technical issue usually raised during the feasibility stage of the investigation includes the following:

- Does the necessary technology exist to do what is suggested?

- Do the proposed equipment's have the technical capacity to hold the data required to use the new system?

- Will the proposed system provide adequate response to inquiries, regardless of the number or location of users?

- Can the system be upgraded if developed?

- Are there technical guarantees of accuracy, reliability, ease of access and data security?

➢ Operation Feasibility

User-friendly

Customer will use the forms for their various transactions i.e. for adding new routes, viewing the routes details. Also, the Customer wants the reports to view the various transactions based on the constraints. These forms and reports are generated as user friendly to the Client.

Reliability

The package wills pick-up current transactions on line. Regarding the old transactions, User will enter them in to the system.

Security

The web server and database server should be protected from hacking, virus etc.

Portability

The application will be developed using standard open source software (Except Oracle) like Java, tomcat web server, Internet Explorer Browser etc. this software will work both on Windows and Linux o/s. Hence portability problems will not arise.

Availability

This software will be available always.

Maintainability

The system uses the 2-tier architecture. The 1st tier is the GUI, which is said to be frontend and the 2nd tier is the database, which uses My-SQL, which is the back-end. The front-end can be run on different systems (clients). The database will be running at the server. Users access these forms by using the user-ids and the passwords.

➢ Economic Feasibility

The computerized system takes care of the present existing system's data flow and procedures completely and should generate all the reports of the manual system besides a host of other management reports.

It should be built as a web-based application with separate web server and database server. This is required as the activities are spread throughout the organization customer wants a centralized database. Further some of the linked transactions take place in different locations.

CHAPTER 5

SYSTEM DESIGN

System design phase is one of the most important phases in development. This phase will tell how the system would look like and how it works. During this phase, the software's overall structure is defined and logical system of the product is developed in this phase. It helps in specifying hardware and application requirements and also helps in defining overall application architecture. It describes desired features and operations in detail and may include system architecture model, ER diagram, dataflow diagram, use-case diagram activity diagram, sequence diagram and other documentation to know more details about the flow of the system.

5.1 SYSTEM ARCHITECTURE

A System architecture is the conceptual model that defines the structure, behavior, and more views of a system. An architecture description is a formal description and representation of a system, organized in a way that supports reasoning about the structures and behaviors of the system. A system architecture can consist of system components and the sub-systems developed, that will work together to implement the overall system.
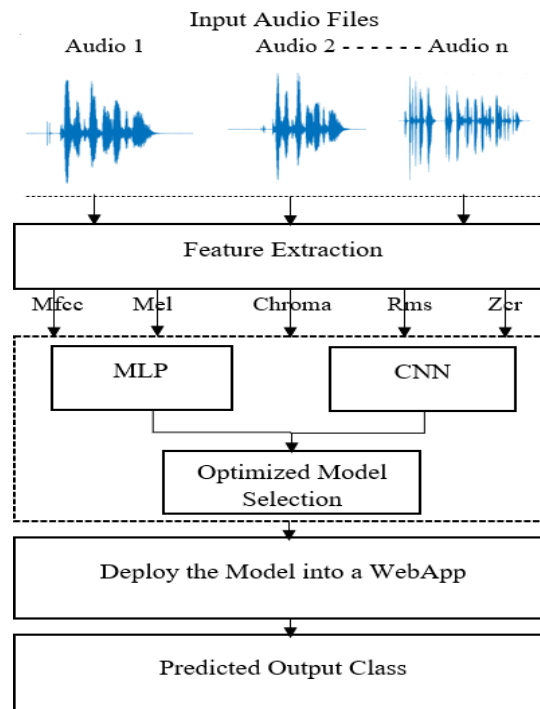
Fig.5 System Architecture diagram.

The system architecture is shown in Fig. 5, the emotion detection process involves two phases in the system they are; The training phase and the testing phase. The left side shows the training phase, and the right side shows the testing phase. So, firstly we need to divide our dataset into two steps to train and test our model. After splitting our dataset, we need to load the data and perform two processes they are extracting the features from the dataset and then applying the various classifiers to detect the exact emotion from the given input voice signal. After completing training and testing by performing feature extraction and then applying the classifier, we need to check the accuracy of our model. Within this which classifier gives more accuracy, we need to save that particular model so that we can deploy that model into a WebApp using Flask framework. By using this WebApp, the user can give the input audio file. This web app will interact with the proposed model and detect the exact emotion from the given input audio file, and returns that prediction to the user.

5.2 ACTIVITY DIAGRAM

Activity diagrams are graphical representations of workflows of stepwise activities and actions with support for choice, iteration and concurrency. In the Unified Modeling Language, activity diagrams can be used to describe the business and operational step-by- step workflows of components in a system. An activity diagram shows the overall flow of control

The activity diagram (Fig. 6) shows that the user first gives a voice input file using the web app. If the user gives the correct voice input file, he can request emotion from the web app by clicking to make prediction button. If the user's input file is not correct or has chosen some other voice file that is not supported, the user must choose another voice file that was supported and relevant. After giving the correct input file, the user can request the emotion of that particular voice input file from the web app. Web app saves the given input voice file in the system and extracts features from it, and gets the saved model path by interacting with the system. Using these selected features and saved model web app maps the selected features with particular emotion by using the saved model for making a prediction, and then finally displays the output for the given input voice file.
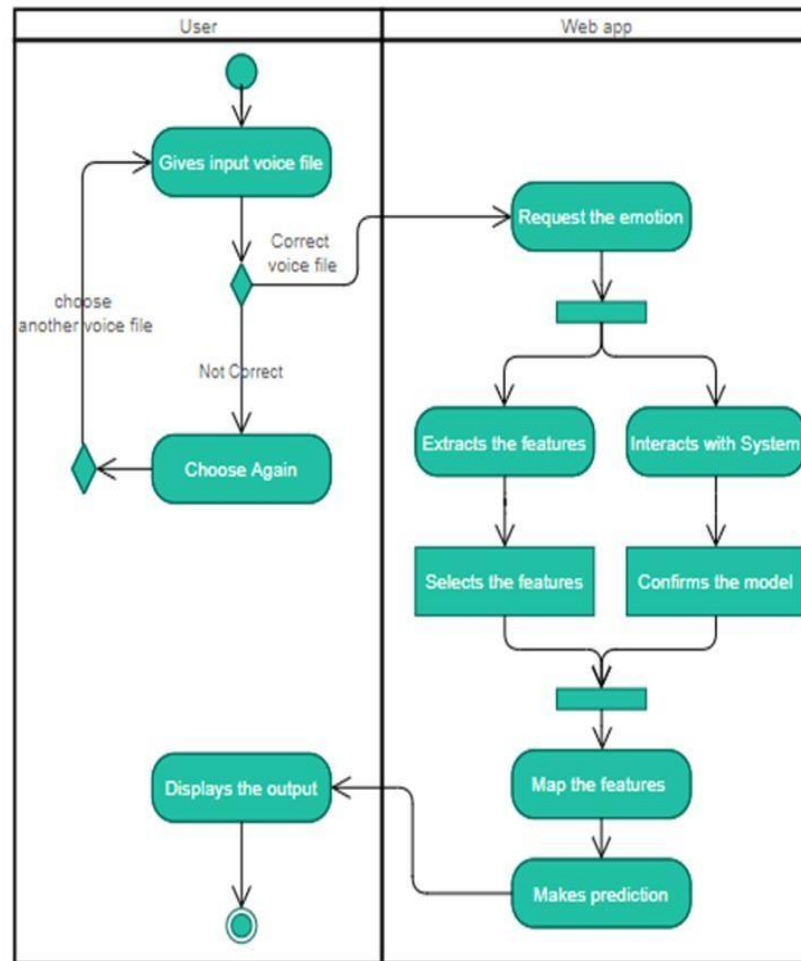
Fig. 6   Activity diagram

5.3 RUNTIME PIPELINE

Fig.7 shows the runtime pipeline of the system where firstly, the user gives the voice input file by requesting emotion of that voice file to the web app. Web app extracts the features from the given input voice file and selects some features from that voice file. Using the best model among two models, MLP or CNN maps these features with a specific emotion, which gives the maximum probability for that emotion, makes the prediction, and finally displays the result in the front-end.
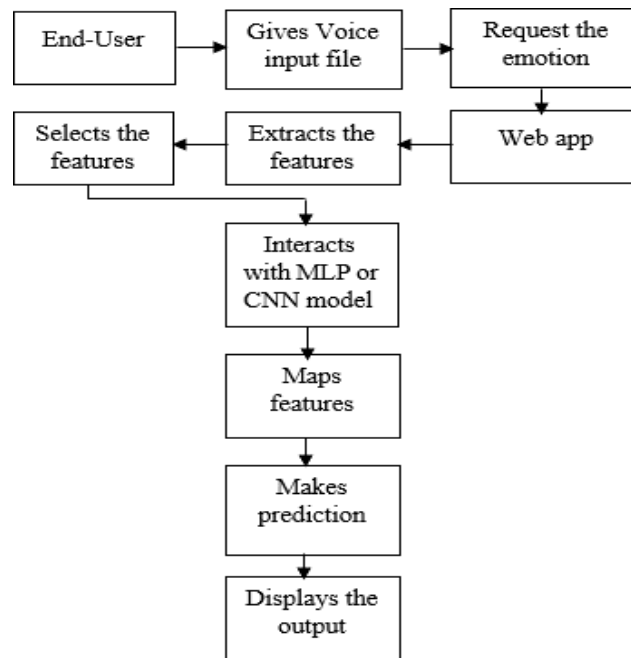
Fig. 7 Runtime Pipeline of the System.

CONCLUSION

In the paper brief introduction about speech emotion recognition is given along with the speech emotion recognition system block diagram description. In the field of affect detection, a very important role is played by a suitable choice of speech database. For good emotion recognition system mainly three databases are used. On the basis of ability, they have to recognize a speech recognition system can be separated in different classes are isolated, connected, spontaneous and continuous words. Relevant emotional features extraction from the speech is the second important step in emotions recognition. To classify features there is no unique way but preferably acoustic and linguistic features taxonomy is considered separately. This technique further helps us in implementing the process of detecting speech in various machines such as cars where the car get automatically gets stopped when a person is facing anger issues. We can attach this technique with various industries such as cyber security, call centers, military security purposes, robotics etc. With this emotion detection techniques the psychiatric department can grow very vast where the doctors can easily detecting the people's emotions.

REFERENCES

1) https://www.sciencedirect.com/science/article/pii/S1877050920318512

2) https://www.researchgate.net/publication/348834462_Machine_Learning_ Based_Speech_Emotions_Recognition_System

3) https://www.analyticsinsight.net/speech-emotion-recognition-ser-through-machine- learning/

4) https://en.wikipedia.org/wiki/Emotion_recognition

5) https://www.sciencedirect.com/topics/computer-science/emotion- recognition

6) https://recfaces.com/articles/emotion-recognition