Table of Contents

List of Figures

Abstract

This project aims to develop a speech emotion recognition system that can accurately detect the emotion of a speaker from their speech. The system will use machine learning algorithms to analyze the audio signal and extract features such as pitch, intensity, and spectral characteristics. These features will then be used to classify the emotion of the speaker. The system will be evaluated using a dataset of labeled audio recordings of different emotions. The results of this project will provide insight into how well machine learning algorithms can detect emotions from speech and could potentially be used in applications such as virtual assistants or customer service systems. With an emotional counselor, the chatbot will provide support and advice to users who are struggling with their emotions. The chatbot will be able to detect the user's emotional state and provide appropriate advice and support. It will also be able to detect any signs of depression or anxiety and refer the user to a professional counselor if necessary. The chatbot will use natural language processing (NLP) techniques to understand the user's input and generate appropriate responses. The chatbot will also be able to remember previous conversations with the user so that it can provide more personalized advice in future interactions.

Chapter 1: Introduction

Speech emotion recognition (SER), which is crucial for human-computer interaction and emotional health, has attracted increasing attention in recent years. In a number of industries, including healthcare, entertainment, customer service, and human-robot interaction, the capacity to reliably detect and interpret emotions from voice signals has important ramifications. The Flask web framework and Python programming language are used to implement a thorough study on Speech Emotion Recognition using Artificial Neural Networks (ANN) in this project report. The objective of this research is to create an effective and trustworthy system that can automatically detect emotions from voice signals. The project makes use of Artificial Neural Networks, a subfield of deep learning that imitates brain function, to assemble intricate characteristics and patterns from speech data. Python is used because of its extensive ecosystem of libraries and tools for machine learning, signal processing, and web development. The SER system's web interface is made with the aid of Flask, a compact web framework. The purpose of speech emotion recognition is discussed in the report's opening, along with its importance in a number of practical applications. Following that, it provides a summary of the pertinent literature while looking at the current SER strategies, methodology, and problems. Understanding the advancements made in this project is based on this. The project's many phases, such as data collection and preprocessing, feature extraction, training, and optimization, are described in the methodology section. Modelling an artificial neural network and assessing the effectiveness of the system. The selection and creation of pertinent elements that capture emotional cues in voice signals are given particular consideration. The Python-based development approach is described in detail in the implementation section, including how libraries like TensorFlow, Keras, and NumPy were used to create and train the ANN model. The deployment of the SER system is facilitated by the integration of the Flask web framework, which gives users a simple web interface through which to interact with the model. Additionally, the evaluation part displays the findings from testing the created system on benchmark datasets, demonstrating its precision, recall, and accuracy in identifying various emotions. It also talks about the system's performance indicators and contrasts them with cutting-edge SER methods. Last but not least, this project report proposes an ANN-based method that is put into practice in Python using the Flask web framework with the intention of advancing the field of speech emotion recognition. Emotional chatbot counsellor, the purpose of it is to provide a compassionate and empathetic space for individuals to express their emotions and seek guidance. It understands emotions that are sometimes complex and overwhelming. The main purpose of the chatbot listen without judgment. Whether one is feeling sad, anxious, angry, or confused, the chatbot is ready to offer a listening ear and help the user to navigate through their emotions. Through the conversations between the user and the chatbot, the aim is to provide support, offer coping strategies, and guide the user towards a greater understanding of yourself and your emotional well-being. Advancements in a variety of applications that call for human-like interaction may be possible because to the system's capacity to reliably identify emotions from speech signals. Future research and development in this field can be built on the conclusions and insights acquired from this study.

Chapter 2: Related Work

There has been a lot of discussion about the use of facial expression recognition in social science and human-computer interaction. Deep learning developments have led to gains in this area that surpass the accuracy of a human. This article examines several popular deep learning techniques for identifying emotions while employing the eXnet library to increase precision. Contrarily, memory and computation still pose challenges. A concern with large models is overfitting. Bringing down the generalisation error is one way to overcome this problem. We use a brand-new CNN called eXnet to build a new CNN model that makes use of concurrent feature extraction. Despite having a lot fewer parameters, the most recent eXnet (Expression Net) model outperforms the inaccurate previous model. strategies for data augmentation with the generalized eXnet, techniques that have been in use for years are being used. It makes use of efficient techniques to lessen overfitting while keeping overall size under control.

The primary form of communication for the world's deaf and dumb population is sign language. Communication between a verbally challenged person and a normal person has, however, never been easy. A breakthrough in communication for deaf-mute people is sign language recognition. The commercialization of a cost-effective and precise recognition system is a current global problem for researchers. Due to their higher accuracy and simplicity, image processing and neural network-based sign language recognition systems are chosen over gadget methods. The purpose of this research is creating a neural network-trained sign language recognition system that is both user-friendly and accurate, generating both text and speech from the input gesture. In addition, a model for text to sign language production that permits two-way communication without the use of a translator is presented in this study.

Researchers are currently very interested in the subject of voice emotion recognition. Using machine learning algorithms such as k-nearest neighbor, multi-layer perceptron, convolutional neural network, and random forest, this study offers many techniques for extracting emotions from audio inputs. From the Berlin database of emotional speech, short-term Fourier transform spectrograms and mel frequency cepstral coefficients were retrieved. Spectrograms were fed into CNN as input. While MLP, random forest, and k-NN each received MFCC features. Each categorizer categorized seven emotions (happy, sad, angry, neutral, disgust, boredom, and fear) with satisfactory results, but the MLP classifier stood out with an overall accuracy of 90.36%. Also offered is a comparison of how well certain categorization methods perform.

In this study, it is recommended to use a Gaussian Probabilistic Linear Discriminant Analysis (GPLDA) back-end to classify emotions at the utterance level using i-vectors that capture the distribution of MFCC features at the frame level. The GPLDA back-end outperforms an SVM-based back-end while being less susceptible to i-vector dimensionality, according to experimental results based on the IEMOCAP corpus. This makes the proposed framework more robust to parameter tuning throughout system development.

Reconstruction-error-based (RE-based) method is used for continuous emotion identification from speech in order to improve performance. Recurrent neural networks (RNN) with memory enhancement for learning. The framework adopts two successive RNN models, the first of which serves as an autoencoder to rebuild the original characteristics and the second of which is utilized to forecast emotions. As a complementary description, the RE of the original characteristics is combined with the original features and provided to the second model. This paradigm makes the premise that the system can recognize its "drawback," which is represented by the RE. The suggested framework greatly outperforms the baseline systems without any RE information in terms of Concordance Correlation Coefficient (.729 vs..710 for arousal,.360 vs..237 for valence), according to experimental results using the RECOLA database. surpasses other state-of-the-art techniques by a substantial margin.

In order to demonstrate the dispersion of the MFCC's effectiveness, Gamage et al. suggested using a Gaussian inquiry that can differentiate discussion emotion level based on i-vectors. An evaluation based on the IEMOCAP corpus reveals that the GPLDA's foundation is stronger than the SVM's foundation and less sensitive to the i-vector, making the normal level more potent for changing rules during framework improvement.

Han et al. suggested that the focus is on the ability to continuously recognize emotions from speech, and supporting an organization that recursively produces memory as well as an instructive error-based approach to learning. In this way, the basic model is used as a computerized tool with the aid of two continuous RNN (Recurrent Neural Networks) techniques. In order to forecast with passion, code is used to recover the original substance. As an additional resource, RE (Reconstruction-error-based) of the primary resource is used, matched to the primary resource, and placed in the next class.

SVM is used in the current system to identify the speaker's emotion. Support Vector Machine (SVM) is a type of supervised machine learning that has been applied to both classification and regression issues.

Chapter 3: Problem Statement and Objectives

The creation of an integrated system for speech emotion recognition (SER) and a chatbot is the issue this project attempts to solve. The goal is to reliably identify and decipher emotions from voice signals and use this knowledge to build an intelligent chatbot that can converse with users in a natural and sensitive way.

Construction of a powerful Speech Emotion Recognition (SER) system. The main goal is to create and put into use a system that can recognize and categorize emotions as they are communicated in voice signals. To attain high accuracy and reliability, the system should make use of the proper signal processing techniques, feature extraction techniques, and machine learning algorithms. The chatbot need to make use of the identified emotions to customize its responses and have heartfelt, sympathetic dialogues with users. Create an intuitive user interface: Both the SER system and the chatbot should have user-friendly interfaces created to guarantee user accessibility and simplicity of interaction. Users should be able to receive feedback on emotion recognition, and have intuitive dialogues with the chatbot through the UI. Natural language processing techniques are used to train the chatbot: Natural language processing techniques should be used to educate the chatbot to comprehend user inquiries, produce acceptable responses, and sustain coherent and situationally relevant discussions. It ought to be able to manage a variety of emotions that users convey. Analyze and improve performance of the system: The system's performance will be assessed in terms of user satisfaction, chatbot response, and the accuracy of voice emotion recognition. To enhance the system's effectiveness, responsiveness, and overall user experience, performance optimization approaches should be used. By attaining these goals, this project hopes to advance the study of human-computer interaction by creating an empathetic and intelligent chatbot. The integrated system may improve user experiences, offer sympathetic support, and enable more interesting and interactive interactions between people and technology. Developing an emotional counsellor chatbot to provide support and guidance to individuals experiencing emotional distress, anxiety, and other mental health issues. The chatbot should offer a compassionate and understanding virtual environment where users feel comfortable discussing their emotions and receiving guidance to improve their emotional well-being.

The chatbot's objectives are to provide a safe and supportive environment by creating a chatbot interface that encourages users to open up about their emotions and fosters a sense of trust and safety, emotion recognition and understanding by developing the ability for the chatbot to accurately recognize and understand various emotions expressed by users through text or contextual cues, active listening and empathy by training the chatbot to actively listen to user concerns, express empathy, and provide non-judgmental responses to create a supportive and compassionate interaction, tailored guidance and resources by offering personalized guidance and coping strategies based on individual emotional needs, using evidence-based techniques and resources for emotional well-being, referral and crisis management by incorporating protocols to identify users in crisis or in need of immediate professional help, and provide appropriate referrals to mental health professionals or helpline services when necessary, continuous learning and improvement by continuously updating and enhancing the chatbot's knowledge base by analyzing user interactions, feedback, and incorporating new research and best practices in emotional counselling, user privacy and data security by ensuring the chatbot adheres to strict privacy policies and data security measures to protect user information and maintain confidentiality, user-friendly interface and accessibility by designing a user-friendly and intuitive interface that is accessible across various platforms and devices, enabling easy and convenient access for individuals seeking emotional support, monitoring and evaluation by implementing mechanisms to monitor the chatbot's effectiveness in providing emotional support, gather user feedback, and conduct regular evaluations to improve its performance and user satisfaction and ethical considerations by ensuring the chatbot adheres to ethical guidelines, such as maintaining user confidentiality, providing transparent information about its capabilities and limitations, and avoiding harm or inappropriate advice. By addressing these objectives, the emotional counsellor chatbot aims to provide individuals with a reliable and accessible resource for emotional support, helping them manage their emotions and improve their overall well-being.

Recent advances in machine learning and natural language processing have made it possible to create chatbots that can understand and react to user input. Contrarily, traditional chatbots are unable to determine the user's emotional state, which limits their ability to provide intelligent and customized responses. The intention is to use this project to develop a chatbot system that can recognize a user's emotional state and respond appropriately. The project's aim is to design a webpage with both speech emotion recognition and chatbot. After recording an audio, input it in the Speech Emotion Recognition module and get the predicted emotion. This helps the user to know what he/she is feeling at the moment. Converse with the chatbot according to what you feel. This can make you feel better as the chatbot aims to be the user's companion and may also suggest ways to make you feel better.

The project calls for a selection of speeches that demonstrate various emotional states (such as happy, sad, etc.). The dataset is used to build a machine-learning model to recognize spoken emotional states. Once the model has been trained, it is incorporated into a proper webpage. The emotional chatbot with the help of an API is able to examine user input after receiving it in order to determine the user's emotional state. The chatbot then responds in a way that considers the user's emotional condition.

The chatbot's ability to understand the user's emotional state and reply properly and the proper detection of emotion from speech input will be key to the project's success. The project aims to improve user experience and engagement. As a tool for communicating with customers, chatbots are becoming more and more popular. A project on Speech Emotion Recognition (SER) and Chatbots was created to give users who communicate with chatbots a more tailored and sympathetic experience and to help them find what they feel. The SER model might increase customer satisfaction and engagement when used as an application in the industry. The project with more improvisations and work could possibly be advantageous for the offering of mental health care. This could be quite helpful for folks who might be reluctant to seek therapy or counselling from a real person. With the use of chatbot technology, SER hopes to enhance user experience and offer more personalized and empathetic interactions between users and chatbots.

Emotions	Pitch	Intensity	Speaking rate	Voice quality
Anger	abrupt on stress	much higher	marginally faster	breathy, chest
Disgust	wide, downward inflections	lower	very much faster	grumble chest tone
Fear	wide, normal	lower	much faster	irregular voicing
Happiness	much wider, upward inflections	higher	faster/slower	breathy, blaring tone
Joy	high mean, wide range	higher	faster	breathy; blaring timbre
Sadness	slightly narrower	downward inflections	lower	resonant

Figure 1: Speech Variations with Emotions

Chapter 4: Project Analysis and Design

4.1 Hardware and Software Requirement Specifications

Hardware Requirements:-

System	:	Pentium i3 Processor.
Hard Disk	:	500 GB.
Monitor	:	15’’ LED
Input Devices	:	Keyboard, Mouse
Ram	:	4 GB

Software Requirements:-

Operating system	:	Windows 10.
Coding Language	:	Python
Web Framework	:	Flask.

Software Environment

Python is a powerful, interpreted, object-oriented, and interactive scripting language. Python has been created to be very readable. It has fewer syntactical structures than other languages and typically employs English keywords rather than punctuation.

Framework for Flask

Python is used to create the Flask web application framework. It is created by Armin Ronacher, the president of Pocco, a global organization of Python aficionados. The Werkzeug WSGI toolkit and Jinja2 template engine form the foundation of Flask. Each are Pocco initiatives.

Python 3 is the most recent major version even though Python 2 continues to be used widely despite receiving only security updates. That is tutorial A text editor will be used to write Python. Python can be written in an Integrated Development Environment (IDE), such as Thonny, Pycharm, Netbeans, or Eclipse. These IDEs are especially helpful for organizing huge collections of Python files. Python was created with readability in mind and, thanks to mathematical influence, shares certain parallels with the English language. In contrast to other programming languages, which frequently employ semicolons or parentheses, Python uses new lines to finish a command. Indentation, which utilizes whitespace, is how Python defines scope, including the scope of loops, functions, and classes. Curly brackets are frequently used in other computer languages for this reason.

Financial viability

The project is financially viable and can be conducted with minimal investment and resources in today’s time. The software and hardware requirements are mentioned above in the report. Apart from those requirements nothing else is needed for the fulfilment of the project on “Speech Emotion Recognition and Chatbot”.

Technical Capabilities

This study is being done to evaluate the system's technical requirements or technical feasibility. Any system created must not place a heavy burden on the technical resources at hand. The number of technological resources available will be heavily strained. As a result, the client will face high expectations. The created system must have a low demand because its implementation merely necessitates little or no adjustments.

Social Feasibility

The goal of the study is to determine how much the user accepts the system. This includes the instruction needed for the user to operate the system effectively. The user must not perceive the system as a danger and must instead accept it as a requirement. The techniques used to inform and acquaint the user with the system are the only factors that affect the level of acceptance by the users. As the system's ultimate user, his confidence must be increased so that he may offer some helpful criticism, which is encouraged.

4.2 Use Case Diagrams, Flow Chart/ Activity Diagram etc.

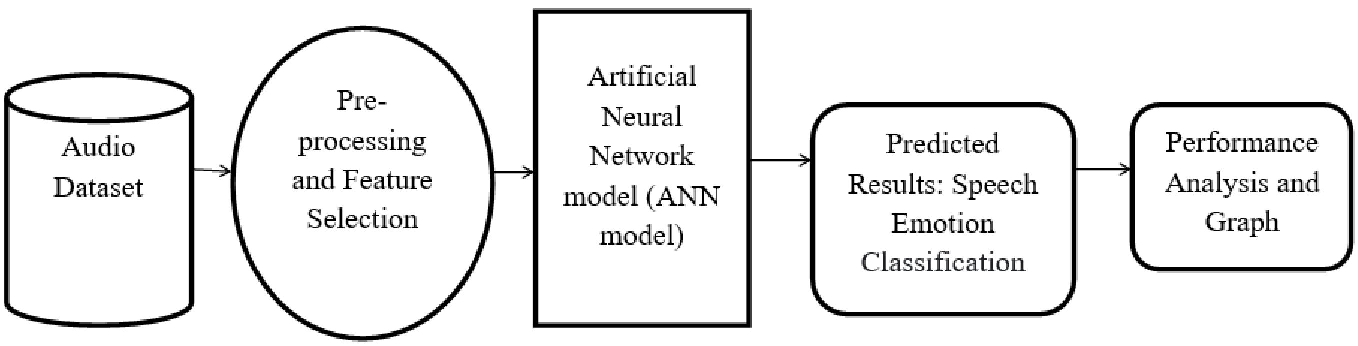

Figure 2: Proposed Model

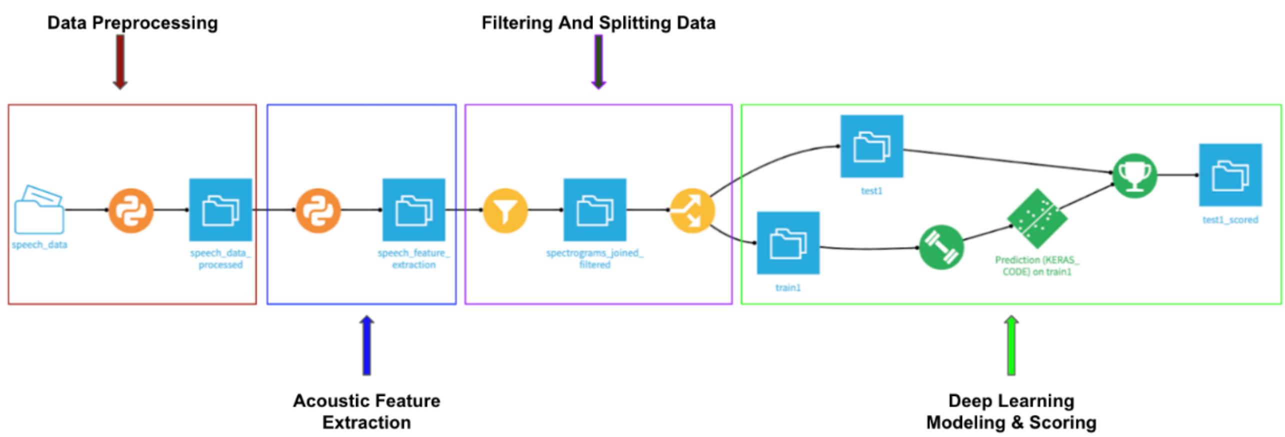

Figure 3: Machine Learning Model

The elements of the human emotion detection system are input speech signal, preprocessing, acoustic feature extraction and selection, filtering and splitting data, and deep learning model and scoring. The emotion recognizer system analyses the incoming speech signal for emotional states and displays the relevant feelings for that specific utterance.

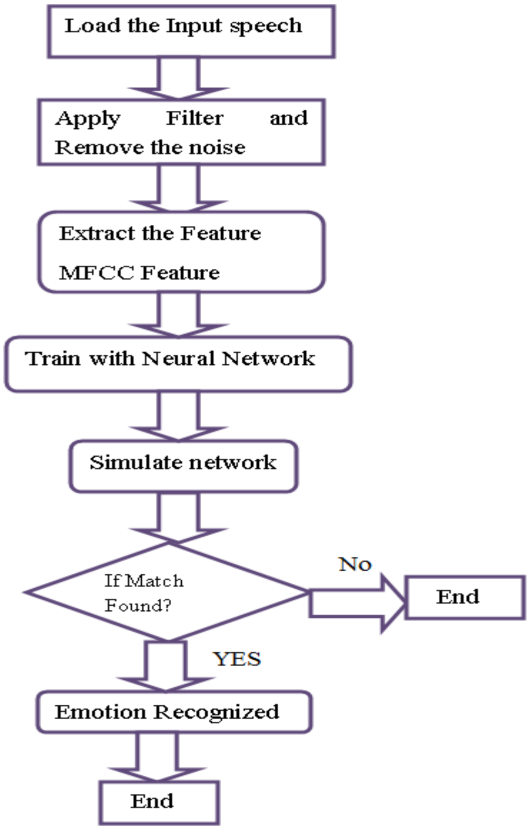

Figure 4: SER flowchart

First, load the signal and input the signal that needs to be verified. The input and the data sample ought to be of the same frequency and time period. The noise has been reduced in this step. Here, noise removal only smooths the signal. The recorded or input sample data don't need to be changed. High pass filtering is employed in the suggested system to take the noise out of the input signal. After that, the voice sample's feature must be retrieved. So, to extract the features, use the Mel Cepstral Coefficient. A neural network is used to train on the input and loaded data. The results from the simulated data are checked in this step using the fuzzy theory approach. With neural networks, network simulations are carried out. If the output matches the target, the emotion is recognized, displayed, and the voice is also played.

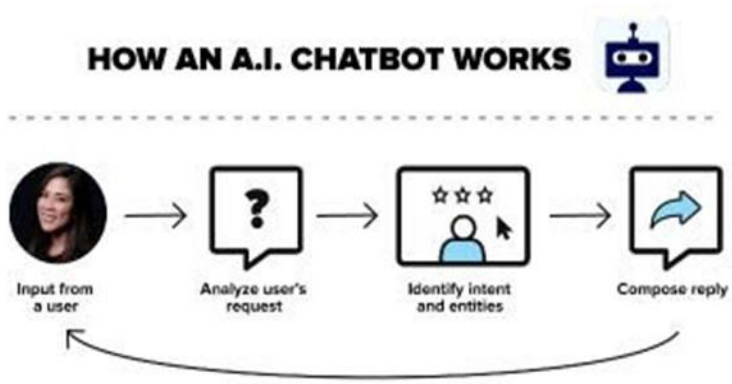

Figure 5: Working of AI chatbot

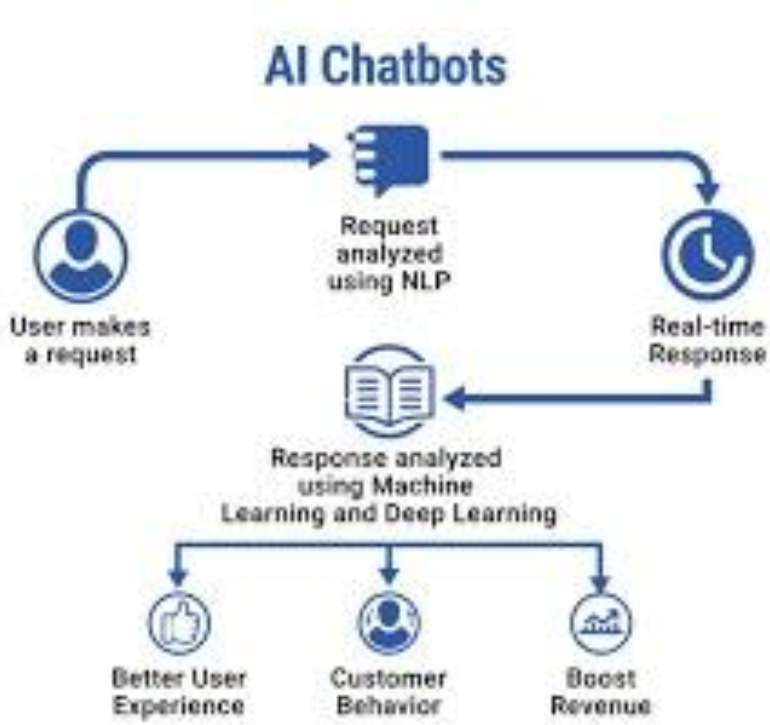

Figure 6: Chatbot

Chapter 5: Proposed Work and Methodology Adopted

Enabling effective and natural human-computer interaction is the aim of computing. Making it feasible for computers to understand the emotional states that individuals display so that customized solutions can be offered is one of the most important goals. Practical applications are hindered because the majority of studies in the literature focus solely on identifying emotions from brief, isolated utterances. In the proposed system (ANN model), vocal emotion recognition is implemented using artificial neural networks. The suggested system, which consists of seven distinct emotion categories, was developed based on Kaggle trials utilizing pre-recorded datasets. The target test accuracy is 99%, and the target train accuracy is 100% for the suggested system.

The system asks training data, such as expression labelling and weight training within that network. There is an audio file input. After that, the audio is normalized for intensity. The ANN is trained using a normalized audio to avoid having the presentation order of the samples have an impact on the training results. When combined with the learning data, the weight collections generated by this training technique yield the best results. Based on the final network weights learned and the system's pitch and energy during testing, the dataset returns the emotion during identification.

Advantages of Proposed System:-

Artificial neural networks have the capacity to give data for parallel processing, allowing them to tackle multiple tasks at once. Resistance to artificial neural networks has existed. This implies that the performance of artificial neural networks is affected when one or more cells, or neural networks, are lost. Artificial neural networks are used to store data so that even in the absence of a data pair, the network is still capable of producing results. Artificial neural networks are progressively dissipating, so they won't suddenly stop functioning. Because of this, these networks are gradually dissipating. ANNs are taught such that they can draw lessons from the past events to take decisions.

Methodology for SER model:-

Dataset:

A system was created in the first module to obtain the input dataset for training and testing. The model folder contains the dataset. 2,800 Speech Emotion audio datasets make up the dataset. The dataset includes categories such as anger, contempt, fear, joy, neutrality, sadness, and surprise. The Kaggle website is where the dataset is located.

Importing the necessary libraries:

The required libraries were imported for voice emotion recognition system in the second module. Librosa is a significant and outstanding library that facilitates audio and music analysis. To install the library, just use the Pip command. It offers the necessary building components to create a music-based information retrieval paradigm. TensorFlow is a fantastic package that is used for deep learning models. Hence, it is loaded.

Exploratory Data Analysis of Audio data:

Under the dataset folder, there are other folders. To comprehend how to load audio files and how to view them as waveforms before any preprocessing is done. Use the IPython module and explicitly provide an audio file path to load the audio file and play it. The folder's first audio file was taken. Now load audio data via librosa. Librosa provides with 2 things when any audio file is loaded. Sample rate is one, and a two-dimensional array is the other.

Use librosa to load the aforementioned audio file and plot the waveform. The sample rate is a measure of how there are several samples taken per second. Librosa reads the file by default at a sample rate of 2,800. Depending on the library you select, the sampling rate changes. The first axis of a 2-D array represents recorded amplitude samples. The number of channels is represented on the second axis. There are various kinds of channels, including stereo (which has two channels) and monophonic (which has one channel). The data is loaded into librosa, which normalizes it all before attempting to deliver it at a single sample rate. The Scipy Python Library allows to accomplish the same thing. Additionally, it will provide with two pieces of data and one item of information & sample rate. It is not the same as librosa when printing the sample rate. One crucial distinction between the two is that while librosa data can be normalized when printed, the same cannot be said for audio files when read using Scipy. The following three factors explain why librosa is currently becoming more and more popular for audio signal processing.

1. The signal is attempted to be converted into mono (one channel).

2. A regular pattern can be seen because it can represent the audio signal between -1 and +1 (in normalized form).

3. It can also observe sample rates, which are by default converted to 22 kHz, unlike other libraries where sample rates are displayed in accordance with different values.

<u>Imbalance Dataset check:</u>

Understand what audio files are and how to view them in audio format. Load the CSV data file provided for each audio track while moving from format to data exploration and count the number of records each class has. Examine the records for each class using the value counts function. While looking at the results, it is noticed that the data is balanced and that most classes have roughly equal numbers of records.

<u>Data Preprocessing:</u>

Certain audio files are being recorded at different rates, such as 44 or 22 kHz. It will be at 22 KHz when using librosa, making the data in a normalized pattern observable. Now that the data is in the form of independent (extracted features from the audio signal) and dependent (class labels) characteristics, the objective is to extract some key information. Extract independent characteristics from audio streams using Mel Frequency Cepstral coefficients.

MFCCs - The frequency distribution throughout the window size is summarized by the MFCC. Therefore, it is feasible to examine the sound's frequency and time properties. Locate features for categorization using this audio representation. Therefore, based on time and frequency, it will attempt to transform audio into some kind of characteristics that facilitate categorization.

First apply MFCC to a single audio file that is already in use in order to show how it is used in practice. Now construct the data frame and extract features from each audio sample. Therefore, write a function that accepts the filename (or file path, if applicable). Obtain 2 details after the file has been loaded using librosa. The mean of the transpose of an array will be found in order to determine scaled features once the MFCC for the audio data is determined. Now, in order to extract every characteristic for each audio file, utilize a loop through each row in the data frame. To track the development, additionally use the TQDM Python package. The method to extract MFCC features will be called inside the loop, establish a unique file path for each file before calling it to append features and accompanying labels to a freshly created data frame.

Splitting the dataset:

Split the dataset into train and test. 80% train data and 20% test data.

Audio Classification Model Creation:

From the audio sample and splitter in the train and test sets, characteristics are retrieved. Now use the Keras sequential API to create an ANN model. The output shape (number of classes) is 7, and build an ANN with three dense layers. The architecture is described below.

1. There are 100 neurons in the top layer. According to the number of features with Relu's activation function, the input shape is 40, and utilize the Dropout layer at a rate of 0.5 to prevent any overfitting.

2. There are 200 neurons in the second layer that activate in a Relu manner and drop out at a rate of 0.5.

3. There are 100 neurons in the third layer with Relu activation and the drop out at a rate of 0.5.

Compile the Model:

Define the loss function, which is categorical cross-entropy, the accuracy metrics, which is accuracy score, and the optimizer, which is Adam, in order to build the model.

Train the Model:

The model will be trained and saved in HDF5 format. A model will be trained across 100 epochs with a batch size of 32. Utilize a callback, a checkpoint, to determine how long it took to train on the data.

Check the Test Accuracy:

Assess the model using test data. On the training dataset, an accuracy close to 100% was achieved, while on the test data, 99% accuracy was achieved.

Saving the Trained Model:

The first thing to do is store the trained and tested model into a.h5 or.pkl file using a library like pickle once it is ready to be used in a production-ready setting. Verify that Pickle is set up in the environment. The model will now be imported into the module and dumped into an.h5 file.

Methodology for chatbot:-

Data Collection:

Gather a diverse dataset of emotional expressions and conversations from individuals facing various emotional challenges. Annotate the dataset with labelled emotions to train the chatbot in emotion recognition.

Natural Language Processing (NLP) Techniques:

Implement sentiment analysis algorithms to recognize the emotional tone of user input. Utilize Named Entity Recognition (NER) to identify keywords related to emotions, topics, and experiences.

Dialogue Management:

Design a dialogue flow that ensures a conversational and supportive experience for users. Utilize rule-based and machine learning-based approaches to generate appropriate responses based on identified emotions and user input.

Emotional Intelligence:

Incorporate emotional intelligence principles to foster empathy, active listening, and understanding within the chatbot's responses. Implement techniques such as paraphrasing, reflecting, and providing validation to create a supportive environment.

Integration of Resources:

Incorporate relevant resources, such as guided meditation, breathing exercises, self-help articles, and hotline numbers, to provide users with immediate support and coping strategies.

User Feedback and Learning:

Implement a feedback system to gather user input and evaluate the chatbot's effectiveness. Continuously learn from user interactions to improve the chatbot's responses and expand its emotional understanding.

Evaluation:

Conduct user tests and surveys to evaluate the effectiveness and user satisfaction with the emotional counsellor chatbot. Measure the chatbot's ability to recognize and respond appropriately to different emotions. Assess the impact of the chatbot in promoting emotional well-being and providing valuable support.

Chapter 6: Results and Discussion

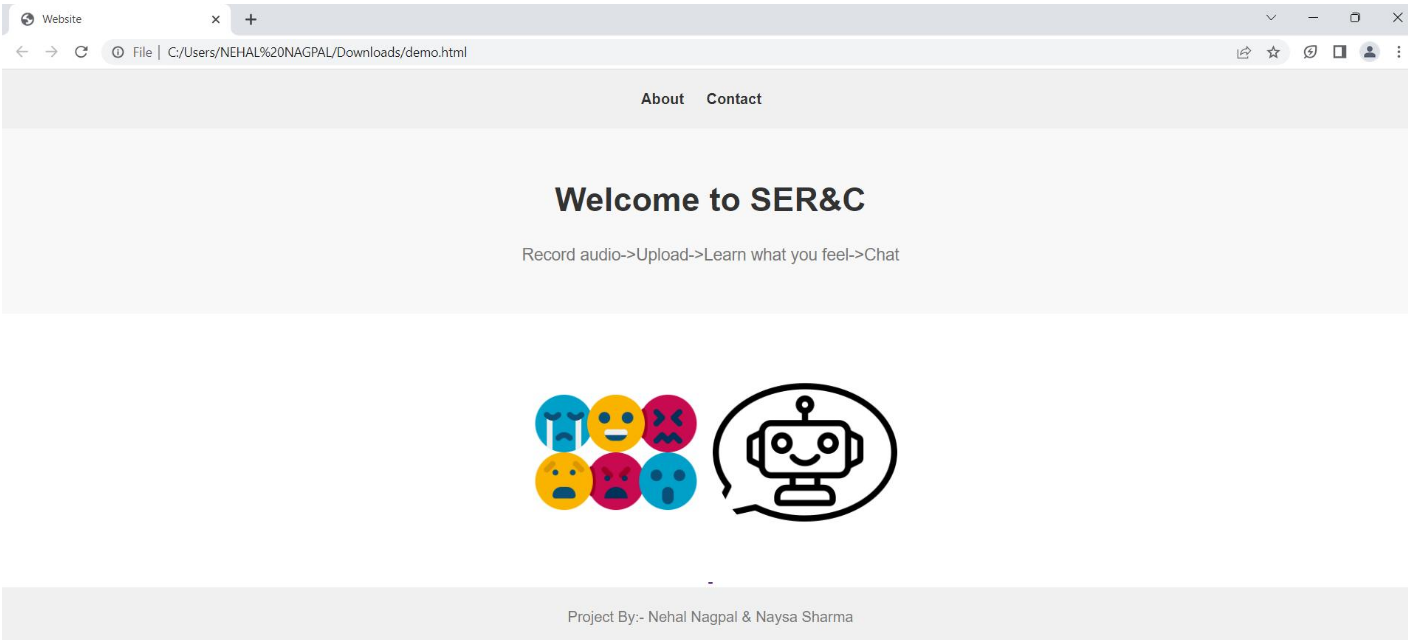

Figure 17: Main Webpage

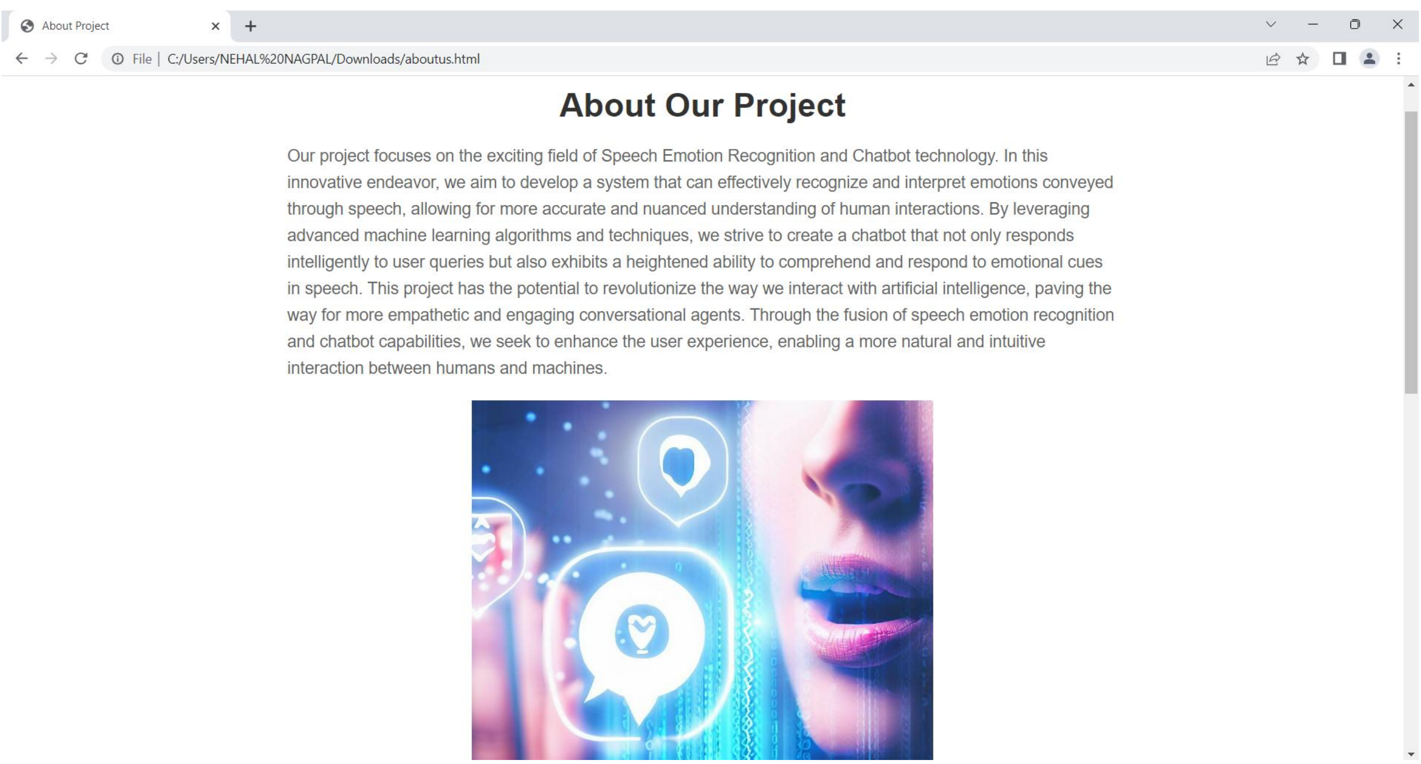

Figure 18: About Our Project Page

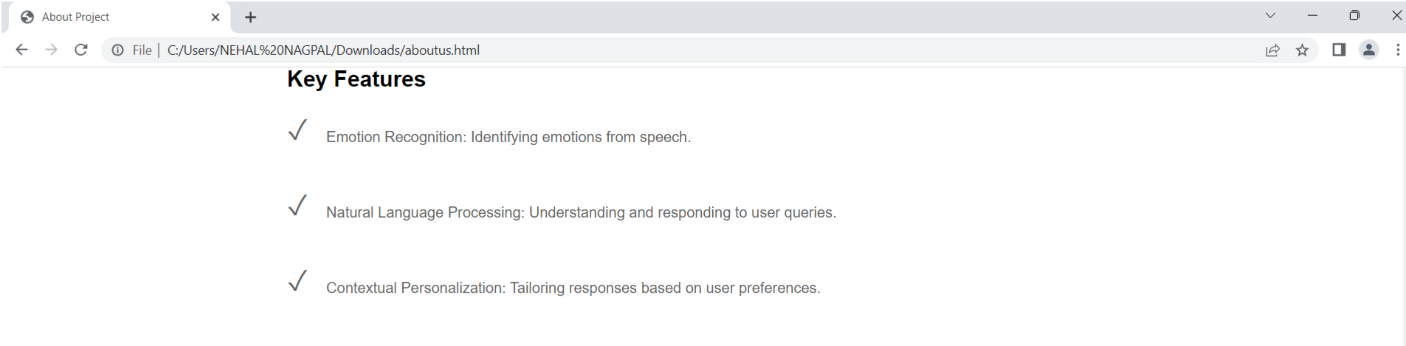

Figure 19: Key Features

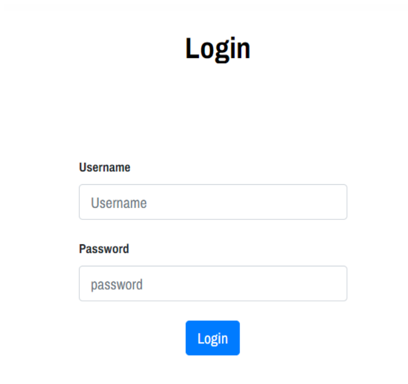

Figure 21: Login

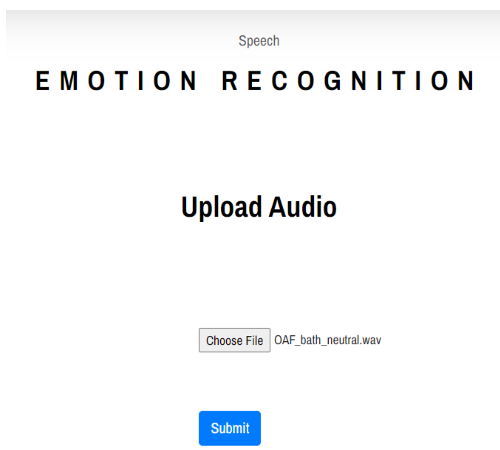

Figure 22: Audio Input

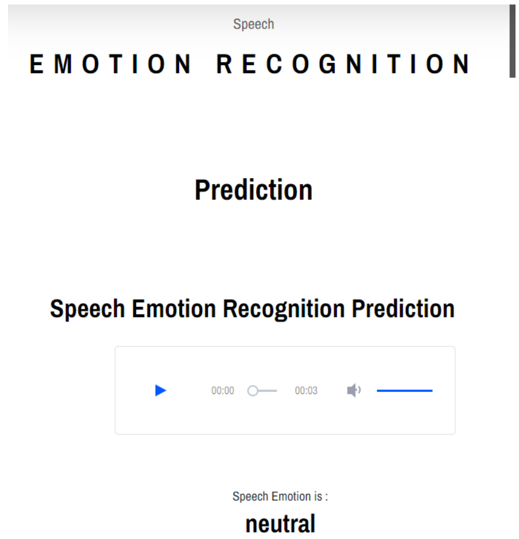

Figure 23: Emotion Detection using SER

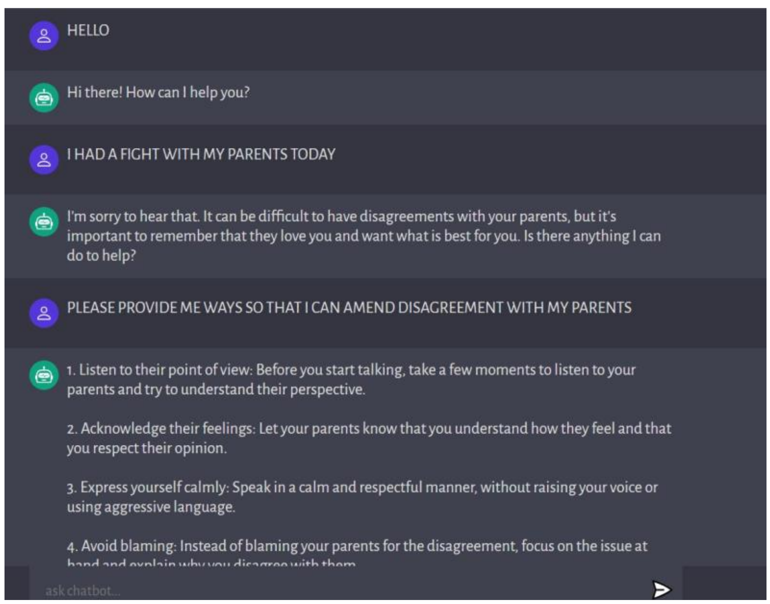

Figure 24: Chatbot Testing 1

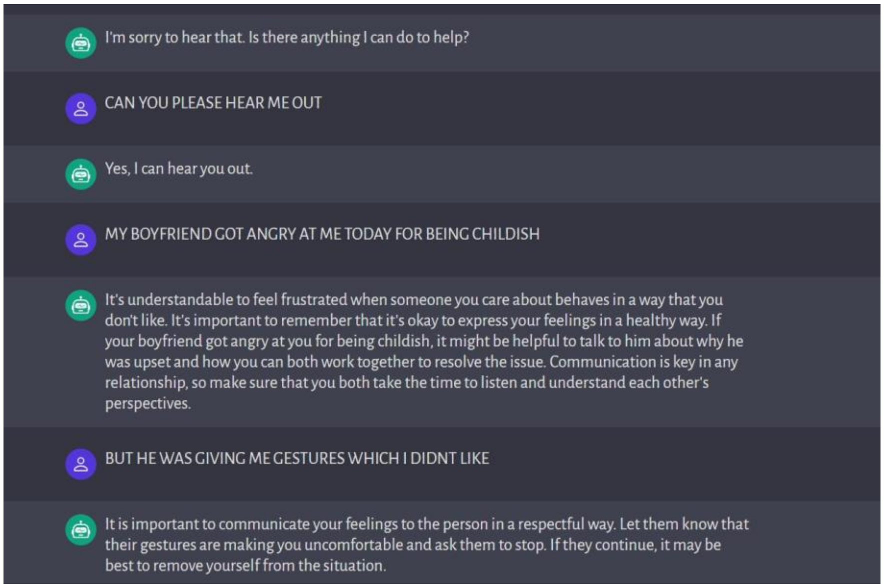

Figure 25: Chatbot Testing 2

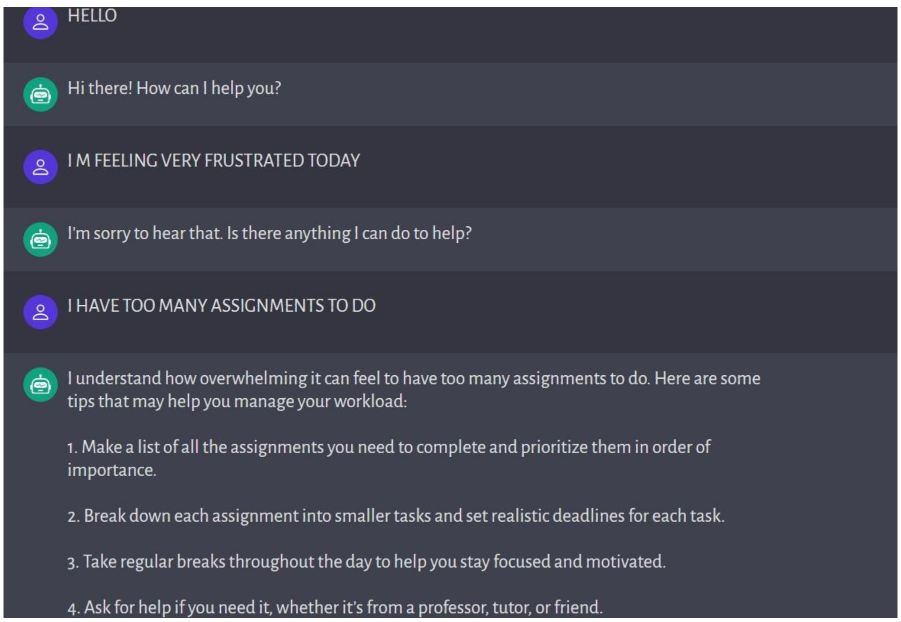

Figure 26: Chatbot Testing 3

Chapter 7: Conclusion

In software engineering research, SER is a fascinating topic. The suggested architecture is similar to how the SER computation is already being done. The suggested architecture might further be expanded to provide multilingual Emotion. Feelings can also be used to describe small scales and design.

The proposed emotional counsellor chatbot aims to provide accessible and effective emotional support to individuals in need. It is expected to demonstrate high accuracy in emotion recognition, offer empathetic and compassionate responses, and provide valuable resources for managing emotional challenges. By developing an emotional counsellor chatbot using natural language processing techniques and incorporating emotional intelligence principles, this project aims to create a supportive environment for individuals seeking emotional guidance. The proposed work and methodology outlined in this document provide a roadmap for the development of an effective and empathetic chatbot capable of positively impacting users' emotional well-being.

Chapter 8: Future Scope of Work

While this project achieved promising results, several limitations and future directions should be considered. First, the speech emotion recognition model's accuracy could be further improved, especially for complex emotions like anger and fear. Additionally, expanding the dataset with more diverse emotional expressions would enhance the model's generalization capabilities.

Furthermore, the chatbot's performance could be enhanced by incorporating sentiment analysis and contextual understanding to capture the subtleties of emotional expressions more effectively. Fine-tuning the responses based on specific emotional contexts and conducting user studies to gather feedback for continuous improvement would be valuable.

Lastly, it is crucial to ensure the ethical use of such systems and address potential biases that may arise from training data or algorithmic decisions. Ongoing efforts should be made to mitigate any unintended consequences and ensure fairness and inclusivity in the development and deployment of emotion recognition and chatbot technologies.

References

Akrami, N. et-al; Speech based emotion recognition and next reaction prediction. In **25th Signal Processing and Communications Applications Conference**, Antalya, pp. 1-6. 2017.

Chourasia, M. et-al; Emotion recognition from speech signal using deep learning**. Intelligent Data Communication Technologies and Internet of Things: Proceedings of ICICI** 2020 (2021): 471-481.

Gamage, K. W. et-al; Ani-vector gplda system for speech based emotion recognition. In 2015 **IEEE Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA)**, pp. 289-292., 2015.

Han, J. et-al; Reconstruct ion-error-based learning for continuous emotion recognition in speech. In 2017 **IEEE international conference on acoustics, speech and signal processing (ICASSP)**, pp. 2367-2371., 2017.

Kaur, J. & Kumar, A.; Speech Emotion Recognition Using CNN, k-NN, MLP and Random Forest. **Springer, In Computer Networks and Inventive Communication Technologies**, pp. 499-509., Singapore, 2021.

Khalil R. A. et-al; Speech emotion recognition using deep learning techniques: A review, **IEEE Access**, vol. 2, no. 7, pp. 117327-117345, 2019.

Kottursamy, K. :A review on finding efficient approach to detect customer emotion analysis using deep learning analysis; **Journal of Trends in Computer Science and Smart Technology**, 3, no. 2 (2021): 95-113.

Livingstone, S. R., & Russo A. F; The Ryerson Audio-Visual Database of Emotional Speech and Song (RAVDESS): A dynamic, multimodal set of facial and vocal expressions in North American English." **PloS** one 13, no. 5 (2018): e0196391

Nwe T. L. et-al; Speech emotion recognition using hidden Markov models, **Speech Communication**, vol. 41, no. 4, pp. 603–623, 2003.

Rieger, S. A. et-al; Speech based emotion recognition using spectral feature extraction and an ensemble of kNN classifiers. In **IEEE the 9th International Symposium on Chinese Spoken Language Processing** (pp. 589-593)., 2015.

Tabatabaei, T. S. & Krishnan Sridhar; Towards robust speech based emotion recognition. 2010 **IEEE International Conference on Systems, Man and Cybernetics**., 2010.

Thakur, A. et-al: Real Time Sign Language Recognition and Speech Generation.; **Journal of Innovative Image Processing**, 2, no. 2 (2020): 65-76.

Vaijayanthi, S. & J. Arunnehru.; Synthesis Approach for Emotion Recognition from Cepstral and Pitch Coefficients Using Machine Learning. In **International Conference on Communication, Computing and Electronics Systems**, p. 515.

Y. Kim et-all; Deep learning for robust feature generation in audio-visual emotion recognition, in Proceedings of the **IEEE International Conference on Acoustics, Speech and Signal Processing** (ICASSP '13), Vancouver, Canada, 2013.

Z. Li: A study on emotional feature analysis and recognition in speech signal, **Journal of China Institute of Communications**, vol. 21, no. 10, pp. 18–24, 2000.

Links:-

https://en.wikipedia.org/wiki/44,100_Hz

https://www.researchgate.net/publication/270899631_Emotion_Recognition_for_Instantaneous_Marathi_Spoken_Words

https://www.researchgate.net/publication/297345239_Emotion_Recognition_through_Speech_Using_Neural_Network

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

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

https://blog.dataiku.com/speech-emotion-recognition-deep-learning

https://www.collabraspace.com/data-science-blog/what-is-chatbot/

https://www.kaggle.com/datasets/jayaprakashpondy/speech-emotion-dataset

.