SPEECH EMOTION RECOGNITION USING LSTM AND CNN

K. Shobhan Babu[1], P. Divya[2], T. Akhila[3], B. Bindu[4], J. Varshitha[5], Ramdas Vankdothu[6]

[2345]BTech Student, Department of CSE, Balaji Institute of Technology and Science, Laknepally, Warangal, India

[1,6] Assistant Professor, Department of CSE, Balaji Institute of Technology and Science, Laknepally, Warangal, India

ABSTRACT

Speech Emotion Recognition (SER) can be an important area in terms of artificial intelligence as it puts deep learning systems into working for the recognition of emotions from speech. Developments in Speech Emotion Recognition (SER) are becoming a flourishing branch of research in human-computer speech interaction enhancement. A robust SER model is expected, entraining deep learning on Convolutional Neural Networks (CNNs) and Long Short-Term Memory networks (LSTMs). The model is trained such that the primary features or emotions it will recognize would be happiness, sadness, anger, and neutral emotive states. Deep learning methods, in contrast to classic methods, outperform such approaches because they are ideally suited to capturing temporal and spatial patterns of speech. Many applications such as a customer service or virtual assistant can take advantage of this developed system and are capable of responding in real-time to the user regarding their emotions.

1.INTRODUCTION

Human speech is more than mere words; it incorporates within its meaning the non-verbal features of tone, rhythm, and intensity, all of which capture emotion. Indeed, Speech Emotion Recognition (SER) is a very broad field of research that aims to capture such meaningful emotional units and put them into action to automate systems more human-friendly. Finally, SER will determine a completely new paradigm for HMI by being capable of automatically detecting feelings such as joy, sadness, anger, and neutrality through voice signals and speech.

Deep learning has worked a lot in favour of the SER enhancement because it works with patterns deriving from so much data. Unlike traditional techniques which depend on hand-made

doi: 10.48047/ijiee.2025.15.4.34

elements, deep learning architecture such as Convolutional Neural Networks (CNN) and Long Short-Term Memory (LSTM) networks has outperformed others on emotion-related features of the speech signals and their corresponding features. Changes like these allow virtual assistants to change their voice according to how the users feel, adaptive learning environments to modify teaching approaches according to students' emotional reactions, and customer care services to generate improved policies on how to serve notice to the feelings and concerns of clients[1-25].

Semi-automated tools for SER via deep learning and the promise of advances in interoperability standards can still save the edges of collections of data, annotation discrepancies or inconsistencies, and the particular emotion and expression culture. Ethical issues concerning information were outside that.

2. LITERATURE SURVEY

Optimized Cross-Corpus Speech Emotion Recognition Framework based on Normalized 1D Convolutional Neural Network - Barsainyan, N., Singh, D.K.(2025): This paper presented a Speech Emotion Recognition model validated for multilingual environment with deep learning approaches that were trained on data augmentation and feature selection techniques to make advanced improvements in classification accuracy for emotions recognition. Achieved 80.2% accuracy beyond generic classifiers on cross corpus emotions. The project was made using high computational resources and was not feasible for real time applications on low power devices.

Real-time Speech Emotion Recognition Using Deep Learning and Data Augmentation - Barhoumi, C., BenAyed, Y (2024): This project is concerned with a real-time system for Speech Emotion Recognition (SER) using deep learning methodologies and data augmentation. Feature extraction and data augmentation techniques have been applied to the system using CNN and BiLSTM models. Model performance is compromised by high computational cost, dataset imbalance, real-life problems, and ambiguity in emotions

Speech Emotion Recognition and Deep Learning: An Extensive Validation Using Convolutional Neural Networks - F. A. D. Rí, F. C. Ciardi and N. Conci (2023): A Comprehensive Validation Using Convolutional Neural Networks: This study research paper deals with evaluation of SER performance while using the deep learning technique of convolution neural networks (CNN), between cross-generalization, and development of a CNN-based model equipped with the Attention Module. Here, four datasets were required for development and application of this project, involving CNN with CBAM, data augmentation, and cross-validation. There may be model performance drawbacks associated with this paper due to data variability, emotion ambiguity, need for high computational power, and limited real-world testing.

doi: 10.48047/ijiee.2025.15.4.34

Speech Emotion Recognition System Using Machine Learning - Raja, Prof & Sanghani (2022): Developed an emoting Speech Recognition (SER) model based on machine learning for classifying emotions in speech. Used RAVDESS and TESS datasets for feature extraction and subjected to various classifiers such as SVM, Random Forest, K-NN, Neural Networks, HMM, GMM. Each of the classifiers had its variable performance, with Neural Networks and SVM performing well compared to the rest of the classifiers and showing different effects of dataset diversity on generalization. The limitations induced by using machine learning were in reduction of features, overfitting, dataset restrictions, and challenges in real world applications affecting model performance.

A Comprehensive Review of Speech Emotion Recognition Systems - T. M. Wani, T. S. Gunawan, S. A. A. Qadri (2021): This paper covers a critical review of systems using Speech Emotion Recognition (SER) technology, features extraction, classification techniques, and problems. Provides analysis of several emotions from emotional speech database, feature extraction methods such as MFCC, spectral, prosodic features, and classification models such as SVM, ANN, HMM, CNN, and RNN. Using deep learning models like CNN, RNN, and LSTM, the methodologies described above have been found to have better accuracies than traditional classifiers but demand more training data and computation. Dataset variability, high computational cost, challenges faced in real life when handling noise, difficulty in emotion labelling, and overfitting in deep models.

3.EXISTING SYSTEM

The entire Speech Emotion Recognition (SER) system under consideration employs classical Machine Learning (ML) methods and is created according to a systematic approach consisting of data collection, feature engineering, model development, and evaluation. Since data collection influences the model accuracy due to the quality and variance of the speech data that are collected, it becomes very crucial in this approach. The next very important step is featureengineering, which extracts features of great significance such as spectral contrast, chroma features, and Mel-frequency cepstral coefficients (MFCCs) to capture emotional characteristics. After that, the algorithm addresses all the appropriate patterns in the data in order to learn to predict emotions, where the applicable ones would be Support Vector Machines (SVM), Random Forest (RF), and K-Nearest Neighbors (KNN).

At last, the model will be evaluated and validated on various performance measures, such as accuracy, precision, and recall, to assure maximum rating. The proposed system suffers from serious shortcomings, emanating from its systematic approach. The manual way of feature engineering is very much domain-knowledge-centric and may tend to miss important emotional cues. In addition, traditional machine learning models tend to work poorly in real-life scenarios because they are unable to capture complex emotional patterns and suffer from higher sensitivity to background clutters,

doi: 10.48047/ijiee.2025.15.4.34

speaker variations, and noisy data. This would definitely call for more advanced and trendy methods such as deep learning that would allow identification of complex patterns right from the raw speech data.

Limitations:

Dependence on Feature Engineering: Traditional ML models are greatly dependent on manual feature extraction. Effective feature extraction is a time-consuming and error-prone process. Ineffective feature extraction leads to the loss of important information, thus degrading the model performance.

Limited Capacity to Handle Complex Patterns: Emotions are conveyed through subtle variations of pitch, tone, and rhythm, which traditional ML models cannot capture effectively. There is the possibility of these models not being able to detect complex temporal dependencies in words.

Sensitivity to Data Quality: The performance is much affected by odd working conditions such as the presence of background noise, speaker accents, permutations, and noisy data. Achieving reliable accuracy in real-world settings is still fraught with difficulties.

Handling Class Imbalances Are Ghastly: SER datasets are typically heavily imbalanced in the sense that certain emotions, such as neutral or happiness, tend to be overrepresented while others such as fear or disgust are rare. An imbalance of this nature can lead to the development of biased models and poor performance concerning the less frequent emotions.

Limited Real-Time Performance: Conventional machine learning models are processing load heavy for feature extraction and classification and thus cannot operate efficiently in real-time.

4. PROBLEM STATEMENT

Speech Emotion Recognition (SER) recognizes emotional cues in a spoken language, as it aims to facilitate human-computer interaction. Neural networks embedded into SER learn to identify complex patterns from raw data representing tone, pitch, and rhythm. This methodology provides a higher degree of accuracy and flexibility than classic methodologies. SER's ability to sense and respond to human emotions finds tremendous applications in health care, education, customer service, and entertainment. SER is changing the way machines communicate with humans.

5.PROPOSED METHODOLGY

The new Speech Emotion Recognition (SER) sets forth a deep learning approach based on LSTM networks and CNN architectures. Increasingly, the system is intricate, accurate, capable, and real-time in performance and is therefore considered to be comprised of a process involving stages such as data preprocessing, features extraction, model training, and deployment.

doi: 10.48047/ijiee.2025.15.4.34

Data Preprocessing

In this stage, the speech dataset is prepared for training by extracting essential features such as the Mel-frequency cepstral coefficients (MFCCs), pitch, and energy that able to efficiently capture some of the important emotional characteristics from the audio signal. The dataset is then separated into training, validation, and test sets to guarantee the proper evaluation of the model and performance tracking.

Feature Representation

The extracted features are converted into a format suitable for input into the LSTM model. This may involve reshaping the data into sequential patterns or applying time-frequency representations to improve feature clarity.

Model Architecture

The LSTM-based architecture specifically for emotion recognition is at the center of the system. This architecture comprises stacked LSTM layers connected to fully connected layers that capture short-term and long-term dependencies on the audio data. Configurations such as varying the number of LSTM layers, changing the number of hidden units, introducing dropout layers in the architecture, and selection of activation functions will be thoroughly explored in order to improve model performance.

Model Training

The training of the LSTM model is done based on the prepared dataset with optimization algorithms such as Adam or RMSprop to achieve stable convergence. Early stopping is adopted in training to avoid overfitting and save those models that perform best. In the course of training, validation performance is monitored continuously and hyperparameters are tuned in order to improve accuracy.

Model Evaluation

The trained model is evaluated on the test set with the aid of performance metrics such as accuracy and F1-score, along with a confusion matrix. Particular emphasis is placed on analyzing performances between different emotion classes in order to identify possible biases or shortcomings in the model.

Fine-tuning and Optimization

For further performance enhancement, fine-tuning of the model is done through hyperparameter settings along with trying different architectures. Transfer learning and domain adaptation strategies will be undertaken if pre-trained models are available; this will speed up the learning process and improve accuracy.

doi: 10.48047/ijiee.2025.15.4.34

Deployment and Integration

When the performance of the model is satisfactory, it is then deployed into the real-world application. Integration is done through methods such as APIs or embedded systems to connect the model with platforms like virtual assistants, healthcare systems, and customer support applications. Special consideration was put on computational efficiency to ensure the system effectively runs on resource-limited devices like smartphones or IoT systems.

LSTM Model

The LSTM (Long Short-Term Memory) model is a type of RNN (recurrent neural network) architecture that suits well for sequence prediction and sequence-divide to sequence tasks, such as time series forecasting, natural language processing (NLP), and speech recognition. LSTM works on the principle of enhancing RNNs that suffer from capturing long-distance dependencies in sequential data due to conventional design. LSTM introduces specialized memory cells that can store the information for long periods of time.

LSTM Cell:

The basic building block of an LSTM is an LSTM cell, which has several gates and memory units to control the information flow.

The key components of an LSTM cell are:

- The **input gate** governs what new input is allowed into the memory cell.
- **Forget gate** governs what information to retain or forget from the previous memory cell state.
- **Output gate** assesses the information that will come out of the cell based on current input and internal state.
- **Cell state** (or memory cell): Internal state of the LSTM cell, which theoretically stores information for some time.

LSTM Architecture

The LSTM architecture is equipped with memory cell that is gated by three controls, namely, input gate, forget gate and output gate. These gates decide which information to add into the memory cell, which to remove from it and which to emit out as output.

Input gate: It will control what data goes into the memory cell.

Forget gate: It determines which information will be forgotten in the memory cell.

doi: 10.48047/ijiee.2025.15.4.34

Output gate: It determines which information will be read out of the memory cell.

This allows LSTM networks to learn to keep or forget some information as it progresses down the network and thereby learn long-term dependencies. Internally, the network maintains hidden states that resemble short-term memories. It updates these memories utilizing the current input, the previous hidden state, and the current state of the memory cell.

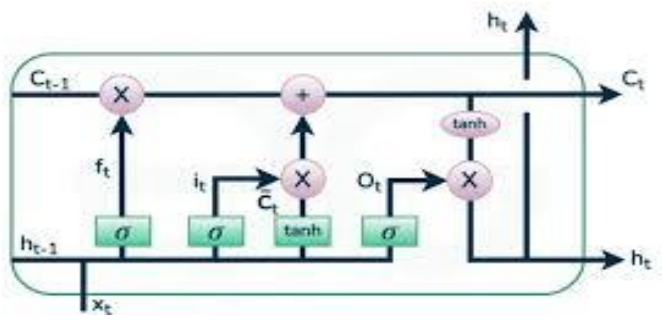

Fig 5.1 LSTM Architecture

MFCC (Mel-Frequency Cepstral Coefficients)

MFCC is short for Mel-Frequency Cepstral Coefficients, an important feature extraction technology used in speech and audio processing for tasks like speech recognition, speaker verification, or emotion detection. The performing block is described below:

1. Pre-emphasis: Most often, the pre-emphasis is the process that adds gain to jetting high frequencies so as to balance a frequency spectrum. A high-pass filter is applied to the signal.

2. Frame blocking: Instead Audio signal is divided into short overlapping frames. Typical frame lengths are about 20-40 milliseconds with overlap between adjacent frames.

3. Windowing: Each frame is windowed to lessen spectral leakage. The most common window function used is the Hamming window.

4. FFT (Fast Fourier Transform): The Fourier Transform is made from the signal with respect to the time for every frame to obtain the signal transformed from the time domain into the frequency domain, which is the magnitude spectrum of each frame.

5. Mel scaling: The Mel filter bank is applied onto the frame's magnitude spectrum. The Mel scale is a psychophysical pitch scale that approximates how the human auditory system would respond at different frequencies. The filter bank is defined as a set of triangular filters spaced along the Mel scale.

doi: 10.48047/ijiee.2025.15.4.34

6. Logarithm: The logarithm of filter bank energies takes into account the human listener's perception of loudness.

7. Discrete Cosine Transform (DCT): Finally, the Log filter bank energies are decorrelated through Discrete Cosine Transform and these form the final coefficients. The result is MFCCs**.**

MFCCs are typically used as feature vectors for training machine learning models, such as HiddenMarkov Models (HMMs), Gaussian Mixture Models (GMMs), or neural networks, for tasks likespeech recognition. They effectively capture the essential characteristics of the audio signal in acompact representation, making them well-suited for such tasks.

6.FLOW DIAGRAM

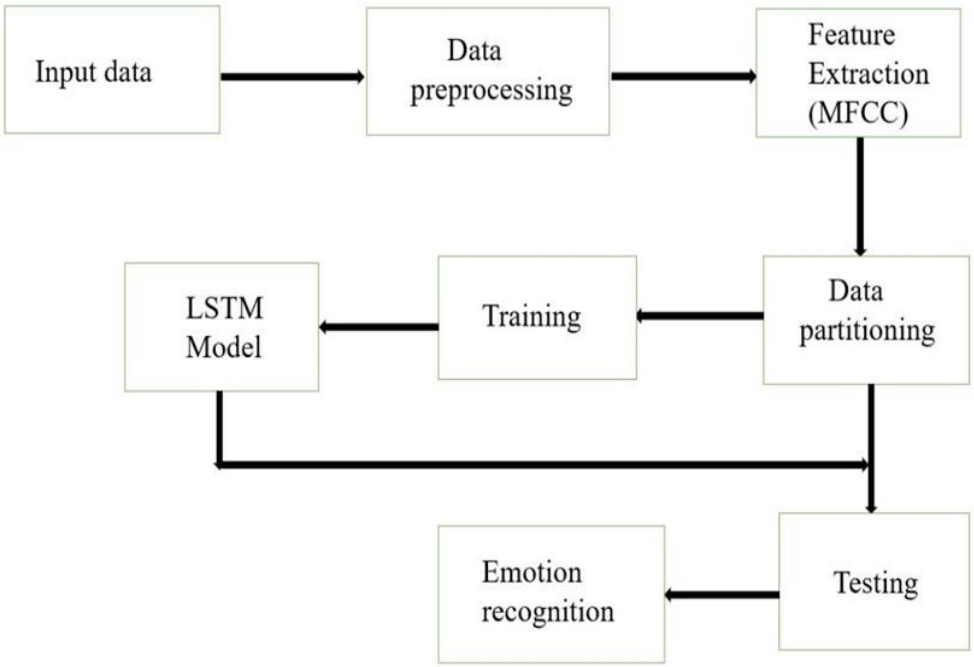

Fig 6.1: Flow Diagram

1. Input data: Input speech is that audio data received into the system for processing. It can be any spoken word, phrase, or sentence coming from a microphone or any kind of audio input device.

2. Data Preprocessing: Data preprocessing refers to that cleaning, transforming, and organizing of raw data so that it may be ready for analysis or modelling. The sequence of preprocessing begins with the treatment of missing values, scaling the numerical features, encoding the categorical variables, and then it goes on to detect the outliers.

3. Feature Extraction: Feature extraction can be defined as identifying and extracting the relevant attributes or characteristics found in the input speech data. These are usually spectral features such as

doi: 10.48047/ijiee.2025.15.4.34

MFCCs (Mel-Frequency Cepstral Coefficients) and other acoustic properties that play an important role in speech recognition and analysis.

4. Data Partitioning: Data partitioning involves the splitting of a dataset into different subsets to be used for training, validating, and testing. Usually, the most significant portion of data is used for training the model, whereas the smaller partition is used for validations to fine-tune the hyperparameters, which are again tested and put aside to evaluate model performance over unseen data. This way, model generalization and overfit assessment are made possible.

5. Training: Training means providing data to a learning algorithm so that it can develop a model expressing patterns and relationships in the given dataset. During training, this model iteratively alters specific internal parameters to minimize the value of an explicit loss function which, in turn, maximizes the model's ability to predict well.

6. Testing: Testing is the process of evaluating the performance of a trained machine learning model on unseen data so that its generalization ability can be estimated. The testing data are generally withheld from the training dataset, and they are typically utilized to evaluate various performance metrics such as accuracy, precision, recall, and F1-score.

7. LSTM Model: Long Short-Term Memory (LSTM) is a type of recurrent neural network (RNN) architecture able to learn long-term dependencies in sequential data. It is characterized by special memory cells that maintain information indefinitely along sequences, making it suitable for tasks like time series prediction, natural language processing, and speech recognition.

8. Emotion Recognition: Emotion recognition aims to understand and classify the emotion manifested by the input speech. These could be transient emotional states such as happiness, sorrow, anger, or neutral sentiment. Emotion recognition algorithms analyze different acoustic features, intonational patterns, and linguistic cues to arrive at a proper interpretation of the actual emotional state.

7.IMPLEMENTATION

The procedure for SER includes several essential steps for carrying out work, namely, data preprocessing, model training, and evaluation. First off is the loading and preparation of the audio data, which generally includes feature extraction and normalization. Once again, some common features include pitch or intensity features and mel-frequency cepstral coefficients, which describe the spectral properties of the speech signal. With this, setting up a deep-learning model including architectures such as Recurrent Neural Networks or specifically Long Short-Term Memory, which efficiently model sequential data, is undertaken. The attributes are then fed into the model. Before

doi: 10.48047/ijiee.2025.15.4.34

constructing the model, the architecture-a number of layers, hidden units, and activation functions-is established.

After the LSTM layer, fully connected layers for classification generally constitute the architecture of LSTM-based models. Next, the model is assembled for the training procedure, which makes use of some optimizers and loss functions. It is then trained using labelled data, iteratively changing its parameters to minimize the loss function. Training is done by splitting the whole dataset into training, validation, and testing sets for the purpose of evaluating the model performance. Early stopping techniques and mini-batch gradient descent are normally employed for better training and preventing overfitting. After training, the performance of the model detecting emotions from unseen data is evaluated on the test dataset.

The models of Long Short-Term Memory (LSTM) have proved immensely valuable to employ for the purpose of characterizing temporal dependencies and nuances in speech signals. Implementing an LSTM model in SER then typically encompasses a few of the essential steps as given. The audio data were first pre-processed by extracting important features such as spectrograms or MFCCs to describe the speech signals. It is then passed to an LSTM network, which consists of multiple LSTM layers such that the sequential patterns lying in the data can be exploited. The LSTM model learns different emotional cues embedded within the speech signals by updating its parameters via the gradient descent and backpropagation methods during training. Finally, the accuracy of classifying emotions by the trained model is validated on a separate test dataset. The LSTM offers a very good option for enhancing speech recognition technology with the consideration of temporal dynamics and long-range relationships in voice data.

SAMPLE CODE

```
import pandas as pd

import numpy as np

import os

import seaborn as sns

import matplotlib.pyplot as plt

import librosa

import librosa.display

from IPython.display import Audio

import warnings
```


doi: 10.48047/ijiee.2025.15.4.34

```
warnings.filterwarnings('ignore')
paths=[]
labels=[]
for dirname, _, filenames in os.walk('/kaggle/input'):
    for filename in filenames:
        paths.append(os.path.join(dirname, filename))
        label = filename.split('_')[-1]
        label = label.split('.')[0]
        labels.append(label.lower())
print('Dataset is loaded.')
emotion = 'fear'
path = df['speech'][df['label']==emotion].iloc[0]
data, sampling_rate = librosa.load(path)
waveplot(data, sampling_rate, emotion)
spectogram(data, sampling_rate, emotion)
Audio(path)
emotion = 'angry'
path = df['speech'][df['label']==emotion].iloc[0]
data, sampling_rate = librosa.load(path)
waveplot(data, sampling_rate, emotion)
spectogram(data, sampling_rate, emotion)
Audio(path)
```


doi: 10.48047/ijiee.2025.15.4.34

8.RESULT

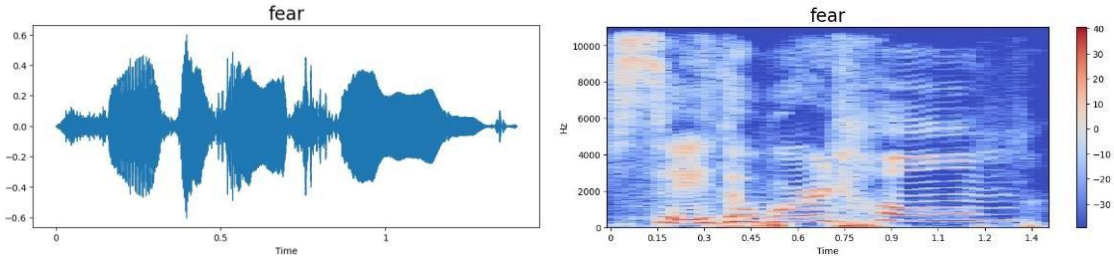

Fig 8.1:Fear

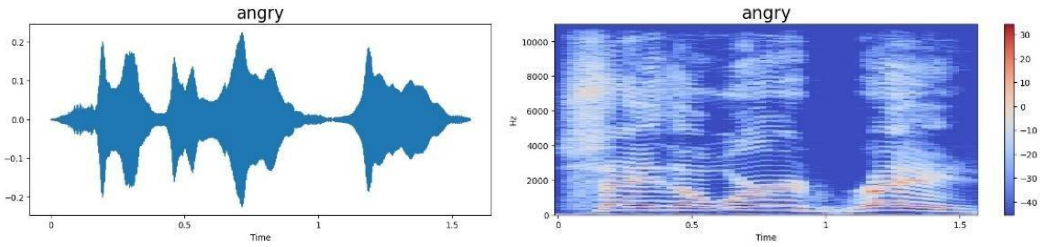

Fig 8.2:Angry

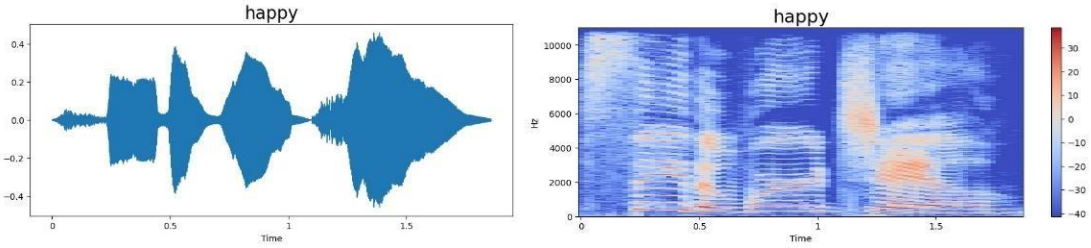

Fig 8.3:Happy

doi: 10.48047/ijiee.2025.15.4.34

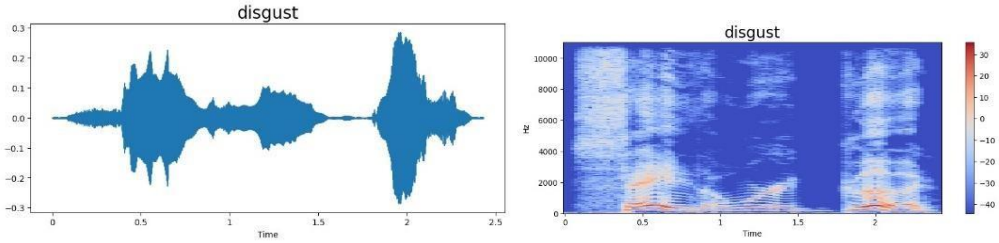

Fig 8.4:Disgust

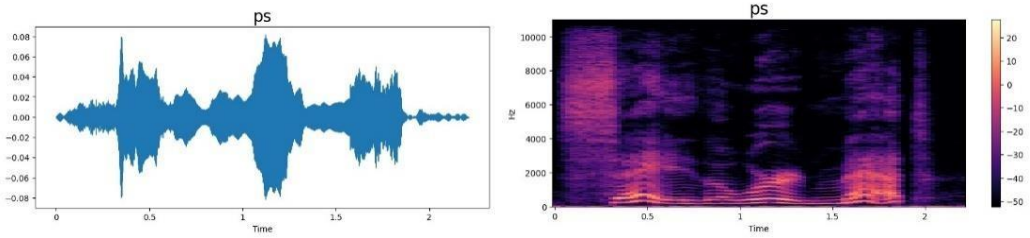

Fig 8.5:Pleasant Surprise

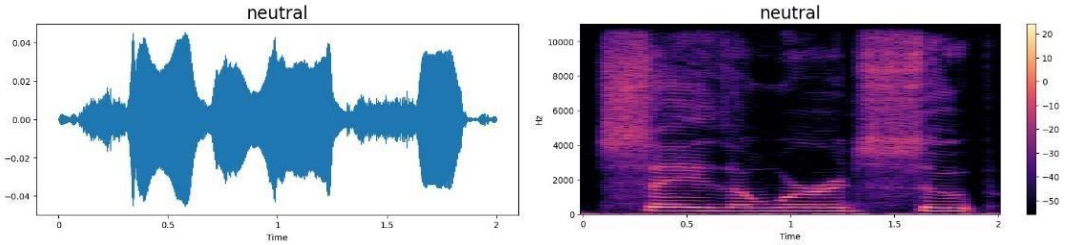

Fig 8.6:Neutral

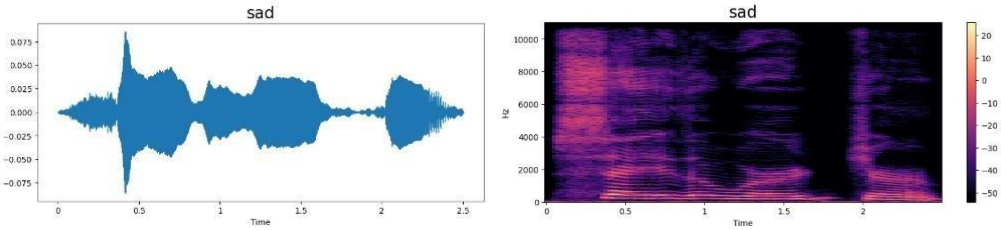

Fig 8.7:Sad

ACCURACY GRAPH

doi: 10.48047/ijiee.2025.15.4.34

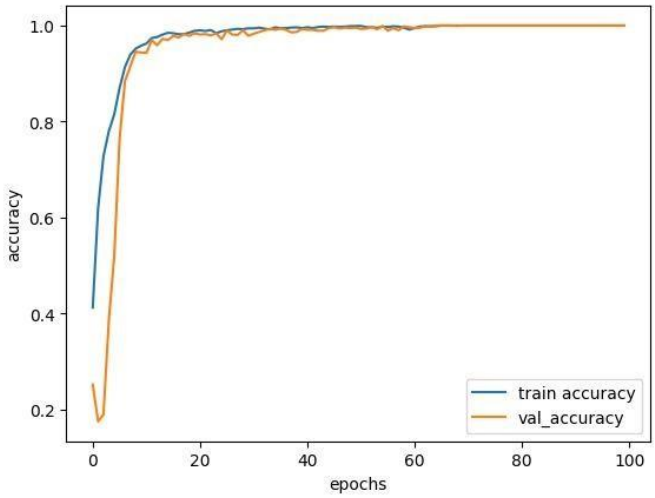

Fig 8.8:Accuracy Graph

9.CONCLUSION

Speech Emotion Recognition is big progress in the direction of making human-computer interaction and emotional intelligence applications thrive. Emotions are recognized in speech by analyzing speech features and applying superior ML or DL models. With the increasing incorporation of such systems into virtual assistants, healthcare, and entertainment, user experiences will be tremendously uplifted. Further improvements in accuracy, reliability, and real-time performance need research; only then will we be close to realizing machines capable of comprehending human emotion and reacting accordingly.

REFERENCES

1. Barsainyan, N., Singh, D.K. Optimized cross-corpus speech emotion recognition framework based on normalized 1D convolutional neural network with data augmentation and feature selection. *Int J Speech Technol* 26, 947–961 (2023). https://doi.org/10.1007/s10772-023-10063-8.
2. Barhoumi, C., BenAyed, Y. Real-time speech emotion recognition using deep learning and data augmentation. *ArtifIntell Rev* **58**, 49 (2025). https://doi.org/10.1007/s10462-024-11065-x.
3. F. A. D. Rí, F. C. Ciardi and N. Conci, "Speech Emotion Recognition and Deep Learning: An Extensive Validation Using Convolutional Neural Networks," in *IEEE Access*, vol. 11, pp. 116638-116649, 2023, doi: 10.1109/ACCESS.2023.3326071.

doi: 10.48047/ijiee.2025.15.4.34

4. Raja, Prof & Sanghani, Prof. (2024). Speech Emotion Recognition Using Machine Learning. Educational Administration Theory and Practices. 30. 10.53555/kuey.v30i6(S).5333.
5. T. M. Wani, T. S. Gunawan, S. A. A. Qadri, M. Kartiwi and E. Ambikairajah, "A Comprehensive Review of Speech Emotion Recognition Systems," in IEEE Access, vol. 9, pp. 47795-47814, 2021, doi: 10.1109/ACCESS.2021.3068045.
6. Abolfazl Younesi, Mohsen Ansari, Mohammadamin Fazli, Alireza Ejlali, Muhammad Shafique, Jörg Henkel, "A Comprehensive Survey of Convolutions in Deep Learning: Applications, Challenges, and Future Trends", IEEE Access, vol.12, pp.41180-41218, 2024.
7. Mehmet Berkehan Akçay, Kaya Oğuz, Speech emotion recognition: Emotional models, databases, features, preprocessing methods, supporting modalities, and classifiers, Speech Communication, Volume116,2020, ISSN0167-6393.
8. Francesco Ardan Dal Rí, Fabio Cifariello Ciardi, Nicola Conci, "Speech Emotion Recognition and Deep Learning: An Extensive Validation Using Convolutional Neural Networks", IEEE Access, vol.11, pp.116638-116649, 2023.
9. Ramdas Vankdothu,Dr.Mohd Abdul Hameed, Husnah Fatima" A Brain Tumor Identification and Classification Using Deep Learning based on CNN-LSTM Method" Computers and Electrical Engineering , 101 (2022) 107960
10. Ramdas Vankdothu,.Mohd Abdul Hameed "Adaptive features selection and EDNN based brain image recognition on the internet of medical things", Computers and Electrical Engineering , 103 (2022) 108338.
11. Ramdas Vankdothu,.Mohd Abdul Hameed,Ayesha Ameen,Raheem,Unnisa " Brain image identification and classification on Internet of Medical Things in healthcare system using support value based deep neural network" Computers and Electrical Engineering,102(2022) 108196.
12. Ramdas Vankdothu,.Mohd Abdul Hameed" Brain tumor segmentation of MR images using SVM and fuzzy classifier in machine learning" Measurement: Sensors Journal,Volume 24, 2022, 100440 .
13. Ramdas Vankdothu,.Mohd Abdul Hameed" Brain tumor MRI images identification and classification based on the recurrent convolutional neural network" Measurement: Sensors Journal,Volume 24, 2022, 100412 .
14. Bhukya Madhu, M.Venu Gopala Chari, Ramdas Vankdothu,.Arun Kumar Silivery,Veerender Aerranagula " Intrusion detection models for IOT networks via deep learning approaches " Measurement: Sensors Journal,Volume 25, 2022, 100641
15. Mohd Thousif Ahemad ,Mohd Abdul Hameed, Ramdas Vankdothu" COVID-19 detection and classification for machine learning methods using human genomic data" Measurement:

doi: 10.48047/ijiee.2025.15.4.34

Sensors Journal,Volume 24, 2022, 100537

16. S. Rakesh [a], NagaratnaP. Hegde [b], M. VenuGopalachari [c], D. Jayaram [c], Bhukya Madhu [d], Mo hdAbdul Hameed [a], Ramdas Vankdothu [e], L.K. Suresh Kumar "Moving object detection using modified GMM based background subtraction" Measurement: Sensors ,Journal,Volume 30, 2023, 100898
17. Ramdas Vankdothu,Dr.Mohd Abdul Hameed, Husnah Fatima "Efficient Detection of Brain Tumor Using Unsupervised Modified Deep Belief Network in Big Data" Journal of Adv Research in Dynamical & Control Systems, Vol. 12, 2020.
18. Ramdas Vankdothu,Dr.Mohd Abdul Hameed, Husnah Fatima "Internet of Medical Things of Brain Image Recognition Algorithm and High Performance Computing by Convolutional Neural Network" International Journal of Advanced Science and Technology, Vol. 29, No. 6, (2020), pp. 2875 – 2881
19. Ramdas Vankdothu,Dr.Mohd Abdul Hameed, Husnah Fatima "Convolutional Neural Network-Based Brain Image Recognition Algorithm And High-Performance Computing", Journal Of Critical Reviews,Vol 7, Issue 08, 2020(Scopus Indexed)
20. Ramdas Vankdothu, Dr.Mohd Abdul Hameed "A Security Applicable with Deep Learning Algorithm for Big Data Analysis",Test Engineering & Management Journal,January-February 2020
21. Ramdas Vankdothu, G. Shyama Chandra Prasad " A Study on Privacy Applicable Deep Learning Schemes for Big Data" Complexity International Journal, Volume 23, Issue 2, July-August 2019
22. Ramdas Vankdothu, Dr.Mohd Abdul Hameed, Husnah Fatima " Brain Image Recognition using Internet of Medical Things based Support Value based Adaptive Deep Neural Network" The International journal of analytical and experimental modal analysis, Volume XII, Issue IV, April/2020
23. Ramdas Vankdothu,Dr.Mohd Abdul Hameed, Husnah Fatima" Adaptive Features Selection and EDNN based Brain Image Recognition In Internet Of Medical Things " Journal of Engineering Sciences, Vol 11,Issue 4 , April/ 2020(UGC Care Journal)
24. Ramdas Vankdothu, Dr.Mohd Abdul Hameed " Implementation of a Privacy based Deep Learning Algorithm for Big Data Analytics", Complexity International Journal , Volume 24, Issue 01, Jan 2020
25. Ramdas Vankdothu, G. Shyama Chandra Prasad" A Survey On Big Data Analytics: Challenges, Open Research Issues and Tools" International Journal For Innovative Engineering and Management Research,Vol 08 Issue08, Aug 2019

doi: 10.48047/ijiee.2025.15.4.34

BIBLIOGRAPHY

I am Pendyala Divya, student at Balaji Institute of Technology and Science, from the department of Computer Science and Engineering. I have completed my research on “Speech Emotion Recognition using Deep Learning”.

I am Thotla Akhila,student at Balaji Institute of Technology and Science, from the department of Computer Science and Engineering. I have completed my research on “Speech Emotion Recognition using Deep Learning”.

I am Bavu Bindu,student at Balaji Institute of Technology and Science, from the department of Computer Science and Engineering. I have completed my research on “Speech Emotion Recognition using Deep Learning”.

I am Jinukala Varshitha, student at Balaji Institute of Technology and Science, from the department of Computer Science and Engineering. I have completed my research on “Speech Emotion Recognition using Deep Learning”.

doi: 10.48047/ijiee.2025.15.4.34