Journal of **N**onlinear **A**nalysis and **O**ptimization
Vol. 15, Issue. 1, No.6 : 2024
ISSN : **1906-9685**

EMOTION RECOGNITION USING SPEECH PROCESSING

Mrs Priyata Mishra Department of Computer Science Shri Shankaracharya Institute of Professional Management and Technology.
Amit Prasad Department of Computer Science Shri Shankaracharya Institute of Professional Management and Technology.
Aalind Shukla Department of Computer Science Shri Shankaracharya Institute of Professional Management and Technology.

Abstract
Speech Emotion Recognition (SER) is the process which helps us to automatically detect the emotional state of a speaker by analyzing their speech signal. This field has garnered notable attention in recent years, driven by its potential applications in areas such as human-computer interaction, customer service, healthcare, and education. This paper presents a comprehensive overview of SER with a focus on machine learning (ML) techniques. The paper begins by discussing the challenges of SER and the various emotional categories typically used in such systems. It then describes the three main stages of an ML-based SER system: data preprocessing, feature extraction, and classification. The ML algorithm used for this project MLP Classifier.

Keywords — Speech emotion recognition, Machine Learning, Deep Learning, Speech Features

1. INTRODUCTION

Human communication is a complex process that extends beyond the mere exchange of words. Embedded within our speech lies a wealth of information, including the emotional state of the speaker. Recognizing this emotional content can be crucial for facilitating effective communication and building meaningful relationships. In the age of intelligent machines, the ability to automatically understand human emotions from speech offers exciting possibilities for a wide range of applications.[1]

This is where Speech Emotion Recognition / (SER) comes. SER falls within the domain of artificial intelligence dedicated to automatically determining the emotional state of a speaker. based on their speech signal. By analyzing features such as pitch, energy, and intonation, SER systems can infer emotions like happiness, sadness, anger, and fear. SER holds immense potential to revolutionize human-computer interaction. Imagine a virtual assistant that can adapt its tone and behavior based on your emotional state, or a chatbot that can provide personalized support based on your emotional needs. In industries like customer service and healthcare, SER can lead to more empathetic and effective interactions. Additionally, SER can be used to analyze the emotional climate in meetings, classrooms, and online communities, providing valuable insights into group dynamics and social trends [2].

The field of SER has witnessed significant progress in recent years, driven by advancements in machine learning (ML) techniques. ML algorithms have proven highly effective in extracting relevant features from speech signals and classifying them into different emotional categories. The objective of this paper is to offer a thorough overview of SER with a focus on ML approaches. We will delve into the challenges of SER, explore the key stages of an ML-based SER system, and discuss the various ML algorithms used for achieving accurate emotion recognition. By understanding the fundamental

principles and advancements in SER with ML, we can unlock the potential of this technology to enhance human-computer interaction and pave the way for a future where machines understand and respond to our emotions with empathy and intelligence.[3]

2. LITERATURE REVIEW

Zheng Lian , Jianhua Tao, Bin Liu, Jian Huang, Zhanlei Yang, Rongjun Li ,2020[1] used domain adversarial neutral network (DANN) to address preserve emotion data during different types of speaker which used as mic. Gang Liu, Shifang Cai and Ce Wang , 2023[2] have designed a human brain-like implicit emotion attribute classification. Babak Joze Abbaschian, Daniel Sierra-Sosa and Adel Elmaghrab,2021[3] has employing algorithm to recognize the human emotion. Hadhami Aouani , and Yassine Ben Ayed,2020[4]. In a dual-phase approach, specifically involving feature extraction and a classification engine., this research suggests an emotion identification system based on voice signals. First, two sets of features are examined. (TEO) are extracted from a 42- dimensional vector Initially, 39 coefficients encompassing the Mel Frequency Cepstral Coefficients, Zero Crossing Rate (ZCR), Harmonic to Noise Rate, and Teager Energy Operator are extracted as audio features. In the second, we suggest selecting relevant parameters from the previously retrieved parameters by using the Auto-Encoder approach. Second, as a classifier technique, we employ Support Vector Machines (SVM). Kunxia Wang; Ning An; Bing Nan Li; Yanyong Zhang; Lian Li[5] have used Fourier parameters model which makes use of speaker-independent speech recognition's primary and secondary orders differences.

S. no	**Title**	**Publisher**	**Author**	**Finding**
1.	Conversational Emotion Recognition Using Self-Attention Mechanisms and Graph Neural Networks.	INTERSPEECH 2020	ZhengLian ,JianhuaTao, Bin Liu, Jian Huang, Zhanlei Yang, Rongju n Li	**Used (DANN) to address preserve emotion data during different types of speaker.**
2.	Speech emotion recognition based on emotion perception.	EURASIP Journal on Audio, Speech, and Music Processing, Springer.	Gang Liu, Shifang Cai andCe Wang	**Designed a human brain like implicit emotion attributeclassification.**
3.	Deep Learning Techniques for Speech Emotion Recognition , from Databases to Models.	MDPI	Babak Joze Abbaschian ,Daniel Sierra-Sosa and Adel Elmaghrab	**Employing algorithm to recognize the human emotion.**
4.	Speech Emotion Recognition	Multimedia Information	Hadha mi Aouani ,and Yassine Ben	**Emotion recognition system based on speech**

	with Deep Learning.	systems and Advanced Computing Laboratory, MIRACL University of Sfax,Tunisia	Ayed	**signals in two -stage approach.**
5.	Speech Emotion Recognition Using Fourier Parameters.	IEEE	Kunxia Wang; Ning An; Bing Nan Li; Yanyon g Zhang; Lian Li	**Used Fourier Parameters model.**

METHODOLOGY

Speech recognition method which uses deep networks for training. The method which are used for this project are Mel- frequency cepstral coefficients, Tonal Centroid, chromogram. These features will be used train DNN Model.

1.1 DATASET

This research will employ the dataset known as the Ryerson Audio-Visual Database of Emotional Speech and Song (RAVDESS). Within the specified section of the RAVDESS, there exist 1440 files, derived from conducting 60 trials for each of the 24 actors. The RAVDESS The dataset consists of 24 professional actors, with an equal distribution of 12 females and 12 males.

1.2 EXISTING METHOD

In existing method, the model used for speech recognition used algorithm for instance, support vector machine and k-nearest neighbors, Nevertheless, the accuracy of these models is notably low and they require a large dataset to work with that's why it not optimal.

1.3 MODULE'S

1) Upload: Use the librosa library to upload the audio dataset (.wav files) to be read.
2) View: You can view the uploaded dataset.
3) Preparing the data: One method for transforming the raw data into a clean data set is pre-processing. Cleaning the data entails eliminating duplicate values, eliminating outliers, eliminating unwanted properties, and replacing in null values with meaningful ones. In the event that the dataset includes any category entries, those variables should be converted to numerical values
4) Identifying Features: Mel-frequency cepstral coefficients, a chromatogram, a melancholic spectrogram along with spectral contrast and tonal centroid features are the extracted features.
5) Train and Test Split: We divided our 1440 audio file dataset into two groups: 1008 audio files were used for training, while 432 audio files were used for testing. In this case, the training dataset contains 70% of the data.
6) Constructing the model: Our proposed solution utilizes deep learning to comprehend audio and forecast emotions. Reduced processing power and improved accuracy are two benefits of deep learning. Deep Neural Networks will be utilized in the model's creation. In deep learning, deep neural networks are frequently used to train models for tasks that are difficult or impossible for typical machine learning algorithms to complete. Neural networks of five layers are used to generate the model. Dropouts have been employed to reduce the issue of overfitting.
7) Prediction: A user uploads an audio file with human speech, and the model is used to anticipate the emotional state of the speaker.
8) User Interface: The web application is designed with the Flask framework to house the model, consisting of two integral parts: the user and the system.

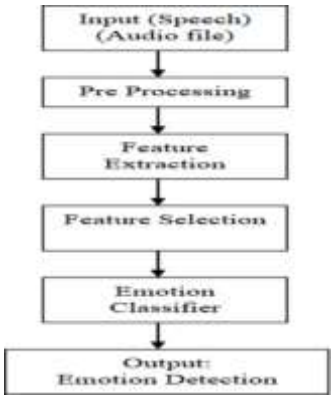

Figure 1 Method Overview

3. RESULT

This study looks at the use of sophisticated voice processing methods for emotion identification. Using a heterogeneous dataset that spans a spectrum of emotional states, we investigate several algorithms and approaches for precisely categorizing and comprehending emotions expressed via speech. Hidden Markov Models, Mel-Frequency Cepstral Coefficients, and deep learning architectures, including Convolutional Neural Networks and Long Short-Term Memory networks, are utilized in this study. The findings highlight the efficacy of deep learning models by showing a high degree of accuracy (73.3 %) in identifying and categorizing emotions. The research yields valuable insights that aid in the creation of emotionally intelligent systems that find use in virtual assistants, affective computing, and human-computer interaction.

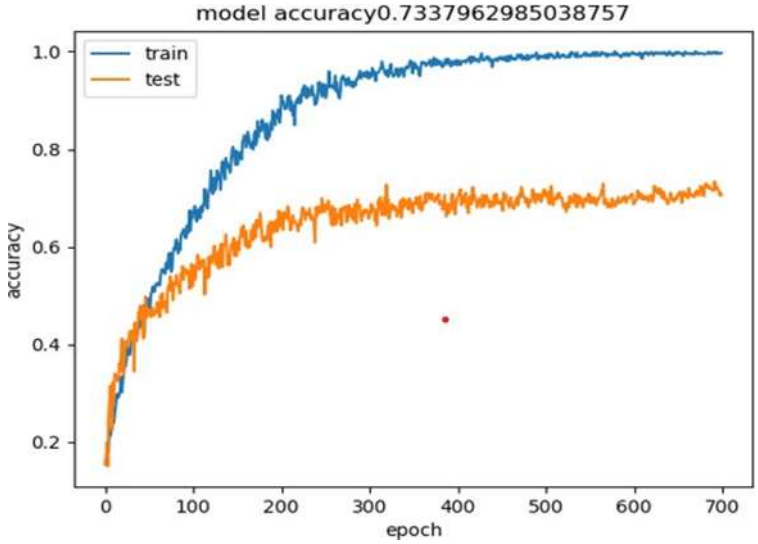

Fig4.1 Model.Accuracy for Speech Recognition

4. CONCLUSION

The suggested plan offered a method for identifying emotions in spoken language.
Neural networks have been used to accomplish this strategy. Using the deep neural network architecture, we have developed a deep learning model with success that predicts the speaker's emotions in an audio file. We have implemented the Flask architecture in our project to create a web-based application. With the trained model, we achieved 73.3 % test accuracy. Please be aware that people's assessments of the same audio's emotions might vary, and that emotion prediction is subjective. This is also the rationale behind the algorithm's training on emotions judged by humans.

REFERENCES

[1] Szegedy, Christian & Toshev, Alexander & Erhan, Dumitru. (2013). Deep Neural Networks for Object Detection. 1-9.

[2] Benk, Sal & Elmir, Youssef & Dennai, Abdeslem. (2019). A Study on Automatic Speech Recognition. 10. 77-85. 10.6025/jitr/2019/10/3/77-85.
[3] Ashish B. Ingale & D. S. Chaudhari (2012). Speech Emotion Recognition. International Journal of Soft Computing and Engineering (IJSCE) ISSN: 2231- 2307, Volume-2 Issue-1, March 2012.
[4] Chenchen Huang, Wei Gong, Wenlong Fu, Dongyu Feng, "A Research of Speech Emotion Recognition Based on Deep Belief Network and SVM", Mathematical Problems in Engineering, vol. 2014, Article ID 749604, 7 pages, 2014. https://doi.org/10.1155/2014/749604
[5] Livingstone SR, Russo FA (2018) 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(5): e0196391.
[6] Zheng Lian ,Jianhua Tao, Bin Liu, Jian Huang, Zhanlei Yang (2020), Conversational Emotion Recognition Using Self-Attention Mechanisms and Graph Neural Networks.
[7] Gang Liu, ShifangCai and Ce Wang ,(2023) , Speech emotion recognition on emotion perception.
[8] Babak Joze Abbaschian , Daniel Sierra-Sosa and Adel Elmaghrab(2021), Deep Learning Techniques for Speech Emotion Recognition,from Databases to Models.
[9] Hadhami Aouani ,and Yassine Ben Ayed(2020), Hadhami Aouani ,and Yassine Ben Ayed.
[10] Kunxia Wang; Ning An; Bing Nan Li; Yanyong Zhang; Lian Li (2015),Speech Emotion Recognition Using Fourier Parameters.