IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 1, February 2024

Emotion Detection via Voice and Speech Recognition

Chethan C V[1], Greeshma K S[2], Dr. Kiran Y C[3]
Undergraduate Students, Department of Information Science and Engineering[1,2]
Professor, Department of Information Science and Engineering[2]
Global Academy of Technology, Bangalore, India

Abstract: *In human-computer interaction, spoken emotion detection is essential yet difficult. Researchers extract emotional information from speech sounds using a variety of analytic techniques. Effective communication requires an understanding of human emotions, which range from happiness to astonishment and beyond. Classification, pre-processing, feature extraction, and pattern recognition are all involved in emotion recognition. While some approaches take into account a continuous spectrum, others concentrate on a single-dimensional representation. The study of emotional cognition benefits from the use of technologies such as fMRI, which facilitate the distinction between positive and negative emotions. The intricacy and significance of identifying emotions in speech are emphasised in this abstract, underscoring the necessity of sophisticated methodologies in the field of human-computer interaction research.*

Keywords: Deep Learning, Emotion Recognition, RAVDESS Dataset, SK learn, Speech Feature, SVM Classification, Real-time emotion detection, Audio feature engineering, Sentiment analysis, emotional intelligent system.

I. INTRODUCTION

The AI community has been actively involved in the advancement of Speech Emotion Recognition (SER), with applications extending to household robots, which could significantly enhance their performance through interactions with owners [6]. Real-time SER models have been developed for diverse applications, including assessing customer satisfaction in call centers, personalizing song recommendations based on emotional states, and enhancing car security systems based on driver emotions. Notably, SER has also found applications in monitoring emotional and mental well-being in vulnerable populations.

Despite the progress in human-machine interaction, recognizing emotions during dialogue remains an understudied area. Incorporating emotional recognition into models of human-machine communication is essential for more comprehensive interactions [3]. This study focuses on developing a robust method for identifying primary emotions conveyed during conversations.

The successful development of a SER system hinges on addressing three key challenges: selecting appropriate emotional speech databases, extracting effective features, and designing reliable classifiers using machine learning algorithms. Emotional feature extraction is particularly crucial, with significant research focusing on features such as energy, pitch, formant frequency, LPCC, MFCC, and MSFs [4][5]. Many researchers advocate for using combined feature sets to capture richer emotional information [6]. This introduction underscores the importance and complexity of SER in human-computer interaction research.

II. LITERATURE REVIEW

2.1 Enhanced speech emotion detection using deep neural networks:

In [1], This paper investigates the effectiveness of perceptual-based speech features for emotion detection, employing features such as MFCCs, PLPC, MFPLPC, BFCC, RPLP, and IMFCC. The algorithm utilizes deep neural networks (DNN) to evaluate auditory cues and identifies predominant features containing significant emotional information. The study validates the algorithm on the Berlin database, considering seven emotions in 1-dimensional categorical space and 2-dimensional continuous space with valence and arousal dimensions. Comparative analysis indicates substantial

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 1, February 2024

improvement in emotion recognition performance using DNN with the identified combination of perceptual features. The importance of emotion expression and perception in human interaction is emphasized, noting their influence on message interpretation and decisionmaking at the cognitive level. The paper highlights the need for man-machine interfaces to recognize user emotional states and react accordingly. The challenging task of emotion detection in the rapidly advancing field of humanmachine interaction (HCI) is acknowledged, with applications extending to household robots, interactive gaming, mobile communication, call centers, and more, integrating emotion recognition as an essential feature.

2.2 A Review on Emotion Detection and Classification using Speech:

The [2] This paper delves into the challenges and recent advances in detecting emotions from vocal audio, primarily focusing on speech signals. Emotion recognition from speech remains a complex issue in artificial intelligence, necessitating the careful selection of a suitable classification model for effective detection. The document reviews various features and extraction techniques related to emotional speech data, providing a comprehensive analysis of different classification methods with a specific emphasis on classifying diverse emotions within the speech detection model.

Significant features like MFCC, fundamental frequencies, and LPCC for emotion classification are discussed, underscoring their importance. The paper highlights potential applications of speech emotion recognition in teaching, entertainment, and clinical diagnosis, emphasizing its role in improving human-to-machine communication. As organizations increasingly demand effective communication interfaces between people and machines, the paper stresses the importance of automating emotion detection from speech for understanding consumers and making informed decisions. Despite significant progress, the review acknowledges the ongoing efforts needed to make machines more human-friendly and adept at accurately understanding emotions, particularly as human dependence on machines is expected to grow in the next 10-20 years.

2.3 Emotion Detection via Voice and Speech Recognition:

[3] This paper tackles the intricate task of emotion detection from voice signals, crucial for improving human-computer interaction (HCI). It surveys existing literature on speech emotion recognition, emphasizing the use of traditional speech analysis methods and the recent emergence of deep learning strategies. The study explores various databases, emotions collected, and contributions to speech emotion recognition, with a specific focus on the research team's Speech Emotion Recognition Project. Employing Python 3.6, the RAVDESS dataset, and Pycharm, the project utilizes unsupervised learning, specifically the MLP-Classifier algorithm, to build a model adept at recognizing frustration or annoyance in speakers' voices. This work contributes valuable insights to the growing field of emotion recognition in voice signals, particularly regarding deep learning applications and their impact on HC

2.4 Speech Emotion Recognition:

This [4] paper underscores the significance of speech in human communication and emphasizes the crucial role of emotion recognition for effective humanmachine interaction (HMI). Recognizing emotions in speech enhances the naturalness of communication and is particularly important for simulating realistic interactions in the virtual world. The paper provides a review of existing methods for speech emotion detection, highlighting the predominant use of features such as MFCC and Energy in current approaches. It acknowledges the historical importance of speech in human communication and discusses the increasing relevance of HMI in both industry and academia. The complexity of human brain analysis of auditory input is noted, emphasizing the conversion of sounds into conceptual ideas for instructions, commands, information, and entertainment. The paper also outlines the three main steps in speech processing: preprocessing, feature extraction, and pattern recognition, with a focus on the informative role of vowels in speech signals. Overall, the review sets the stage for further research in the field of speech emotion detection, recognizing its pivotal role in advancing natural and effective human-machine interactions.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 1, February 2024

2.5 Emotional speech Recognition using CNN and Deep learning techniques.

[5] This paper emphasizes the integral role of emotions in human life, acting as a crucial means of expression and communication about one's well-being. It introduces the concept of Speech Emotion Recognition (SER), a system designed to extract and predict a speaker's emotional tone from audio signals. The paper discusses the categorization of emotions into types such as Anger, Happiness, Sadness, and Neutral, and highlights the significance of finite resources for developing an effective SER system. Spectral and prosodic features, including Mel-frequency Cepstral Coefficients (MFCC) and pitch, loudness, and frequency, are identified as crucial elements for training machine learning models to recognize emotional states in speech signals. The study presents a Speech Emotion Recognition system that surpasses existing models in data, feature selection, and methodology, aiming for improved accuracy (averaging 78%) and fewer false positives in identifying speech-based emotions. Overall, the paper underscores the importance of emotions in human communication and introduces advanced methodologies to enhance the precision of Speech Emotion Recognition systems.

2.6 A voice-based real-time emotion detection technique using recurrent neural network empowered

[6] This paper addresses the increasing importance of mining audio data from human conversations, particularly in the context of the Internet of Things (IoT) and voice-based multimedia applications. Recognizing the substantial big data generated by these applications, reflecting various aspects of human behavior, the paper focuses on extracting emotions from speech as a critical element for next-generation artificial intelligence. Introducing a novel approach using Bag-of-Audio-Words (BoAW) for feature embeddings in conversational audio data, coupled with a state-of-the-art emotion detection model based on Recurrent Neural Networks (RNN), the study reports significant improvements in accuracy. The research highlights the significance of real-time multimedia applications in understanding human behavior and emotions through digital footprints.

2.7 Emotion detection from text and speech:

[7] This survey explores the evolving field of emotion recognition, focusing on text and speech analysis as crucial sources of emotional expression. With people expressing emotions through various mediums such as writing, speech, and social media posts, the paper underscores the importance of efficient emotion detection for diverse purposes, including human-computer interaction and human-robot interaction in the context of growing robotic research. The survey reviews existing research efforts, emotion models, datasets, and detection techniques, emphasizing the multidimensional nature of human emotions. It acknowledges the challenges in accurately detecting emotions from textual content and highlights the increasing relevance of automatic emotion detection systems acr
oss various socioeconomic domains, offering a comprehensive overview of achievements and potential extensions for future advancements in the field.

2.8 English speech emotion recognition method based on speech recognition

[8] This paper addresses the limited focus on English speech emotion recognition in current research efforts in China, emphasizing the importance of understanding emotional content in English texts. It highlights the significance of affective computing in intelligent human-computer interaction technology and its impact on various fields, including auditory and vocal mechanisms, pattern recognition, and artificial intelligence. The paper introduces an English speech emotion recognition method based on speech recognition, acknowledging its effectiveness while recognizing the need for further improvement. The limitations include a focus on experimental research with English databases and the necessity to extend research to encompass multiple languages. Additionally, the paper suggests exploring multi-modal information fusion, considering various human expressions such as facial expressions, words, and brain waves, to enhance overall accuracy in emotion recognition

2.9 Deep features-based speech emotion recognition for smart affective services

[9] This paper delves into speech emotion recognition using a deep convolutional neural network (CNN) with rectangular kernels applied to spectrograms, representing an innovative approach. Unlike traditional CNNs with square kernels designed for 2D image data, spectrograms encode information differently, with time along the x-axis and

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 1, February 2024

frequency along the y-axis. The proposed method introduces rectangular kernels of varying shapes and sizes, along with max pooling in rectangular neighborhoods, to effectively extract discriminative features. The study evaluates its performance on Emo-DB and Korean speech databases, showcasing superior results compared to many state-of-the-art techniques. The experiments involve training CNN models on spectrogram images and exploring a transfer learning approach. The paper highlights the effectiveness of this method in recognizing emotions from speech, particularly with its modification of kernel sizes and pooling strategies for better adaptation to spectrogram data.

2.10 Study on emotion recognition and companion Chatbot using deep neural networ

[10] This research focuses on the development of a communication system incorporating speech emotion recognition and semantic analysis, with applications in companion robots, technology products, and medical purposes. The proposed system employs a Convolutional Neural Network (CNN) based on GoogLeNet to recognize five emotions from speech, achieving a top accuracy of 79.81%. Additionally, semantic analysis is performed using Recurrent Neural Network (RNN) in the Seq2Seq framework. The system comprises client and server components, with the client implemented on the Android system for user interaction and the server on Ubuntu Linux with a web server for backend processing. Users can record voice on the cellphone app, and the system analyzes emotions and conducts semantic analysis to function as a responsive chatting machine. Key contributions include the integration of Chinese word vectors for improved dialogue tolerance, direct emotion classification through CNN without tokenization, and the utilization of an app-based approach for user-friendly interaction and cost-effectiveness. The system's capabilities extend to collecting users' emotional indices for analysis by back-end care organizations. The experiments involve training on a high-performance computing system, and the proposed system demonstrates efficacy in recognizing emotions and conducting semantic analysis on speech data.

2.11 "Emotion Detection Through Speech Analysis"

[11] The use of speech analysis techniques for emotion detection is examined in the research article "Emotion Detection Through Speech Analysis". To extract acoustic features, recognise emotional states from spoken language, and analyse language are some of the techniques it investigates. Machine learning algorithms are also discussed. In a variety of domains, including customer service, human-computer interaction, and mental health monitoring, the study examines the importance of emotion detection. Important discoveries include the ability of deep learning models to reliably identify emotions from voice inputs, such as convolutional neural networks (CNNs) and recurrent neural networks (RNNs). To increase emotion identification ability, the paper also emphasises the need of taking into account both linguistic content and auditory cues. More issues are covered, including bias reduction, interpretability of the model, and shortage of data. As a whole, the study advances

2.12 Emotion Detection from Voice-Based Classified Frame-Energy Signal Using K-Means Clustering".

[12] A technique for identifying emotions from speech signals is presented in the publication "Emotion Detection Through Speech Analysis" The method entails breaking up the speech signal into frames and categorising the individual frames according to their energy levels. These frames are then grouped into groups that correlate to various emotional states using the K-Means clustering technique. The method deduces the main emotion in the speech signal by examining the distribution of frames among clusters. The preparatory stages of feature extraction, the specifics of the K-Means clustering algorithm's implementation, and the assessment of the method's effectiveness using pertinent metrics like accuracy and precision are probably covered in the paper. The study may also compare its methods to those already in use for detecting emotions and explore the possible

2.13 "Speech Emotion Recognition Using a Dual-Channel Complementary Spectrogram and the CNN-SSAE Neural Network."

[13] A innovative method for identifying emotions from voice data is presented in the publication "Speech Emotion Recognition Using a Dual-Channel Complementary Spectrogram and the CNN-SSAE Neural Network." The suggested approach makes use of a dual-channel complementary spectrogram representation, which records the speech signal's phase and magnitude information and offers an all-encompassing view of the audio data. Furthermore, a Convolutional

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 1, February 2024

Neural Network with Stacked Sparse Autoencoder (CNN-SSAE) architecture is presented by the authors to facilitate the extraction of high-level characteristics from the spectrogram representations. Efficient emotion identification is made possible by the CNN-SSAE model's ability to automatically extract discriminative characteristics from the input data. The construction of the dual-channel complementary spectrogram, the CNN-SSAE model's architecture and training process, and the assessment of the suggested strategy on benchmark data are probably covered in the work.

2.14 "Early Threat Warning Via Speech and Emotion Recognition from Voice Calls"

[14] The paper "Early Threat Warning Via Speech and Emotion Recognition from Voice Calls" describes a novel method for using speech and emotion recognition in voice conversations to identify possible threats or crises. The suggested approach examines a number of speech signal characteristics, such as emotional content and acoustic characteristics, in order to spot patterns linked to potentially dangerous circumstances. With the use of machine learning methods like ensemble classifiers and deep neural networks, the system is able to automatically identify voice calls as either normal or suspicious. The method of feature extraction from voice signals, the creation and training of machine learning models, and the assessment of the system's effectiveness with the use of actual voice call data are probably covered in the paper. Outcomes might contain criteria like precision,

2.15 "A Survey of Deep Learning-Based Multimodal Emotion Recognition: Speech, Text, and Face"

[15] The paper "A Survey of Deep Learning-Based Multimodal Emotion Recognition: Speech, Text, and Face" offers a thorough summary of the most current developments in deep learning-based emotion detection spanning voice, text, and facial expressions, among other modalities. The survey addresses several facets of multimodal emotion identification, such as deep learning architectures, feature extraction methods, data gathering strategies, and assessment measures. It talks about how combining data from several modalities might increase the resilience and accuracy of emotion identification. Along with examining these issues, the study also addresses potential remedies and future goals for research on issues including data fusion, model interpretability, and cross-modal inconsistency. The survey also offers a comparative review of current methods, emphasising their advantages and disadvantages. All things considered, the article is an invaluable resource for scholars and

TABLE I. Table of Comparison Results

Sl no	Author/year	Research /Work Paper	Objective	Technology	Achieved	Not Achieved
1	Madanian S, Chen T, Adeleye O,et al	Speech emotion recognition using machine learning — A systematic review	In order to compile current approaches, feature extraction strategies, machine learning algorithms, difficulties, and potential research avenues, a thorough assessment of the literature on voice emotion identification using machine learning techniques is being conducted.	The study includes works using a variety of machine learning algorithms, including Support Vector Machines, Random Forest, CNNs, RNNs, and LSTMs, that were published up to 2023.	identified popular approaches, assessed feature extraction strategies, examined the effectiveness of machine learning algorithms, emphasised difficulties, and suggested lines of inquiry for further study.	Potential gaps in addressing new issues or developments beyond the purview of the study, as well as potential restrictions in covering recent material published after 2023.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 1, February 2024

2	Singh P, Srivastava R, Rana KP, Kumar V. et al.	A Multimodal Hierarchical Approach to Speech Emotion Recognition from Audio and Text	Create a hierarchical method in a multimodal framework to identify emotions from text and audio inputs.	makes use of deep learning architectures for multimodal learning as well as machine learning models that combine text and audio characteristics.	created a multimodal hierarchical method to recognise speech emotions that uses text and audio inputs to increase accuracy.	Potential barriers to generalisation or scalability for a variety of datasets or practical uses.
3	Michael Brown et al.	Gender-Driven Emotion Recognition Through Speech Signals for Ambient Intelligence Applications	analyses voice aspects using machine learning algorithms, taking gender-specific traits into account for emotion identification	analyses voice aspects using machine learning algorithms, taking gender-specific traits into account for emotion identification	. a method for recognising emotions from voice signals that takes gender variations into account and is well-suited for ambient intelligence applications has been created	Potential drawbacks in extrapolating findings to larger datasets or real-world situations beyond the purview of the research.
4	Singh YB, Goel S, et al.	A Systematic Literature Review of Speech Emotion Recognition Approaches.	to thoroughly examine current techniques and developments in voice emotion recognition methods.	examines many deep learning and machine learning methods used in voice emotion identification.	gave a thorough summary of the techniques, innovations, and strategies currently in use for voice emotion recognition.	Possible restrictions on discussing new issues or recent advancements that fall beyond the purview of the study.
5	Avots E, Sapiński T, Bachmann M, Kamińska D,et al.	Audiovisual Emotion Recognition in the Wild	Create an emotion identification system with audiovisual signals that work in a variety of real-world settings.	combines machine learning with methods for audio and visual processing to identify emotions in real environments.	created an effective audiovisual system for recognising emotions that may be used in a variety of real-world settings.	Possible restrictions on the study's scalability or generalisation to larger datasets or real-world applications.
6	Dai W, Han D, Dai Y, Xu D,et al	Emotion Recognition and Affective Computing on Vocal Social Media	to provide techniques for voice social media platforms affective computing analysis and emotion recognition.	Uses natural language processing and machine learning methods to recognise emotions and perform affective computing on vocal social	created techniques for using vocal social media sites to do affective computing and analyse emotions.	developed methods for doing affective computing and emotion analysis on voice social media platforms.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 1, February 2024

				networking platforms.		
7	Vicsi K, Sztahó D,et al	Recognition of Emotions on the Basis of Different Levels of SpeechSegments	Create a system for recognising emotions in speech segments at different levels.	makes use of machine learning to identify emotions from voice segments at various levels.	created a system that can identify emotions from speech samples at different levels.	Possible restrictions on extrapolating findings to different datasets or real-world uses beyond the segments under study.
8	Zhao Z, Bao Z, Zhao Y, Zhang Z, Cummins N, Ren Z, Schuller B.	Exploring Deep Spectrum Representations for Speech Emotion Recognition.	Examine deep spectrum representations to study attention-based recurrent and convolutional neural networks for voice emotion identification.	extracts information from spectrum representations for emotion identification using deep learning architectures that incorporate attention processes.	improved identification of emotions in speech by utilising attention-based recurrent and convolutional neural networks to explore deep spectrum representations.	Potential restrictions on the study's capacity to scale or generalise to different datasets or real-world applications.
9	Kakuba S, Poulose A, Han DS.	Deep Learning-Based Speech Emotion Recognition with Multi-Level Fusion of Concurrent Features.	Use deep learning to recognise emotions in speech by combining features at different degrees of simultaneity.	fuses concurrent features at several layers for voice emotion identification using deep learning models.	enhanced voice emotion identification by deep learning integration of many tiers of concurrent information.	Potential restrictions on the generalisation to different datasets or practical uses other than concurrent feature fusion.
10	Alslaity A, Orji R, et al	Machine Learning Techniques for Emotion Detection and Sentiment Analysis: Current State, Challenges, and Future Directions.	Examine the state, difficulties, and prospects of machine learning methods for sentiment analysis and emotion detection.	examines several machine learning techniques used for sentiment analysis and emotion identification applications.	gives information on the state of the art, difficulties, and potential paths for machine learning in sentiment analysis and emotion detection.	Potential restrictions on discussing new developments or problems that are outside the purview of the review.

III. CONCLUSION

To sum up, research on voice and speech recognition study on emotion detection shows notable progress in using deep learning methods, multimodal fusion, and attention processes for precise emotion categorization. Research highlights the significance of integrating language context with acoustic information, using domain-specific expertise, and correcting class imbalance in order to improve model resilience and performance. It has been demonstrated that transfer

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 1, February 2024

learning and ensemble approaches enhance generalisation and dependability, especially in situations with a dearth of labelled data. Moreover, models must be updated and adjusted continuously to retain performance in dynamic contexts. Interpretability, bias reduction, and real-world application remain obstacles in spite of advancements. In order to allow useful applications in domains like mental health, future research should concentrate on creating more interpretable models, guaranteeing equity across demographic groups, and improving deployment techniques.

REFERENCES

[1]. Lalitha S, Tripathi S, Gupta D. Enhanced speech emotion detection using deep neural networks. International Journal of Speech Technology. 2019 Sep;22:497-510.

[2]. Tripathi A, Singh U, Bansal G, Gupta R, Singh AK. A review on emotion detection and classification using speech. InProceedings of the international conference on innovative computing & communications (ICICC) 2020 May 15.

[3]. Rastogi R, Anand T, Sharma SK, Panwar S. Emotion Detection via Voice and Speech Recognition. International Journal of Cyber Behavior, Psychology and Learning (IJCBPL). 2023 Jan 1;13(1):1-24. 4.

[4]. Vaishnav S, Mitra S. Speech emotion recognition: a review. Int. Res. J. Eng. Technol. IRJET. 2016 Apr;3(04):313-6.

[5]. Hema C, Marquez FP. Emotional speech recognition using cnn and deep learning techniques. Applied Acoustics. 2023 Aug 1;211:109492.

[6]. Chamishka S, Madhavi I, Nawaratne R, Alahakoon D, De Silva D, Chilamkurti N, Nanayakkara V. A voice-based real-time emotion detection technique using recurrent neural network empowered feature modelling. Multimedia Tools and Applications. 2022 Oct;81(24):35173-94.

[7]. Sailunaz K, Dhaliwal M, Rokne J, Alhajj R. Emotion detection from text and speech: a survey. Social Network Analysis and Mining. 2018 Dec;8:1-26.

[8]. Liu M. English speech emotion recognition method based on speech recognition. International Journal of Speech Technology. 2022 Jun;25(2):391-8.

[9]. Badshah AM, Rahim N, Ullah N, Ahmad J, Muhammad K, Lee MY, Kwon S, Baik SW. Deep features-based speech emotion recognition for smart affective services. Multimedia Tools and Applications. 2019 Mar;78:5571-89.

[10]. Lee MC, Chiang SY, Yeh SC, Wen TF. Study on emotion recognition and companion Chatbot using deep neural network. Multimedia Tools and Applications. 2020 Jul;79:19629-57.

[11]. Johnson A. Emotion Detection Through Speech Analysis (Doctoral dissertation, Dublin, National College of Ireland)

[12]. Hossain N, Jahan R, Tunka TT. Emotion detection from voice based classified frame-energy signal using k-means clustering. Int. J. Softw. Eng. Appl. 2018 Jul;9(4):37-44.)

[13]. Li J, Zhang X, Huang L, Li F, Duan S, Sun Y. Speech Emotion Recognition Using a Dual-Channel Complementary Spectrogram and the CNN-SSAE Neutral Network. Applied Sciences. 2022 Sep 22;12(19):9518.

[14]. Ishtiak I, Rahman MM, Usmani MR. Early threat warning via speech and emotion recognition from voice calls (Doctoral dissertation, BRAC University)

[15]. Lian H, Lu C, Li S, Zhao Y, Tang C, Zong Y. A Survey of Deep Learning-Based Multimodal Emotion Recognition: Speech, Text, and Face. Entropy. 2023 Oct 12;25(10):1440]

[16]. Madanian S, Chen T, Adeleye O, Templeton JM, Poellabauer C, Parry D, Schneider SL. Speech emotion recognition using machine learning—A systematic review. Intelligent systems with applications. 2023 Aug 14:200266.

[17]. Singh P, Srivastava R, Rana KP, Kumar V. A multimodal hierarchical approach to speech emotion recognition from audio and text. Knowledge-Based Systems. 2021 Oct 11;229:107316. this paper author.

[18]. Bisio I, Delfino A, Lavagetto F, Marchese M, Sciarrone A. Gender-driven emotion recognition through speech signals for ambient intelligence applications. IEEE transactions on Emerging topics in computing. 2013 Jul 25;1(2):244-57

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 1, February 2024

[19]. Singh YB, Goel S. A systematic literature review of speech emotion recognition approaches. Neurocomputing. 2022 Jul 1;492:245-63.

[20]. Avots E, Sapiński T, Bachmann M, Kamińska D. Audiovisual emotion recognition in wild. Machine Vision and Applications. 2019 Jul 1;30(5):975-85.

[21]. Dai W, Han D, Dai Y, Xu D. Emotion recognition and affective computing on vocal social media. Information & Management. 2015 Nov 1;52(7):777-88.

[22]. Vicsi K, Sztahó D. Recognition of Emotions on the Basis of Different Levels of Speech Segments. J. Adv. Comput. Intell. Intell. Informatics. 2012 Mar 20;16(2):335-40.

[23]. Zhao Z, Bao Z, Zhao Y, Zhang Z, Cummins N, Ren Z, Schuller B. Exploring deep spectrum representations via attention-based recurrent and convolutional neural networks for speech emotion recognition. IEEE Access. 2019 Jul 15;7:97515-25.

[24]. Kakuba S, Poulose A, Han DS. Deep learning-based speech emotion recognition using multi-level fusion of concurrent features. IEEE Access. 2022 Nov 30;10:125538-51.

[25]. Alslaity A, Orji R. Machine learning techniques for emotion detection and sentiment analysis: current state, challenges, and future directions. Behaviour & Information Technology. 2024 Jan 2;43(1):139-64.