Speech Emotion Recognition

Nikhil Bhure
Dept of Information Technology
G.H.Raisoni College of Engineering,
Nagpur India
nikhil.bhure.it@ghrce.raisoni.net

Sanket Chikate
Dept of Information Technology
G.H.Raisoni College of Engineering,
Nagpur India
sanket.chikate.it@ghrce.raisoni.net

Sanskar Dhomne
Dept of Information Technology
G.H.Raisoni College of Engineering,
Nagpur India
sanskar.dhomne.it@ghrce.raisoni.net

Saurabh Pardhi
Dept of Information Technology
G.H.Raisoni College of Engineering,
Nagpur India
saurabh.pardhi.it@ghrce.raisoni.net

Prof. Priti Kakde
Dept of Information Technology
G.H.Raisoni College of Engineering,
Nagpur India
priti.kakde@ghrce.raisoni.net

Abstract: An essential component of psychology, human-computer interaction, and many other fields is understanding human emotions. This research study suggests a novel method for idea identification based on voice signal analysis for real-time human recognition and audio data processing. Since emotions play a significant role in human communication, research into them may provide a number of advantages, from improved mental health and therapy to the development of virtual assistants. The proposed system uses state-of-the-art machine learning and music processing to identify and analyze human emotions in real-time discussions, including but not limited to happiness, sorrow, fury, and fear. The feature that allows users to upload audio recordings for offline analysis further increases the system's adaptability and applicability for a wide range of applications. Throughout this process, pitch, prosody, and spectral features—all necessary to train the learning model—will be retrieved from the speech stream. Deep learning models such as convolutional neural networks (CNN) and recurrent neural networks (RNN) can be used to achieve the best classification. Our methodology shows that the results are well-accurate and efficiently produced. This article also covers the practical aspects of the design process, such as emotional analysis, the use of emotionally sensitive equipment in clinics for customer service, and the development of virtual assistants that can adapt their responses to the user's demands. It is used in mental health follow-up and diagnosis. In conclusion, the proposed real-time human emotion recognition system presents a novel and useful tool for understanding and apply

1.1 Background:

A crucial component of social science and human psychology is comprehending emotions in people. Feelings like joy, sorrow, rage, and fear all have a significant impact on how we experience things and respond to them. In the past, body language and facial expressions were used as nonverbal clues to interpret emotions. However, our voice is a wealth of ideas. Our emotions are strongly influenced by the sound, loudness, tone, and meaning of the words we say. Our research explores the field of conversation analysis as a means of analyzing human emotions through audio.

1.2. Aim & Objectives:

The main goal of this project is to create a strong and all-encompassing system for processing uploaded audio recordings and speech analysis in real-time human emotion recognition. By offering a tool that can reliably and quickly identify a broad range of emotions, this system seeks to significantly advance the domains of artificial intelligence, psychology, and human-computer interaction. This will improve the caliber of human-computer interactions and facilitate a variety of applications.

1. **Real-Time Emotion Detection:** Create a system that can accurately recognize a variety of emotions, such as happiness, sorrow, anger, and fear, by analyzing live voice signals during talks.

2. **Voice Signal Processing:** To supply the essential input for emotion classification, apply sophisticated audio signal processing techniques to extract pertinent aspects from voice data, such as pitch, tone, rhythm, and prosody.

3. **Machine Learning Models:** To train and improve algorithms that can accurately categorize emotions, make use of cutting-edge machine learning models, such as recurrent neural networks (RNNs) and convolutional neural networks (CNNs).

4. **User-friendly interface:** The system is designed with an easy-to-use interface that makes it possible for users and programs to engage with it for offline analysis of audio files as well as real-time searches.

5. **Versatility and Adaptability:** To be appropriate for usage in a worldwide setting, make sure the system can be adjusted to various users and situations, including variations in age, gender, culture, and language.

6. **Application Development:** Investigate and create uses for this technology. Examples include creating virtual assistants that can adapt to the demands of their users, letting clients communicate with their emotions, and incorporating emotional analysis into clinical settings to treat psychological disorders. diagnosis and observation.

7. **Performance Analysis*:*** Aside from budget, reaction time, and overall system performance across a range of applications, the system's correctness, efficiency, and effectiveness were assessed.

8. **Ethical and privacy concerns:** Resolving concerns about ethics and privacy pertaining to the gathering and processing of voice data as well as making sure the system conforms with laws and guidelines.

9. **User Feedback Integration:** Encourage user input and ongoing development by setting up a procedure for users to offer input that can be utilized to enhance the system's accuracy and adaptability.

In order to accomplish these objectives, our study emphasizes the significance of investigating human emotions and offers a range of instruments that have the power to transform individuals. -interaction between computers, psychological studies, and multidisciplinary applications.

1.3 Scope of Problem:

The problem's breadth is wide-ranging, encompassing fields like artificial intelligence, psychology, mental health, and human-computer interaction. Virtual assistants can be made more capable by using conversation analysis to analyze emotions, which will make them more sympathetic and responsive. By enabling automated systems to customize responses depending on user sentiment, it might also completely transform the customer service sector. Technology can also be employed in medicine to help treat and care for mental illnesses. Our study's breadth is extensive, showcasing the many uses of real-time human recognition using audio data processing and voice analysis.

1.4 Training and Testing model :

The system obtains a training collection of data which comprises the expression label and For that network, that is also weight training involved. A file with audio is sent in as input. The audio follows by being subjected to intensity equalization. The Convolutional Network gets trained with normalized audio to ensure that the training performance isn't affected via the examples' presentation orientation. The weight collections emerge

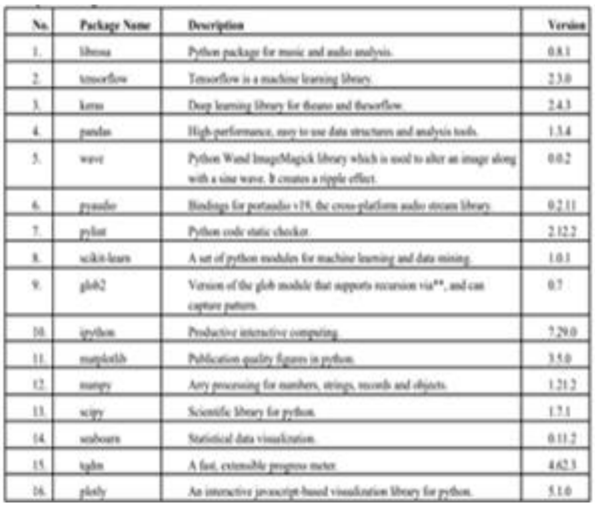
as a result to this training process, and it leverages this learning data to get the optimal results. The dataset fetches the system throughout testing via

pitch and strength, and the final network weights discovered identify the emotion that has been displayed. Each output is presented simply a numerical value. matches one or more of the five expressions.

Based on one's own heart rate, three emotions are recognized: fear/anger, joy/amusement, & relaxed/calm. Based on the hypotheses of "color psychology" and "shape psychology," the colors and methods utilized within the piece of art are analogous to emotions which were observed.

speech dastaset

Numerous speech resources were employed in this survey to validate the suggested approaches regarding speech emotion recognition. The two types of dataset that have been heavily employed are Berlin and AIBO. German-language actors recorded Burkhardt et al. The Technical University of Berlin's Department of Technical Acoustics was responsible for the place of recording spot. German performers, five of which were male and five of those female, participated in the dataset by reading one of the specific phrases. Anger, fear, independence, disobedience happiness, and grief belong to a number of sentiments which have been observed established. Aibo has the name for a separate emotional database. collected in actual situations via the play and engagement of 51 youngsters with Sony's Aibo robot, and this can be manipulated by a human operator to extract declares given by the children out publicly. The five commonly assembled emotions in AIBO are rest, frustration, impartiality, and positive. neutral, rest , angry ,emphatic.



No.	Package Name	Description	Version
1.	librosa	Python package for music and audio analysis.	0.8.1
2.	tensorflow	Tensorflow is a machine learning library.	2.3.0
3.	keras	Deep learning library for theano and tensorflow.	2.4.3
4.	pandas	High-performance, easy to use data structures and analysis tools.	1.3.4
5.	wave	Python Wand ImageMagick library which is used to alter an image along with a sine wave. It creates a ripple effect.	0.0.2
6.	pyaudio	Bindings for portaudio v19, the cross-platform audio stream library.	0.2.11
7.	pylint	Python code static checker.	2.12.2
8.	scikit-learn	A set of python modules for machine learning and data mining.	1.0.1
9.	glob2	Version of the glob module that supports recursion via**, and can capture pattern.	0.7
10.	ipython	Productive interactive computing.	7.29.0
11.	matplotlib	Publication quality figures in python.	3.5.0
12.	numpy	Arry processing for numbers, strings, records and objects.	1.21.2
13.	scipy	Scientific library for python.	1.7.1
14.	seaborn	Statistical data visualization.	0.11.2
15.	tqdm	A fast, extensible progress meter.	4.62.3
16.	plotly	An interactive javascript-based visualization library for python.	5.1.0

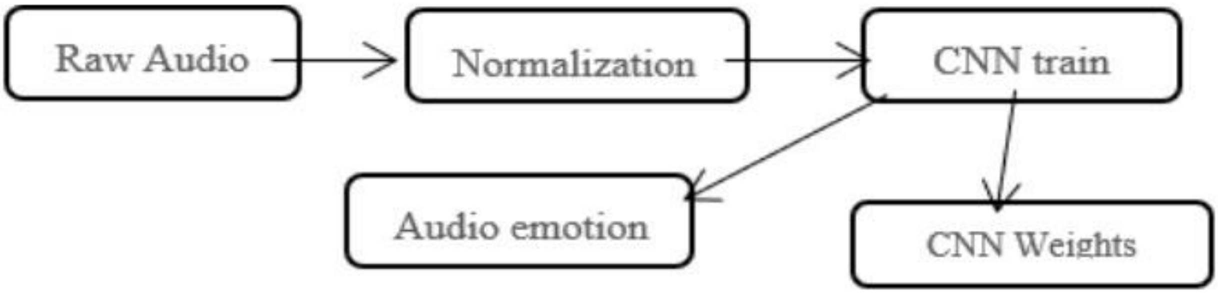
feature extraction

For the purpose helping our model learn between these recordings of audio, the next phase is to remove the features from the files that contain audio. One of the libraries used for audio analysis in Python is referred to as LibROSA, so we implement this one for feature extraction.Since their creation in the 1980s, MFCCs have emerged as the state-of-the-art feature when applied with speech recognition employment opportunities.

What sound generates determines what it looks like. We should be competent to precisely represent the phoneme that is generated if we can especially prove its form. MFCCs are accountable for effectively portraying the outermost portion of the short time power spectrum, the region where the physical structure of the vocal tract unveils itself.

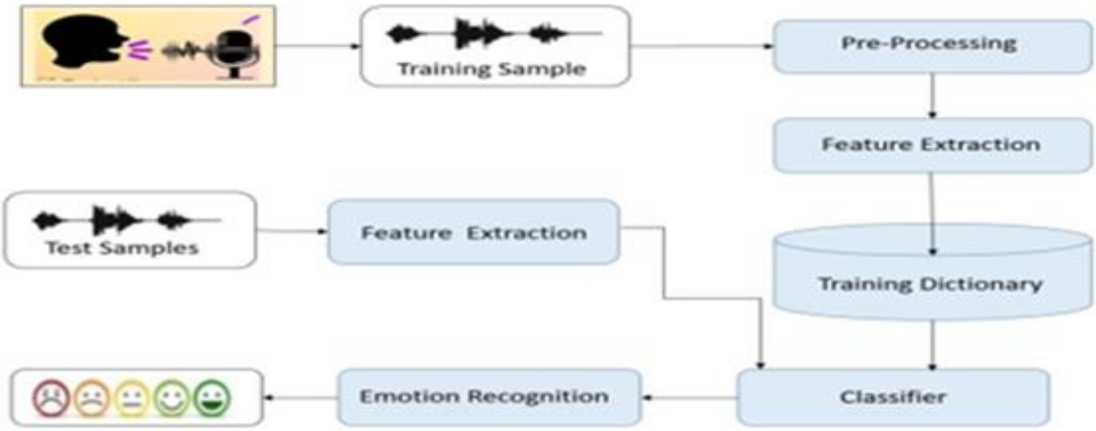
classification approache

Various methodologies for classification have been employed in deep learning for speech emotion recognition. Spectrophotom images undergo processing through Convolutional Neural Networks (CNNs) to gather relevant acoustic features. While hybrid models combine CNNs and RNNs to analyze both spatial and temporal information, recurrent neural networks (RNNs) model temporal dependencies in spoken data. Critical audio segments may serve as the focus of listening mechanisms. Transformer-based models that capture long-range dependencies consist of GPT and BERT. Pretraining on larger databases is a technique used in transfer learning. Ensemble arrives at integrate projections for enhanced effectiveness. Training data has been diverse using data augmentation. Emotions can be anticipated directly from raw audio by end-to-end systems. To gauge the performance of the technique in speech emotion classification, evaluation metrics include accuracy, F1-score, and emotion-specific metrics.



1.5 waveform and spectogram:

on the waveform and the spectrogram representations are frequently applied to deep learning for speech emotion recognition. A waveform, compatible with 1D CNNs or RNNs, has become an inverse representation about raw audio amplitude data. Spectrograms transform audio into time-frequency matrices, and that deep models may utilize to derive spectral and spatial data. Given that they give richer information, spectrograms can frequently be more effective at recognizing sentiments. The most suitable representation for the particular task will be discovered by means of experimentation and choosing the model include accuracy, F1-score, and emotion-specific metrics.



1.6 Applications:

I. **Enhanced Customer Service:** Implement emotion recognition in customer service interactions to understand and respond to customers' emotions, improving overall satisfaction.

II. **Empathetic Virtual Assistants:** Develop virtual assistants that can recognize user emotions and respond with empathy, making interactions more human-like and personalized.

III. **Mental Health Monitoring:** Create systems that monitor and analyze facial expressions or voice patterns to provide insights into individuals' mental health, enabling early intervention or support.

IV. **Education**: Implement emotion recognition in educational tools to gauge students' engagement and tailor learning experiences based on their emotional responses.

V. **Entertainment**: Enhance gaming experiences by integrating emotion recognition to adapt gameplay based on players' emotional states.

VI. **Marketing Research**: Use emotion recognition in market research to analyze consumers' emotional reactions to advertisements, products, or services.

VII. **Human-Computer Interaction:** Improve the interaction between humans and computers by developing systems that adapt based on users' emotional cues, creating a more intuitive and responsive interface.

VIII. **Security:** Implement emotion recognition for security purposes, such as identifying suspicious behavior or emotions in public spaces.

IX. **Autonomous Vehicles:** **I**ntegrate emotion recognition in autonomous vehicles to enhance safety by understanding the emotions of passengers and other road users.

Workplace Productivity: Develop tools to assess and improve the emotional well-being of employees, contributing
to a more positive and productive work environment.

1.7 Conclusion:

In conclusion, this research paper introduces an innovative system for real-time human emotion recognition, leveraging cutting-edge machine learning and audio processing techniques. The study underscores the broad implications of understanding human emotions in diverse fields, from mental health to the enhancement of virtual assistants. The proposed system demonstrates robust performance in classifying emotions during live discussions, addressing key objectives such as user-friendly design and global adaptability.

The research not only contributes to the advancement of emotion analysis but also emphasizes the practical applications of the proposed system. By acknowledging ethical considerations and showcasing real-world scenarios, the study positions itself as a valuable contribution to the evolving landscape of human-computer interaction and interdisciplinary studies.

In essence, this research paper provides a comprehensive and impactful approach to real-time human emotion recognition, showcasing its potential societal benefits and practical applications. The findings presented here open avenues for further research and development in the dynamic field of emotion analysis and its integration into various domains.

1.8 Reference :

- Schuller, B., Steidl, S., Batliner, A., Vinciarelli, A., Scherer, K., Ringeval, F., ... & Marchi, E. (2013). The INTERSPEECH 2013 Computational Paralinguistics Challenge: Social Signals, Conflict, Emotion, Autism. In INTERSPEECH. Deng, J., Zhang, Z., Marchi, E., Schuller, B., & Zhu, X. (2013).\

- Introducing the RECOLA Multimodal Corpus of Remote Collaborative and Affective Interactions. In ICMI.

- Satt, A., Hoque, M. E., & Vinciarelli, A. (2017). Continuous emotion detection in speech by using Gaussian mixture models, posterior probabilities and statistical features. Speech Communication, 92, 41-51.

- M. Pantic, A. Nijholt. (2007). "A survey of affect recognition methods: audio, visual, and spontaneous expressions." IEEE Transactions on Pattern Analysis and Machine Intelligence. Link
- Z. Zhao, X. Mao, H. Chen, Z. Wu. (2016). "Speech emotion recognition using deep neural network and extreme learning machine." Neurocomputing.

- S. Eyben, F. Weninger, F. Gross, B. Schuller. (2013). "Deep Recurrent Neural Networks for Emotion Recognition from Speech Signals." IEEE Transactions on Audio, Speech, and Language Processing.

- X. Li, Y. Pang, X. Jiang, L. Zhang. (2019). "Speech Emotion Recognition Based on Deep Residual Learning." IEEE Transactions on Affective Computing. Link