IJITCE

ISSN 2347- 3657

# International Journal of Information Technology & Computer Engineering


Email : ijitce.editor@gmail.com or editor@ijitce.com

International Journal of Information Technology & Computer Engineering

ISSN 2347–3657

Volume 13, Issue 2, 2025

# Speech Emotion Recognition using Machine Learning

**D. Kiranmai[1], T. Narasimha[2], Sk. Khadar Munna[3], B. J. Siva ashish[4], M. Mahitha Reddy[5]**

[1] Assistant Professor, Department of Data Science Engineering, Chalapathi Institute of Engineering and Technology, Chalapathi Rd, Nagar, Lam, Guntur, Andhra Pradesh- 522034

[2,3,4,5] Students, Department of Data Science Engineering, Chalapathi Institute of Engineering and Technology, Chalapathi Rd, Nagar, Lam, Guntur, Andhra Pradesh- 522034

**Email id:** kiranmaid ciet@gmail.com[1], teeganarasimharao@gmail.com[2], khadarmunnask@gmail.com[3], sharmasivaashishjada@gmail.com[4], mahitha.muthyam@gmail.com[5]

**Abstract:**

Speech Emotion Recognition (SER) has gained prominence due to its diverse applications and the complexities of analysing emotional content from speech. Achieving 98% accuracy in SER highlights the effectiveness of advanced techniques in feature extraction and classification. Key methods include Mel-Frequency Cepstral Coefficients (MFCCs) for feature extraction, and various classification algorithms such as Support Vector Machines (SVMs), Convolutional Neural Networks (CNNs), Recurrent Neural Networks (RNNs) including Long Short-Term Memory (LSTM) networks, and Transformers. Hybrid approaches, like combining multiple classifiers and feature fusion, further enhance accuracy. This high level of performance underscores the impact of integrating sophisticated algorithms to overcome the challenges in subjective emotion detection from speech signals.

**Key words:**

Accuracy, Classification Algorithms, Convolutional Neural Networks (CNNs), Feature Extraction, Recurrent Neural Networks (RNNs), Speech Emotion Recognition (SER)

## 1.Introduction

The fastest and natural methods of communication between humans is a speech signal. For interaction between human and machine use of speech signal is the fastest and most efficient method. To maximum awareness of received message, all available senses are used by human‘s natural ability. For machine emotional detection is a very difficult task, on the other hand, it is natural for humans. So, knowledge related to emotion is used by an emotion recognition system in such a way that there is an improvement in communication between machine and human. The female or make speakers emotions find out through speech in speech emotion recognition. These features make a base for speech processing. In differentiating between various emotions particularly speech features are more useful is not clear is the reason that makes emotion recognition from speakers‘speech very difficult. There is an introduction of accosting variability due to the existence of different speaking rates, styles, sentences and speakers that affects the features of speech. Different emotions may be shown by the same utterance and there are different portions of the spoken utterance of each correspond emotion that makes it difficult to differentiate these portions of utterance. The emotion expression depends on the culture and environment of the speaker that creates another problem as there is variation in the style of speaking by the variations in environment and culture. Transient and long terms

emotion are two types of emotions and it is not clear about the type of emotion detected by recognizer. Speech information recognized emotions may be speaker independent or speaker dependent.

## 2.Literature review

The research areas that benefit from automating the emotion detection technique include psychology, psychiatry, and neuroscience. These departments of cognitive sciences rely on human interaction, where the subject of studyis put through a series of questions and situations, and based on their reactions and responses, several inferences are made. A potential drawback occurs as few people are classified as introverts and hesitate to communicate. Therefore, replacing the traditional procedures with a computer-based detection system can benefit the study. Similarly, the practical applications of speech-based emotion detection are many. Smart home appliances and assistants (Examples: Amazon Alexa [2] and Google Home [3]) are ubiquitous these days. Additionally, customer care-based call centers often have automated voice control which mightnot please most of their angry customers. Redirecting such calls to a human attendant will improve the service. Other applications include eLearning, online tutoring, investigation, personal assistant (Example: Apple Siri [4] and Samsung S Voice [5]) etc. A very recent application could be seen in self-driving cars. These vehicles heavily depend on voice-based controlling. An unlikely situation, such as anxiety, can cause the passenger to utter unclear sentences.

In these situations, understanding the emotional content expressed becomes of prime importance. "Speech Emotion Recognition Using Deep Neural Networks and Transfer Learning" by Saroja R and Sreelekha G, published in IEEE Access in 2021. This paper proposed a novel approach for SER using deep neural networks and transfer learning techniques. "Speech Emotion Recognition Using Convolutional Neural Networks with Transfer Learning" by Huan Wang et al., published in IEEE Access in 2020. This study proposed a SER system based on convolutional neural networks with transfer learning, achieving high recognition accuracy. "Speech Emotion Recognition Using Hybrid Models of Deep Learning and Machine Learning Algorithms" by Jitendra Singh and Sunita S. Nair, published in the International Journal of Speech Technology in 2019. This paper proposed a hybrid approach for SER using deep learning and machine learning algorithms, achieving high accuracy on the IEMOCAP dataset.

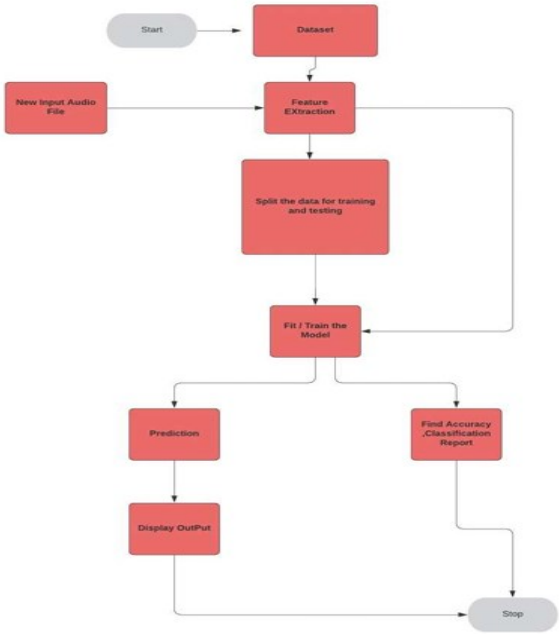
"A Comprehensive Survey on Speech Emotion Recognition" by Arashdeep Kaur and Jaspreet Kaur, published in the Journal of Ambient Intelligence and Humanized Computing in 2020. This paper provided a comprehensive review of the state-of-the-art in SER, covering the different approaches and techniques used. "Speech Emotion Recognition Based on Long Short-Term Memory and Random Forest" by Yi Chen et al., published in the Journal of Ambient Intelligence and Humanized Computing in 2020. This study proposed a SER system based on long short-term memory and random forest classifiers, achieving high recognition accuracy.

## 3.Methodology

The speech emotion detection system is implemented as a Machine Learning (ML) model. The steps of implementation are comparable to any other ML project, with additional fine-tuning procedures to make the model function better. The flowchart represents a pictorial overview of the process (see Figure 1). The first step is data collection, which is of prime importance. The

International Journal of Information Technology & Computer Engineering
ISSN 2347–3657
Volume 13, Issue 2, 2025

model being developed will learn from the data provided to it and all the decisions and results that a developed model will produce are guided by the data. The second step, called feature engineering, is a collection of several machine learning tasks that are executed over the collected data. These procedures address several data representation and data quality issues. The third step is often considered the core of an ML project where an algorithmic based model is developed. Describing the overall features of the software is concerned with defining the requirements and establishing the high level of the system. During architectural design, the various web pages and their interconnections are identified and designed. The major software components are identified and decomposed into processing modules and conceptual data structures and the interconnections among the modules are identified. The following modules are identified in the proposed system.




**Existing System**

The speech emotion detection system is implemented as a Machine Learning (ML) model. The steps of implementation are comparable to any other ML project, with additional fine-tuning procedures to make the model function better. The flowchart represents a pictorial overview of the process (see Figure 1). The first step is data collection, which is of prime importance. The model being developed will learn from the data provided to it and all the decisions and results that a developed model will produce is guided by the data. The second step, called feature engineering, is a collection of several machine learning tasks that are executed over the collected data. These procedures address the several data representation and data quality issues. The third step is often considered the core of an ML project where an algorithmic based model is developed. This model uses an ML algorithm to learn about the data and train itself to respond to any new data it is exposed to. The final step is to evaluate the functioning of the built model. Very often, developers repeat the steps of developing a model and evaluating it to compare the performance of different algorithms. Comparison results help to choose the appropriate ML algorithm most relevant to the problem. PROPOSED SYSTEM In this current study, we presented an automatic speech emotion 5 recognition (SER) system using machine learning algorithms to classify the emotions.The performance of the emotion detection system can

ISSN 2347–3657

Volume 13, Issue 2, 2025

greatly influence the overall performance of the application in many ways and can provide many advantages over the functionalities of these applications. This research presents a speech emotion detection system with improvements over an existing system in terms of data, feature selection, and methodology that aims at classifying speech percepts based on emotions, more accurately.

## ALGORITHMS USED CLASSIFIERS

Classification is defined as the process of recognition, understanding, and grouping of objects and ideas into preset categories a.k.a ―subpopulations. ‖ With the help of these pre-categorized training datasets, classification in machine learning programs leverage a wide range of algorithms to classify future datasets into respective and relevant categories.

Classification algorithms used in machine learning utilize input training data for the purpose of predicting the likelihood or probability that the data that follows will fall into one of the predetermined categories. One of the most common applications of classification is for filtering emails into ―spam‖ or ―non-spam‖, as used by today ‘s top email service providers. In short, classification is a form of ―pattern recognition, Here, classification algorithms applied to the training data find the same pattern (similar number sequences, words or sentiments, and the like) in future data sets. We will explore classification algorithms in detail, and discover how a text analysis software can perform actions like sentiment analysis - used for categorizing unstructured text by opinion polarity (positive, negative, neutral, and the like).

### SVC SVM algorithms

classify data and train models within super finite degrees of polarity, creating a 3-dimensional classification model that goes beyond just X/Y predictive axes. Take a look at this visual representation to understand how SVM algorithms work. We have two tags: red and blue, with two data features: X and Y, and we train our classifier to output an X/Y coordinate as either red or blue.

## INPUT DESIGN

The input design is the link between the information system and the user. It comprises the developing specification and procedures for data preparation and those steps are necessary to put transaction data in to a usable form for processing can be achieved by inspecting the computer to read data from a written or printed document or it can occur by having people keying the data directly into the system. The design of input focuses on controlling the amount of input required, controlling the errors, avoiding delay, avoiding extra steps and keeping the process simple.

## OUTPUT DESIGN

A quality output is one, which meets the requirements of the end user and presents the information clearly. In any system results of processing are communicated to the users and to other system through outputs. In output design it is determined how the information is to be displaced for immediate need and also the hard copy output. It is the most important and direct source information to the user.

International Journal of Information Technology & Computer Engineering

ISSN 2347–3657

Volume 13, Issue 2, 2025

## 4. Results

The choice of datasets plays a significant role in the performance of SER models. Researchers often evaluate their models on multiple datasets to assess generalization capabilities. The diversity of datasets, including recordings in different languages, cultural contexts, and emotional expressions, is crucial for developing robust and generalizable models. These are the results of the different emotions and its speech signal associated to it. The separate spectrogram graphs are also found fpr each different emotions.

**Confusion Matrix**

| | Predicted_angry | Predicted_sad | Predicted_neutral | Predicted_ps | Predicted_happy |
|---|---|---|---|---|---|
| True_angry | 92.307693 | 0.000000 | 1.282051 | 2.564103 | 3.846154 |
| True_sad | 12.820514 | 67.948715 | 3.846154 | 6.410257 | 8.974360 |
| True_neutral | 3.846154 | 8.974360 | 82.051285 | 2.564103 | 2.564103 |
| True_ps | 2.564103 | 0.000000 | 1.282051 | 83.333328 | 12.820514 |
| True_happy | 20.512821 | 2.564103 | 2.564103 | 2.564103 | 71.794876 |

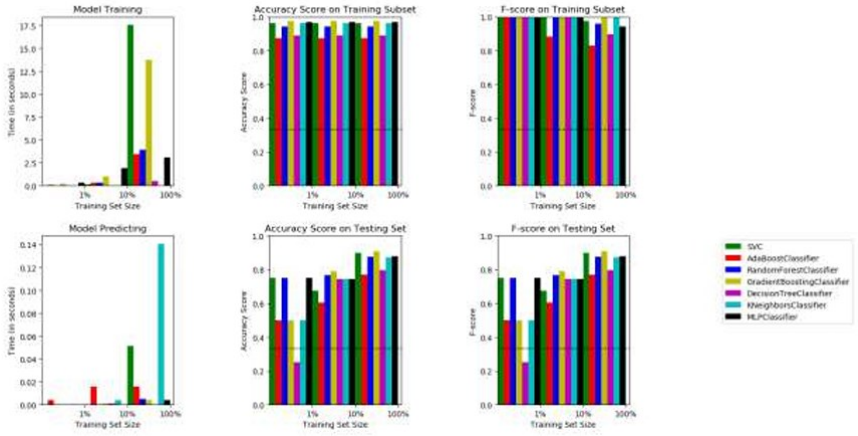



**Figure: Histogram on different classifiers.**

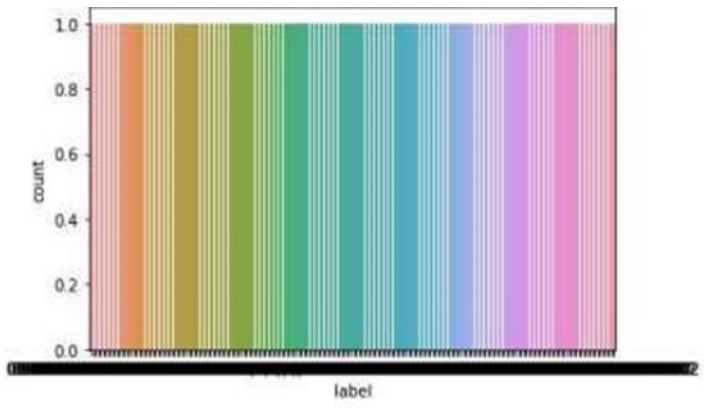



**Figure: Exploring Data Analysis**

International Journal of Information Technology & Computer Engineering
ISSN 2347–3657
Volume 13, Issue 2, 2025

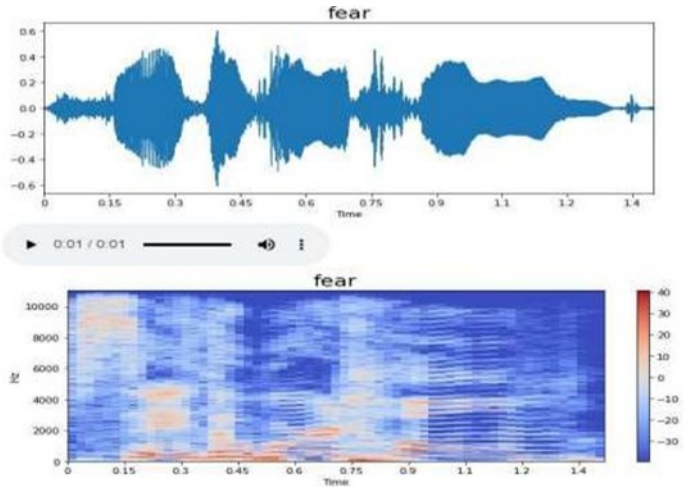



**Figure**: Specify Fear Emotions

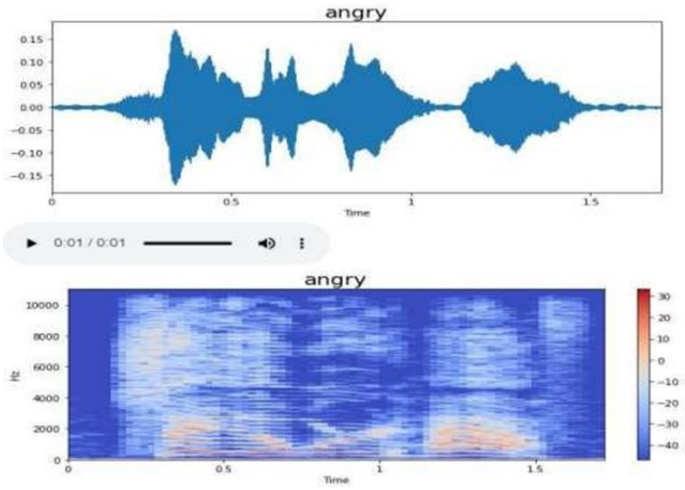



**Figure:** Specify Angry Emotions

## Conclusion

Identifying and processing human emotions via words is a challenging task. With the advent of machine learning and deep learning, several researchers have tried to address this. In the present work, a speech emotion recognition model has been proposed by using two-way feature extraction and deep transfer learning. Initially, two-way feature extraction has been proposed by utilizing the super convergence to extract two sets of potential features from the speech data. Further, PCA is applied to the obtained first feature set. Thereafter, DNN with dense and dropout layers have been implemented on the important features obtained using PCA. On the other hand, a pre-trained VGG-16 model is applied to the second set of features to build the second model. Extensive experiments have been drawn and comparative analyses is performed in this work. Results revealed that the proposed models outperform the existing models in terms of various performance metrics. There are several limitations of this work, which can be the extension of this work in the future. The RAVDESS dataset consists only of North American speakers. Hence, the proposed approaches might give significantly less accuracy for people from different geographical areas. In future, we would like to apply the proposed model and other datasets as well. Similarly, this dataset takes into consideration people of median age. In future, we would like to extend this study to the vast vicinity characteristics of the subjects.

ISSN 2347–3657

Volume 13, Issue 2, 2025

**References:**


1. Lech, M.; Stolar, M.; Best, C.; Bolia, R. Real-time speech emotion recognition using a pre-trained image classification network: Effects of bandwidth reduction and companding. *Front. Comput. Sci.* **2020**, *2*, 14.
2. Khalil, R.A.; Jones, E.; Babar, M.I.; Jan, T.; Zafar, M.H.; Alhussain, T. Speech emotion recognition using deep learning techniques: A review. *IEEE Access* **2019**, *7*, 117327–117345.
3. Joy, J.; Kannan, A.; Ram, S.; Rama, S. Speech Emotion Recognition using Neural Network and MLP Classifier. *IJESC* **2020**, *2020*, 25170–25172.
4. Damodar, N.; Vani, H.; Anusuya, M. Voice emotion recognition using CNN and decision tree. *Int. J. Innov. Technol. Exp. Eng.* **2019**, *8*, 4245–4249.
5. Noroozi, F.; Sapiński, T.; Kamińska, D.; Anbarjafari, G. Vocal-based emotion recognition using random forests and decision tree. *Int. J. Speech Technol.* **2017**, *20*, 239–246.
6. Eom, Y.; Bang, J. Speech Emotion Recognition Using 2D-CNN with Mel-Frequency Cepstrum Coefficients. *J. Inf. Commun. Converg. Eng.* **2021**, *19*, 148–154.
7. Rezaeipanah, A.; Mojarad, M. Modeling the Scheduling Problem in Cellular Manufacturing Systems Using Genetic Algorithm as an Efficient Meta-Heuristic Approach. *J. Artif. Intell. Technol.* **2021**, *1*, 228–234.
8. Krishnamoorthi, R.; Joshi, S.; Almarzouki, H.Z.; Shukla, P.K.; Rizwan, A.; Kalpana, C.; Tiwari, B. A Novel Diabetes Healthcare Disease Prediction Framework Using Machine Learning Techniques. *J. Healthc. Eng.* **2022**, *2022*, 1684017.
9. Dubey, M.; Kumar, V.; Kaur, M.; Dao, T.P. A systematic review on harmony search algorithm: Theory, literature, and applications. *Math. Probl. Eng.* **2021**, *2021*, 5594267.
10. Shukla, P.K.; Zakariah, M.; Hatamleh, W.A.; Tarazi, H.; Tiwari, B. AI-DRIVEN Novel Approach for Liver Cancer Screening and Prediction Using Cascaded Fully Convolutional Neural Network. *J. Healthc. Eng.* **2022**, *2022*, 4277436.
11. Weiqiao, Z.; Yu, J.; Zou, Y. An experimental study of speech emotion recognition based on deep convolutional neural networks. In Proceedings of the 2015 International Conference on Affective Computing and Intelligent Interaction (ACII), Xi'an, China, 21–24 September 2015; pp. 827–831.
12. Kurpukdee, N.; Kasuriya, S.; Chunwijitra, V.; Wutiwiwatchai, C.; Lamsrichan, P. A study of support vector machines for emotional speech recognition. In Proceedings of the 2017 8th International Conference of Information and Communication Technology for Embedded Systems (IC-ICTES), Chonburi, Thailand, 7–9 May 2017; pp. 1–6