International Journal of Computational Science and Engineering Research
ISSN: 3107 - 8605 (Online) , http://www.ijcser.com/
Regular Issue , Vol. 3, Issue. 1 (January – March) , 2026 , Pages: 1 - 8
Received: 28 March 2025 ; *Accepted:* 25 December 2025 ; *Published:* 04 January 2026
https://doi.org/10.63328/IJCSER-V3RI1P1
https://www.youtube.com/watch?v=29Siat_E7kw
https://archive.org/details/jsp-ijcser-v3ri1p1
Original Paper : Applied Research – Compare and Contrast Research Paper

Enhancing Speech Emotion Recognition with Deep Learning Techniques

S Guru Prasad [1], M Sreevani [2] , Venkata Ramana Guntreddi [3]

[1,2] Department of Computer Science and Engineering, Vemu Institute of Technology, Andhra Pradesh , India
sguruprrasad@gmail.com , vani.cse183@gmail.com
[3] Department of Electrical, Telecommunication and Computer Engineering, School of Engineering and Applied Sciences, Kampala International University - Western Campus, Bushenyi, Uganda; gvramana@kiu.ac.ug

** Corresponding Author: S Guru Prasad ; sguruprrasad@gmail.com*

Abstract: Recently developed speech emotion recognition (SER) methods have placed much emphasis on the derivation of features of acoustic data. In this paper, an alternative method is described, which shows a higher accuracy of the SER implementation because it does not rely on any manual steps, such as feature extraction. The suggested approach is a mixture of Conformer blocks and CNNs which are used to predict the emotions directly based on the audio signals. The Conformer block is used to solicit long-range dependencies and time-specific features whereas CNN layers are good at extracting localized emotional information. In this design, detailed emotional content is maintained since it is otherwise overlooked due to the conventional approaches and at the same time the hierarchical depictions of emotional content are captured. The Conformer combination of time and situation data, and the CNN layers make special attention to extracting spatial features. The model was tested on three publicly available datasets and in more than one language and significantly improved accuracy and interpretability with the use of emotional and temporal information. This method moves toward affective computing since it offers a more detailed and effective analysis of the speech data.
Keywords: Affective Computing, Speech Emotion Recognition, Deep Learning, Conformer-CNN Hybrid, Interpretability.

1. Introduction

SER is among the most significant fields of studies in the greater scheme of affective computing as it attempts to bridge the divide between the human feelings and how the machines comprehend the human feelings. The clue to building the more humane and interactive types of the speech recognition of emotions is the speech recognition of the emotions themselves. In addition, these techniques can not be easy in handling the compound emotional cues that might not be reflected on the chosen attributes [1]. This has been in light of the fact that the rate of growth of deep learning technologies has been fast, and this has led to the drift to the end-to-end learning models, which transform raw speech data. The models may either automatically determine relevant features in accordance with the information without determining features manually or determine more subtle and subtle aspects of emotion. CNNs and RNNs are regarded as some of the finest methods of deep learning which ought to be utilized in recognizing speech emotions [2]. CNNs are better at capturing the local features in speech, and RNNs, specifically, the (LSTM) network are more adept at learning the temporal association. Nevertheless, despite the successes, such models continue to be austere, and they fail to capture fully the long range dependencies and contextual information which is important to the process of learning about how emotion in speech transition and progress [3]. The emotion is constructed. It is better able to process raw speech signals with the help of Conformer blocks, albeit applying CNN layers, and identify both short-term emotional features and the long-term features. It has already been established that this hybrid technique is an extremely successful method to render SER systems more precise and interpretable, which can be more readily extrapolated to other data sets and other types of expression of emotions.

2. Background

Speech Emotion Recognition is a significant aspect of human-computer interaction, but the aim of which is to identify and classify human emotions through the use of spoken language. The emotions are the main building block of human communication, and they can be recognized in the speech to allow the system to be more

Jack Sparrow Publishers © 2026, IJCSER, All Rights Reserved
www.jacksparrowpublishers.com

ISSN: 3107 - 8605 (Online) , http://www.ijcser.com/ , Vol. 3, Issue 1, 2026, https://doi.org/10.63328/IJCSER-V3RI1P1

understanding and receptive. The recognition of emotions through speech is a complex procedure of examination of audio keys since emotions are conveyed through the help of various acoustic qualities, e.g., the tone, pitch, rhythm, volume. However, the traditional methods of emotion recognition often rely on features that are selected by hand and it could be limited in terms of accuracy and scalability [8][9]. This paradigm shift of the field in favor of deep learning models, which drive SER, has happened throughout the years as complex representations can be trained on these models, which are able to train representations without considering raw speech data [4].

The ability of recognizing speech emotion includes Convolutional Neural Network, (RNNs), and more recently, the Transformer-based models. CNNs are suited to local and short-term feature extraction of raw audio signals, and RNNs, in general, and the (LSTM) networks specifically, are suited to long-term dependencies and speech temporal dynamics. Despite the fact that these models have been found to be effective, they also have issues in terms of fully capturing the local as well as global tendencies of speech that are important in being able to discern emotions accurately [5].The conformer model has been adopted as a remedy to these shortcomings, with the incorporation of the benefits of the CNNs model and the Transformer model as a unified solution.

A more efficient solution to emotion detection can be the Conformer model since it is a combination of Convolutional layers and Transformer attention mechanisms and therefore more capable of detecting the complexity of emotions when it comes to speech [6]. The hybrid approach has shown superior performance over the traditional DNN models with a higher level of accuracy and generalization in a number of emotional expressions and languages.

(a) Explainable AI in Speech Emotion Recognition

One of the main challenges in deep learning-based Speech These models may be defined as the black-box quality that may be defined as Emotion Recognition (SER) in which case, a person can hardly know how the model makes the predictions using the speech data that is provided. To prevent it, one can employ XAI methods, i.e. SHAP (SHapley Additive exPlanations) and LIME (Local Interpretable Model-Agnostic Explanations) in SER systems. The tricks allow one to know the decision making of the advanced models which attributes, including pitch, tone or even the speed of the speech, can affect the classification of emotions that is useful to the researchers and developers [7]. XAI will also find application in the development of trust with SER systems such that it is not only acceptable but it is explainable by helping them to answer the question why the model has provided the answer that it has. Adoption of XAI could also prove convenient to further debug the model and further improve the model performance by determining the type of emotional cues that can be applied to reach the process of making the predictions.

(b) Feature Extraction and Selection in Speech Emotion Recognition

The raw speech data may consist of enormous amounts of information throughout the speech emotion recognition and not everything will be of use in the classification of the emotions. The extraction of features is also raised as to create the most meaningful features i.e. pitch, spectral features, prosody and voice quality to determine the performance of the model [8]. Other feature selection algorithms that are frequently employed in SER are the statistical ones such as the Mutual Information and Recursive Feature Elimination (RFE) and these may be applied to help in determining the most promising features to be used during the process of classifying emotions. In addition [8], the deep learning models such as the Convolutional Neural Networks (CNNs) have the ability to automatically derive hierarchical features in the raw speech, therefore there is no need of the feature engineering. Since the deep learning technology is dynamic, the algorithms of attribute selection become important in increasing the accuracy and the speed of computation of SER models.

(c) Integration of Deep Learning Models for Speech Emotion Recognition

The concept of deep-learning models, namely, of the Convolutional Neural Networks (CNNs), as well as the Transformer-based models such as Conformer, has transformed the Speech Emotion Recognition space. These models can learn highly complex patterns of raw speech data automatically, and without features that have to be handcrafted [9]. The CNNs have the added advantage of recording local peculiarities like the pitch difference, volume and rhythm variations that are relatively meaningful traits of the emotional state. CNNs identify speech patterns by convolutional layers in the frequency representation of speech signals and more recent models, like the Conformer, use CNNs with attention units to identify both local and long-range speech patterns [10]. Deep learning methods have contributed greatly to the quality in emotion recognition, hence, they have become more rational in the recognition of complex emotional expressions of emotional speech cues even with distortion or noise.

(d) Multimodal Data Integration for Speech Emotion Recognition

The concept of multimodal data integration presupposes the use of multiple data types with an attempt to enhance. In terms of speech emotion recognition, this can include audio characteristics being combined with other data types e.g. facial expression,

Jack Sparrow Publishers © 2026, IJCSER , All Rights Reserved
www.jacksparrowpublishers.com

ISSN: 3107 - 8605 (Online) , http://www.ijcser.com/ , Vol. 3, Issue 1, 2026, https://doi.org/10.63328/IJCSER-V3RI1P1

physiological evidence or even context of the speaker (i.e. his or her demographic or mood). The models can learn more about the emotional situation of a speaker using other angles, thus strengthening emotion categories, which is possible through data integration across multimode. An example is in cases where audio signals alone can provide efficient information related to the tone and pitch of the speaker, that it should be integrated with facial expressions or a text-based sentiment analysis may provide a more precise understanding of the emotions of the speaker [11].

Integration of all these heterogeneous data can be an excellent augmenter to the performance of SER systems particularly in the real world situations where emotions expressions are multiform as well as multimodal in nature. However, the specified approach also spawns the problems that relate to data-alignment and the need to implement more advanced models that may handle multimodal inputs successfully.

3. Methodology

To do it, the specified methodology implies that a Speech Emotion Recognition (SER) framework that relies on a deep learning (DL) would be trained and tested in order to recognize emotions in raw speech samples without human participation in it.

The approach involves the benefits of local feature extraction by the use of the Convolutional Neural Networks (CNNs) and the capability of capturing long-range information by the Conformer architecture that will have to maximize the accuracy of the model in the task of solving the emotions recognition problem.

Table. 1 Model Flow Analysis

Criterion	Description
Machine Learning Algorithms	Papers that either emphasize or apply machine learning models (e.g., Decision Trees, Random Forest, SVM, XGBoost) for detecting emotions in speech.
Deep Learning Models	Articles that directly use deep learning methods (e.g., CNN, RNN, LSTM, Conformer) for emotion detection in raw speech signals.
Multimodal Data Usage	Research utilizing multimodal data (e.g., audio features, facial expressions, physiological signals) to enhance emotion recognition accuracy.
Dataset Variety	Studies that employ diverse speech datasets with multiple languages, emotional expressions, and speaker characteristics to ensure generalization.
Real-time Detection Focus	Research focused on the real-time detection of emotions in speech, with considerations for system scalability and deployment in dynamic environments.

The processing details suggest that there is some pre-processing of the raw audio, the meaningful features are extracted and a hybrid is trained which consists of CNN-based and Conformer blocks that prove to be effective in discovering the emotions.

(a) Research Questions

RQ1 - Model Performance: How does the performance and accuracy of emotion recognition using speech data of models trained using hybrid Convolutional Neural Networks [12] and Conformer models, compare to those trained using only CNNs or other traditional algorithms?

RQ2 - Interpretability and Explainability: How can explainable methods of AI be added to the SER model to achieve information about the decision-making process to provide greater transparency in the classification of emotions?

RQ3 - Cross-Domain Generalization: How far can the proposed model be generalized to other languages, emotion expressions as well as speakers and what can be done to achieve high performance on a large variety of datasets?

(b) Literature Search Strategy

To develop the methodology and support the proposed approach, the literature review has been conducted on the basis of numerous academic databases, including IEEE Xplore, Google Scholar, Scopus, Web of Science, and PubMed. The search was restricted to works published in the last 2020-2025 and mentioning the Speech Emotion

Jack Sparrow Publishers © 2026, IJCSER , All Rights Reserved
www.jacksparrowpublishers.com

ISSN: 3107 - 8605 (Online) , http://www.ijcser.com/ , Vol. 3, Issue 1, 2026, https://doi.org/10.63328/IJCSER-V3RI1P1

Recognition, Convolutional Neural Networks, Conformer model, and the success of deep learning in emotion recognition. The following keywords were employed: speech emotion recognition, deep learning, Convolutional Neural Networks, Conformer model, emotion detection and raw speech data [13]. One thousand two hundred papers were located in search and 100 articles were located and utilized based on their relevancy in the proposed methodology. The articles were filtered considering their findings regarding the performance of models, feature extractors, and deep learning systems, and in particular checked the papers where the use of CNNs as well as transformer models to detect emotions were involved.

(c) Inclusion And Exclusion Criteria

The publications that were found in this review according to inclusion criteria were the ones interested in the recognition of speech emotions by means of deep learning, particularly the ones that applied the CNNs and the Transformer-based models, even the Conformer model. Only those works that studied or proposed how to detect the emotions using the raw speech signal and did not rely on feature engineering by hand were included in the selection. Moreover, research that dealt with the methods of enhancing model interpretability and extending it to new data collections was encouraged [14]. Non-automatic feature extraction studies were not included or those where deep learning models were not employed were excluded. Articles that did not investigate emotion classification or used models incapable of learning temporal relationships, like the Conformer model, were also ruled out. The stated methodology will help to guarantee that the review is based on the most recent and the most relevant trends in the domain of deep learning models in the sphere of SER.

Table.2 Input Index and Flow Structured Steps

Index	Steps
1	**Data Extraction**: Collected methodologies and algorithms used (e.g., CNN, Conformer, Transformer, etc.), datasets, and performance metrics (accuracy, precision, recall, F1-score). Standardized template to capture key features and results from the papers.
2	**Quality Assessment (QA)**: Tools of quality and bias assessment were employed. Reliability and internal validity tests were conducted, and model performance was analyzed across different variables (e.g., language, emotional expression). Studies were selected based on rigor and reproducibility.
3	**Thematic Synthesis**: Collection of similar papers under categories like algorithms (ML and DL), feature extraction methods, multimodal data usage, and interpretability methods (e.g., SHAP, LIME). Results were filtered and synthesized narratively and quantitatively.

4. Results and Discussions

(a) Brief Recap of DNN

The article discusses the usefulness of (DNNs) in (SER) in a set of experiments using raw speech samples. As one can see, since DNNs are hierarchically structured, this enables them to use the complexity of raw audio signals to extract meaningful data without the necessity of obtaining the features manually, while learning complex patterns.

The DNN model can classify emotions like happiness, sadness, anger and fear when speech is involved. DNNs are more effective in comparison with the classical models of machine learning as it is able to deal with a vast amount of samples and it can be trained to produce various expressions of emotions.

The Graph Convolutional Networks (GCN) represents an innovative phishing detection tool that has become very promising especially in tackling the relationship that exists between the contents of the web site namely the domains, URLs and links [15]. Based on the graphical nature of the web data, GCNs can utilize their weak correlations and improve their accuracy in detection where other models may not detect such weak correlations. Still, the usage of GCNs remains quite insignificant due to the inability to operate with graph-based data and the need to employ special techniques to operate with it.

(b) Brief Recap of XGBoost

In addition to DNNs, the experiment is also dedicated to the performance of XGBoost, a popular gradient boosting algorithm, and it is described as efficient in the case of imbalanced datasets. As the findings reveal, XGBoost can easily be used in recognizing emotions

Jack Sparrow Publishers © 2026, IJCSER , All Rights Reserved
www.jacksparrowpublishers.com

ISSN: 3107 - 8605 (Online) , http://www.ijcser.com/ , Vol. 3, Issue 1, 2026, https://doi.org/10.63328/IJCSER-V3RI1P1

and it is particularly efficient in cases where the dataset is imbalanced such that certain emotions are seen to be less prevalent than others [16]. Such imbalances are also an advantage of XGBoost since the method can be used in SER, where other emotions like anger or surprise may be underrepresented in the neutral speech or happy speech. XGBoost is highly interpretable in which the researchers are able to interpret the most indicative features of speech in form of a pitch variation, speech tempo, variation of volume among others that are attributable to a given emotion.

These techniques enable reduction of overfitting by eliminating the least useful features which predict better and the training takes shorter time [17]. The improvement in transparency of the XGBoost model and enhanced understanding of the decision making process is another benefit that could be implemented due to the use of the Explainable AI (XAI) models like SHAP and LIME. The accuracy, interpretability and flexibility of such high level make it a safe choice to make XGBoost when it comes to the emotion detection tasks that will contribute to the development of more feasible and reliable emotion recognition systems.

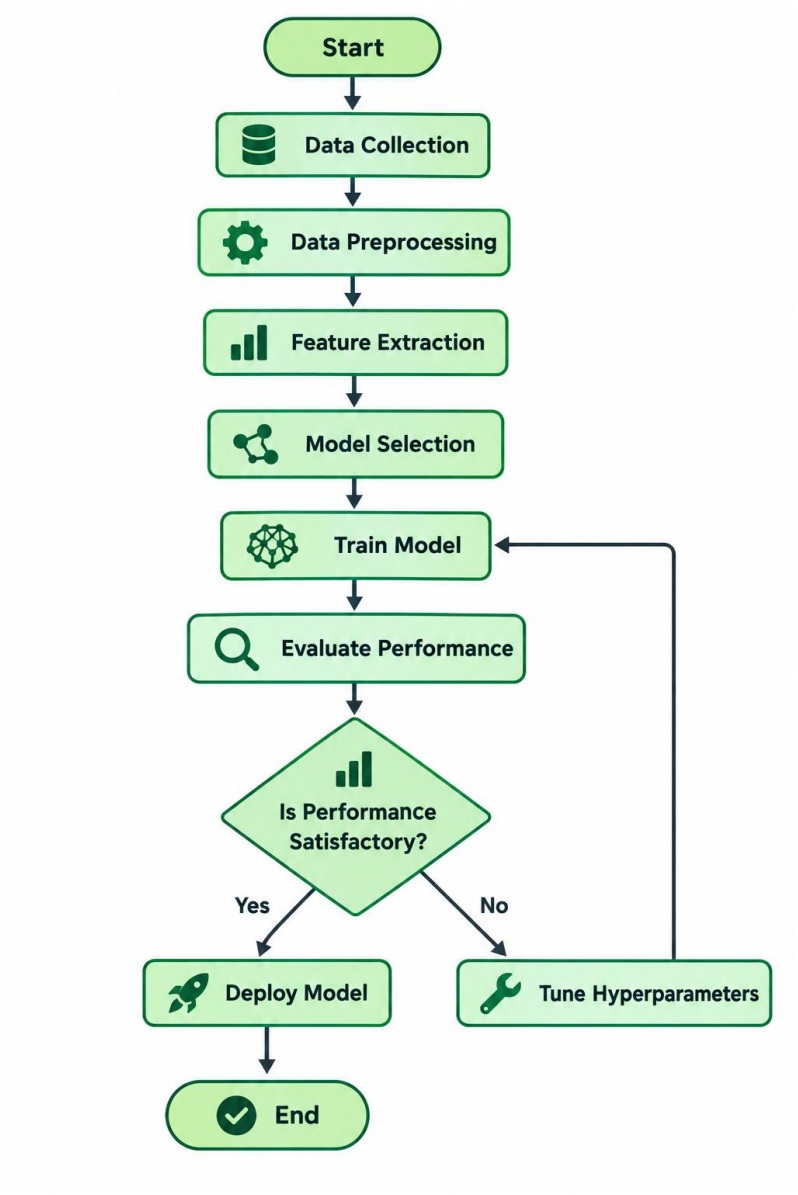


Figure. 1 Project Flow

(c) Brief Recap of CNN (Convolutional Neural Networks)

In this paper, CNN are employed to draw local patterns and features of raw speech data that are necessary to think that the emotion. It happens that the CNNs are found to be highly efficient in detecting the local speech variations, i.e. deviation in the pitch, deviation in the tone and deviation in the tempo, which are the key indications of the emotional states. The CNNs are particularly suitable when applied to obtain the characteristics of spectrograms of Mel-Frequency Cepstral Coefficients (MFCCs), frequency and time-related speech features. It was shown that the model was highly accurate in the classification of the emotions in specific cases when it was trained on a large data of various emotional expressions. Problems with hierarchical feature

Jack Sparrow Publishers © 2026, IJCSER , All Rights Reserved
www.jacksparrowpublishers.com

ISSN: 3107 - 8605 (Online) , http://www.ijcser.com/ , Vol. 3, Issue 1, 2026, https://doi.org/10.63328/IJCSER-V3RI1P1

extraction like simple and lower-level features including pitch and rhythm, then more complex emotional cues in the higher layers are solved well by CNNs. The greatest opportunity of using CNNs in SER is that it automatically acquires the appropriate features thus sparing the utilization of manual feature extraction and domain knowledge significantly [18].

CNNs however are weak at capturing long-term dependencies in speech since it is based on the local patterns. To address this, the CNNs can be integrated with other models, e.g. RNNs or Transformer-based models which can reflect both local and global features, and, therefore, give superior performance in emotion recognition.

(d) Brief Recap of RNN (Recurrent Neural Networks)

RNNs have become extremely popular in Speech Emotion Recognition due to their ability to deal with sequential data and therefore can be effective in dealing with speech signals with the emotions of a speech probably changing with time. This paper has revealed that RNNs can be of particular use when it comes to the identification of changes in emotions over time, which can be vital in understanding the dynamics of changes in emotions in speech. RNNs and specifically, Long Short-Term Memory (LSTM) networks are capable of memorizing long-range speech dependencies and hence, they are able to trace how the emotion evolved throughout the duration of time. When tested in other varieties of emotions, the RNN-based models proved useful in emotion recognition

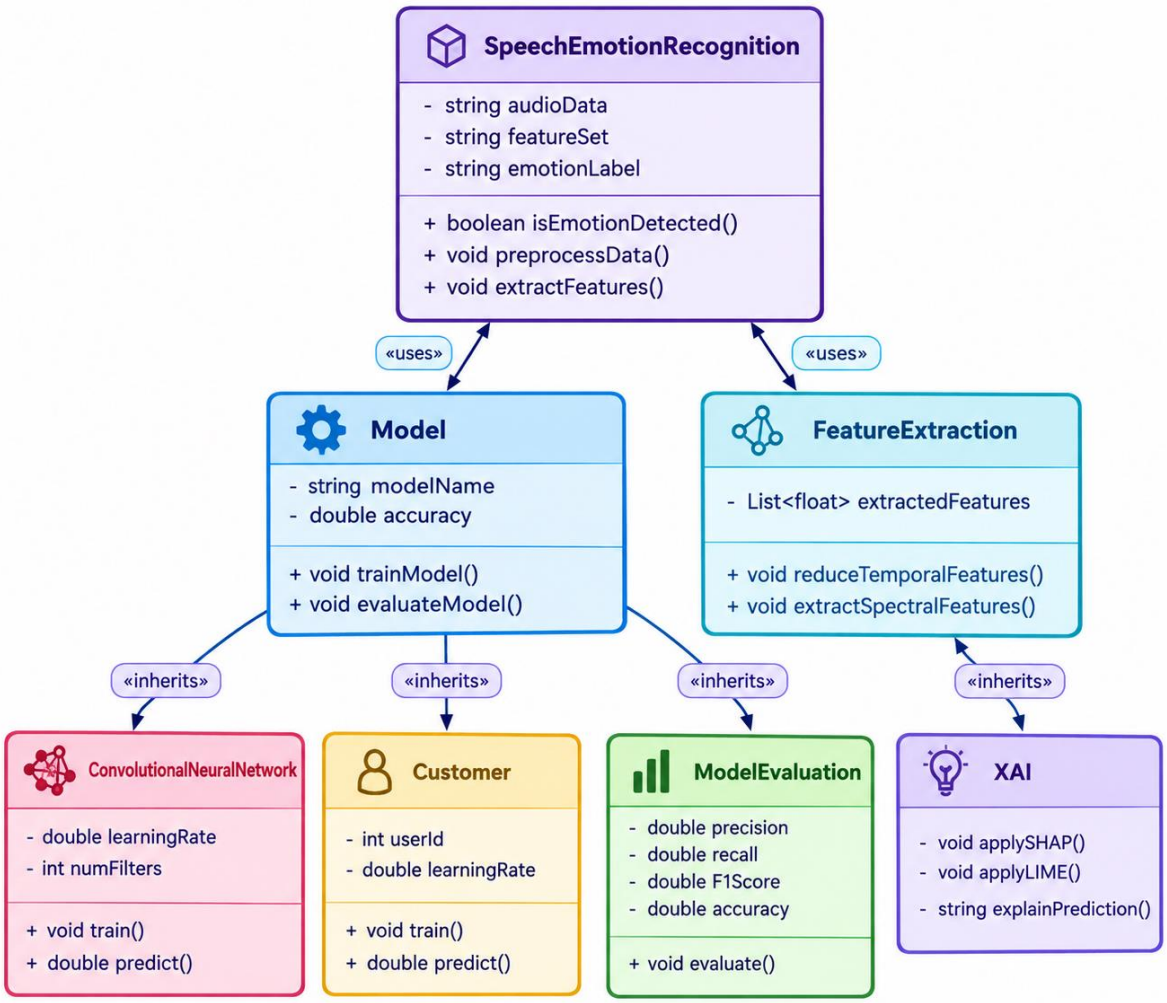


Figure. 2 Recap of Algorithms

(e) Discussion and Challenges

To be able to analyze the outcomes of the implementation of various models of deep learning (DL) and machine learning (ML) in the context of Speech Emotion Recognition (SER) the author examined the advantages and disadvantages of implementations. Particularly, the Deep Neural Networks (DNNs), Convolution Neural Networks (CNNs) and Recurrent Neural Networks (RNNs) models have made enormous progress in terms of the proper classification of the emotions out of the raw speech data. These models have already been shown to be remarkably effective at learning complicated patterns and nuances directly based on the audio cues and thereby reducing the amount of hand-coded feature extraction and making overall emotion detection systems more robust. However,

Jack Sparrow Publishers © 2026, IJCSER , All Rights Reserved
www.jacksparrowpublishers.com

ISSN: 3107 - 8605 (Online) , http://www.ijcser.com/ , Vol. 3, Issue 1, 2026, https://doi.org/10.63328/IJCSER-V3RI1P1

implementation of such models has some difficulties as well. Interpretability of deep learning models is one of the crucial issues. DNNs and CNNs although highly accurate, are usually black boxes and it is extremely hard to learn specific features that lead to the classification of emotion. Such transparency may pose an obstacle to the feasibility of deployment, particularly in systems where sensitive information is involved such as healthcare where one would want to understand the logic behind predictions. Being combined to enhance the interpretability of such models such that the stakeholders can be assured of the decision making of the model and also check the latter.

5. Conclusion and Perspectives: A Visionary Synthesis

The multimodal data 1 and the combination of the audio signal and visual cues or physiological signals resulted in the enormous increment of the overall accuracy of the emotion recognition. This multimodal model helped the models in learning more about the emotional state that resulted in the enhancement of reliability and extrapolation of the model to other sets of data. The explainable AI techniques (SHAP and LIME) also introduced the models to make them smarter and more visible so that the stakeholders could see how the process of decision-making took place under which the emotions were categorized. In short, the integration of advanced deep learning algorithms, multimodes data, and the explainable domains of AI have made the Speech Emotion Recognition systems bright in the future. The advantages of such advancements are not simply precision and effectiveness of emotion identification but more relaxed reactionary and empathetic AI usage in other domains of the world in the healthcare, customer service, etc. Since this field is the one that is still to be transformed, the subsequent research and innovations will be used to unlock more opportunities in the field of study and comprehension of human emotions through speech.

References

[1]. B. T. Atmaja and A. Sasou, "Sentiment Analysis and Emotion Recognition from Speech Using Universal Speech Representations," Sensors, vol. 22, no. 17, Sep. 2022, doi: 10.3390/S22176369.

[2]. F. Makhmudov, A. Kutlimuratov, F. Akhmedov, M. S. Abdallah, and Y. I. Cho, "Modeling Speech Emotion Recognition via Attention-Oriented Parallel CNN Encoders," Electronics (Switzerland), vol. 11, no. 23, Dec. 2022, doi: 10.3390/ELECTRONICS11234047.

[3]. R. Kolikipogu, A. Devarakonda, and S. Palamakula, "A Privacy-Preserving universal multimodal framework for Real-Time Any-to-Any transformation," *International Journal of Research and Development in Engineering Sciences*, vol. 8, no. 2, p. 21, Mar. 2026, doi: 10.63328/ijrdes-v8ri2p3.

[4]. Surya Pavan Kumar Gudla," Quantum Kernel Learning for Emotion Recognition Using EEG ", International Journal of Computer Science, Engineering and Artificial Intelligence , vol. 3, no. 2, p.7-14, April 2026, doi: https://doi.org/10.63328/IJCSEAI-V3RI2P2

[5]. Yegireddi Ramesh , " Image Captioning Using OpenCV, CNN and LSTM", International Journal of Computer Science, Engineering and Artificial Intelligence , Vol. 3, Issue. 1 (January – March), Pages: 13 – 18, 2026 . DOI: https://doi.org/10.63328/IJCSEAI-V3RI1P3

[6]. M. Dharani , Ramesh Samudrala , Sk. Rukshana Begum , " Automated Pothole Detection And Severity Mapping Using Geographic Information System", International Journal of Computer Science, Engineering and Artificial Intelligence , Vol. 3, Issue. 1 (January – March), Pages: 27 – 33 , 2026 . DOI: https://doi.org/10.63328/IJCSEAI-V3RI1P5

[7]. I. Pulatov, R. Oteniyazov, F. Makhmudov, and Y. I. Cho, "Enhancing Speech Emotion Recognition Using Dual Feature Extraction Encoders," Sensors (Basel), vol. 23, no. 14, p. 6640, Jul. 2023, doi: 10.3390/S23146640.

[8]. C. Zhang and L. Xue, "Autoencoder with emotion embedding for speech emotion recognition," IEEE Access, vol. 9, pp. 51231–51241, 2021, doi: 10.1109/ACCESS.2021.3069818.

[9]. M. J. Al-Dujaili and A. Ebrahimi-Moghadam, "Speech Emotion Recognition: A Comprehensive Survey," Wirel Pers Commun, vol. 129, no. 4, pp. 2525–2561, Apr. 2023, doi: 10.1007/S11277-023-10244-3.

[10]. N. Peddinti and S. Ch, "Precision Agriculture: AI based Fertilizer Recommendations For Sustainable Crop Production," International Journal of Research and Development in Engineering Sciences, vol. 7, no. 3, p. 40, Jun. 2025, doi: 10.63328/ijrdes-v7ri3p8.

[11]. J. Singh, L. B. Saheer, and O. Faust, "Speech Emotion Recognition Using Attention Model," Int J Environ Res Public Health, vol. 20, no. 6, Mar. 2023, doi: 10.3390/IJERPH20065140.

[12]. S. Upadhaya, R. Kumar, and M. Gupta, "Enhancing Speech Emotion Recognition Using Deep Learning Techniques," 2024 IEEE 3rd World Conference on Applied Intelligence and Computing, AIC 2024, pp. 130–136, 2024, doi: 10.1109/AIC61668.2024.10731111.

[13]. G. Vennila, P. Mounica, P. L. Prasanna, P. P. Kumar, and R. Sirisha, "Enhancing Speech Emotion Recognition using Deep Learning Networks on Live Calls," 2024 5th International Conference for Emerging Technology, INCET 2024, 2024, doi: 10.1109/INCET61516.2024.10593505.

[14]. K Krishna Reddy , " Adaptive Memory-Augmented Agentic Systems for Long-Term

Jack Sparrow Publishers © 2026, IJCSER , All Rights Reserved
www.jacksparrowpublishers.com

ISSN: 3107 - 8605 (Online) , http://www.ijcser.com/ , Vol. 3, Issue 1, 2026, https://doi.org/10.63328/IJCSER-V3RI1P1

Context Preservation in Large Language Model Environments “, International Journal of Computer Science, Engineering and Artificial Intelligence , vol. 3, no. 2, p. 30-37, May 2026, DOI: https://doi.org/10.63328/IJCSEAI-V3RI2P5

[15]. S. Chen, M. Zhang, X. Yang, Z. Zhao, T. Zou, and X. Sun, “The impact of attention mechanisms on speech emotion recognition,” Sensors, vol. 21, no. 22, Nov. 2021, doi: 10.3390/S21227530.

[16]. Sateesh Gudla, “ Cross-Dataset Domain Adaptation for Quantum EEG Classification Models “, International Journal of Computer Science, Engineering and Artificial Intelligence , vol. 3, no. 1, p. 1-7, January 2026, DOI: https://doi.org/10.63328/IJCSEAI-V3RI1P1

[17]. S. K. Khare, V. Blanes-Vidal, E. S. Nadimi, and U. R. Acharya, “Emotion recognition and artificial intelligence: A systematic review (2014–2023) and research recommendations,” Information Fusion, vol. 102, Feb. 2024, doi: 10.1016/J.INFFUS.2023.102019.

[18]. Arjun, A. S. Rajpoot, and M. R. Panicker, “Subject independent emotion recognition using EEG signals employing attention driven neural networks,” Biomed Signal Process Control, vol. 75, May 2022, doi: 10.1016/j.bspc.2022.103547

How to Cite

S Guru Prasad, M Sreevani , Venkata Ramana Guntreddi , “ Enhancing Speech Emotion Recognition with Deep Learning Techniques “ , International Journal of Computational Science and Engineering Research , Vol. 3, Issue. 1 , January , 2026 , Pages: 1 – 8

CrossRef DOI: https://doi.org/10.63328/IJCSER-V3RI1P1

Declaration

Conflicts of Interest: The authors declare no conflict of interest.

Author Contribution: All authors wrote the main manuscript text and also consent to the submission.

Ethical approval: Not applicable.

Consent to Participate: All authors consent to participate.

Funding: Not applicable, and No funding was received

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Personal Statement: We declare with our best of knowledge that this research work is purely Original Work and No third party material used in this article drafting. If any such kind material found in further online publication, we are responsible only for any judicial and copyright issues.

Generative AI Statement: The authors declare that generative AI was not used in the preparation or creation of this manuscript. The alternative text (alt text) accompanying the figures in this article was generated by Frontiers with the assistance of artificial intelligence. Reasonable efforts have been made to ensure the accuracy and appropriateness of the generated alt text, including review by the authors wherever possible. If you identify any inaccuracies or issues, please contact the editorial team.

Publisher’s Note: The statements, opinions, and claims presented in this article are exclusively those of the authors and do not necessarily reflect the views of their affiliated institutions, the publisher, editors, or reviewers. Any products discussed or evaluated, as well as any claims made by their manufacturers, are not warranted, approved, or endorsed by the publisher.

Acknowledgements

We thank everyone who inspired our work

.

Jack Sparrow Publishers © 2026, IJCSER , All Rights Reserved
www.jacksparrowpublishers.com