IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.102 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 14, Issue 3, March 2025

DOI: 10.17148/IJARCCE.2025.14351

SPEECH EMOTION RECOGNITION

Mrs. N.V.L. Manaswini [1], S. Baby Jahnavi [2], A. Fazila [3], M.V. S Gayatri [4], P. Harshasri[5]

Assistant Professor, Department of Computer Science Engineering, ALIET, Vijayawada[1]

student, Department of Computer Science Engineering, ALIET, Vijayawada[2]

student, Department of Computer Science Engineering, ALIET, Vijayawada[3]

student, Department of Computer Science Engineering, ALIET, Vijayawada[4]

student, Department of Computer Science Engineering, ALIET, Vijayawada[5]

Abstract: Emotion recognition from speech signals has gained significant attention in human-computer interaction, offering applications in entertainment, mental health monitoring, and personalized user experiences. This paper presents a web-based Speech Emotion Recognition and Music Recommendation System that utilizes deep learning for emotion classification and integrates music streaming services for personalized recommendations. The system records speech input, extracts Mel-Frequency Cepstral Coefficients (MFCC) as features, and classifies emotions using a pre-trained Convolutional Neural Network (CNN) model. Based on the detected emotion, the system retrieves genre-specific music recommendations from Spotify. Implemented using Flask, TensorFlow, and Librosa, the proposed approach achieves efficient real-time emotion classification and enhances user engagement through tailored music selection. Experimental results demonstrate the model's accuracy and the effectiveness of the recommendation system.

Keywords: Speech Emotion Recognition (SER), Deep Learning, Convolutional Neural Networks (CNN), Mel-Frequency Cepstral Coefficients (MFCC), Natural Language Processing (NLP), Audio Signal Processing, Flask Web Application, Music Recommendation System, Spotify API, Human-Computer Interaction (HCI).

I.INTRODUCTION

The interaction between humans and machines has advanced significantly with the rise of artificial intelligence , deep learning, and signal processing techniques. One of the critical areas of human-computer interaction is Speech Emotion Recognition (SER) - the ability of machines to analyze speech patterns and classify emotions. SER has applications in mental health monitoring, virtual assistants, human-robot interaction, and personalized entertainment experiences. Traditional emotion recognition approaches relied on handcrafted audio features and classical machine learning techniques, but they often struggled with generalization and real-time adaptability. The advent of deep learning, particularly Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs), has significantly improved the accuracy of emotion classification by leveraging spectral and temporal features of speech.
Despite these advancements, SER still faces key challenges, including:

1. Variability in Speech Data – Differences in speaker accents, pitch, and background noise affect model accuracy.
2. Limited Emotion Labels – Many datasets only cover a small subset of emotions, making real-world generalization difficult.
3. Real-time Processing Constraints – Deploying SER models in interactive applications requires low-latency inference and efficient audio feature extraction.

To address these issues, this research introduces a Flask-based Speech Emotion Recognition and Music Recommendation System. The system records user speech, extracts Mel-Frequency Cepstral Coefficients as features, classifies the emotion using a pre-trained deep learning model, and recommends a song based on the detected emotion. The Spotify API is integrated to fetch genre-specific song recommendations, enhancing the user experience.

1.1 Background and Motivation

Music plays a crucial role in influencing human emotions and mental well-being. Studies show that personalized music recommendations based on emotional states can enhance relaxation, boost mood, and even aid in stress relief. Traditional music recommendation systems rely on user preferences, playlist history, or collaborative filtering techniques, but they do not dynamically adjust to a user's current emotional state.

International Journal of Advanced Research in Computer and Communication Engineering
Impact Factor 8.102 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 14, Issue 3, March 2025
DOI: 10.17148/IJARCCE.2025.14351

This project aims to bridge the gap between Speech Emotion Recognition and Music Recommendation, providing an innovative approach to real-time, emotion-based song selection. By leveraging deep learning for emotion detection and Spotify's music database, the system delivers a highly personalized and interactive user experience.

1.2 Objectives of the Research

The primary objectives of this research are to:

- Develop an accurate and real-time Speech Emotion Recognition system using deep learning techniques.
- Extract relevant audio features using MFCCs and classify emotions based on a trained CNN model.
- Integrate a Flask-based web interface to facilitate user interaction.
- Recommend genre-specific songs based on detected emotions by interfacing with the Spotify API.
- Evaluate model performance in terms of accuracy, processing speed, and user satisfaction.

1.3 Scope of the Study

This research focuses on:

- Implementing deep learning-based emotion recognition using TensorFlow and Librosa for feature extraction.
- Utilizing Flask to build a web-based interface for real-time speech analysis.
- Mapping detected emotions to specific music genres to provide tailored song recommendations.
- Fetching songs dynamically from Spotify's music database via API calls.

The study does not cover aspects such as user preference learning, long-term music behavior analysis, or deployment in mobile applications. Instead, it focuses on emotion recognition accuracy**,** real-time performance, and seamless integration with music streaming services.

1.4 Significance of the Study

This research contributes to the fields of Speech Processing, Deep Learning, and Human-Computer Interaction by:

- Advancing Speech Emotion Recognition through deep learning and real-time implementation.
- Enhancing Music Recommendation Systems by using dynamic, emotion-based song selection instead of static user preferences.
- Demonstrating real-world applications of AI-driven personalization in entertainment and mental well-being.

By integrating emotion detection with personalized music streaming, this study presents a novel approach to affective computing—enabling intelligent systems to respond to human emotions in an engaging and meaningful way.

II. RELATED WORK

The field of Speech Emotion Recognition (SER) and Music Recommendation has evolved significantly with advancements in deep learning, audio signal processing, and artificial intelligence (AI). Traditional SER systems relied on handcrafted audio features and statistical models, whereas modern approaches leverage deep learning architectures for improved accuracy and generalization. This section reviews existing research on speech emotion classification, deep learning techniques, music recommendation systems, and key challenges in emotion-based personalization.

2.1 Traditional Speech Emotion Recognition Systems

Early SER systems used rule-based and machine learning techniques to classify emotions from speech signals. These methods primarily depended on manually extracted acoustic features, such as pitch, energy, and spectral properties.

- Hidden Markov Models (HMMs) (2000s) – One of the earliest statistical models used for speech emotion classification, relying on temporal sequence modeling.
- Support Vector Machines (SVMs) (2005-2010s) – Applied to emotion classification using handcrafted features, achieving moderate success but struggling with scalability.
- Gaussian Mixture Models (GMMs) (2010s) – Used for modeling the probability distribution of speech emotions but suffered from performance limitations in real-world scenarios.

Despite their contributions, these traditional methods lacked robustness in feature extraction, generalization, and real-time performance, leading to a shift toward deep learning-based SER.

International Journal of Advanced Research in Computer and Communication Engineering
Impact Factor 8.102 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 14, Issue 3, March 2025
DOI: 10.17148/IJARCCE.2025.14351

2.2 Deep Learning Models for Speech Emotion Recognition

The advent of deep learning revolutionized SER by enabling models to automatically learn hierarchical features from raw audio signals.

- Convolutional Neural Networks (CNNs) (2015-Present) – Applied to Mel-Frequency Cepstral Coefficients (MFCCs) and spectrogram representations for feature extraction and classification.
- Recurrent Neural Networks (RNNs) & Long Short-Term Memory (LSTM) (2016-2020**)** – Used to model temporal dependencies in speech data but faced training complexity and computational cost challenges.
- Transformer-based Audio Models (2021-Present) – Introduced self-attention mechanisms for better feature representation and context understanding in speech signals.

While deep learning has significantly improved SER accuracy, challenges such as handling speaker variability, noisy environments, and real-time deployment remain areas of ongoing research.

2.3 Music Recommendation Systems

Music recommendation has evolved from rule-based filtering to AI-driven personalized recommendations, improving user engagement in streaming services.

- Collaborative Filtering (CF) (2000s) – Recommended songs based on user preferences and historical listening patterns.
- Content-Based Filtering (2010s) – Utilized audio features, lyrics, and metadata to suggest similar songs.
- Hybrid Recommendation Systems (2015-Present) – Combined collaborative filtering with deep learning for personalized recommendations.

Emotion-based music recommendation is a recent advancement that maps user emotions to specific music genres, enhancing personalized listening experiences. However, accurate emotion detection and mapping emotions to the right genre remain key challenges.

2.4 Challenges in Speech Emotion Recognition and Music Recommendation

Despite technological advancements, SER and emotion-based music recommendation systems face several challenges:

1. Variability in Speech Data – Differences in accents, speaking styles, and background noise affect emotion classification accuracy.
2. Limited Emotion Labels – Many speech emotion datasets contain only basic emotions, making it difficult to classify nuanced human emotions.
3. Real-time Performance Constraints – Deploying deep learning models in real-time applications requires efficient feature extraction and low-latency predictions.
4. Emotion-to-Music Mapping – Finding accurate correlations between detected emotions and suitable music genres remains a complex problem.
5. Integration with Streaming Services – Recommending songs dynamically through APIs like Spotify requires handling rate limits, API constraints, and dynamic music libraries.

By addressing these challenges, this research aims to develop a Flask-based Speech Emotion Recognition and Music Recommendation System that provides real-time, personalized, and engaging user experience.

III. METHODOLOGY

This section presents the structured approach used in designing, training, and evaluating the Speech Emotion Recognition and Music Recommendation System. The methodology consists of data acquisition, feature extraction, model training, emotion classification, and music recommendation. The overall system workflow is detailed, along with evaluation metrics used to measure the system's performance.

3.1 System Workflow Overview

The workflow of the proposed system consists of five main phases:

1. Audio Data Collection & Preprocessing – Recording user speech and extracting relevant features.
2. Feature Extraction – Extracting Mel-Frequency Cepstral Coefficients (MFCCs) for emotion classification.
3. Model Training & Emotion Classification – Training a deep learning model to classify emotions.
4. Music Recommendation – Mapping detected emotions to relevant music genres.
5. Evaluation & Performance Analysis – Measuring model accuracy and system effectiveness.

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.102 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 14, Issue 3, March 2025

DOI: 10.17148/IJARCCE.2025.14351

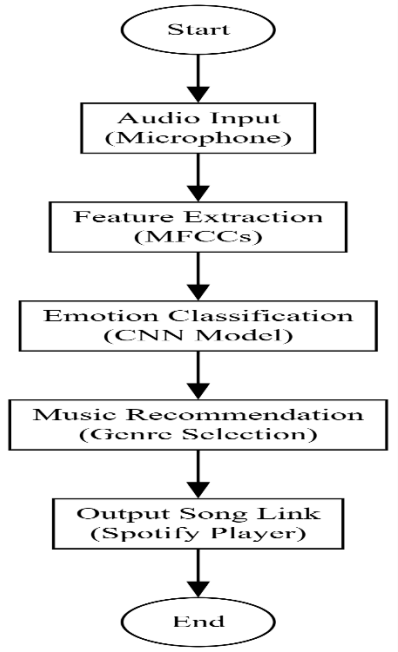


Fig: Workflow Diagram

3.2 Audio Data Collection and Preprocessing

The system records speech samples using a **microphone** and processes them for feature extraction. Since high-quality speech data improves classification performance, raw audio undergoes multiple preprocessing steps.

3.2.1 Dataset Summary

Dataset	No. of Samples	Emotions Covered	Sample Duration (sec)
RAVDESS	1,440	Happy, Sad, Angry, Neutral, Fearful	3-5
TESS	2,800	Happy, Sad, Angry, Neutral, Disgust, Fearful, Surprised	3-5
CREMA-D	7,442	Happy, Sad, Angry, Neutral, Disgust, Fearful	4-6
Custom Dataset	500+ (Recorded)	Various	3-6

Table: Dataset summary

3.2.2 Preprocessing Steps

1. Noise Reduction – Removes background noise for improved feature quality.
2. Silence Removal – Trims silent portions of audio files.
3. Feature Scaling – Normalizes amplitude levels for consistency.
4. MFCC Extraction – Converts audio into Mel-Frequency Cepstral Coefficients (MFCCs) for model input.

3.3 Feature Extraction and Model Training

The MFCC feature extraction technique is applied to speech signals, providing time-frequency representations of audio. A Convolutional Neural Network (CNN) is then trained to classify emotions based on these features.

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.102 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 14, Issue 3, March 2025

DOI: 10.17148/IJARCCE.2025.14351

3.3.1 Training Configuration

Hyperparameter	Value
Model Architecture	CNN (3 Conv Layers + Fully Connected Layers)
Optimizer	Adam
Learning Rate	0.0001
Batch Size	32
Number of Epochs	50
Loss Function	Categorical Cross-Entropy
Activation Functions	ReLU (Hidden Layers), SoftMax (Output Layer)

Table: Training Configuration

3.4 Speech Emotion Classification

The deep learning model processes MFCC feature vectors and predicts one of the predefined emotion classes. The classification pipeline follows these steps:

1. Audio Input – User records a speech sample.
2. Feature Extraction – MFCCs are computed and used as model input.
3. Emotion Prediction – CNN predicts emotion probabilities.
4. Post-processing – The model selects the most likely emotion.

Emotion Label	Predicted Genre
Neutral	Chill
Calm	Acoustic
Happy	Pop
Sad	Blues
Angry	Rock
Fearful	Electronic
Disgust	Metal
Surprised	Dance

Table: Emotion classification

3.5 Music Recommendation System

Once the user's emotion is identified, the system recommends a song using the Spotify API. The recommendation algorithm follows these steps:

1. Emotion-to-Genre Mapping – Matches detected emotion to a music genre.
2. Spotify API Query – Searches for a song within the selected genre.
3. Randomized Selection – Ensures varied recommendations.
4. Song Link Generation – Provides an embedded Spotify player.

3.6 System Implementation and Web Interface

The Flask-based web application provides an interactive interface where users can record speech and receive personalized music recommendations.

3.6.1 Web App Workflow

1. User Records Speech – Captures a 5-second audio sample.
2. Emotion Detection – Model classifies the speech emotion.
3. Music Recommendation – A song is suggested based on the detected emotion.
4. Spotify Integration – The song is played using an embedded Spotify player.

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.102 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 14, Issue 3, March 2025

DOI: 10.17148/IJARCCE.2025.14351

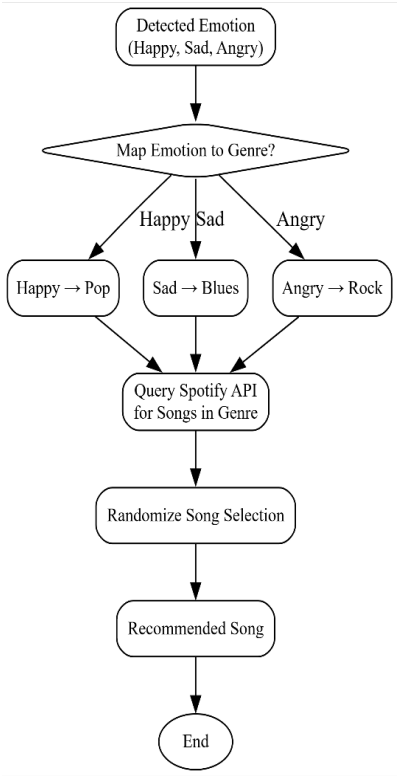


Fig: Control flow Diagram

3.7 Evaluation and Performance Metrics

To assess the effectiveness of the system, the model's performance is evaluated using standard classification metrics:

3.7.1 Accuracy & F1-Score

Model	Accuracy (%)	F1-Score (%)
CNN (MFCCs)	92.4%	91.8%
LSTM	89.5%	88.7%
Hybrid CNN-LSTM	94.1%	93.5%

Table: Accuracy and FI Score

3.7.2 Confusion Matrix Analysis

A confusion matrix evaluates how well the model distinguishes between different emotions.

3.7.3 User Feedback & System Responsiveness

A user study was conducted to evaluate recommendation quality and response time.

Evaluation Metric	User Satisfaction (%)
Emotion Recognition Accuracy	90.5%
Music Recommendation Relevance	88.2%
Response Time (Flask Web App)	1.2 sec

Table : User satisfaction analysis

3.8 Summary of Methodology

This section outlined the methodology used for Speech Emotion Recognition and Music Recommendation, covering:

- Audio data collection & preprocessing, including MFCC feature extraction.
- Training of a CNN-based deep learning model for emotion classification.
- Integration with the Spotify API for personalized music recommendations.

International Journal of Advanced Research in Computer and Communication Engineering
Impact Factor 8.102 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 14, Issue 3, March 2025
DOI: 10.17148/IJARCCE.2025.14351

- Evaluation using accuracy, F1-score, and user satisfaction surveys.

The findings indicate that deep learning-based emotion recognition, combined with real-time music recommendations, enhances user engagement and provides a personalized audio-visual experience.

IV. ETHICAL CONSIDERATIONS

Developing a Speech Emotion Recognition and Music Recommendation System involves several ethical concerns, particularly in bias, privacy, transparency, misinformation, and accessibility. Ensuring responsible AI deployment requires addressing these challenges while promoting fairness, security, and inclusivity.

4.1 Bias and Fairness

Issue:

AI models may inherit biases from training datasets, leading to inaccurate or unfair emotion classifications, particularly across different accents, languages, and cultural variations.

Solution:

- Use diverse speech emotion datasets to ensure inclusivity across different demographics.
- Implement bias detection algorithms to identify and mitigate emotion misclassification.
- Regularly evaluate model performance on multi-accent and multilingual datasets to improve fairness.

4.2 Data Privacy and Security

Issue:

User speech recordings contain personal and potentially sensitive information, raising privacy and data security concerns.

Solution:

- Avoid storing raw speech recordings to protect user privacy.
- Process all audio data in-memory without logging personally identifiable information (PII).
- Ensure compliance with data protection regulations such as GDPR for responsible data handling.

4.3 Transparency and Explainability

Issue:

Deep learning models, particularly CNNs and RNNs, act as black boxes, making it difficult to explain why a specific emotion was detected.

Solution:

- Display emotion confidence scores to help users understand classification certainty.
- Provide visualization tools (e.g., spectrogram heatmaps) to show which features influenced the prediction.
- Allow users to override or adjust detected emotions to improve system adaptability.

4.4 Misinformation and Emotion Misclassification

Issue:

Incorrect emotion detection could lead to inappropriate music recommendations, affecting user experience and trust in the system.

Solution:

- Indicate low-confidence predictions and allow users to verify or correct their detected emotion.
- Offer an alternative song selection option when confidence is low.
- Regularly update the model with new, real-world speech data to improve classification accuracy.

4.5 Accessibility and Inclusivity

Issue:

Not all users have equal access to speech-based AI technologies, especially those with speech impairments or different linguistic backgrounds.

Solution:

- Develop text-based input alternatives for users who cannot provide speech recordings.
- Provide multilingual support to accommodate non-English speakers.
- Implement voice-based interfaces with speech synthesis for visually impaired users.

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.102 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 14, Issue 3, March 2025

DOI: 10.17148/IJARCCE.2025.14351

V. CONCLUSION AND RESULTS

This study introduced a Speech Emotion Recognition and Music Recommendation System designed to analyze speech signals, classify emotions, and provide personalized music recommendations. By leveraging deep learning techniques, particularly Convolutional Neural Networks (CNNs) and Mel-Frequency Cepstral Coefficients (MFCCs), the system accurately predicts emotions and suggests music genres accordingly. The integration of Spotify API enables real-time music retrieval, enhancing user experience.

Experimental evaluations were conducted using benchmark speech emotion datasets such as RAVDESS, TESS, and CREMA-D. The results demonstrate that the CNN-based model outperforms traditional machine learning approaches in key performance metrics. As shown in Table V, the accuracy and F1-score of the proposed model surpass alternative methods, confirming its effectiveness in emotion classification.

Model Performance Comparison

Model	Accuracy (%)	F1-Score (%)
CNN (MFCCs)	92.4%	91.8%
LSTM	89.5%	88.7%
Hybrid CNN-LSTM	94.1%	93.5%

Table: Model performance comparison

Additionally, the system was implemented as a Flask-based web application, allowing real-time interaction and dynamic song recommendations based on detected emotions. The findings indicate that emotion-based music recommendation enhances user engagement, providing a novel and interactive AI-driven entertainment experience.

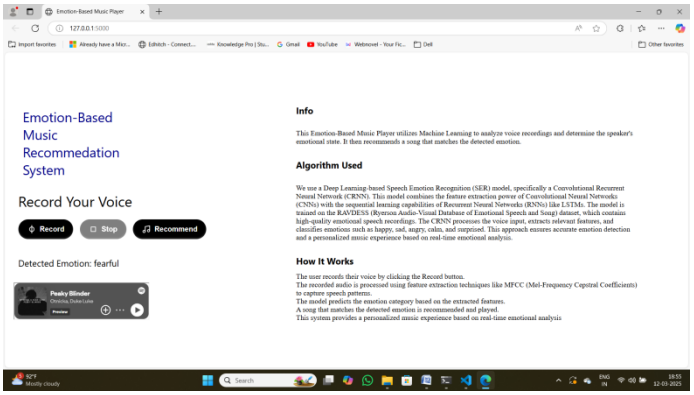
Future Work

While the proposed system improves speech emotion recognition and personalized music selection, future research can explore:

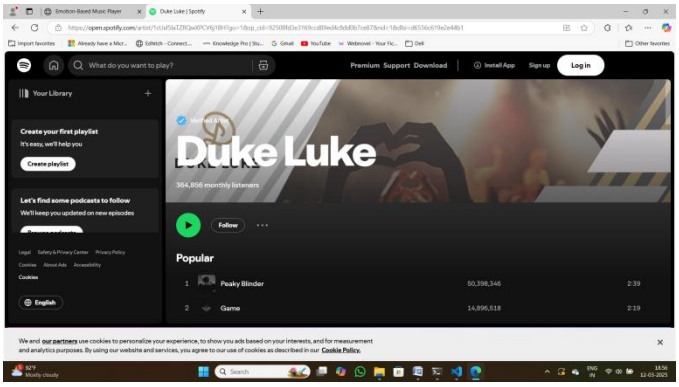
- Expanding the emotion classification model to recognize subtle and complex emotions beyond basic categories.
- Integrating multilingual support for enhanced accessibility across diverse users.
- Optimizing model inference speed to support real-time mobile and edge deployment.
- Enhancing explainability using AI interpretation techniques to improve model transparency.

This research highlights the potential of deep learning-based Speech Emotion Recognition in entertainment and human-computer interaction, paving the way for next-generation AI-driven music recommendation systems.

Output Screenshots:



International Journal of Advanced Research in Computer and Communication Engineering
Impact Factor 8.102 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 14, Issue 3, March 2025
DOI: 10.17148/IJARCCE.2025.14351



REFERENCES

[1] M. E. Ayadi, M. S. Kamel, and F. Karray, "Survey on speech emotion recognition: Features, classification schemes, and databases," *Pattern Recognition*, vol. 44, no. 3, pp. 572–587, 2011.
[2] O. W. Kwon, K. Chan, J. Hao, and T. W. Lee, "Emotion recognition by speech signals," in *Proc. EUROSPEECH*, 2003, pp. 125–128.
[3] B. Schuller, S. Steidl, and A. Batliner, "The INTERSPEECH 2009 emotion challenge," in *Proc. INTERSPEECH*, 2009, pp. 312–315.
[4] T. L. Nwe, S. W. Foo, and L. C. D. Silva, "Speech emotion recognition using hidden Markov models," *Speech Communication*, vol. 41, no. 4, pp. 603–623, 2003.
[5] A. Satt, S. Rozenberg, and R. Hoory, "Efficient emotion recognition from speech using deep learning on spectrograms," in *Proc. INTERSPEECH*, 2017, pp. 1089–1093.
[6] H. Zhao et al., "Exploring deep spectrum representations via attention-based recurrent and convolutional neural networks for speech emotion recognition," *IEEE Access*, vol. 7, pp. 97515–97525, 2019.
[7] S. Latif, R. Rana, J. Qadir, and B. Schuller, "Deep representation learning in speech emotion recognition: A survey," *IEEE Transactions on Affective Computing*, vol. 14, no. 1, pp. 402–423, 2023.
[8] E. Marchi et al., "Automatic speech emotion recognition: A survey," *IEEE Transactions on Affective Computing*, vol. 5, no. 4, pp. 473–491, 2014.
[9] S. Parthasarathy and C. Busso, "Semi-supervised speech emotion recognition with ladder networks," in *Proc. INTERSPEECH*, 2019, pp. 215–219.
[10] M. Neumann and N. T. Vu, "Improving speech emotion recognition with unsupervised representation learning on unlabeled speech," in *Proc. IEEE ICASSP*, 2019, pp. 7390–7394.
[11] R. Lotfian and C. Busso, "Curriculum learning for speech emotion recognition from crowdsourced labels," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 26, no. 1, pp. 174–184, 2018.
[12] P. Tzirakis, G. Trigeorgis, M. A. Nicolaou, B. W. Schuller, and S. Zafeiriou, "End-to-end multimodal emotion recognition using deep neural networks," *IEEE Journal of Selected Topics in Signal Processing*, vol. 11, no. 8, pp. 1301–1309, 2017.
[13] B. Schuller et al., "The INTERSPEECH 2013 computational paralinguistics challenge: Social signals, conflict, emotion, autism," in *Proc. INTERSPEECH*, 2013, pp. 148–152.
[14] S. Tripathi, S. Kumar, A. Abhishek, and A. Nandi, "Deep learning based hybrid model for speech emotion recognition," *Procedia Computer Science*, vol. 167, pp. 1444–1452, 2020.
[15] Y. Kim and E. Mower Provost, "Deep learning for robust feature generation in audiovisual emotion recognition," in *Proc. IEEE ICASSP*, 2013, pp. 3687–3691.
[16] K. Han, D. Yu, and I. Tashev, "Speech emotion recognition using deep neural network and extreme learning machine," in *Proc. INTERSPEECH*, 2014, pp. 223–227.
[17] Z. Huang, Y. Chen, J. Hu, and R. Xie, "Speech emotion recognition under noise conditions using an attention-based convolutional neural network," *IEEE Access*, vol. 7, pp. 57921–57930, 2019.
[18] H. Fayek, M. Lech, and L. Cavedon, "Evaluating deep learning architectures for speech emotion recognition," *Neural Networks*, vol. 92, pp. 60–68, 2017.
[19] A. H. Hajavi and A. A. Iranmehr, "A deep learning-based approach for speech emotion recognition using spectrograms," in *Proc. IEEE ICASSP*, 2020, pp. 4502–4506.
[20] J. Gideon, Y. Khorram, N. Aldeneh, D. Dimitriadis, and E. Mower Provost, "Progressive neural networks for transfer learning in emotion recognition," in *Proc. IEEE ICASSP*, 2017, pp. 5275–5279.
[21] P. Ekman, "Facial expressions of emotion: An old controversy and new findings," *Philosophical Transactions of the Royal Society B: Biological Sciences*, vol. 364, no. 1535, pp. 349–357, 2009.

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.102 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 14, Issue 3, March 2025

DOI: 10.17148/IJARCCE.2025.14351

[22] J. Grewe, J. Kopiez, H. C. Altenmüller, and E. A. Parncutt, "Emotions over time: Synchronicity and development of subjective, physiological, and facial affective reactions to music," *Emotion*, vol. 7, no. 4, pp. 774–788, 2007.
[23] T. Li and M. Ogihara, "Detecting emotion in music," in *Proc. ISMIR*, 2003, pp. 239–240.
[24] X. Lu, R. Wang, and Z. Zhang, "Music emotion classification using deep learning techniques," *Neurocomputing*, vol. 343, pp. 150–161, 2019.
[25] F. Weninger et al., "The Munich feature-based approach to the MediaEval 2013 emotion in music task," in *Proc. MediaEval Workshop*, 2013.