Asian Journal of Computer Science and Technology
ISSN: 2249-0701 (P); 2583-7907 (E)
Vol.13 No.2, 2024, pp.41-47

www.crijournals.org
DOI: https://doi.org/10.70112/ajcst-2024.13.2.4284

Comparative Analysis of Spectrogram and MFCC Representations for Speech Emotion Recognition Using Machine Learning

Rexcharles Enyinna Donatus[1*], B. L. Pal[2], Ifeyinwa Happiness Donatus[3] and Ubadike Osichinaka Chiedu[4]
[1&2]Department of Computer Science and Engineering, Mewar University, Rajasthan, India
[1&4]Africa Centre of Excellence on Technology Enhanced Learning (ACETEL), National Open University of Nigeria, Nigeria
[4]Department of Aerospace Engineering, Air Force Institute of Technology, Kaduna, Nigeria
[3]Department of Computer Science, Kaduna State University, Kaduna, Nigeria
*Corresponding Author: charlly4eyims@yahoo.com
(Received 17 September 2024; Revised 15 October 2024, Accepted 7 November 2024; Available online 10 November 2024)

Abstract **- Emotion recognition is a key area of research within human-computer interaction, addressing the growing need for systems that can respond to human emotional states. While advancements have been made, challenges remain, particularly in selecting appropriate datasets, identifying effective audio features, and optimizing classification models. This study explores how different audio feature representations, specifically Mel-Frequency Cepstral Coefficients (MFCC) and spectrograms, influence the accuracy of emotion classification. By extracting these features from the Ryerson Audio-Visual Database of Emotional Speech and Song (RAVDESS) and applying Random Forest (RF) and Support Vector Machine (SVM) classifiers, the research compares the performance of each feature-classifier pairing. Results indicate that RF and SVM classifiers with MFCC features achieved 50% accuracy, while spectrogram features led to 45% and 54% accuracy, respectively. These findings suggest that simpler models, when combined with appropriate features, can offer promising performance, contributing to more responsive and adaptive human-computer interaction applications.**
Keywords: **Emotion Recognition, Human-Computer Interaction, Mel-Frequency Cepstral Coefficients (MFCC), Support Vector Machine (SVM), Random Forest (RF).**

I. INTRODUCTION

Emotion is described as a conscious mental reaction, deeply felt as an intense sensation and typically paired with changes in both behavior and physiological state [1]. To detect a user's emotional state, machines can employ various methods, including analyzing speech [2], [3], interpreting facial expressions [4], [5], or monitoring bodily signals such as ECG [6], [7]. A deeper understanding of human emotions is crucial for enhancing human-machine interface (HMI) systems [8], [9]. Advances in technology have increased the demand for more natural and intelligent human-machine interaction [10]. In recent years, the use of personal assistants, intelligent chatbots, and smart speakers has grown significantly as key tools for communication, leading to a higher demand for more intuitive interaction methods. However, achieving seamless communication with machines remains a challenge [7].

Research has demonstrated that emotion in humans plays a crucial role in shaping decision-making processes. Consequently, the ability for machines to detect emotions within speech signals has become increasingly essential [11], [8]. The outcomes of this research hold potential applications in areas such as automated customer service, psychological well-being evaluation, and interactive human-machine systems, where understanding emotional states through speech can enhance user experience and decision-making [12].

Numerous approaches have been suggested in the literature to tackle the challenge of recognizing emotions from speech. These systems are generally categorized into two main approaches. Earlier efforts in emotion recognition focused on feature engineering, with researchers emphasizing the importance of specific audio features. Commonly cited features include duration, pitch, intensity, spectral energy distribution, average zero-crossing rate, MFCCs, and filter-bank energy parameters, all of which play significant roles in capturing emotional cues from speech [13], [14].

Efficient feature extraction plays a crucial role in identifying relevant acoustic traits, with widely used tools like Librosa, pyAudioAnalysis, and openSMILE being commonly employed for this purpose [15], [16]. Librosa, in particular, is well-suited for music and audio analysis, offering a range of spectral features such as log mel-spectrogram, MFCCs, tonnetz representation, spectral contrast, and chromagram [17], [8].

In this study, we propose a system capable of recognizing emotions by comparing the efficacy of machine learning classifiers. In [18], a comprehensive survey was presented on speech emotion recognition, utilizing techniques such as MFCC and various classifiers, including SVM and K-Nearest Neighbors (KNN), achieving accuracy rates of up to 84% with GMM and 68% with a Three-Stage SVM classifier.

H. S. Kumbhar *et al.,* [3] proposed a speech emotion recognition (SER) system utilizing MFCC features and a Long Short-Term Memory (LSTM) model, achieving an accuracy of 84.81% and a receiver operating characteristic

(ROC) area of 0.55, indicating room for improvement in reducing false-positive rates. The study concludes that while MFCC is effective for emotion detection, further optimization of the model and exploration of additional features are necessary to enhance performance, suggesting that the reliance on a single feature extraction method may limit the system's robustness across diverse emotional expressions and speaker variations.

M. Hao, *et al.,* [13] proposed a bimodal emotion recognition framework that employs multi-task and ensemble learning techniques, utilizing features from audio (Mel-spectrograms and IS10) and visual data (facial images and LBP) to enhance recognition accuracy, achieving speaker-independent accuracy rates of 56.33% for MTCNN and 54.57% for CNN. The results demonstrate the effectiveness of integrating multiple features and classifiers, suggesting that the proposed method significantly outperforms traditional single-modality approaches.

In their study, M. Ghai *et al.,* [11] focused on recognizing and classifying seven emotions from speech signals using MFCC and energy features, employing classifiers such as SVM, Random Forest (RF), and Gradient Boosting, with RF attaining the highest accuracy of 81.05%. The findings emphasize the ability of machine learning approaches to advance human-machine interaction by facilitating emotion recognition.

A. Mel Frequency Cepstral Coefficients (MFCC)

MFCCs represent the short-term power spectrum of audio by first mapping the log power spectrum onto a Mel frequency scale and then applying a linear cosine transform. MFCCs are designed to align with human auditory perception of frequencies [11], [19]. Additionally, MFCC (Mel-Frequency Cepstral Coefficients) is renowned for its effectiveness in capturing the nuances of the human voice and achieving high recognition accuracy [20], [21], [9]. This analytical tool is widely used in speech recognition due to its power and efficiency [18]. MFCC simplifies the process of feature extraction by compressing the frequency details of speech signals into a smaller, more manageable set of coefficients [20], [21]. Figure 1 illustrates the process of MFCC feature extraction.

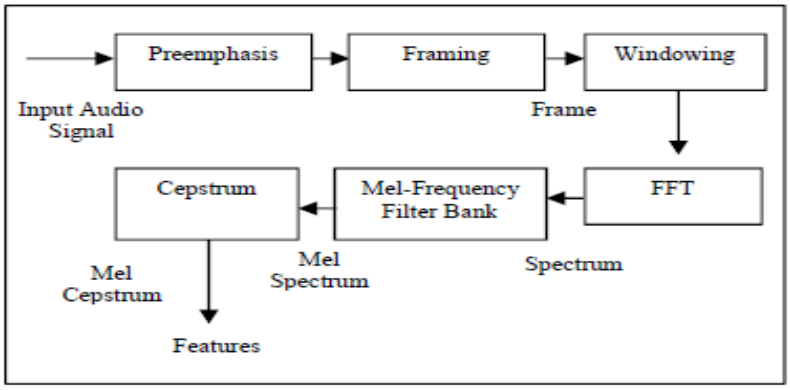

Fig. 1 Diagram illustrating the MFCC extraction process [3]

In Figure 1, the pre-emphasis step involves applying a filter to the speech signal to enhance its spectral characteristics. During the frame-blocking phase, the audio signal is divided into several overlapping frames. This approach allows the audio signal to be processed in smaller segments, enabling the extraction of features that capture temporal characteristics and variations within the audio. Windowing is then employed to analyze portions of longer signals, helping to mitigate aliasing effects. Subsequently, to obtain a frequency spectrum, the time-domain signal is transformed using the Fast Fourier Transform (FFT) [19].

The Mel-frequency filter bank transforms the linear frequency scale into the Mel-frequency scale, which aligns with human auditory perception [19]. This scale is logarithmic, making it more sensitive to lower frequencies compared to higher ones. In the final step of the cepstrum process, a Discrete Cosine Transform (DCT) is applied to revert the Mel spectrum back to the time domain, producing the MFCC. The conversion from Hertz (f) to the Mel scale can be expressed using the following equation [17], [9].

In Figure 1, the pre-emphasis step involves applying a filter to the speech signal to enhance its spectral characteristics. During the frame-blocking phase, the audio signal is divided into several overlapping frames. This approach allows the audio signal to be processed in smaller segments, enabling the extraction of features that capture temporal characteristics and variations within the audio. Windowing is then employed to analyze portions of longer signals, helping to mitigate aliasing effects. Subsequently, to obtain a frequency spectrum, the time-domain signal is transformed using the Fast Fourier Transform (FFT) [19].

The Mel-frequency filter bank transforms the linear frequency scale into the Mel-frequency scale, which aligns with human auditory perception [19]. This scale is logarithmic, making it more sensitive to lower frequencies compared to higher ones. In the final step of the cepstrum process, a Discrete Cosine Transform (DCT) is applied to revert the Mel spectrum back to the time domain, producing the MFCC. The conversion from Hertz (f) to the Mel scale can be expressed using the following equation [17], [9].

$$Mel\ (f) = 295\ log_{10}(1 + \frac{f}{700}) \quad (1)$$

B. Spectrogram

The discrete Short-Time Fourier Transform (STFT) is the most widely used approach for generating a spectrogram and is represented by the following formula [12].

$$STFT\{x[n]\}(m,k) = \sum_{m=-\infty}^{\infty} x[m] \cdot w[n-m] e^{-j\frac{2\pi}{Nx}kn} \quad (2)$$

where N_x is the number of samples.

Over time, various methods for computing spectrograms have been developed. One such technique is the Wigner-Ville distribution, which provides insights into signals that change over time [22]. Wavelet analysis, which applies the continuous wavelet transform using various wavelet bases, can also achieve this objective [12]. While these methods provide a time-frequency representation of the signal's energy density, their resolution accuracy is limited by the time-frequency resolution trade-off, which restricts the ability to obtain precise energy density representations in both time and frequency simultaneously [22]. This study focuses on employing the Short-Time Fourier Transform (STFT) to compute the spectrogram. The original audio signal is analyzed using the discrete STFT, as previously outlined. Figure 3 illustrates examples of spectrograms and MFCCs derived from audio signals that convey different emotions.

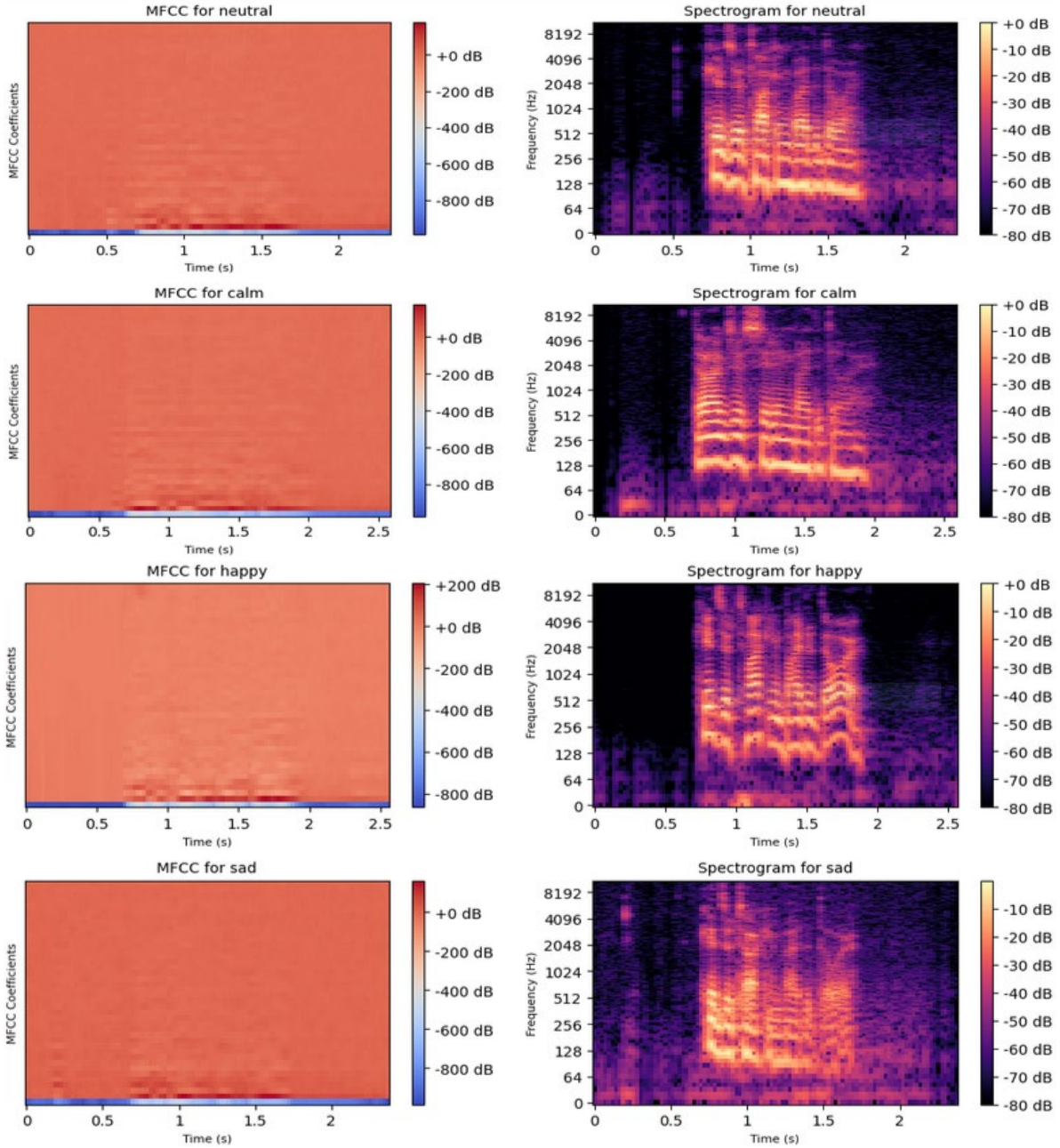

Fig. 2 Spectrogram and MFCC Representations of Emotional Audio Signal

C. Support Vector Machine (SVM) Classifier

SVMs are supervised learning models equipped with algorithms for classification and regression analysis. They offer a straightforward approach for linear classification [2], [23]. However, their effectiveness with non-linearly separable data largely depends on the selected kernel. SVMs are particularly efficient and effective in training large datasets, and their accuracy is generally superior to that of many other techniques [15], [24]. Support Vector Machines (SVMs) utilize kernel functions to manage both non-linear and linear separations. Key advantages include adaptability, scalability to high-dimensional spaces, and the capacity to model complex functions through convex optimization. Despite these strengths, SVMs face challenges such as longer training times, kernel selection sensitivity, noise vulnerability, limited probability estimates, and reduced model interpretability [25].

D. Random Forest (RF) Classifier

The RF classifier is an ensemble learning algorithm that enhances prediction accuracy and mitigates overfitting [26]. It builds several decision trees by randomly selecting both the training data and features at each branching point [27]. The final prediction is obtained by aggregating the outputs from each decision tree, which helps to minimize variance and encapsulate various characteristics of the data [23]. This method maintains a low bias by combining the diverse outputs of the decision trees [28].

$$F(x) = argmax_i \left\{ \sum_{b=1}^{B} T\ \left(A(B, \theta_b)\right) \right\} \qquad (3)$$

Here, $F(x)$ represents the random forest model, B is the number of decision trees, and $\theta_b$characterizes the parameters of each individual tree in the forest [26]

In this paper, Section II outlines the methodology of the implemented system, Section III presents the observed results from the comparative study, and Section IV provides the conclusion.

II. METHODOLOGY

A. Speech Representation

This study evaluates the use of spectrograms as time-frequency representations and compares their effectiveness with Mel-Frequency Cepstral Coefficients (MFCCs) for speech emotion classification. Spectrograms and MFCCs were generated using Python 3, Matplotlib, and Librosa libraries. The spectrograms were produced using the Short-Time Fourier Transform (STFT), which applies the Fast Fourier Transform (FFT) to overlapping portions of the speech signal. Specifically, a 1024-point FFT was utilized with Hanning window functions and a center frequency of 2.442 GHz, covering a range from 2.402 GHz to 2.482 GHz. The images were standardized to 224x224 pixels to ensure uniformity, regardless of audio signal length, facilitating consistent feature extraction.

MFCCs, crucial for capturing speech emotion features, were extracted by dividing the audio signal into overlapping 20 ms frames, with a step size of 20 ms. To transform the signal to the frequency domain, a 1024-point FFT was used, after which the Mel scale was mapped to the linear frequency scale through a Mel-frequency filter bank. Finally, the Mel spectrum was transformed into MFCCs using a Discrete Cosine Transform (DCT).

B. Dataset

This study employs the Ryerson Audio-Visual Database of Emotional Speech and Song (RAVDESS) [2]. It consists of 7,356 audio files featuring speech by professional actors. However, only a subset of this data was used for the experiment. These actors, using a standard North American accent, deliver two statements expressing different emotions. Each emotion is expressed at both typical and elevated emotional intensities, in addition to a neutral expression. The audio files are recorded in 16-bit, 48 kHz WAV format.

C. Machine Learning Classifiers

The MFCC and spectrogram features were directly fed into two distinct machine learning classifiers for evaluation. Support Vector Machine (SVM), a linear model, was selected under the assumption that the feature representations (MFCCs and spectrograms) were sufficiently processed to enable effective classification using linear models. Random Forest (RF) was included for comparative analysis. All models were implemented in Python 3 utilizing the Scikit-learn library. A linear kernel was employed for SVM, while RF was configured with 1,000 trees, allowing the internal optimizer to determine the optimal depth for each tree. The formula for the linear kernel function is expressed as:

$$\text{Kernel}\ (x, y) = (x.y)$$

This formula indicates that the vector operation for the linear kernel involves calculating the inner product between x and y. It is used in SVMs when the data is assumed to be linearly separable.

Random Forest is an ensemble learning approach widely employed for applications such as regression, classification, and more. It functions by generating multiple decision trees during training and making predictions by averaging the outputs for regression tasks or taking the mode for classification tasks. By using several trees, Random Forest helps to minimize the likelihood of overfitting, commonly associated with single decision trees, thus improving model generalization [11]. The classifiers were trained to recognize and categorize eight distinct emotions.

III. RESULTS AND DISCUSSION

This section presents the results of the performance evaluations for the proposed approach, along with a comparison to existing models. The experimental procedures have been detailed in the previous section. The performance of the SVM and Random Forest models was compared using two distinct feature representation methods, MFCC and spectrogram, for classifying speech emotions. The experiment aimed to identify which combination of classifier and feature representation would be more effective in recognizing distinct emotions.

Figure 3 shows the results obtained from the RF and SVM classifiers using the MFCC feature representation, while Figure 4 depicts the results obtained for RF and SVM using the spectrogram feature representation. The accuracy scores across the four models were moderate, with the following outcomes:

1. SVM with spectrogram achieved the highest accuracy at 54%.
2. SVM with MFCC and Random Forest with MFCC both yielded 50% accuracy.
3. Random Forest with spectrogram achieved the lowest accuracy at 45%.

These results indicate that the SVM paired with spectrogram features performed best overall, suggesting that spectrograms may capture emotional characteristics in speech more effectively when combined with a linear classifier like SVM. In contrast, the Random Forest classifier struggled with both MFCC and spectrogram features, particularly with spectrograms, where it achieved the lowest overall performance.

Random Forest Classification Report (MFCC):

	precision	recall	f1-score	support
0	0.50	0.18	0.27	22
1	0.52	0.77	0.62	44
2	0.40	0.44	0.42	36
3	0.43	0.41	0.42	32
4	0.57	0.63	0.60	43
5	0.47	0.26	0.33	31
6	0.49	0.57	0.53	37
7	0.53	0.47	0.49	43
accuracy			0.50	288
macro avg	0.49	0.47	0.46	288
weighted avg	0.49	0.50	0.48	288

SVM Classification Report (MFCC):

	precision	recall	f1-score	support
0	0.48	0.50	0.49	22
1	0.62	0.73	0.67	44
2	0.33	0.42	0.37	36
3	0.37	0.41	0.39	32
4	0.60	0.67	0.64	43
5	0.54	0.45	0.49	31
6	0.61	0.38	0.47	37
7	0.47	0.40	0.43	43
accuracy			0.50	288
macro avg	0.50	0.49	0.49	288
weighted avg	0.51	0.50	0.50	288

Fig. 3 Performance results of the Random Forest and SVM models using MFCC feature representation

Random Forest Classification Report (Spectrogram):

	precision	recall	f1-score	support
0	0.57	0.18	0.28	22
1	0.54	0.75	0.63	44
2	0.22	0.19	0.21	36
3	0.37	0.34	0.35	32
4	0.48	0.56	0.52	43
5	0.56	0.45	0.50	31
6	0.55	0.46	0.50	37
7	0.40	0.49	0.44	43
accuracy			0.45	288
macro avg	0.46	0.43	0.43	288
weighted avg	0.46	0.45	0.44	288

SVM Classification Report (Spectrogram):

	precision	recall	f1-score	support
0	0.46	0.55	0.50	22
1	0.67	0.75	0.71	44
2	0.45	0.56	0.50	36
3	0.33	0.34	0.34	32
4	0.68	0.60	0.64	43
5	0.47	0.48	0.48	31
6	0.60	0.49	0.54	37
7	0.56	0.47	0.51	43
accuracy			0.54	288
macro avg	0.53	0.53	0.53	288
weighted avg	0.54	0.54	0.54	288

Fig. 4 Performance results of the Random Forest and SVM models using spectrogram feature representation

A. Precision, Recall, and F1-Score Analysis

To further analyze the performance of the models, key metrics were examined for each emotion category.

1. SVM with Spectrogram demonstrated the highest precision, recall, and F1-scores across multiple emotions, particularly for emotions such as *Calm* (precision: 67%, recall: 75%, F1-score: 71%) and *Happy* (precision: 68%,

recall: 60%, F1-score: 64%). This suggests that the combination of SVM and spectrogram features captures nuanced emotional cues, enhancing the classifier's ability to identify *Calm* and *Happy* emotions with reasonable accuracy.

2. Random Forest with MFCC exhibited moderate performance, with its highest recall of 77% for *Calm*, but lower precision and F1-scores, especially for emotions such as *Disgust* and *Fearful*. This variability in performance could be attributed to the inherent limitations of Random Forest in handling highly correlated features in MFCCs, which may obscure emotion-specific details.

The classification reports reveal that both models consistently recognized *Calm* and *Happy* emotions more effectively, which might be attributed to the distinctive acoustic patterns of these emotions. Emotions such as *Disgust* and *Fearful* were less accurately identified across models and feature representations, possibly due to subtle tonal differences that are harder to discern, especially with a limited dataset size and default model parameters.

IV. CONCLUSION

This study presented a speech emotion recognition (SER) system using machine learning models with MFCC and spectrogram features. The results indicate that, although MFCC is widely used, spectrograms provide better accuracy for speech emotion detection, particularly when using Support Vector Machine (SVM) classifiers. The SVM with spectrogram features demonstrated promise in recognizing specific emotions such as Calm and Happy. Future research should focus on fine-tuning model parameters, incorporating additional audio features, and expanding the dataset to enhance the model's generalizability and overall performance across all emotion categories. The evaluation of the model achieved commendable accuracy using default parameters; however, there is significant potential for further improvement by exploring a variety of audio features.

REFERENCES

[1] M. A. H. Akhand, S. Roy, N. Siddique, M. A. S. Kamal, and T. Shimamura, "Facial emotion recognition using transfer learning in the deep CNN," *Electronics*, vol. 10, no. 9, p. 1036, 2021, doi: 10.3390/electronics10091036.

[2] J. de Lope and M. Graña, "An ongoing review of speech emotion recognition," *Neurocomputing*, vol. 528, pp. 1-11, 2023, doi: 10.1016/j.neucom.2023.01.002.

[3] H. S. Kumbhar and S. U. Bhandari, "Speech emotion recognition using MFCC features and LSTM network," in *Proc. 2019 5th Int. Conf. Comput. Commun. Control Autom. ICCUBEA 2019*, vol. 1, pp. 1-3, 2019, doi: 10.1109/ICCUBEA47591.2019.9129067.

[4] R. E. Donatus, I. H. Donatus, and U. O. Chiedu, "Exploring the impact of convolutional neural networks on facial emotion detection and recognition," *Asian Journal of Electrical Sciences,* vol. 13, no. 1, pp. 35-45, 2024.

[5] S. Mp and S. A. Hariprasad, "Facial emotion recognition using a modified deep convolutional neural network based on the concatenation of XCEPTION and RESNET50 V2," *Electronics*, vol. 10, no. 6, pp. 94-105, 2023.

[6] D. Shin, D. Shin, and D. Shin, "Development of emotion recognition interface using complex EEG/ECG bio-signal for interactive contents," *Multimed. Tools Appl.*, vol. 76, no. 9, pp. 11449-11470, 2017, doi: 10.1007/s11042-016-4203-7.

[7] Q. Wang, M. Wang, Y. Yang, and X. Zhang, "Multi-modal emotion recognition using EEG and speech signals," *Comput. Biol. Med.*, vol. 149, p. 105907, 2022, doi: 10.1016/j.compbiomed.2022.105907.

[8] Z. Yang, Z. Li, S. Zhou, L. Zhang, and S. Serikawa, "Speech emotion recognition based on multi-feature speed rate and LSTM," *Neurocomputing*, vol. 601, p. 128177, 2024, doi: 10.1016/j.neucom.2024.128177.

[9] A. Bhavan, P. Chauhan, Hitkul, and R. R. Shah, "Bagged support vector machines for emotion recognition from speech," *Knowledge-Based Syst.*, vol. 184, p. 104886, 2019, doi: 10.1016/j.knosys.2019.104886.

[10] S. S. Chandurkar, S. V. Pede, and S. A. Chandurkar, "System for prediction of human emotions and depression level with recommendation of suitable therapy," *Asian Journal of Computer Science and Technology*, vol. 6, no. 2, pp. 5-12, 2017, doi: 10.51983/ajcst-2017.6.2.1787.

[11] M. Ghai, S. Lal, S. D. L., and S. Manik, "Emotion recognition on speech attributes using machine learning," in *Proc. 2024 IEEE Int. Conf. Inf. Technol. Electron. Intell. Commun. Syst. ICITEICS 2024*, pp. 22-27, 2024, doi: 10.1109/ICITEICS61368.2024.10624904.

[12] S. Madanian et al., "Speech emotion recognition using machine learning — A systematic review," *Intell. Syst. with Appl.*, vol. 20, p. 200266, 2023, doi: 10.1016/j.iswa.2023.200266.

[13] M. Hao, W. Cao, Z. Liu, M. Wu, and P. Xiao, "Visual-audio emotion recognition based on multi-task and ensemble learning with multiple features," *Neurocomputing*, vol. 391, pp. 42-51, 2020, doi: 10.1016/j.neucom.2020.01.048.

[14] F. Noroozi, T. Sapiński, D. Kamińska, and G. Anbarjafari, "Vocal-based emotion recognition using random forests and decision tree," *Int. J. Speech Technol.*, vol. 20, no. 2, pp. 239-246, 2017, doi: 10.1007/s10772-017-9396-2.

[15] A. V. Geetha, T. Mala, D. Priyanka, and E. Uma, "Multimodal emotion recognition with deep learning: Advancements, challenges, and future directions," *Inf. Fusion*, vol. 105, p. 102218, 2024, doi: 10.1016/j.inffus.2023.102218.

[16] D. Issa, M. F. Demirci, and A. Yazici, "Speech emotion recognition with deep convolutional neural networks," *Biomed. Signal Process. Control*, vol. 59, p. 101894, 2020, doi: 10.1016/j.bspc.2020.101894.

[17] M. M. R. Mashhadi and K. Osei-Bonsu, "Speech emotion recognition using machine learning techniques: Feature extraction and comparison of convolutional neural network and random forest," *PLoS One*, vol. 18, no. 11, pp. 1-13, 2023, doi: 10.1371/journal.pone.0291500.

[18] S. B. Jagtap, K. R. Desai, and M. J. K. Patil, "A survey on speech emotion recognition using MFCC and different classifiers," in *8th Natl. Conf. Emerg. Trends Engg. Technol.*, pp. 502-509, 2018.

[19] T. Arikrishnan and C. P. Darani, "A pathological voices assessment using classification," *Asian Journal of Engineering and Applied Technology*, vol. 3, no. 1, pp. 5-8, 2014, doi: 10.51983/ajeat-2014.3.1.710.

[20] S. Suke *et al.,* "Speech emotion recognition system," *Int. J. Adv. Res. Sci. Commun. Technol.*, vol. 4, no. 3, pp. 156-159, 2021, doi: 10.48175/ijarsct-v4-i3-024.

[21] R. Panda, R. Malheiro, and R. P. Paiva, "Audio features for music emotion recognition: A survey," *IEEE Trans. Affect. Comput.*, vol. 14, no. 1, pp. 68-88, 2023, doi: 10.1109/TAFFC.2020.3032373.

[22] N. C. Ristea, L. C. Dutu, and A. Radoi, "Emotion recognition system from speech and visual information based on convolutional neural networks," in *Proc. 2019 10th Int. Conf. Speech Technol. Human-Computer Dialogue, SpeD 2019*, pp. 1-6, 2019, doi: 10.1109/SPED.2019.8906538.

[23] G. C. Jyothi, C. Prakash, G. A. Babitha, and G. H. Kiran Kumar, "Comparison analysis of CNN, SVC and random forest algorithms in segmentation of teeth X-ray images," *Asian Journal of Computer Science and Technology*, vol. 11, no. 1, pp. 40-47, 2022, doi: 10.51983/ajcst-2022.11.1.3283.

[24] R. S. Agrawal and U. N. Agrawal, "A review on emotion recognition using hybrid classifier," in *Spec. Issue Natl. Conf. Recent Adv. Technol. Manag. Integr. Growth 2013 (RATMIG 2013)*, no. Icicct, 2017.

[25] A. Hussain, N. Saikia, and C. Dev, "Advancements in Indian sign language recognition systems: Enhancing communication and accessibility for the deaf and hearing impaired," *Asian Journal of Electrical Sciences,* vol. 12, no. 2, pp. 37-49, 2023.

[26] S. Shankaracharya, S. S. S. K. R. Kumar, S. L. Y. G. Varma, and D. S. R. Reddy, "The accuracy analysis of different machine learning classifiers for detecting suicidal ideation and content," *Int. J. Intell. Eng. Syst.*, vol. 12, no. 1, pp. 46-56, 2023.

[27] S. Sathurthi, R. Kamalakannan, and T. Rameshkumar, "Study of ensemble classifier for prediction in health care data," *Asian Journal of Computer Science and Technology*, vol. 8, no. S1, pp. 36-37, 2019, doi: 10.51983/ajcst-2019.8.s1.1963.

[28] J. Wei, X. Yang, and Y. Dong, "User-generated video emotion recognition based on key frames," *Multimed. Tools Appl.*, vol. 80, no. 9, pp. 14343-14361, 2021, doi: 10.1007/s11042-020-10203-1.