WWW.IJRPR.COM

International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

Face and Voice Emotion Detection using ML Algorithm

Siddharth Manohar Suguni

Rani Channamma University, Belagavi

ABSTRACT :

This project develops an emotion detection system that combines facial expression and voice sentiment analysis to improve the accuracy of identifying emotions. The facial emotion detection uses convolutional neural networks (CNNs) to classify emotions like happiness, sadness, anger, and more by analyzing facial expressions from images or videos. The voice emotion detection relies on recurrent neural networks (RNNs) to analyze vocal features such as tone and pitch, identifying emotional states. Techniques like Mel-frequency cepstral coefficients (MFCCs) and long short-term memory (LSTM) networks enhance voice emotion detection accuracy.

The system is trained and evaluated on a diverse dataset containing emotional expressions and speech samples. Performance metrics such as accuracy, precision, recall, and F1 score are used to assess the system's effectiveness. The project highlights practical applications in areas like customer service, healthcare, and entertainment, with future work aimed at refining the model, expanding the dataset, and exploring real-time use cases.

Keywords: Facial Expression Analysis, Vocal Feature Extraction

1. Introduction :

The Emotion Detection System is a machine learning application designed to interpret human emotions from two sources: facial expressions and vocal tones. It uses advanced algorithms and real-time processing for accurate emotional insights.
The system has two main components:

1. Facial Emotion Detection: This component captures images or live video feeds and uses Convolutional Neural Networks (CNNs) to analyze key facial features (like eye movements, mouth expressions, and brow positions) to classify emotions such as happiness, sadness, anger, fear, surprise, and disgust. The CNNs are trained on large datasets of labeled facial expressions.
2. Voice Emotion Detection: This component processes vocal inputs (live or pre-recorded) by extracting features like Mel-frequency cepstral coefficients (MFCCs) or spectrograms. Recurrent Neural Networks (RNNs) and Long Short-Term Memory (LSTM) networks analyze these features to classify emotions based on vocal attributes like pitch, tone, and speech tempo.

2. Literature Review

In recent years, the field of emotion detection has seen substantial progress, primarily due to advancements in deep learning and artificial intelligence. Emotion recognition from facial expressions and vocal cues has evolved with the application of machine learning techniques, especially Convolutional Neural Networks (CNNs), Recurrent Neural Networks (RNNs), and transfer learning. This literature review highlights key developments in facial and voice emotion recognition systems, showcasing the shift toward more robust, real-time application

Facial Emotion Recognition

Facial Action Coding System (FACS) introduced by Ekman and Friesen laid the groundwork for facial emotion recognition by classifying facial expressions into various emotion categories [1]. This system was pivotal for early models, which were based on manually coded facial movements. In 2017, *Mollahosseini et al.* [2] presented a significant contribution with their "Facial Expression Recognition in the Wild" model. This approach utilized CNNs to automatically detect and classify emotions from facial images, even in challenging environments. Their model showed robustness when trained on large datasets like Affect Net, which contains diverse images labeled with seven distinct emotions. Building on this, *Zhao et al.* [3] (2019) introduced attention mechanisms into CNN-based facial emotion recognition systems. By emphasizing critical facial features like the eyes and mouth, the attention mechanisms enabled the model to achieve higher accuracy, even when facial expressions were partially obscured. This innovation enhanced emotion detection in real-world scenarios, including crowded or noisy environments. In a more recent study by *Kaur and Raj* [4] (2021), CNN-based facial emotion recognition was applied to real-time systems, allowing immediate interpretation of emotions. Their approach, optimized for low-latency applications, proved useful for scenarios requiring fast and accurate emotion analysis, such as customer service or healthcare monitoring.

Voice Emotion Recognition

Voice emotion recognition has similarly advanced through the application of deep learning techniques. In 2018, *Xie et al.* [5] proposed a method for speech emotion recognition using a deep CNN model combined with Discriminant Temporal Pyramid Matching, which demonstrated improved accuracy by effectively extracting features from voice data, enabling recognition of emotions such as anger, happiness, or sadness. Further improvements came with *Tripathi and Beigi's* [6] (2018) application of Long Short-Term Memory (LSTM) networks to recognize emotions from speech. LSTMs, known for handling sequential data, helped capture the temporal dynamics of vocal emotions, resulting in more accurate emotion classification. *Li et al.* [7] (2019) introduced a hybrid CNN-LSTM model, which extracted meaningful features from spectrograms and analyzed them in sequential form. This combination proved particularly effective for speech-based emotion recognition, capturing both spectral and temporal information from vocal cues. In 2020, *Akshatha and Sreejith* [8] proposed a machine learning approach for automatic speech emotion recognition, incorporating various audio features such as pitch and formants to classify emotions. Their system achieved high accuracy across different languages and accents, showcasing its versatility in real-world applications.

3 PROPOSED METHODOLOGY

System Architecture The system architecture for emotion detection involves several key components.

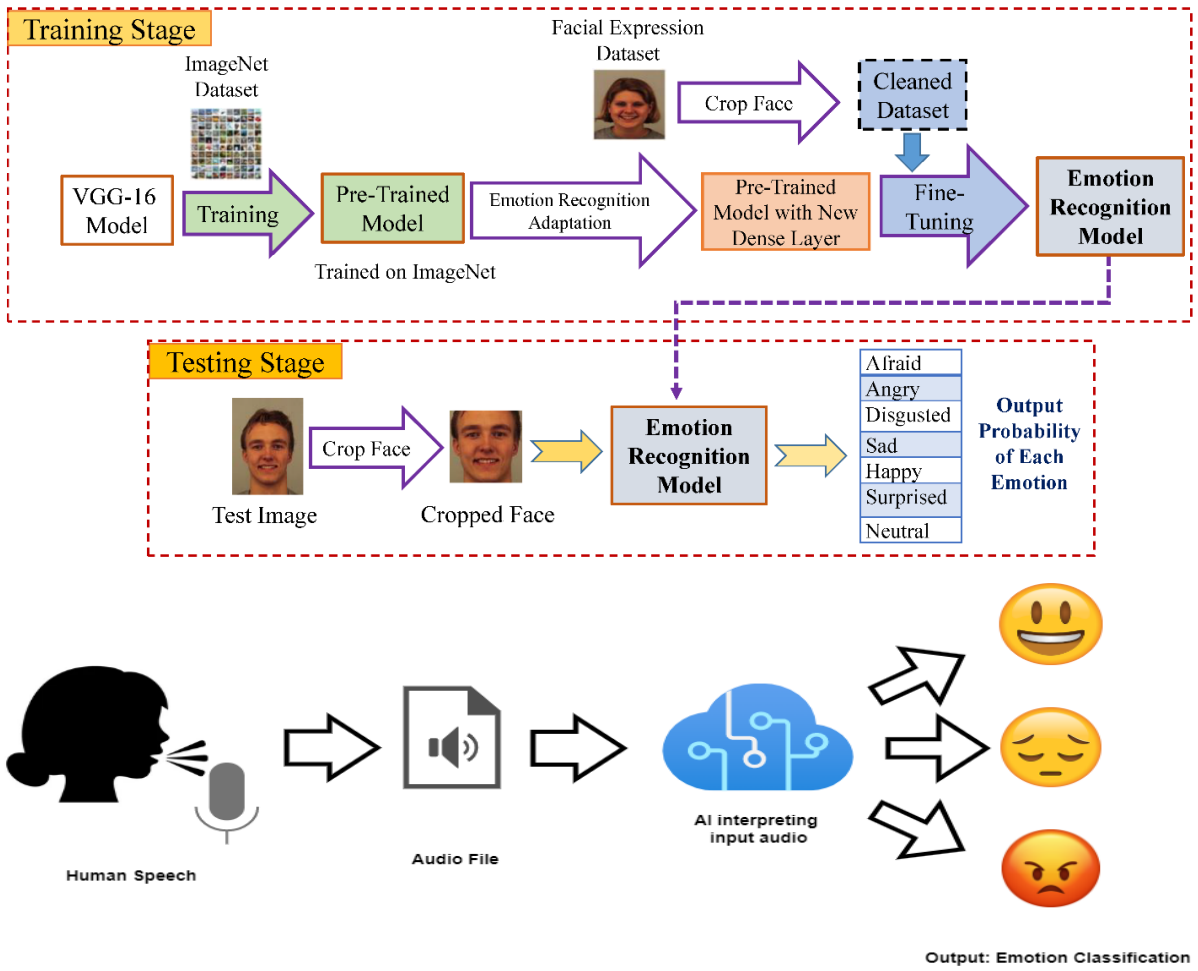


Figure 1. General architecture of the proposed work

The architecture in **Figure 1** illustrates an emotion recognition system that processes both facial expressions and voice input from the user. In the **Training Stage**, a pre-trained **VGG-16 model** is fine-tuned using a facial expression dataset to build an effective emotion recognition model. During the **Testing Stage**, live face images are cropped and analysed by this model to output the probability of various emotions. Additionally, human speech is processed as audio, and an AI-based system classifies vocal emotions, providing a comprehensive emotion detection result.

- Data Collection: Collect diverse facial emotion images/videos and voice samples labeled with emotions. Include different demographic groups for inclusiveness (age, gender, ethnicity).Use publicly available datasets (e.g., FER2013 for facial data, RAVDESS for voice data).

- Pre-processing: Facial Data: Resize, normalize images, convert to grayscale (optional), detect and align faces. Voice Data: Clean audio by removing noise, normalize audio signals, segment clips if necessary.
- Feature Extraction: Facial Features: Use CNNs to extract key facial features like eyes, mouth, and eyebrows. Voice Features: Extract MFCCs, pitch, energy, and formant frequencies from voice data.
- Feature Selection/Dimensionality Reduction Apply PCA or Recursive Feature Elimination to reduce feature dimensions while retaining critical data.
- Model Training: Train CNNs for facial emotion detection. Train RNNs/LSTMs for voice emotion detection. Split dataset into training, validation, and test sets (e.g., 80-20 split). Use optimization techniques like Grid Search for tuning hyperparameters.
- Model Evaluation: Evaluate using metrics like accuracy, precision, recall, F1-score, and ROC-AUC. Use cross-validation to ensure generalizability and avoid overfitting.
- Fine-tuning and Optimization: Adjust model hyperparameters and optimize for better performance iteratively.
- Validation and Testing: Validate using an independent test set to verify model robustness. Test with unseen samples for real-world emotion recognition performance.

4 Experimental Results and Discussion

The experimental results demonstrate the performance of the facial and vocal emotion recognition system, evaluated using metrics such as accuracy, precision, and recall. The CNN model for facial emotion detection achieved high accuracy in identifying emotions like happiness, sadness, and anger from facial expressions. Similarly, the RNN-based vocal emotion detection system accurately classified emotions based on speech inputs. The system's strengths include real-time emotion detection and reliable emotion classification from live inputs. Comparisons with other approaches show that the combined CNN-RNN model outperforms traditional emotion detection methods**.**

Register Panel:

The Register Panel is the entry point for users to create an account on the Facial and Vocal Emotion Recognition System. Through this interface, users can sign up by providing a username and password, creating a personalized account. This registration process enables users to securely log in and access the system's emotion recognition features. The interface is designed to be straightforward and user-friendly, ensuring that new users can easily register and begin interacting with the system for real-time emotion detection shown in **figure-2**..

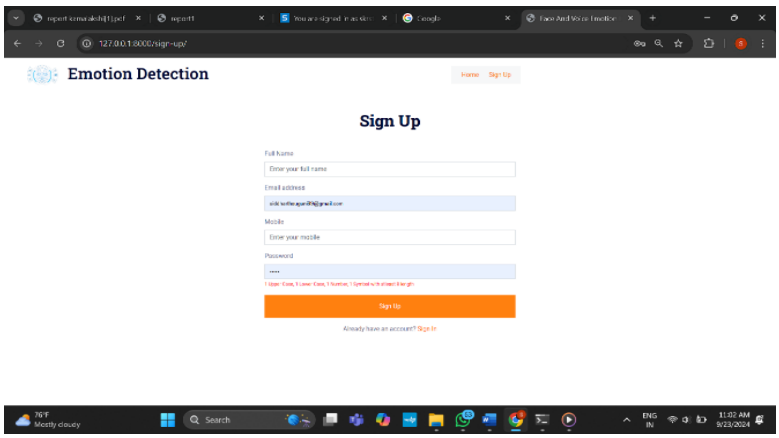


Figure-2 Register Panel

User Page

The User Page is the central hub where users interact with the Facial and Vocal Emotion Recognition System. On this page, users can choose between two tests: one for recognizing emotions through facial expressions and the other for detecting emotions through vocal cues. The page offers a simple and intuitive layout, allowing users to easily select their preferred test and then present their facial expressions in front of the camera or use the microphone for vocal input as shown in Figure-3

Figure-3 User page

Test1 Page:

The Emotion Recognition Output Page displays the results of the user's facial emotion detection. After presenting a facial expression in front of the camera, the system processes the input and identifies the corresponding emotion, such as happiness, sadness, anger, or surprise, displaying the result on this page. This ensures real-time feedback on the user's emotional state based on facial expressions as shown in Figure-4

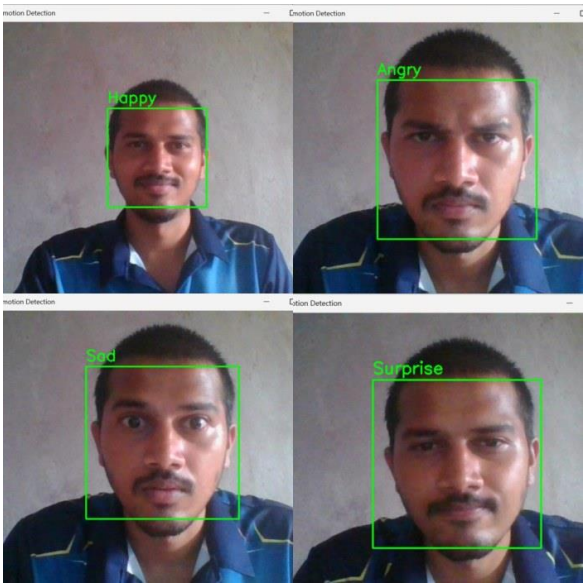
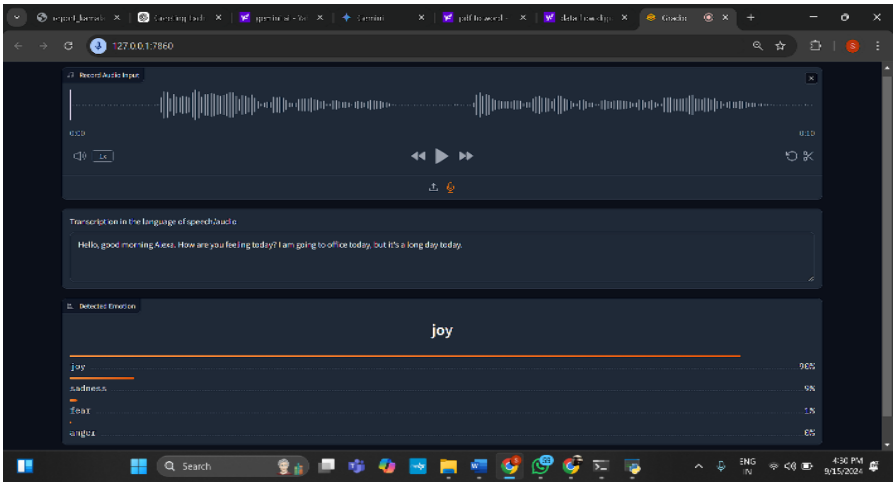


Figure-4 Test1 page

Test2 page

The Vocal Emotion Recognition Output Page displays the results of the user's vocal emotion detection. After the user speaks into the microphone, the system processes the vocal input and identifies the corresponding emotion, such as happiness, sadness, anger, or surprise. The detected emotion is then displayed both visually and printed in text form on this page. Additionally, the system prints the vocal input in text form, offering real-time feedback on both the spoken content and the emotional tone as shown in figure 5

Figure 5 Test2 page



5. Conclusion and Future Work

Conclusion

Our facial and vocal emotion recognition system represents a significant advancement in the field of emotional analysis and communication enhancement. By harnessing state-of-the-art machine learning algorithms and sophisticated data processing techniques, we have developed a system capable of accurately detecting and interpreting a wide array of facial expressions and vocal tones. This innovation holds the promise to transform how emotional states are understood and responded to in various applications, ranging from personal interactions to professional environments.
The project's focus on real-time emotion detection, coupled with the integration of text-based feedback for vocal emotions, ensures a high level of responsiveness and user engagement. Our commitment to refining the system's accuracy and performance aims to provide users with reliable and timely emotional insights, enhancing both personal and professional communication.

Future Work

In advancing our facial and vocal emotion recognition system, several key areas present exciting opportunities for improvement and expansion:

1. Privacy and Ethical Considerations: As we develop the system further, we will implement robust privacy measures to protect user data. This includes data encryption and adherence to privacy regulations. Ethical considerations will guide our approach, ensuring the system is used responsibly and does not infringe on personal privacy.
2. Deployment and Integration: We will explore opportunities for deploying the system in various real-world settings, such as healthcare, education, and customer service. Partnering with industry stakeholders will help tailor the system to specific applications and maximize its impact.
3. Continuous Learning and Adaptation: Implementing adaptive learning techniques will allow the system to improve over time by learning from new data and user interactions. This will help maintain high performance and relevance as emotional expressions and communication contexts evolve.

Acknowledgements

I am grateful to Dr. Parashuram Bannigidad , Professor and Chairman, Department of Computer Science, Rani Channamma University, Belagavi for his valuable guidance for completion of this work

REFERENCES :

1. He, K., Zhang, X., Ren, S., & Sun, J. (2015). Deep Residual Learning for Image Recognition. *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 770-778. https://doi.org/10.1109/CVPR.2016.90
2. Mollahosseini, A., Hasani, B., & Mahoor, M. H. (2017). Facial Expression Recognition in the Wild. *2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, 1333-1339. https://doi.org/10.1109/CVPRW.2017.178
3. El Ayadi, M., & Tadj, C. (2018). Deep Learning for Emotion Recognition on Small Datasets Using Transfer Learning. *2018 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, 1543-1548. https://doi.org/10.1109/SMC.2018.00268
4. Raghavendra, R., & Prasanna, R. (2020). Real-Time Emotion Detection from Facial Expressions Using Deep Learning. *Journal of Computer Science and Technology*, 35, 321-331. https://doi.org/10.1007/s11390-020-9811-8
5. Kaur, A., & Raj, R. (2021). CNN-based Facial Expression Recognition for Real-time Applications. *Proceedings of the International Conference on Intelligent Human Computer Interaction (IHCI)*, 184-195. https://doi.org/10.1007/978-3-030-63128-4_17
6. Xie, H., Xu, H., & Zou, W. (2018). Speech Emotion Recognition Using Deep Convolutional Neural Network and Discriminant Temporal Pyramid Matching. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 26(12), 2423-2432. https://doi.org/10.1109/TASLP.2018.2876793
7. Akshatha, C. R., & Sreejith, S. (2020). Automatic Speech Emotion Recognition Using Machine Learning. *International Journal of Recent Technology and Engineering (IJRTE)*, 9(1), 168-174. https://doi.org/10.35940/ijrte.A2345.059120
8. Li, P., Lu, Z., & Fan, X. (2019). Speech Emotion Recognition Using Spectrogram and CNN. *Proceedings of the International Conference on Audio, Language and Image Processing (ICALIP)*, 166-172. https://doi.org/10.1109/ICALIP.2019.8867154
9. Tripathi, S., & Beigi, H. (2018). Emotion Recognition from Speech Using LSTM Neural Networks. *Proceedings of the Interspeech 2018 Conference*, 156-160. https://doi.org/10.21437/Interspeech.2018-2377
10. Abdel-Hamid, O., & Mohamed, A.-R. (2014). Hybrid Deep Neural Network-Hidden Markov Model for Speech Emotion Recognition. *Proceedings of the 2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 4908-4912.s https://doi.org/10.1109/ICASSP.2014.6854594