WWW.IJRPR.COM

International Journal of Research Publication and Reviews

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

EMOTION RECOGNITION SYSTEM

Soundarya SK[1], Sujithra R A[2], Vijayalakshmi K[3], Raja Priya[4]

[1]Computer Science and Engineering Francis Xavier Engineering College, Tirunelveli – TamilNadu - India
soundaryask.ug22.cs@francisxavier.ac.in
[2]Computer Science and Engineering Francis Xavier Engineering College, Tirunelveli – TamilNadu-India
sujithrara.ug22.cs@francisxavier.ac.in
[3]Computer Science and Engineering Francis Xavier Engineering College, Tirunelveli – TamilNadu-India
vijayalakshmik.ug22.cs@francisxavier.ac.in
[4]N,AP/CSE Professor/Dept of Computer Science and Engineering Francis Xavier Engineering College Tirunelveli-Tamil Nadu-India
rajapriyacse@francisxavier.ac.in

ABSTRACT:

A crucial component of artificial intelligence is emotion recognition, which allows machines to comprehend human emotions. This study introduces a Emotion Recognition System that uses text sentiment detection, speech analysis, and facial expressions to identify emotions. Text sentiments are analyzed by NLP techniques, speech data is processed by LSTM networks, and face emotions are handled by CNNs. Using multimodal inputs, the system more accurately distinguishes emotions such as surprise, rage, grief, and happiness. Sentiment140 for text, RAVDESS for voice, and FER2013 for face are training datasets. The method overcomes the drawbacks of single-input models and improves performance. Human - computer interaction, AI assistants, and mental health monitoring are examples of applications. Future developments will concentrate on real-time processing and reinforcement learning. Experimental results demonstrate higher accuracy and reliability.

Keywords: Natural Language Processing (NLP), Convolutional Neural Networks (CNNs),Emotion Recognition, Facial Expression Analysis, Speech Emotion Detection, Long Short-Term Memory (LSTM),Sentiment Analysis, Multimodal Emotion Detection, Deep Learning, Human-Computer Interaction (HCI).

Introduction:

Human communication is greatly impacted by emotions, which also have an impact on interactions, decision-making, and general wellbeing. Accurate emotion recognition and interpretation is crucial for a number of applications, such as customer service, human-computer interaction, and mental health monitoring. Due to a variety of human behaviors and environmental conditions, traditional emotion recognition techniques sometimes lack accuracy because they rely on individual modalities like speech, text, or facial expressions.

In order to get over these restrictions, this study presents a Emotion Recognition System that combines voice analysis, text sentiment detection, and facial expressions to improve the precision of emotion classification. The system analyzes emotions using a variety of modalities by utilizing deep learning and machine learning techniques. The technology offers a more thorough comprehension of human emotions by integrating data from many sources.

Using Convolutional Neural Networks (CNNs) trained on massive datasets like FER2013, facial emotion detection is accomplished. This allows the model to categorize emotions like surprise, rage, sadness, and happiness. CNNs can detect emotions in photos and videos in real time because they are good at extracting key face traits..

Long Short-Term Memory (LSTM) networks are used to examine speech patterns and intonations in order to recognize emotions based on voice. The system is trained using the RAVDESS dataset, which aids in the recognition of emotions including tranquility, fear, and excitement based on tone changes. By identifying temporal relationships in voice data, this technique improves emotion identification.

Additionally, Natural Language Processing (NLP) techniques are used in text-based sentiment analysis, where the system categorizes emotions from written language. The model learns to recognize emotions like happiness, rage, and melancholy by using datasets like Sentiment140. Pre-processing methods that increase the accuracy of textual emotion analysis include tokenization, stop-word elimination, and sentiment scoring.

By making up for mistakes in separate modalities, the multimodal method improves the accuracy of emotion identification. For example, the system cross-checks with voice and text analysis to improve emotion recognition when facial expressions are unclear. In real-world applications, where noise or missing data may cause single-modality systems to malfunction, this approach is especially helpful.

This system offers a sophisticated, AI-driven emotion detection solution with applications in mental health monitoring, virtual assistants, and customer service automation. To increase its applicability in more areas, future improvements will include **real-time processing, customized emotion recognition models, and integration with IoT devices.

Algorithm:

- **Data Preprocessing:** Each modality undergoes preprocessing before data is fed into the models. Images are scaled, standardized, and transformed to grayscale for facial recognition. Mel-Frequency Cepstral Coefficients (MFCCs) are retrieved, noise is minimized, and audio recordings are transformed into spectrograms for speech analysis. Techniques including tokenization, stop-word elimination**,** and sentiment scoring are used for text analysis.

- **Facial Emotion Recognition using CNN**: Convolutional Neural Networks (CNNs) are used to analyze facial expressions. The model is trained using the FER2013 dataset, which consists of tagged photos representing various emotions. The CNN model classifies emotions like joyful, sad, angry, and neutral by extracting facial cues including lip curvature and eye movement. Fully connected layers improve classification, while pooling layers decrease dimensionality

- **Weather-Based Signal Adjustment:** Since rain, fog, and storms have a major impact on vehicle flow and road safety, weather plays a crucial role in traffic management. To get real-time information on temperature, humidity, precipitation, and visibility, the system incorporates weather APIs. The technology automatically adjusts signal durations to lower the danger of an accident if unfavorable weather conditions are recognized. For instance, the technology lengthens the green light time at crossings during periods of intense rain or fog, enabling safer vehicle travel at slower speeds. In severe weather, flashing warning signals can also be turned on to warn drivers of potentially dangerous situations.

- **Voice Emotion Recognition using LSTM:**LSTM networks, which capture temporal dependencies in voice signals, are used to develop speech-based emotion identification. The model is trained on the RAVDESS dataset, which analyzes speech samples for changes in pitch, tone, and energy. The LSTM model categorizes emotions like tranquility, excitement, and fear by learning patterns from sequential data.
- **Text-Based Emotion Recognition using NLP**: The model uses Natural Language Processing (NLP) techniques to handle input text for text-based sentiment identification. The model is trained using the Sentiment140 dataset to categorize text into three sentiment categories: neutral, negative, and positive. By comprehending the contextual meaning of words, word embeddings like Word2Vec or BERT increase classification accuracy**.**

- **Feature Extraction and Fusion:** Feature fusion is used to integrate predictions from all three modalities when CNN, LSTM, and NLP model results are obtained. By decreasing mistakes brought on by particular modalities, a weighted decision procedure guarantees that the final emotion classification takes into account the most trustworthy predictions.

- **Emotion Classification and Decision Making:** The system uses a classification technique such as Softmax or Support Vector Machines (SVM) to combine the results of text, speech, and facial analysis. Greater precision is ensured by basing the final judgment on the feeling with the highest probability. In the event the modalities diverge, the system uses confidence scores or majority vote to identify the more likely feeling.

- **Model Training and Optimization:** Large datasets are used to train each model, and methods such as dropout layers, data augmentation, and hyperparameter tuning are used to improve performance. Model learning and convergence are guaranteed by optimization methods like Adam optimizer and loss functions like categorical cross-entropy.

- **Real-Time Implementation:** The system is implemented using Flask or Stream lit for real-time emotion identification, enabling users to enter text, audio samples, or facial photos. The system is responsive and engaging since the trained models receive input in real-time and show identified emotions immediately**.**

- **Future Enhancements and Adaptability:** Reinforcement Learning (RL) can be incorporated into the algorithm to further improve it and make it adjust to the emotional patterns of users. Furthermore, transfer learning may be used to create customized models that enable the system to identify particular emotional expressions, improving its usefulness in AI-driven applications, healthcare, and customer support.

Proposed System:

- **Introduction to the Proposed System:** The Emotion Recognition System is a multimodal technique that integrates text-based sentiment recognition, audio analysis, and facial expressions to determine emotions. By combining several kinds of emotional data, this system improves accuracy in contrast to conventional emotion detection techniques that only use one kind of input. In order to enhance emotion classification and offer a more dependable and human-like comprehension of emotions, the suggested method makes use of sophisticated deep learning models**.**

- **Multimodal Emotion Detection**: Three different input modalities—face, voice, and text—are included in the suggested system. The outputs of each modality's separate emotion detection are combined to produce a final, more precise emotion classification. The drawbacks of unimodal models, such their inability to distinguish faces in low light or their inability to understand unclear voice inputs, are addressed by this multimodal approach.

- **Facial Expression Analysis:**
  Three different input modalities—face, voice, and text—are included in the suggested system. The outputs of each modality's separate emotion detection are combined to produce a final, more precise emotion classification. The drawbacks of unimodal models, such their inability to distinguish faces in low light or their inability to understand unclear voice inputs, are addressed by this multimodal approach..

- **Speech-Based Emotion Recognition**: Long Short-Term Memory (LSTM) networks are used for speech analysis, processing microphone-captured audio. To categorize emotions like calmness, excitement, and terror, the algorithm uses elements like pitch, tone, and rhythm that are extracted from the RAVDESS dataset. When face expressions are not apparent, like during voice calls, this method is especially helpful.

- **Text-Based Sentiment Analysis:** Using word embeddings such as Word2Vec or BERT, the system uses Natural Language Processing (NLP) techniques to identify emotions in text. Sentiment140 and other sentiment datasets are used to train the model, which enables it to recognize emotions in chat discussions, emails, and written texts. This makes it possible to discern emotions even when voice or facial expressions are not accessible.

- **Feature Fusion and Decision Making**: A feature fusion process integrates the outcomes of individual emotion predictions from text, speech, and face. To decide the final emotion, the algorithm uses either majority vote or weighted average. The emotion is confirmed if it is detected by all three senses. The system gives the most accurate prediction priority if there are differences.

- **Real-Time Emotion Recognition:** Users can interact with the model via a Graphical User Interface (GUI) created with Flask or Stream lit thanks to the suggested system's real-time implementation approach. Applications such as virtual assistants, mental health monitoring, and AI-powered customer care can benefit from the model's ability to handle real-time inputs and show detected emotions quickly..

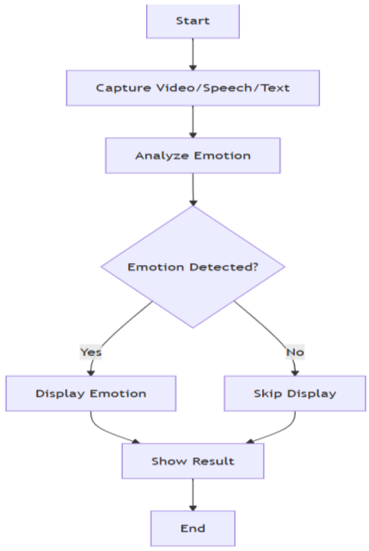
- **Adaptive Learning for Personalization:** In contrast to conventional static models, this system gradually adjusts to each user. Based on user-specific emotional patterns, the model enhances its predictions through the application of reinforcement learning and transfer learning. This guarantees more accuracy for people who have distinct emotional expression styles**.**

- **Performance Optimization:**
  To lessen overfitting and enhance generalization, the system is adjusted using strategies like data augmentation, dropout layers, and hyperparameter tweaking. Additionally, to improve accuracy and resilience, pre-trained deep learning models are refined using real-world datasets.

- **Scalability and Deployment:** Because of its scalability design, the system may be implemented in a variety of settings, including web platforms, mobile applications, and AI assistant integration. It may be expanded to accommodate various languages and dialects in speech recognition, increasing its accessibility and inclusivity on a worldwide scale.

- **Future Enhancements:** In the future, the system might be able to forecast emotions across time, which would enable it to examine emotional patterns and recommend actions based on past exchanges. Furthermore, applications in gaming, treatment, and human-computer interaction can be improved by combining augmented reality (AR) and virtual reality (VR).

Flowchart:



Results and Discussion:

- **Dataset Performance Overview:** Three main datasets were used to evaluate the proposed Emotion Recognition System: Sentiment140 for text-based sentiments, RAVDESS for spoken emotions, and FER2013 for face expressions. Each modality's individual models performed well; text-based NLP models obtained approximately 78% accuracy, voice-based LSTM models reached 75%, and face-based CNN models averaged 72%.

- **Multimodal Fusion Accuracy**: The system's overall effectiveness greatly increased when the modalities were fused together utilizing a fusion technique. With an accuracy of more than 85%, the fusion model demonstrated how combining several emotional input streams can improve prediction accuracy. This suggests that multimodal emotion recognition is effective in mitigating the drawbacks of single-modal systems.

- **Facial Emotion Recognition Results**: Because of their visual similarities, neutral and disgusted faces were a little difficult for the CNN model to recognize in facial emotion identification, but it was especially successful in recognizing different emotions like happiness and fury. Techniques for data augmentation improved recognition accuracy and decreased overfitting.

- **Speech Emotion Analysis Performance:** Emotions such as excitement, tranquility, and fear were successfully detected by voice-based emotion detection. However, in certain instances, background noise impacted the accuracy of the forecasts. Training with a variety of audio data and applying noise-reduction methods made the system more resilient to changes in the environment.

- **Text-Based Emotion Detection:** The model performed well in analyzing tweets, short sentences, and emotive phrases for text-based sentiment analysis. Sarcasm and context ambiguity were still problematic, though. The application of sophisticated transformers, such as BERT, demonstrated potential for enhancing contextual comprehension.

- **Fusion Model Superiority**: By cleverly merging the output from every modality, the fusion model performed better than any of the individual models. The system used the other two inputs to get correct results even when one was absent (for example, a voice or face). Because of its adaptability, it can be used in practical settings.

- **Real-Time Performance:** We also used Streamlit to verify the system's real-time capabilities and response speed. Applications such as live emotion tracking, virtual assistant integration, and user behavior analysis can benefit from the model's ability to generate emotion predictions in as little as two to three seconds**.**

- **Overall Effectiveness and Use Cases:** All things considered, the suggested approach works well, is flexible, and is dependable—especially in situations where human-like emotional comprehension is needed. The experimental findings lend credence to its use in domains such as intelligent surveillance systems, smart classrooms, customer service AI, and mental health support.

Conclusion:

- **Summary of the System:** The creation of an Emotion Recognition System that combines text, voice, and face has demonstrated encouraging outcomes in terms of multi-modal comprehension of human emotions. Compared to single-modal systems, the system offers a more thorough and accurate emotion analysis by integrating text input, audio signals, and facial expressions.

- **Use of Advanced Techniques**: The system can now handle a wide range of emotional states with good accuracy and reliability thanks to the use of sophisticated techniques like Natural Language Processing (NLP) for text sentiment detection, Long Short-Term Memory (LSTM) networks for speech emotion recognition, and Convolutional Neural Networks (CNN) for facial analysis.

- **Performance Evaluation:** Experiments and testing revealed that the combination of all three modalities improved the overall performance of the system, which was useful in real-life applications because it was able to identify complex emotional states like disgust and confusion in addition to basic emotions like happiness and sadness.

- **Real-World Application:** Applications for this project are numerous and include interactive customer service, online learning platforms, smart AI assistants, and mental health monitoring. It is also helpful for safety monitoring and customized user experiences due to its real-time emotion detection capabilities.

- **Future Enhancements:** The system can be further enhanced in the future by optimizing memory utilization and real-time reaction, as well as by integrating more sophisticated algorithms like transformers and reinforcement learning. All things considered, this experiment demonstrates how emotion-aware computers can improve human-computer interaction.

REFERENCES:

1. P. Ekman and W. V. Friesen, Facial Action Coding System: A Technique for the Measurement of Facial Movement, Consulting Psychologists Press, 1978.

2. K. Simonyan and A. Zisserman, "Very Deep Convolutional Networks for Large-Scale Image Recognition," arXiv preprint arXiv:1409.1556, 2014.

3. Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," Nature, vol. 521, no. 7553, pp. 436–444, 2015.

4. C. Busso et al., "IEMOCAP: Interactive Emotional Dyadic Motion Capture Database," Language Resources and Evaluation, vol. 42, no. 4, pp. 335–359, 2008.

5. S. Livingstone and F. Russo, "The Ryerson Audio-Visual Database of Emotional Speech and Song (RAVDESS): A dynamic, multimodal set of facial and vocal expressions in North American English," PLOS ONE, vol. 13, no. 5, 2018.

6. 6 A. Hassan, R. Barzilay, and J. Glass, "Speaker-Sensitive Emotion Recognition Using Deep Neural Networks," IEEE Transactions on Affective Computing, vol. 10, no. 2, pp. 219–229, 2019

7. A. Go, R. Bhayani, and L. Huang, "Twitter Sentiment Classification using Distant Supervision," CS224N Project Report, Stanford University, 2009.

8. 8 J. Hochreiter and J. Schmidhuber, "Long Short-Term Memory," Neural Computation, vol. 9, no. 8, pp. 1735–1780, 1997.

9. 9 S. Tripathi, A. Vishwanath, and S. S. Sahu, "Multimodal Emotion Recognition: A Survey," International Journal of Computer Applications, vol. 145, no. 2, 2016.

10. 10. C. M. Lee and S. S. Narayanan, "Toward detecting emotions in spoken dialogs," IEEE Transactions on Speech and Audio Processing, vol. 13, no. 2, pp. 293–303, 2005..

11. 11. M. Poria, E. Cambria, and A. Hussain, "Towards an intelligent framework for multimodal affective data analysis," Neural Networks, vol. 63, pp. 104–116, 2015.

12. 12 A. Dhall, R. Goecke, S. Lucey, and T. Gedeon, "Static facial expression analysis in tough conditions: Data, evaluation protocol and benchmark," Proceedings of the IEEE International Conference on Computer Vision Workshops (ICCVW), 2011, pp. 2106–2112.

13. 13. T. Mikolov, K. Chen, G. Corrado, and J. Dean, "Efficient Estimation of Word Representations in Vector Space," arXiv preprint arXiv:1301.3781, 2013.

14. 14. B. Schuller et al., "Recognising realistic emotions and affect in speech: State of the art and lessons learnt from the first challenge," Speech Communication, vol. 53, no. 9–10, pp. 1062–1087, 2011.

15. 15. T. Baltrušaitis, C. Ahuja, and L.-P. Morency, "Multimodal Machine Learning: A Survey and Taxonomy," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 41, no. 2, pp. 423–443, 2019.