



International Journal of Innovative Research in Science Engineering and Technology (IJIRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)



Impact Factor: 8.699

Volume 14, Issue 3, March 2025

Emotion Recognition from Facial Expressions and Speech Deep Learning.

Shashivardhan Reddy Jukkanolla, Pradeep Reddy, Kalyan Reddy Gaddam, Raparathi Rohith,

Asst. Prof. Tejas Rana

Department of Computer Science and Engineering, Parul University, Vadodara, Gujarat, India

ABSTRACT: Hospital patients commonly display psychological problems like anxiety, melancholy, pessimism, and eccentricity. However, hospitals typically lack the resources and infrastructure necessary to regularly assess patients' psychological well-being. It is preferable to detect depression in patients so that improved therapy can be administered right away. Improvements in machine learning for image processing, with noteworthy uses in the field of emotion recognition, can make this feasible. facial expressions. In this research, we have suggested two distinct approaches to predict emotions: voice analysis and facial expression detection. We have employed two methods for face expression recognition: convolutional neural networks (CNNs) and Gabor filters for feature extraction with support vector machines for classification.

KEYWORDS: Facial emotion · Speech · Expressions · Deep learning · CNN

I. INTRODUCTION

In the modern world, it is crucial to comprehend human psychology. An expression can be helpful in interpreting the human thoughts, according to studies. When individuals suffer from neuropsychiatric conditions that make it difficult for them to recognize and express their emotions, facial expressions may play a critical role in the assessment of those conditions. Psychologists have identified seven key methods for studying human behaviour: self-observation, direct observation, experimental approaches, surveys, clinical analysis, genetic studies, and testing procedures. Each of these techniques is applicable in different aspects of human life. With advancements in technology, psychological assessment methods can be enhanced by incorporating facial expression and speech analysis. Facial expressions play a vital role in understanding an individual's emotional and psychological state and serve as a key form of non-verbal communication.

II. BACKGROUND AND RELATED WORK

- A. Emotion recognition plays a crucial role in human-computer interaction, affective computing, and psychological analysis. Traditionally, emotion recognition relied on unimodal approaches, such as facial expression analysis using computer vision or speech emotion recognition through acoustic signal processing. However, unimodal systems often struggle with accuracy due to variations in individual expression, environmental factors, and cultural differences. Comparative Analysis: Digital Platforms: Some countries have implemented digital platforms to streamline the process of accessing government schemes.
- B. However, unimodal systems often struggle with accuracy due to variations in individual expression, environmental factors, and cultural differences. To overcome these limitations, researchers have explored multimodal emotion recognition, integrating facial and speech cues to enhance reliability and robustness.

III. METHODOLOGY

THIS STUDY ADOPTS A QUALITATIVE RESEARCH METHODOLOGY BY ANALYZING MULTIPLE CASE STUDIES FROM DIFFERENT INDUSTRIES. THE FOLLOWING APPROACH WAS USED:

The process starts with data collection, where facial expression datasets (e.g., FER-2013, CK+, or AffectNet) and speech emotion datasets (e.g., RAVDESS, CREMA-D, or IEMOCAP) are gathered. The facial images and audio recordings are preprocessed to enhance quality. Facial preprocessing involves face detection, cropping, resizing, and

normalization using techniques like OpenCV and MTCNN. Speech preprocessing includes noise reduction, audio segmentation, Mel spectrogram extraction, and feature normalization using Librosa.

IV. CASE STUDIES AND FINDINGS

A. CHALLENGES IN THE CURRENT STRATEGY

A study was conducted to evaluate the effectiveness of multimodal emotion recognition by integrating facial expressions and speech analysis using deep learning techniques. The dataset used comprised videos from publicly available emotion recognition datasets, such as RAVDESS and CREMA-D, featuring individuals expressing various emotions like happiness, sadness, anger, and neutrality. The system employed a hybrid deep learning model combining Convolutional Neural Networks (CNNs) for facial expression analysis and Recurrent Neural Networks (RNNs) with Mel-Frequency Cepstral Coefficients (MFCCs) for speech-based emotion recognition.

Benefits:

- **Improved Accuracy with Multimodal Fusion** – The combination of facial and speech features significantly outperformed unimodal approaches, achieving an overall accuracy of 85.6%, compared to 76.3% for facial-only and 72.8% for speech-only models. With this web application, people can check their eligibility quickly and easily without having to fill out forms or wait in long lines
- **Cultural and Individual Variability** – Some emotions, such as surprise and fear, were misclassified more frequently due to individual differences in expression and cultural variations in emotional display.

V. CHALLENGES AND LIMITATIONS

Multimodal emotion recognition through face and speech analysis faces several challenges, including dataset bias, variability in emotional expressions, and fusion complexities. Facial and speech expressions can be asynchronous, making it difficult to align emotional cues effectively. Moreover, external factors such as occlusions, lighting conditions, accents, and background noise significantly impact recognition accuracy. Computational constraints also pose a challenge, as deep learning models require extensive training data and resources, often leading to overfitting on small datasets. Additionally, ethical concerns related to privacy, bias, and potential misuse of emotion recognition systems must be addressed to ensure fairness and responsible deployment in real-world applications.

VI. FUTURE TRENDS

Advancements in multimodal emotion recognition will focus on improving model accuracy, robustness, and ethical considerations. Emerging deep learning architectures, such as transformer-based models, will enhance the ability to capture temporal and contextual dependencies between facial and speech modalities. Self-supervised and few-shot learning techniques will reduce reliance on large labeled datasets, enabling models to generalize across diverse populations. Personalized and cross-cultural emotion recognition systems will adapt to individual differences, improving real-world applicability. Real-time processing and deployment on edge devices will make emotion recognition more accessible, while privacy-preserving techniques like federated learning will address security and bias concerns. These developments will drive progress in human-computer interaction, mental health monitoring, and affective computing.

VII. CONCLUSION

Multimodal emotion recognition through face and speech analysis has made significant advancements, but challenges such as data variability, fusion complexities, and ethical concerns remain. Addressing these issues requires improvements in deep learning models, adaptive recognition techniques, and privacy-preserving methods to ensure accuracy, fairness, and security. As research progresses, integrating robust, real-time, and unbiased emotion recognition systems will enhance applications in human-computer interaction, healthcare, and affective computing, ultimately leading to more intuitive .

REFERENCES

- [1] WOA-TLBO: Whale optimization algorithm with Teaching-learning-based optimization for global optimization and facial emotion recognition, Applied Soft Computing, ScienceDirect, 2021
- [2] An Improved Method to Recognize Hand-over-Face Gesture based Facial Emotion using Convolutional Neural Network
Niti Naik; Mayuri A. Mehta 2020 IEEE International Conference on Electronics, Computing and Communication Technologies (CONECCT).
- [3] Mollahosseini, A., & Chan, D. (2016). Going deeper in facial expression recognition using deep neural networks. In Proceedings of the 2016 ACM on Multimedia Conference (pp. 507-511).
- [4] Livingstone, S. R., & Russo, F. A. (2018). 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, 13(5), e0196391.
- [5]. Panning A, Al-Hamadi AK, Niese R, Michaelis B. Facial expression recognition based on Haar-like feature detection. Pattern Recognit Image Anal. 2008;18(3):447–52.
- [6]. Tarnowski P, Kołodziej M, Majkowski A, Rak RJ. Emotion recognition using facial expressions. Proc Comput Sci. 2017;108:1175–84.



INTERNATIONAL
STANDARD
SERIAL
NUMBER
INDIA



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

 9940 572 462  6381 907 438  ijirset@gmail.com



www.ijirset.com

Scan to save the contact details