Journal of Advanced Zoology

ISSN: 0253-7214

Volume 44 Issue S-8 Year 2023 Page 33-36

Multimodal Emotion Recognition System Using Machine Learning Classifier

R. Prabha[1]*

[1]**Assistant Professor, Department of Computer Science, Changu Kana Thakur Arts, Commerce and Science College, New Panvel, rashmiryan22@gmail.com +91-7666980624*

***Corresponding Author:** *R. Prabha*

**Assistant Professor, Department of Computer Science, Changu Kana Thakur Arts, Commerce and Science College, New Panvel, rashmiryan22@gmail.com +91-7666980624*

Abstract

Multi-modal emotion recognition refers to the process of identifying human emotions using information from multiple sources, such as facial expressions, voice intonation, EEG signals etc. Ultimately, emotion recognition is poised to play a pivotal role in healthcare, education, customer service etc. As we progress, it is imperative to address privacy concerns associated with this technology in a responsible manner. Challenges in multi-modal emotion recognition include aligning data from different modalities in time, dealing with noisy or incomplete information. In this paper, we aim to address this issue by employing the SVM as our machine learning classifier. Here we use IEMOCAP for speech and video and DEAP dataset for EEG signals. After applying SVM we got 76.22 % accuracy for IEMOCAP and 68.89 % accuracy for DEAP dataset.

Keywords: Emotion recognition, SVM, EEG Signals, Machine Learning, Multimodal system

Introduction:

The significance of emotion gained prominence with the emergence of cognitive psychology in the mid-twentieth century, as cognitive theories highlighted the importance of both emotion and thinking processes in our daily functioning [1]. Mental health monitoring entails the ongoing evaluation and monitoring of an emotional well-being and mental state of individual. Emotion detection methods involve the collection of EEG signals, providing valuable insights into an individual's mental health [2]. Physiological signals like Electroencephalogram (EEG) are directly generated by the central nervous system and are closely linked to human emotions. As a result, EEG signals can objectively and promptly reflect the real-time emotional state of individuals. Through monitoring stress levels, healthcare providers can provide personalized interventions and coping strategies, assisting individuals in more effectively managing stress. In recent years, the advancement of brain-computer interface technology has led to increased maturity in the acquisition and analysis of human EEG signals. Consequently, more researchers are employing EEG-based research methods to investigate emotion recognition [3].

The main objective is to develop advanced fusion techniques to effectively combine features from different modalities, enabling the proposed model to effectively process a variety of input signals. This endeavor not only improves the accuracy of emotion recognition systems but also pushes the field towards a more refined comprehension of human emotions in diverse contexts and applications. This emotionally intelligent device can communicate directly with individuals anytime and anywhere, thereby reducing the dependence on psychologists.

Literature Review:

The most important speech characteristics for determining the emotions in speech are [4] targeted feature extraction for emotion recognition by creating an unsupervised technique and carrying out tests on the IEMOCAP dataset. Classification accuracy was 78.67%; however, no feature selection algorithms were used in this study, and only one dataset was examined. According to [5] states that decision tree SVM with Fisher feature selection is constructed and applied in the Berlin speech corpus and the CASIA Chinese emotion speech corpus in order to evaluate the effectiveness of this system. The experimental results show that the average emotion identification rate of the proposed approach is 8.26% higher on the Berlin speech corpus and 9% higher on CASIA than the traditional SVM classification method. It has been shown that the recommended method of reducing emotional confusion and raising the rate of emotion detection works. According to [6] divided the SER procedure into three stages for the RAVDESS dataset exclusively: preprocessing, feature extraction, and MLP classifier classification. This study only looks at one dataset, but its best-case accuracy is 85.12%. Because multimodal affective computing systems provide better classification accuracy than unimodal ones, they are studied alongside them. The amount of emotions noticed, the features retrieved, the classification scheme, and the consistency of the database all affect accuracy [3].

Proposed Method:

In this section, we suggest a ML based algorithm that can process multimodal input signals in an efficient way. Figure 1 shows schematic depictions of the steps. The entire model is broken down into four components: data collection, feature extraction, feature fusion, and classification of precise emotion. IEMOCAP and DEAP are two datasets on which this method is applied. Here we proposed an algorithm for multimodal emotion recognition.

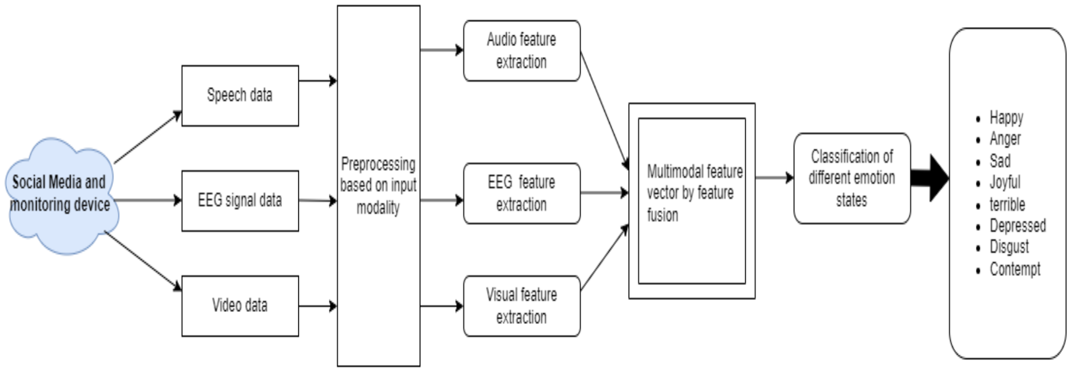

Fig. 1 Proposed system architecture

Algorithm 1

Input: Speech signal, EEG signals, Video signals
Output: Emotional classes (Em)
1. **Start**
2. Pre-process input signal $S_{Train} = (t_i, u_i)$
3. Extract features from different input signals (S_f, E_f, V_f)
4. Perform feature fusion. $F_f[x, y, z] = \left(w_x . S_f + w_y . E_f + w_z . V_f\right)$
5. Generate single feature representation matrix.
6. Pass fused feature to SVM classifier.
7. **Return Em**

Where t_i is feature vector$\{R_1, R_2, R_3, R_4 \dots \dots \dots \dots R_n\}$ and u_i is multiple emotions where$F_f[x, y, z]$ is fused feature w_x , w_y , w_z determine the contribution of each feature to the fused representation. S_f represents Speech feature, E_f represents EEG features, V_f represents voice features.

Result and Discussion:

In this section, we will explore the achievements of the model. Figure 2 illustrates the results obtained after training and testing these datasets using the described SVM classifier (Table 1). We achieved 82.16% and 75.33% accuracy for IEMOCAP and DEAP training dataset and 76.22% and 68.89% accuracy for IEMOCAP and DEAP test dataset. We divide the train and test dataset in 80:20 ratio.

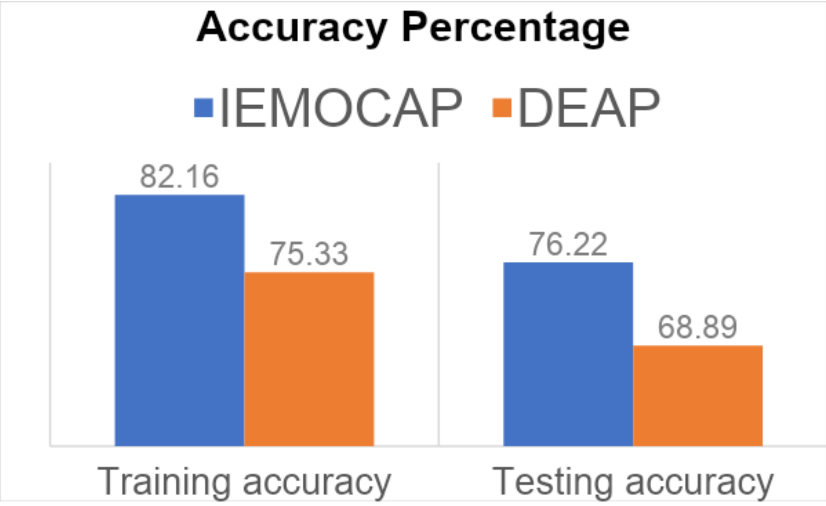

Fig. 2 Accuracy Percentage

Table 1 Result comparison with Training accuracy v/s Testing Accuracy

Sr. No.	Dataset	Classifier	Accuracy (%)
1. Training Accuracy	IEMOCAP	SVM	82.16
	DEAP		75.33
2. Testing Accuracy	IEMOCAP	SVM	76.22
	DEAP		68.89

Conclusion:

In this paper, we used machine learning to analyze and discuss the multimodal system of human emotions. The results show that a multimodal approach from EEG signals to identify emotional states can do emotion recognition more accurately and effectively. Human emotions are prone to change, are not stable, and can be easily controlled in response to a range of social situations. We tested this method on two distinct datasets, and the outcomes are promising. In the future, we hope to broaden this research to encompasses more input data.

References:

1. A. Tripathi and P. Rani, "PCA-Based Feature Selection and Hybrid Classification Model for Speech Emotion Recognition," *Lecture notes in networks and systems*, pp. 347–353, Jan. 2023, doi: https://doi.org/10.1007/978-981-99-3315-0_26.
2. N. Ahmed, Z. A. Aghbari, and S. Girija, "A systematic survey on multimodal emotion recognition using learning algorithms," *Intelligent Systems with Applications*, vol. 17, p. 200171, Feb. 2023, doi: https://doi.org/10.1016/j.iswa.2022.200171.
3. S. M. A. Abdullah, S. Y. A. Ameen, M. A. M. Sadeeq, and S. Zeebaree, "*Multimodal Emotion Recognition using Deep Learning*", JASTT, vol. 2, no. 02, pp. 52-58, Apr. 2021.
4. Ying Y, Tu Y, Zhou H (2021) Unsupervised feature learning for speech emotion recognition based on auto encoder. MDPI

5. Sultana S (2021) SUST Bangla emotional speech corpus. PLOS One
6. Prasanth S, Roshni Thanka M, Bijolin Edwin E, Nagaraj V (2021) Speech emotion recognition based on machine learning tactics and algorithms. Mater Today Proc