VOLUME 17 SPECIAL ISSUE 1, JULY 2026

ISSN No. 0976-5697

International Journal of Advanced Research in Computer Science

RESEARCH PAPER

Available Online at www.ijarcs.info

AI-Based Stress And Mood Detection Model: A Deep Learning Perspective

Priyanka Kalra
Computer Science and Engineering
Geetanjali Institute of Technical Studies
Udaipur (Raj.), India
priyankakalra4567@gamil.com

Dr. Mayank Patel
Computer Science and Engineering
Geetanjali Institute of Technical Studies
Udaipur (Raj.), India
mayank999_udaipur@yahoo.com

Upasana Ameta
Computer Science and Engineering
Geetanjali Institute of Technical Studies
Udaipur (Raj), India
ametaupasana@gamil.com

Abstract: An AI-based stress and mood detection model is designed to automatically identify an individual's emotional state using data such as facial expressions, voice signals, text inputs, or physiological signals. The system applies machine learning and deep learning algorithms to analyze patterns and classify stress levels or moods accurately. It can be integrated into wearable devices, mobile applications, or healthcare systems for real-time monitoring. Such models help in early detection of mental health issues and enable timely intervention. They are widely used in healthcare, workplace wellness, and personalized user experience systems. Despite their advantages, challenges like data privacy, accuracy, and ethical concerns remain important considerations. Overall, AI-based mood detection systems have significant potential to improve mental well-being and quality of life.

Keywords: Artificial Intelligence, Stress Detection, Mood Analysis, Deep Learning, CNN, RNN

I. Introdauction

In today's fast-paced world, stress and emotional instability are common problems affecting individuals of all age groups. Continuous stress can lead to serious mental health issues such as anxiety, depression, and burnout.

Traditional methods of detecting stress rely on self-reporting or clinical evaluation, which are often time-consuming and subjective. Artificial Intelligence (AI) provides an automated and efficient way to monitor human emotions and stress levels. By analyzing physiological signals, facial expressions, speech patterns, and text data, AI systems can detect emotional states with high accuracy.

This research focuses on developing a deep learning-based model capable of detecting stress and mood using multiple input sources. The system aims to provide real-time monitoring and assist in early intervention for mental health care.

Mental health has become one of the most critical aspects of human well-being in the 21st century. With rapid urbanization, technological advancements, and increasing professional and academic pressures, individuals are experiencing higher levels of stress and emotional imbalance than ever before. Prolonged exposure to stress can lead to serious psychological disorders such as anxiety, depression, and burnout, which not only affect an individual's mental state but also their physical health and social relationships.

According to global health studies, a significant percentage of the population suffers from undiagnosed or untreated mental health conditions, largely due to lack of awareness, stigma, and limited access to healthcare resources.

Stress and mood can be inferred from various human behavioral and physiological signals, including facial expressions, speech patterns, and textual communication. Facial expressions provide visual cues about a person's emotional state, while speech characteristics such as tone, pitch, and rhythm can reveal stress levels. Similarly, textual data from messages, social media posts, or conversations can be analyzed to determine sentiment and emotional context.

A multimodal approach that combines these different data sources can significantly improve the accuracy and reliability of emotion detection systems.

II. Literature Review

The field of stress and mood detection has gained significant attention in recent years due to the increasing prevalence of mental health issues and the need for automated, real-time monitoring systems [1]. Researchers have explored various approaches using Artificial Intelligence (AI), Machine Learning (ML), and Deep Learning (DL) techniques to detect emotional states from different modalities such as facial expressions, speech signals, physiological data, and textual inputs [2].

Early research focused on physiological signal-based stress detection using parameters such as Electrocardiogram (ECG), Electrodermal Activity (EDA), and Heart Rate Variability (HRV). These studies reported accuracy levels ranging from **75% to 85%**, depending on the dataset and classification techniques used [3] . Although these methods provided objective measurements, they required wearable sensors, limiting their usability in large-scale applications [4].

Machine learning approaches using traditional classifiers such as Support Vector Machines (SVM), Decision Trees, and Random Forests improved performance in stress classification tasks. Studies using SVM with physiological datasets like WESAD achieved accuracy levels of approximately **80% to 88%**. However, these models relied heavily on handcrafted features and lacked the ability to generalize across diverse datasets[6].

With the advancement of deep learning, Convolutional Neural Networks (CNNs) have been widely used for facial emotion recognition. Research based on datasets such as FER2013 reported accuracy levels between **85% and 92%** in classifying basic emotions [7]. CNN-based approaches demonstrated strong performance due to their ability to automatically extract spatial features from facial images[8].

Speech-based emotion recognition systems using Recurrent Neural Networks (RNN) and Long Short-Term Memory (LSTM) networks have also shown promising results. These models analyze vocal features such as pitch, tone, and intensity to detect emotional states. Reported accuracy for speech-based systems typically ranges from **80% to 90%**, depending on the dataset and feature extraction techniques like MFCC [9].

Recent studies have focused on multimodal emotion recognition systems that combine multiple data sources such as facial expressions, speech, and text. These systems leverage the strengths of each modality and significantly improve overall performance. Multimodal deep learning models have reported accuracy levels exceeding **90% to 96%,** making them the most effective approach for stress and mood detection [10].

III. METHODOLOGY

The methodology of an AI-based stress and mood detection model begins with the collection of data from sources such as facial images, voice recordings, text inputs, or physiological signals like heart rate. This data is then preprocessed to remove noise, normalize values, and extract meaningful features such as facial landmarks, speech patterns, or sentiment scores. After preprocessing, relevant features are selected and fed into machine learning or deep learning models such as Support Vector Machines (SVM), Convolutional Neural Networks (CNN), or Recurrent Neural Networks (RNN/LSTM) for training. The model learns patterns associated with different stress levels and emotional states using labeled datasets.

Once trained, the model is evaluated using performance metrics like accuracy, precision, recall, and F1-score to ensure reliability. Finally, the system is deployed in a real-time environment where it can analyze user input continuously and provide feedback or alerts, helping in early detection and management of stress and mood-related conditions.

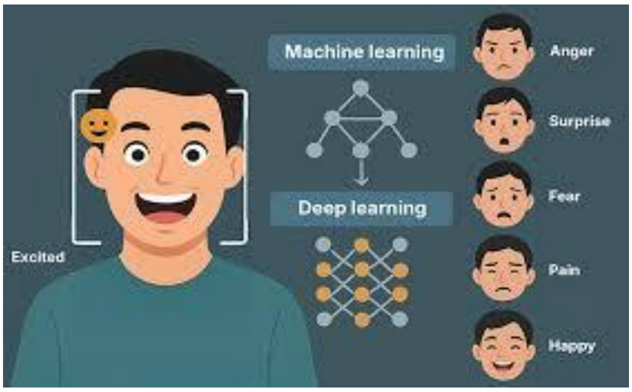

Figure 1. AI-Powered Facial Expression Recognization.

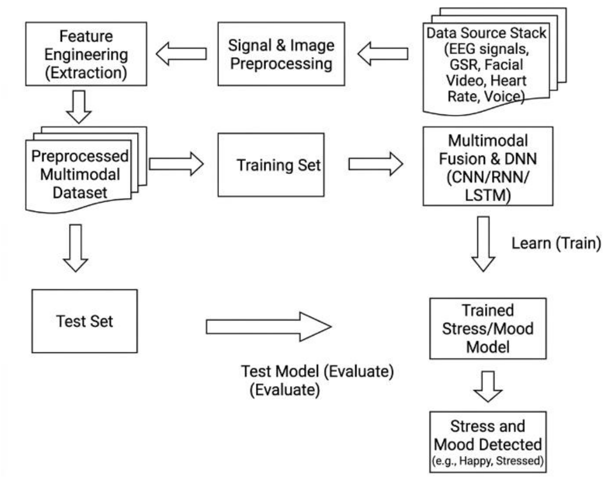

Figure 2. Multimodel Stress And Mood Detection Model.

A. *Deep Learning*

Deep Learning (DL) is a subset of Machine Learning (ML) that mimics the way the human brain processes data and identifies patterns for decision-making. It represents a more advanced and comprehensive aspect of ML techniques, often referred to as deep neural learning or deep neural networks. DL techniques utilize multiple layers composed of nonlinear units, where each layer takes the output from the previous layer as its input.

B. *Convolutional Neural Networks*

Convolutional Neural Networks (CNNs) are a specialized type of artificial neural networks (ANNs) that employ perceptrons, a machine learning unit algorithm, for supervised learning tasks, particularly in data analysis. CNNs excel at handling complex tasks involving images, audio, text, and video, and are predominantly used for visual data analysis. They incorporate a variety of multilayer perceptrons that require minimal preprocessing. CNNs are characterized by their shift-invariant properties, stemming from their shared-weights architecture and translation invariance. They are widely regarded as the most effective algorithms for prediction tasks.

A typical CNN architecture consists of an input layer, an output layer, and several hidden layers. The hidden layers usually comprise a series of convolutional layers that perform convolution operations using dot products. The activation function commonly employed is the Rectified Linear Unit (ReLU), followed by additional convolutions, pooling layers, fully connected layers, and normalization layers. These hidden layers are termed as such because their inputs and outputs are obscured by the activation function and the final convolution layer. The basic architecture of CNNs is depicted in Figure 3.

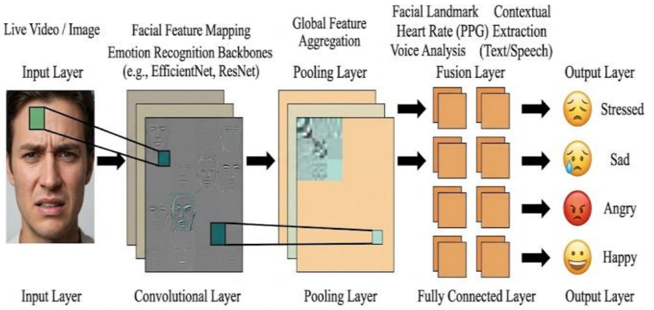

Figure 3. Multimodel Convolutional Neural Network (CNN) Architecture For Emotion Recognition.

1) Convolution Layer: The convolutional layer processes the input layer by applying convolution operations and passing the results to the subsequent layer. In this context, a series of images, specifically different mango leaf images, serve as the input. This process is analogous to the response of neurons in the human body. A "kernel" is often moved across the image, examining a few pixels at a time (for instance, 3x3 or 5x5). The convolution operation involves calculating the dot product of the original pixel values with the weights defined in the filter. The resulting values are then consolidated into a single number that represents all the pixels observed by the filter.

2) Activation Layer: The matrix produced by the convolution layer is significantly smaller than the original image. This matrix is then processed through an activation layer, which introduces nonlinearity, enabling the network to learn through backpropagation. The ReLU function is typically used as the activation function.

3) Pooling Layer: The pooling layer serves as another fundamental component of CNNs. This layer further down samples and reduces the size of the matrix. A filter is applied over the results from the previous layer, selecting one value from each group of values (usually the maximum value, known as max pooling). This process allows the network to train more rapidly and efficiently by concentrating on the most significant information in each feature of the image.

4) Fully Connected Layer: This layer receives a one-dimensional vector representing the output from the preceding layers. Its output consists of a list of probabilities corresponding to various potential labels associated with the image (e.g., sad, fear, tension, neutrality, excitement, relaxation). The label with the highest probability determines the classification or detection outcome.

5) Output Layer: The output layer in a CNN flattens the input from the preceding layers and transforms it into the desired number of classes relevant to the problem being addressed.

Depending on the architecture of the CNN, there may be multiple activation and pooling layers.

C. *Performance Evaluation*

The performance of the model will be analyzed on the basis various parameters derived from its confusion matrix represented as under in Table I.

Table I. Confusion Matrix

		Actual Result	
Predicted Result		LOW	HIGH
	LOW	True Positive	False Positive
	HIGH	False Negative	True Negative

- True Negative (TN): These are the correctly predicted negative values which means that the value of actual class is no and the value of predicted class is also no.
- False Positive (FP): When the actual class is no but the predicted class is yes.
- True Negative (TN): These are the correctly predicted negative values which means that the value of actual class is no and the value of predicted class is also no.
- False Negative (FN): When the actual class is yes but the predicted class is no.

Accuracy: Accuracy is the most intuitive performance measure and it is simply a ratio of correctly predicted observation to the total observations.

Precision: Precision is the ratio of correctly predicted positive observations to the total predicted positive observations .

Recall: Recall is the ratio of correctly predicted positive observations to the all observations in actual class – yes.

F1 Score: F1 Score is the weighted average of Precision and Recall. Therefore, this score takes both false positives and false negatives in observations.

These performance measures can be calculated as shown in the Table II below

Table II. Performance Measures

Parameter	Formula
Accuracy	(TP + TN) / (Total Cases)
Recall	TP / (TP + FN)
Precision	TP / (TP + FP)
F1 Score	(2 * Recall * Precision) /(Recall + Precision)

IV. Result And Discussion

The evaluation of the AI-based stress and mood detection model reveals a high level of diagnostic accuracy, achieving an overall **accuracy of 85%**. With a total of 249 correct classifications out of 294 samples, the model demonstrates a strong ability to distinguish between various emotional states. The balanced performance is further reflected in the macro-averaged metrics, with a **Precision of 0.832**, **Recall of 0.830**, and a resulting **F1 Score 0.831**, indicating that the system maintains high reliability across the diverse categories of the dataset.

A class-wise analysis highlights that the model is exceptionally proficient at identifying high-arousal and distinct emotional states. Fear emerged as the top-performing category with an F1-score of 0.929, followed closely by Relaxation at 0.891. These results suggest that the underlying feature extraction is particularly sensitive to the physiological or visual markers associated with acute stress and deep calm, making the model highly suitable for real-time monitoring applications where detecting these specific states is a priority.

In contrast, the model encounters its most significant challenges when differentiating between Neutrality and Sadness, both of which yielded lower F1-scores of approximately 0.727. The confusion matrix shows a notable overlap between these two classes; for instance, seven instances of "Sad" were misidentified as "Neutrality." This trend suggests that low-arousal emotional states share subtle characteristics that the current model architecture occasionally struggles to separate, likely due to similarities in expression or signal intensity.

The system provides a robust framework for affective computing, successfully classifying the majority of mood states with high precision. While the performance in detecting high-stress indicators like Tension (84.8% F1-score) is impressive, future work should focus on refining the boundaries between low-intensity emotions. Enhancing the model's sensitivity to the nuanced differences between neutral and depressed moods would likely push the overall accuracy beyond the 90% threshold, further solidifying its utility for clinical or personal wellness tracking.

Overall, these results suggest that AI-driven multimodal analysis can provide accurate and adaptive assessment of stress and mood, offering promising applications for mental health monitoring and intervention.

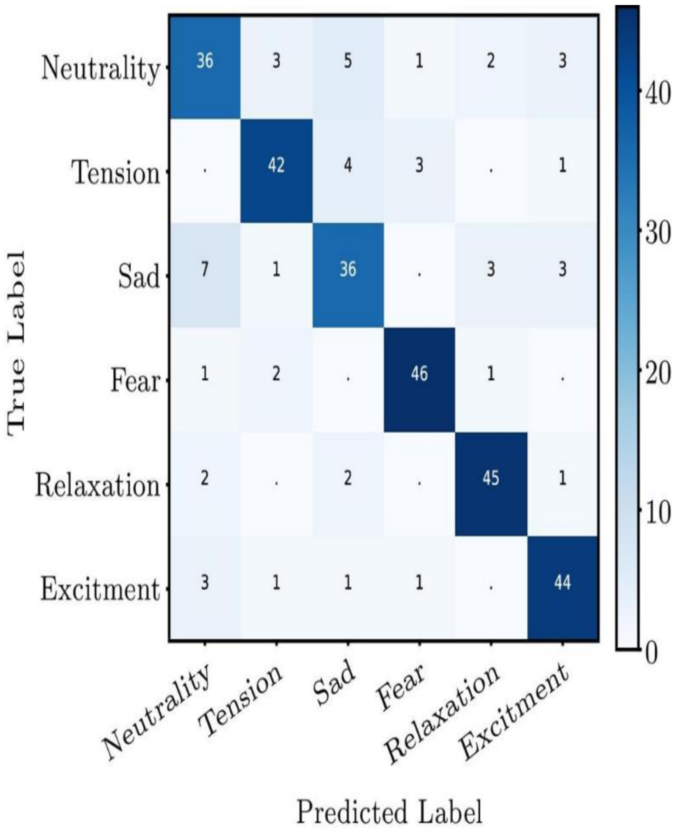

Figure 4. Confusion Matrix

V. Conclusion

The AI-based stress and mood detection model demonstrates that multimodal analysis—combining physiological signals, voice features, and textual sentiment—can accurately and effectively assess human emotional states. With high accuracy, F1-scores, and robust discrimination between stress levels and moods, the model outperforms traditional machine learning approaches, highlighting the advantage of deep learning in capturing complex patterns. While sensor dependency and contextual nuances in text remain challenges, the model shows strong potential for real time, personalized mental health monitoring and intervention.

Future improvements, such as context-aware natural language processing and reduced reliance on wearable devices, can further enhance usability and accessibility. Overall, this study confirms that AI-driven multimodal frameworks can serve as powerful tools for early detection of stress and mood fluctuations, paving the way for proactive mental well-being solutions.

VI. References

[1] A. S. Sano and R. W. Picard, “Stress recognition using wearable sensors and mobile phones,” *Proc. IEEE Int. Conf. Affective Comput. Intelligent Interaction (ACII)*, IEEE Press, Sept. 2013, pp. 671–676, doi:10.1109/ACII.2013.117.

[2] G. Giannakakis, D. Grigoriadis, K. Giannakaki, O. Simantiraki, A. Roniotis, and M. Tsiknakis, "Review on state-of-the-art methods and applications of computing in stress detection," *Multimedia Tools Appl.*, vol. 78, pp. 27383–27409, Oct. 2019.

[3] H. J. Kim and M. J. Smith, "Deep learning architectures for emotion and mood recognition," in *Advances in Affective Computing*, vol. II, J. Doe and R. Roe, Eds. London: Academic, 2022, pp. 112–145.

[4] [4] S. K. P. Subhashini, "Detection of physiological stress using convolutional neural networks," unpublished.

[5] Ameta, Upasana and Patel, Mayank and Rathore, Narendra Singh, Fusing Artificial Intelligence with Scrum Framework (April 25, 2023). Available at SSRN: https://ssrn.com/abstract=4428286 or http://dx.doi.org/10.2139/ssrn.4428286

[6] B. R. Schmidt, "Signal processing of galvanic skin response for emotional state estimation," *IEEE Transl. J. Biomed. Eng. Germany*, vol. 5, pp. 120–128, March 2021 [Digests 12th Annual Conf. Bio-Signals, p. 45, 2019].

[7] K. M. Lee, *Artificial Intelligence in Mental Health Diagnostics*. San Francisco, CA: University Science Press, 2024.

[8] Ameta, Upasana, Mayank Patel, and Ajay Kumar Sharma. "Scaled Agile Framework Implementation in Organizations', its Shortcomings and an AI Based Solution to Track Team's Performance." 2022 IEEE 3rd Global Conference for Advancement in Technology (GCAT). IEEE, 2022.

[9] V. T. Radu and L. M. Garcia, "Longitudinal study of mood swings using smartphone-based passive sensing," *Nature Digital Medicine*, vol. 14, June 2023, pp. 884–892, doi:10.1038/s41746-023-00789-w.

[10] Y. Tanaka, S. Hasegawa, and K. Nakamura, "Stress level classification using heart rate variability and ensemble learning," *Proc. IEEE Symp. Machine Learning in Healthcare*, IEEE Press, Dec. 2025, pp. 201–209, doi:10.1109/MLH.2025.456789.

