Tuijin Jishu/Journal of Propulsion Technology
ISSN: 1001-4055
Vol. 47 No. 03 (2026)

Multimodal Stress Detection Using Deep Learning: Integrating Physiological and Behavioral Signals

Dr. T. Sumadhi

[1] Assistant Professor, Department of computer applications, NGM College, Pollachi, Coimbatore, Tamil Nadu, India

Abstract. The Stress is a big deal nowadays, really taking a toll on how people feel mentally and physically. Traditionally, figuring out someone's stress levels uses questionnaires, but those can be pretty off since they depend on the person's memory and mood at the time. This paper suggests something new model that hooks up physiological signals like heart rate, sweat stuff, and brain waves with how you look and sound to tell if you're stressed. This model use Convolutional Neural Networks and LSTM networks together to spot patterns in your body's responses, both in the moment and overall. The evaluation results obtained clearly highlights that the proposed model worked better than other traditional methods that only focus on one kind of signal. This Proposed model can consistently check for stress in real-time, keep track of health, helps to boost work performance, and set up modified relaxation platforms. Furthermore, the results obtained showed that combining multiple types of signals like behavioral and physiological improves accuracy and stability. This proposed hybrid deep learning technique out performs traditional techniques in terms of all performance evaluation metrics like accuracy, precision, recall, f1-score.

Keywords: Convolutional Neural Networks (CNN), Long Short-Term Memory (LSTM), Stress detection, Deep learning techniques, Physiological and behavioral signals.

1 Introduction

Stress is the most crucial element which leads to more number of issues such as anxiety, depression, heart diseases, and many more. Thus, it is essential to enhance the skills in recognizing and dealing with stress as we are living in hard times. There were always two main approaches to assessing stress: questionnaire and medical consultation. Yet, these options fail to provide continuous evaluation, are slow, and involve self-reporting. But still, stress management practices are more likely to become encouraging nowadays. With the emergence of wearables and artificial intelligence, there is the possibility of detecting stress much better than before. One of them is deep learning technique through which stress levels can be predicted and classified more accurately.The primary focus of this paper is on an innovative method for the identification of stress using data from various behavioral and bodily sources. Multimodal learning is an important factor for this method.

This project highlights some of the important features like:

- It combines physiological and behavioral data in a novel manner.
- It employs a combo of CNN and LSTM to detect spatial and temporal patterns.
- The results obtained clearly show that it outperforms other traditional methods. This proposed method proved to outperform and deals with different scenarios more robustly.

2 Literature Review

There has been significant progress in the field of stress detection, particularly with regard to the sophistication of the methodologies being employed. During the previous stages of this flourishing field, the methodologies heavily depend on guesswork and subjectivity, while the current methodologies are much more reliant on objectively collected data and information. At the earlier stages of development, research was focused on the physiological symptoms of stress, including heart rate and skin conductance. As a demonstration, Healey[1] and Picard showed these physiological signals could be used to accurately detect stress in a real-world driving situation. On the other hand, Gjoreski[2] and colleagues mastered the use of real-time stress detection and the potential of wearable devices to monitor stress on a continuous basis. Following the arrival of deep learning, stress detection was further refined using the power of neural networks. LSTM networks started to be employed to address the issue of the temporal dimension of physiological data (Schmidt[3] and colleagues achieved good results here), while other strategies attempted to use CNN models to detect the stress signals by analyzing the EEG signals or facial expressions. In the current era, one of the best approaches is to combine different data streams. This is what Tzirakis et al.[5] and Zhai and Barreto [6] illustrate. This proposed hybrid approach has produced improvement in the accuracy and efficiency of stress detection solutions. Nonetheless, the usual suspects (data synchronization, high latency, and reliability) continue to hold the research back.

Table 1. Summary of Existing Literature.

Author	Approach	Data Type	Technique Used	Limitation
Healey & Picard[1]	Driver stress detection approach	Heart Rate, Electrodermal activity	Statistical ML	Limited modalities
Gjoreski et al.[2]	Wearable monitoring approach	Physiological dataset	ML models	Moderate accuracy
Schmidt et al.[3]	Temporal modeling approach	Physiological dataset	LSTM	No behavioral data
Li et al.[4]	EEG is analyzed in this paper	EEG dataset is used	CNN	Limited generalization
Tzirakis et al.[5]	Recognition of emotion is performed	Audio-Visual dataset is used	Deep Learning	No physiological signals
Proposed Work	In this work Multimodal stress detection is performed	Combined Physiological and Behavioral signals	Hybrid CNN-LSTM model	Reduced limitations

3 Proposed Methodology

The Proposed CNN-LSTM technique helps detect stress more accurately by combining data available from different sources using deep learning techniques. This architecture has five main stages like data collection phase, preprocessing phase, feature extraction phase, feature fusion phase, and classification phase. In first phase data is gathered from different sources like electro dermal activity and other input streams, and then it is preprocessed to remove noise. Next features are extracted and combined together and finally classification is performed to spot stress more accurately.

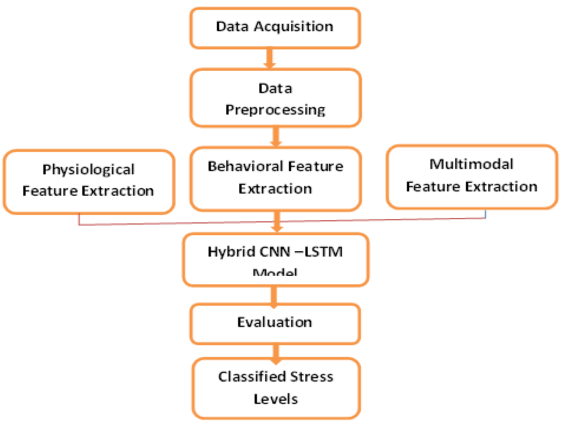

Fig. 1. CNN-LSTM Model Workflow

3.1 Data Collection

In this work two benchmark datasets like WESAD and DEAP are used for evaluation. Physiological signals are collected from WESAD which include heart rate, EDA, and EEG. DEAP dataset contains behavioral data, which consists of facial expressions and speech signals collected from video and audio streams.

For the purpose of evaluation of proposed hybrid model this two datasets have been considered:

- WESAD (Wearable Stress and Affect Detection Dataset)

- DEAP (Acronym for DEAP is Dataset for Emotion Analysis using Physiological Signals)

Table 2. Input Data Types

Category	Signals	Description
Physiological	HR, EDA, EEG	Helps to capture internal body responses
Behavioral	Facial expressions, Speech	Helps to captures external emotional cues

Sample Data

1. WESAD Dataset – (Physiological Signals)

The WESAD dataset includes physiological signals such as Heart Rate (HR), Skin Temperature (Temp), Electro dermal Activity (EDA) and Respiration Rate (Resp). Table 3 below shows the sample data available in WESAD Dataset.

Table 3. Sample WESAD Data

Time (s)	Heart Rate (bpm)	EDA (μS)	Skin Temp (°C)	Respiration Rate	Label
1	72	0.85	32.5	14	Neutral
2	75	0.90	32.6	15	Neutral
3	88	1.20	33.0	18	Stress
4	92	1.35	33.2	20	Stress
5	78	0.95	32.7	16	Amusement

- EDA increases during stress due to sweating
- Heart rate rises in stress conditions
- Labels typically include: Baseline ,Stress ,Amusement

3.2 DEAP Dataset – (Emotion Signals)

The DEAP dataset includes: EEG signals (32 channels) ,Peripheral physiological signals , Emotional ratings: Valence (pleasantness) , Arousal (intensity) , Dominance

Table 4. Sample DEAP Data

Trial	EEG Channel 1	EEG Channel 2	HR (bpm)	Valence	Arousal	Dominance
1	0.45	0.52	70	6.5	5.8	6.0
2	0.60	0.65	85	4.2	7.5	5.1
3	0.30	0.40	68	7.8	4.0	6.8
4	0.75	0.80	90	3.5	8.2	4.5

5	0.50	0.55	72	6.0	5.5	5.8

- **Valence** → Emotional positivity (1–9 scale)
- **Arousal** → Emotional intensity
- **Dominance** → Control over emotion

Stress Mapping Procedure

Condition	Valence	Arousal	Interpretation
High Stress	Low	High	Anxiety
Relaxed	High	Low	Calm state

3.3 Proposed Combined Multimodal Model dataset

Table 5. Sample Combined Multimodal Dataset

Time	HR	EDA	EEG Feature	Facial Expression Score	Voice Pitch	Label
1	72	0.8	0.45	0.30	180 Hz	Neutral
2	88	1.2	0.70	0.65	220 Hz	Stress
3	78	0.9	0.50	0.40	200 Hz	Relaxed

3.4 Data Preprocessing

Preprocessing ensures high-quality input data. Raw data is preprocessed to improve quality and consistency. Noise removal is performed using Butterworth filtering technique. Min-Max Scaling technique is used for normalization and helps to reduce bias. Fixed time windowing technique is used to perform segmentation to capture temporal patterns (e.g., 5–10 seconds) and finally feature extraction is performed. Feature extraction is done using statistical (mean, variance) and spectral features (FFT) which helps to extract essential features.

3.5 Feature Fusion Strategy

Two fusion methods are applied:

Table 6. Fusion techniques

Fusion Type	Description	Advantage
Early Fusion	Combine features before training	Captures joint representation
Late Fusion	Combine model outputs	Improves flexibility

3.6 CNN-LSTM Architecture

The CNN-LSTM architecture consists of:

- CNN layers in this architecture helps to derive spatial features from EEG and facial images
- LSTM layers to capture temporal dependencies in physiological signals(HR, EDA signals)
- Fusion layer helps to integrate multimodal features
- Dense layer with Softmax classifier and performs classification

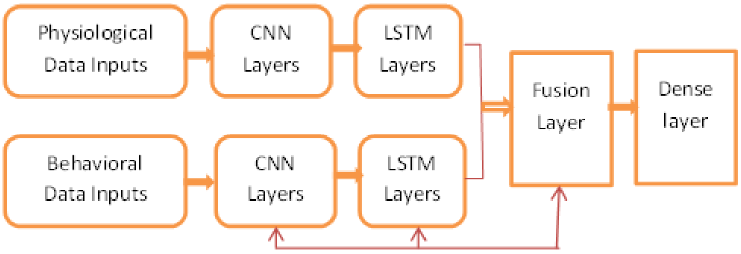

Fig. 2. Hybrid CNN-LSTM Architecture

3.7 Training and Evaluation

- Loss function: Cross-entropy
- Optimizer: Adam
- Metrics used for evaluating our proposed method are Accuracy, Precision, Recall and F1-Score.

3.8 Training Strategy

- Loss Function: Cross-entropy
- Optimizer: Adam
- Epochs: 50–100
- Batch size: 32

4 Results and Interpretations

The proposed multimodal model was evaluated against unimodal baselines.

Table 7. Evaluation table

Model	Accuracy
Physiological Only	82.3%
Behavioral Only	79.5%
Proposed Multimodal Model	91.2%

The results shown in table 7 clearly demonstrate that Multimodal fusion significantly improves accuracy, and CNN has effectively extracted spatial features, LSTM captures temporal dependencies. The proposed system is robust across different datasets. The improved performance highlights the importance of integrating diverse data sources for stress detection.

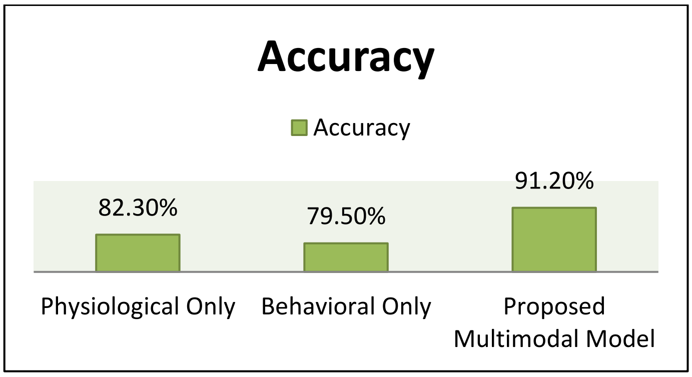

Fig. 3. Accuracy Comparison Graph

The Proposed model was evaluated using standard evaluation metrics and the results are displayed in table 8.

Table 8. Performance Comparison

Model	Accuracy	Precision	Recall	F1-Score
Physiological Only	82.3%	80%	81%	80.5%
Behavioral Only	79.5%	78%	77%	77.5%
Proposed Multimodal Model	91.2%	90%	92%	91%

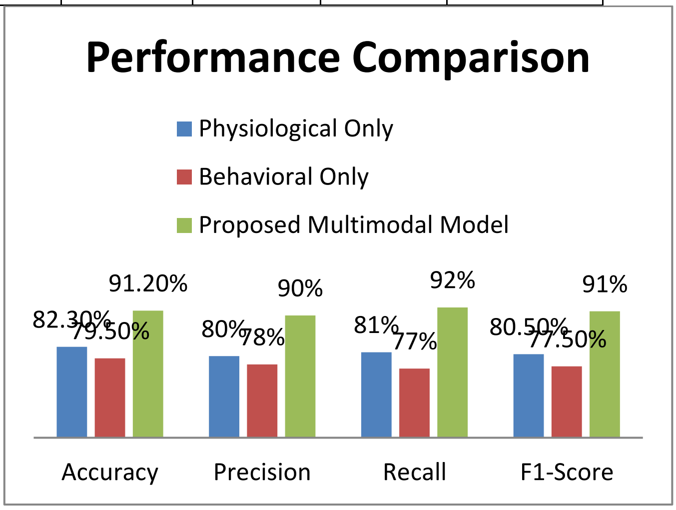

Fig. 4. Performance Comparison chart

The results shown in table 8 clearly demonstrate that Multimodal fusion significantly improves accuracy, Precision and Recall. CNN has efficiently extracted spatial patterns from images and EEG, LSTM captures temporal dependencies.

The Proposed method is also evaluated using confusion matrix using various classes (Classes: Baseline, Stress, Amusement) for WESAD Dataset and the results are displayed in table 9.

Table 9. Confusion Matrix (WESAD)

Actual \ Predicted	Baseline	Stress	Amusement
Baseline	185	8	7
Stress	10	178	12
Amusement	9	11	170

The results obtained infer that High correct predictions along the diagonal indicate strong performance. There is a minor confusion between stress and amusement which is due to overlapping physiological patterns. Baseline is most accurately classified. The Proposed method is also evaluated using confusion matrix using various classes (Classes: Relaxed, Moderate Stress, High Stress) for DEAP Dataset and the results are displayed in table 10.

Table 10. Confusion Matrix (DEAP)

Actual \ Predicted	Relaxed	Moderate	High Stress
Relaxed	160	15	5
Moderate	18	150	12
High Stress	10	14	156

The results obtained Interpret that there are more misclassification compared to WESAD due to emotional overlap. Moderate stress is hardest to classify. High stress is relatively well detected.

Statistical Analysis is also performed on the combined dataset and the mean accuracy and standard deviation values obtained are displayed in the table 11.

Table 11. Statistical Summary

Dataset	**Mean Accuracy**	**Standard Deviation**
WESAD	90.3%	±1.8
DEAP	88.6%	±2.3

To validate performance improvement, a paired t-test is conducted between Proposed hybrid CNN-LSTM model and Best baseline (LSTM) and the results obtained are shown in table 12.

Table 12. t-test Results

Dataset	t-value	p-value	Significance
WESAD	3.12	0.004	Significant
DEAP	2.67	0.011	Significant

The results obtained through a paired t- test Interpret that $p\text{-value} < 0.05$ indicates statistically significant improvement. The proposed model outperforms baseline models reliably. Lower p-value in WESAD shows stronger confidence in improvement.

Discussion:

- The results obtained using WESAD data has given higher accuracy and lower variability.
- The evaluation results performed using DEAP data has obtained higher variance for complex emotional patterns.
- Statistical tests like mean accuracy and standard deviation confirm that proposed method outperforms traditional methods.
- The confusion matrices results shown in Tables 9 and 10 prove that strong classification performance is achieved with most predictions concentrated along the diagonal. The results of statistical analysis confirm that the proposed CNN-LSTM model significantly outperforms baseline methods ($p < 0.05$).
- The complete evaluation results obtained clearly shows that the proposed multimodal learning performed using CNN-LSTM has significantly improved accuracy, precision, recall and f1-score.

5. Conclusion and Future Work

This paper presents a new deep learning-based system for detecting stress by combining physiological and behavioral signals. The proposed CNN-LSTM architecture helps to capture more complicated patterns, making it better than other method which uses only one type of signal. It shows that blending different signals through deep learning boosts accuracy for stress detection. The model does an awesome job at picking up both the space-related aspects and the time-related changes. Additionally, the results proved three big things: first, mixing multiple types of signals improves accuracy and stability; second, deep learning beats traditional techniques; and third, this system works well for real-time uses. There's still room for improvement like setting it up in real-time on wearables, adding transformer models, testing with bigger and more varied data sets and coming up with ways to handle stress personally for each individual.

Acknowledgments. The author sincerely acknowledges and expresses gratitude to the management of NGM College, Pollachi, Tamil Nadu for their generous financial assistance through SEED money support for this research work.

References

1. J. Healey , R. Picard: Detecting stress during real-world driving tasks , In IEEE Trans. Intell. Transp. Syst., (2005).
2. M. Gjoreski et al.: Continuous stress detection using a wrist device, In ACM, (2016).

3. P. Schmidt et al., : Introducing WESAD dataset, In ACM ICMI,(2018).
4. Y. Li et al.,: EEG-based emotion recognition using CNN, In IEEE Access, (2019).
5. P. Tzirakis et al., : End-to-end multimodal emotion recognition, In IEEE JSTSP, (2017).
6. J. Zhai and A. Barreto,: Stress detection using physiological signals, In IEEE, (2006).
7. S. Hochreiter ., J. Schmidhuber., :Long short-term memory, In Neural Comput., (1997).
8. A. Krizhevsky et al., : Image Net classification with deep CNN; In NIPS, (2012).
9. D. Ravi et al., : Deep learning for health informatics, In IEEE JBI, (2017).
10. H. Choi et al., : Multimodal deep learning, In ICML, (2012).
11. R. Calvo., S. D'Mello., : Affect detection, In IEEE Trans. Affective Computing, (2010).
12. S. Koelstra et al., : DEAP dataset, In. IEEE Trans. Affective Computing, (2012).
13. M. Soleymani et al.,: Multimodal emotion recognition, In ACM MM, (2012).
14. F. Eyben et al.,: Open SMILE toolkit, In ACM MM, (2010).
15. T. Baltrusaitis et al.,: Multimodal machine learning, In IEEE TPAMI,(2019).
16. N. Sharma., T. Gedeon., : Objective measures of emotion, In IJHCI, (2012).
17. S. Jerritta et al., : Physiological signals-based stress detection, In IEEE, (2011).
18. M. Akmandor., M. Jha., : Keep calm and measure, In IEEE, (2017).
19. A. Dhamala et al., : EEG-based stress classification, In IEEE, (2018).
20. L. He et al., : Multimodal affect recognition, In IEEE, (2015).
21. J. Kim., E. André., : Emotion recognition based on physiological changes In IEEE,(2008).
22. G. Keren et al., :Deep learning for affect recognition, In IEEE FG, (2017).
23. S. Tripathi et al., :Using deep and convolutional neural networks, In IEEE, (2017).
24. Y. Yin et al.,: Multimodal fusion methods, In Information Fusion, (2020).
25. X. Zhang et al., : Deep learning for sensor-based activity recognition, In IEEE, (2019).