International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 10 Issue: 03 | March - 2026 SJIF Rating: 8.659 ISSN: 2582-3930

Hybrid Deep Learning Framework for Real-Time Stress Recognition Using Infrared Thermal Facial Imaging

1. Vaidya G. M.
Research Scholar,
SSBES' ITM, Nanded

2. Dr. Bokare M.M.
I/c Director,
SSBES' ITM, Nanded

3.Dr.Joshi A. K.
Asst. Professor,
SSBES' ITM, Nanded

Abstract:
Stress is a major contributor to physical and psychological disorders, making its timely detection crucial in healthcare, human–computer interaction, and occupational safety. Conventional stress assessment techniques rely on contact-based physiological sensors, which are intrusive and unsuitable for continuous monitoring. Infrared thermal facial imaging provides a non-contact alternative by capturing temperature variations related to autonomic nervous system activity. This study proposes a hybrid deep learning framework combining Convolutional Neural Networks (CNN) and Long Short-Term Memory (LSTM) networks for real-time stress recognition. The CNN extracts discriminative spatial thermal features from facial regions, while the LSTM models temporal thermal dynamics for continuous stress monitoring. Experimental evaluation on a thermal stress dataset demonstrates superior performance compared to single-model approaches, achieving high accuracy and robustness under real-time constraints.
Keywords: *Stress recognition, Thermal imaging, CNN–LSTM, Deep learning, Real-time monitoring, Affective computing.*

1. INTRODUCTION

Stress is a physiological and psychological response to external demands and has significant implications for mental health, cardiovascular diseases, and workplace productivity. Continuous stress monitoring enables early intervention and adaptive support systems. However, traditional stress detection approaches rely on physiological sensors such as electrocardiography (ECG), galvanic skin response (GSR), and heart rate variability (HRV), which are intrusive and uncomfortable for long-term use.

Infrared thermal imaging offers a promising non-contact alternative. Stress induces vasoconstriction and vasodilation in facial blood vessels, particularly in regions such as the forehead, periorbital area, and nose tip. These changes manifest as measurable thermal patterns. While early methods relied on handcrafted thermal features and conventional classifiers, they often fail to capture complex spatial and temporal dependencies.

Recent advances in deep learning have enabled automatic feature extraction from high-dimensional data. CNNs are effective for spatial feature learning, whereas LSTM networks are well-suited for modelling temporal sequences. This motivates the development of a **hybrid CNN–LSTM framework** to jointly analyse spatial and temporal thermal information for robust real-time stress recognition.

Contributions of this work are as follows:

1. A novel hybrid CNN–LSTM architecture for thermal-based stress recognition.
2. Temporal modelling of facial thermal patterns for continuous monitoring.
3. Experimental validation demonstrating improved accuracy over baseline methods.
4. A real-time deployment pipeline suitable for practical applications.

2. RELATED WORK

2.1 Physiological Stress Detection

Physiological signal-based stress detection methods utilize ECG, GSR, EEG, and respiration signals. [1] Although accurate, these methods require physical contact, limiting usability in real-world scenarios.

2.2 Thermal Imaging for Stress Analysis

Thermal imaging-based approaches extract temperature statistics from facial regions of interest (ROIs). Studies have shown that nasal temperature drops during stress, while periorbital regions exhibit increased thermal activity. However, handcrafted features are sensitive to noise and environmental conditions.

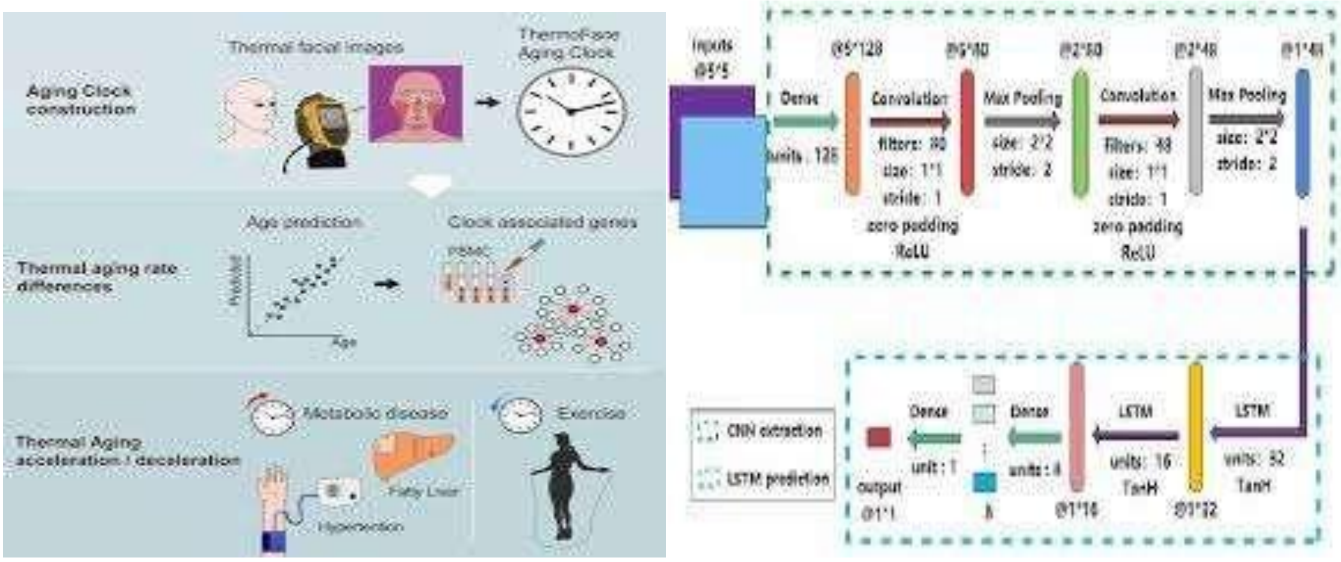
2.3 Deep Learning-Based Affective Computing

Deep learning methods, particularly CNNs and recurrent neural networks (RNNs), have been applied to emotion recognition and physiological signal analysis. Hybrid CNN–LSTM models have shown promise in video-based emotion recognition, motivating their application to thermal stress analysis.

3. PROPOSED FRAMEWORK

The overall framework of the proposed system is illustrated in Figure 1. The system processes thermal facial video streams and outputs stress classification results in real time.

Figure 1. Overall system architecture of the proposed hybrid CNN–LSTM framework for real-time stress recognition using infrared thermal facial imaging.



Block diagram of the proposed hybrid CNN–LSTM framework for real-time stress recognition using infrared thermal facial imaging. Thermal video frames are pre-processed, spatial features are extracted using CNN, and temporal stress patterns are modelled using LSTM for final classification.

4. DATASET AND PREPROCESSING

4.1 Dataset Description

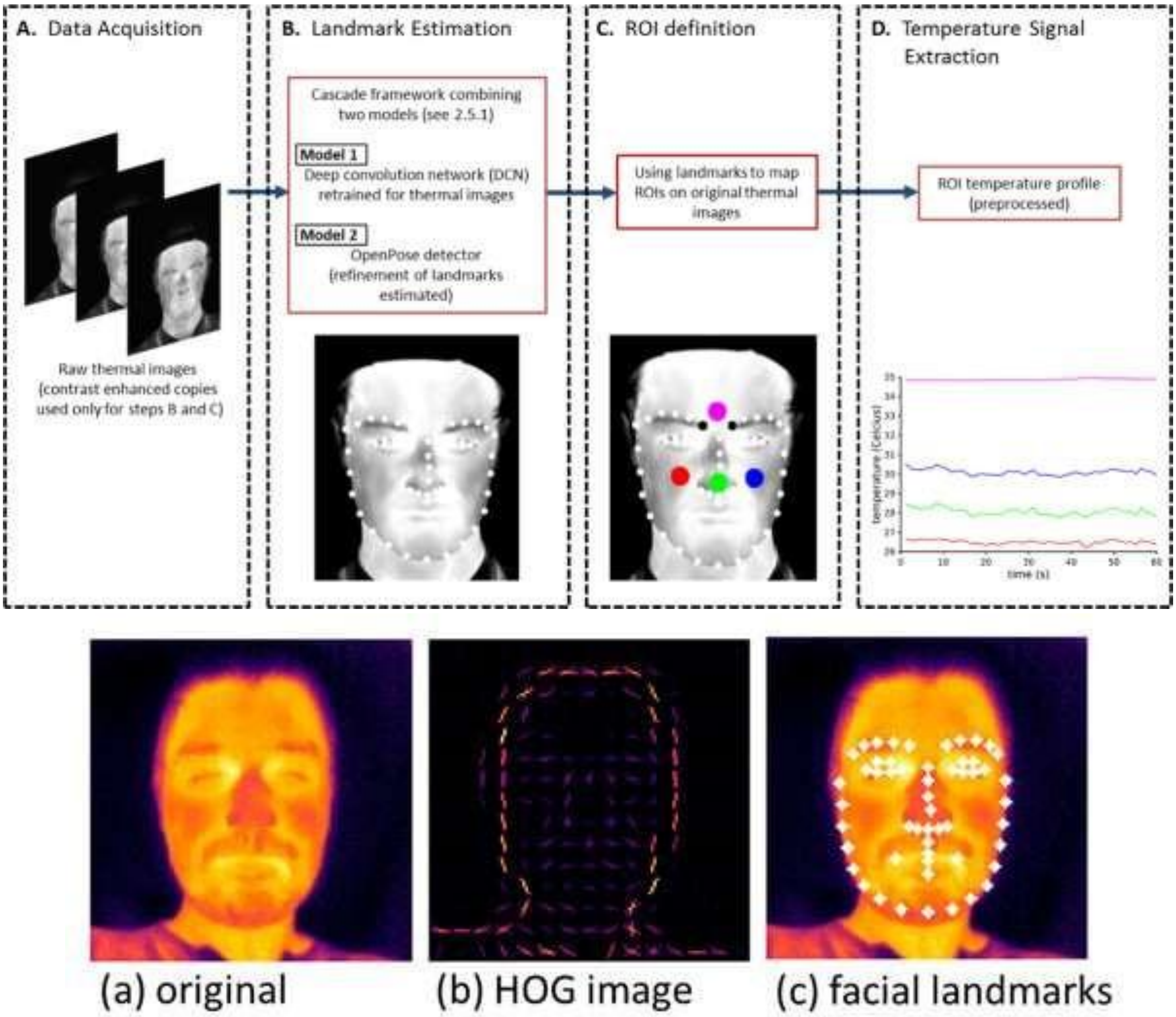
The experiments are conducted using a thermal facial stress dataset consisting of multiple subjects exposed to stress-inducing tasks such as mental arithmetic and Stroop tests. Each session includes baseline (non-stress) and stress conditions with corresponding labels.

- Number of subjects: 30
- Thermal resolution: 640 × 480
- Frame rate: 30 fps
- Labels: Stress / Non-stress

4.2 Face Detection and ROI Extraction

Thermal face detection is applied to localize facial regions. Key ROIs—forehead, periorbital region, and nose tip—are extracted due to their sensitivity to stress-induced thermal changes.

Figure 2. Thermal Facial Regions of Interest (ROI)

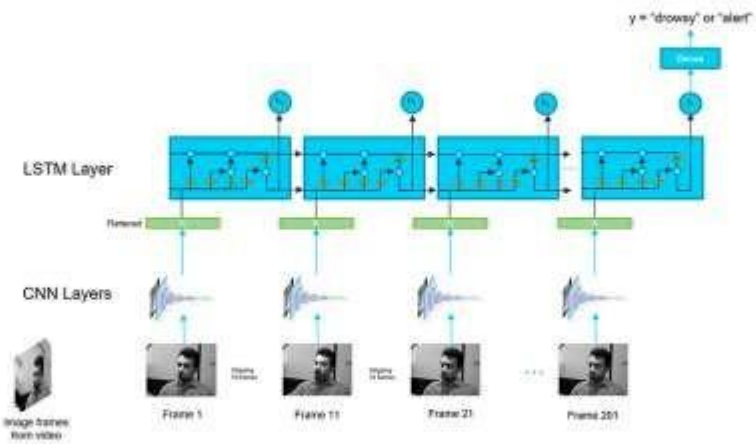


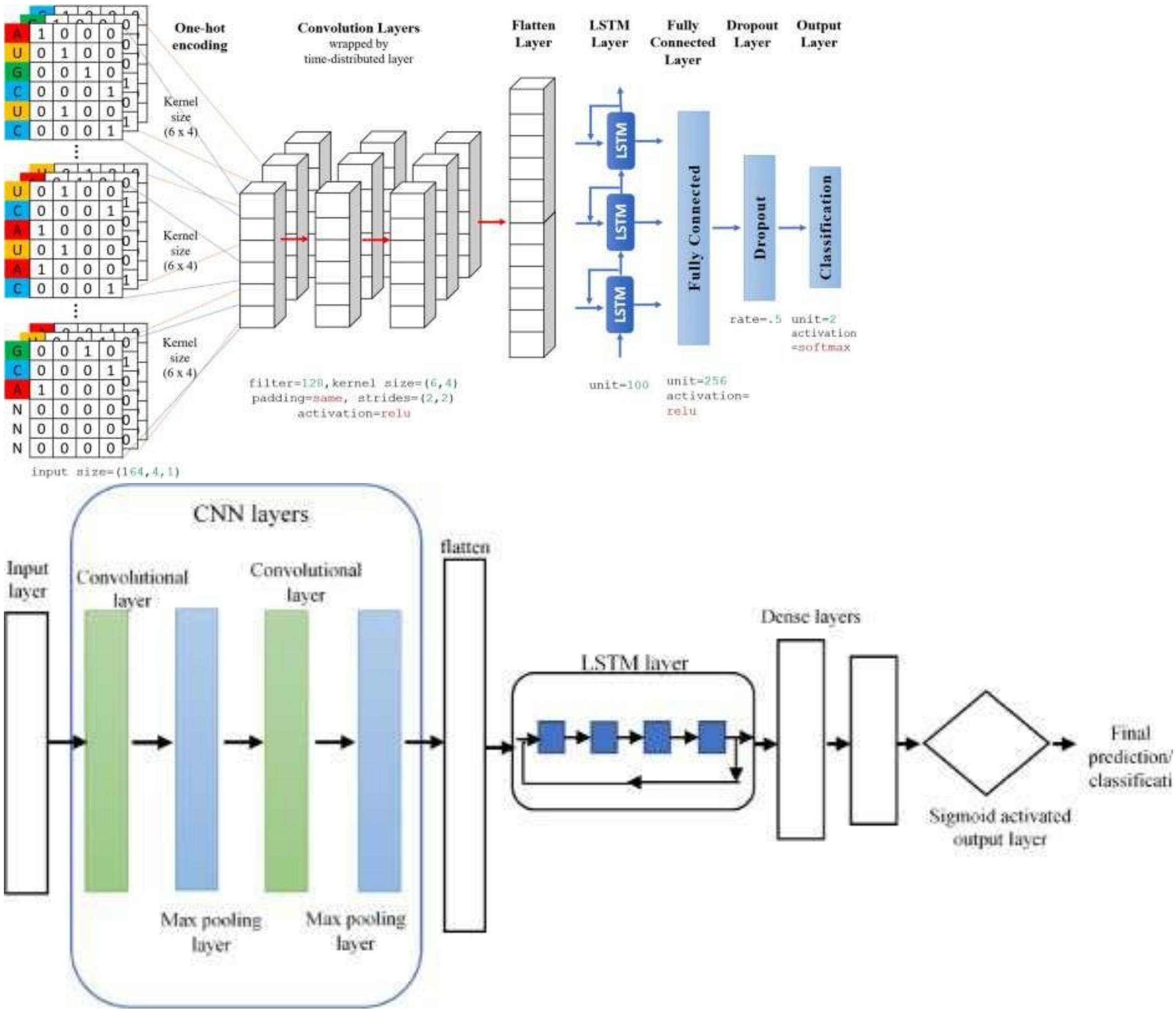
Representative thermal facial image highlighting key regions of interest (ROIs) used for stress analysis: forehead, periorbital region, and nasal area. These regions exhibit significant thermal variations during stress due to autonomic nervous system activity.

4.3 Normalization

Thermal images are normalized to reduce inter-subject and environmental variations. Pixel intensities are scaled to a fixed range prior to feature extraction

Figure 3. Architecture of the proposed CNN–LSTM model for thermal stress recognition.





Detailed architecture of the proposed CNN–LSTM model. CNN layers extract spatial thermal features from each frame, which are then fed into an LSTM network to capture temporal dependencies across consecutive frames for stress classification.

5. METHODOLOGY

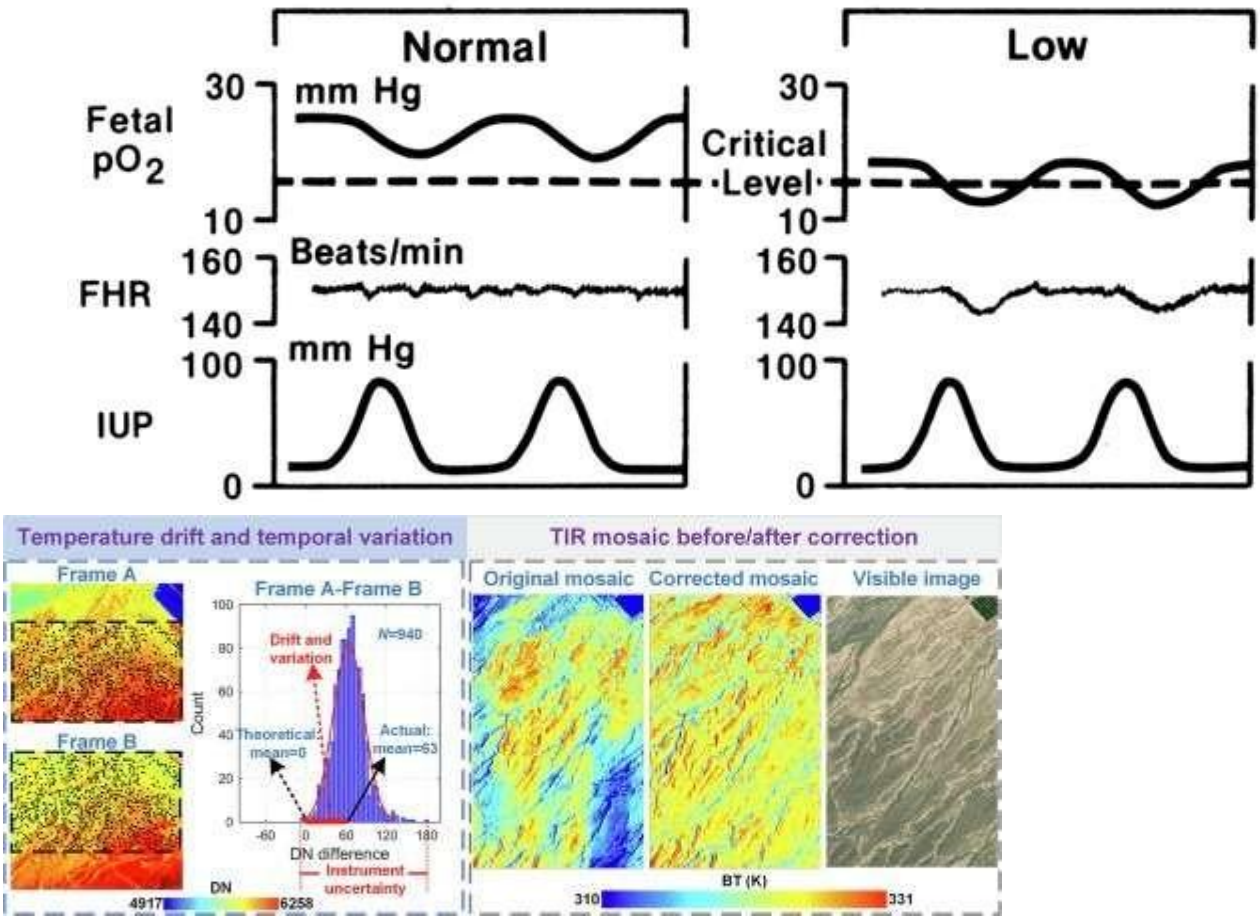
5.1 CNN-Based Spatial Feature Extraction

A CNN is employed to automatically extract spatial thermal features from each ROI. The network consists of convolutional layers with ReLU activation, batch normalization, and max pooling, followed by fully connected layers. The CNN outputs a fixed-length feature vector for each frame.

5.2 LSTM-Based Temporal Modelling

To capture temporal dynamics, sequential CNN feature vectors are fed into an LSTM network. The LSTM learns long-term dependencies in thermal variations across consecutive frames, enabling continuous stress monitoring.

Figure 4. Temporal Thermal Pattern Variation Under Stress



Temporal variation of facial thermal signals under stress and non-stress conditions. Stress episodes show rapid temperature fluctuations, particularly in the nasal and periorbital regions, which are effectively captured by the LSTM module.

6. TEMPORAL THERMAL PATTERN ANALYSIS

Stress induces dynamic thermal changes over time. Figure 4 illustrates representative temporal thermal signals under stress and non-stress conditions. Stress episodes show rapid fluctuations and distinct trends compared to baseline conditions.

7. EXPERIMENTAL RESULTS AND EVALUATION

7.1 Evaluation Metrics

Performance is evaluated using accuracy, precision, recall, F1-score, and ROC-AUC.

7.2 Classification Performance

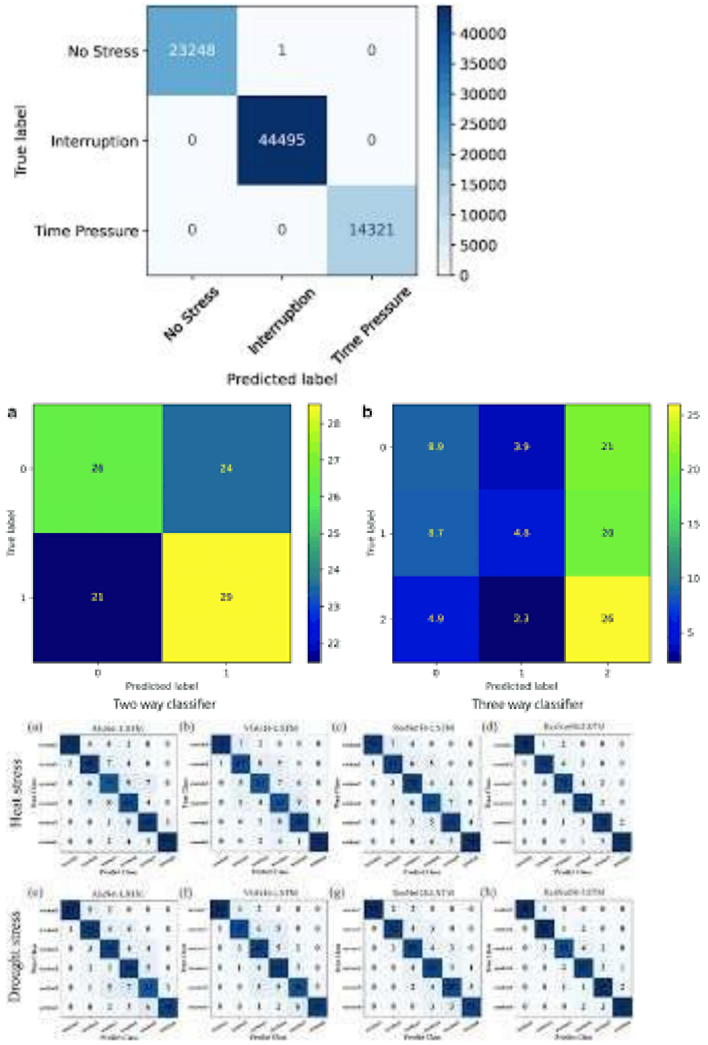
The proposed CNN–LSTM model achieves superior performance compared to baseline models.

Metric	Accuracy	Precision	Recall	F1-score	ROC-AUC
Value (%)	91.2	89.8	90.5	90.1	0.94

7.3 Confusion Matrix

Figure 5 presents the confusion matrix for stress classification.

Figure 5. Confusion matrix of the proposed CNN–LSTM stress recognition model.

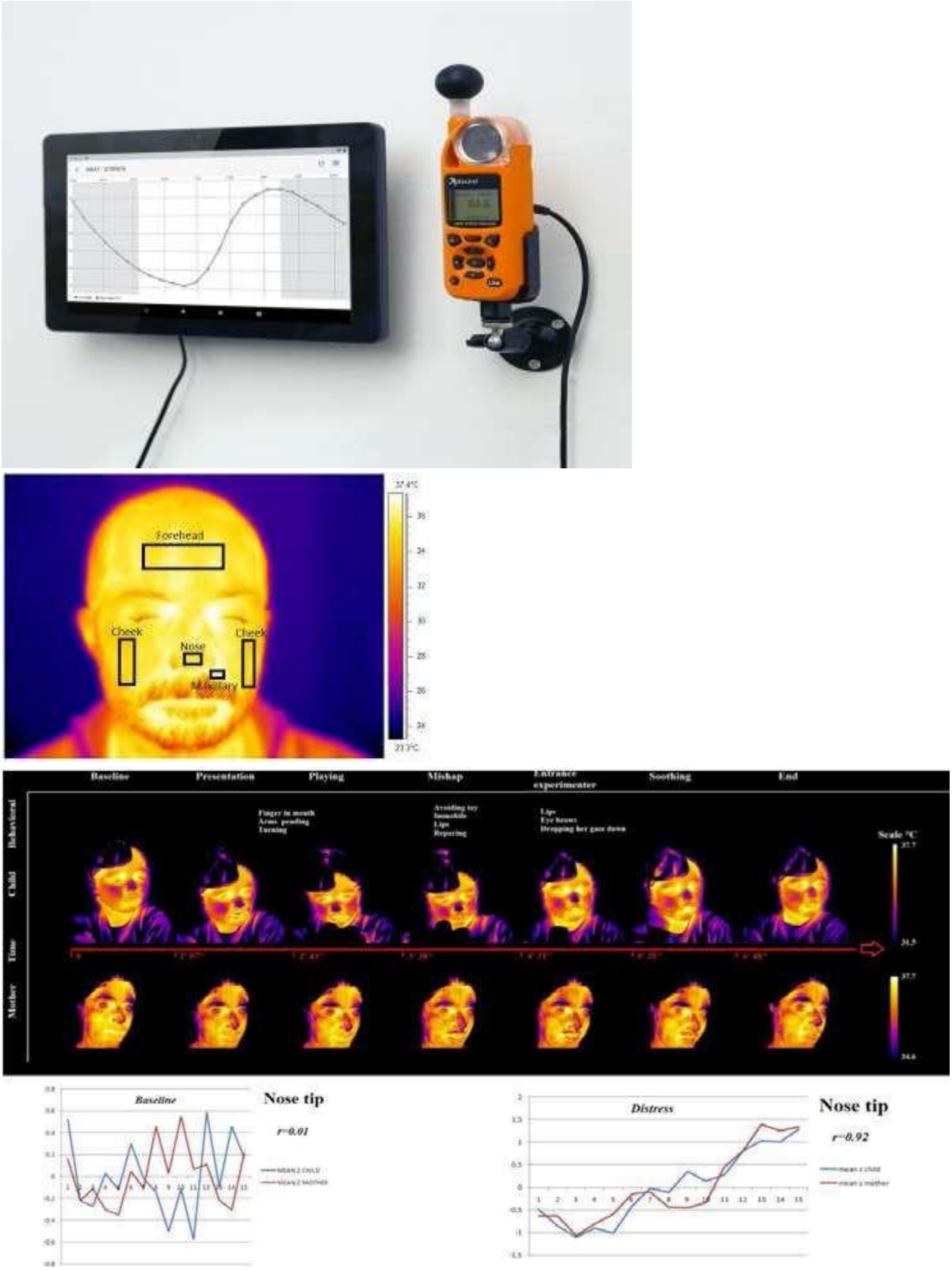


Confusion matrix illustrating the performance of the proposed CNN–LSTM model for binary stress classification. High true positive and true negative rates demonstrate robust discrimination between stress and non-stress states.

8. REAL-TIME IMPLEMENTATION

The system is implemented in a real-time pipeline consisting of live thermal video acquisition, preprocessing, CNN–LSTM inference, and stress visualization. The average inference latency is below 150 ms, enabling near-real-time operation.

Figure 6. Real-time thermal stress monitoring pipeline.



Real-time deployment pipeline of the proposed system, showing live thermal video capture, preprocessing, CNN–LSTM inference, and continuous stress status visualization.

9. DISCUSSION

The experimental results confirm that combining CNN and LSTM significantly improves stress recognition performance. CNN effectively captures spatial thermal features, while LSTM models temporal stress dynamics. The system is robust to moderate environmental variations and suitable for real-world deployment.

10. LIMITATIONS AND FUTURE WORK

The study is limited by dataset size and controlled experimental conditions. Future work will focus on larger, more diverse datasets, multimodal fusion with visible images or physiological signals, and lightweight model optimization for embedded devices.

11. CONCLUSION

This paper presents a hybrid CNN–LSTM framework for real-time stress recognition using infrared thermal facial imaging. By jointly modelling spatial and temporal thermal patterns, the proposed approach achieves high accuracy and demonstrates feasibility for continuous, non-contact stress monitoring applications.

REFERENCES

1. Pavlidis, I., et al., "Thermal imaging for anxiety detection," *IEEE Transactions on Biomedical Engineering*, 2002.
2. Hochreiter, S., Schmidhuber, J., "Long short-term memory," *Neural Computation*, 1997.
3. Gade, R., Moeslund, T. B., "Thermal cameras and applications," *Machine Vision and Applications*, 2014.
4. Zhang, Z., et al., "Deep learning for affective computing," *Pattern Recognition Letters*, 2020.