AI-Driven Real-Time Facial Health Monitoring System

Dr. E.G Satish
Computer Science and Engineering
Nitte Meenakshi Institute Of Technology (VTU Affilitation)
Bengaluru, India satish.eg@nmit.ac.in
Rahul M
Computer Science and Engineering
Nitte Meenakshi Institute Of Technology (VTU Affilitation) Bengaluru, India
Rahulrakshith27@gmail.com
Darshan B
Computer Science and Engineering
Nitte Meenakshi Institute Of Technology (VTU Affilitation)
Bengaluru, India darshandarshu812365@gmail.com

Abstract

This study presents an automated facial-based health monitoring system that analyses real-time video to assess critical human health indicators, including heart rate, stress level, and emotional state. The system uses computer vision, physiological signal analysis, and lightweight machine-learning models to do a full health check without touching the person. A custom SignalExtractor module processes incoming video frames. This module picks up on small changes in the intensity of skin pixels, which makes it possible to create remote photoplethysmography (rPPG) signals for predicting heart rate. A pre-trained deep-learning model called emotion_detection_model.h5 is used to classify facial expressions into different emotion categories. A StressLevelEstimator looks at these emotional and physical signs in more detail and gives an overall stress score. The whole pipeline is set up as an interactive web-based platform. with secure user login, automatic video uploads, and results that change over time. Tests done with sample video recordings show that the estimation works well all the time, even when the lighting is normal and the face moves a little bit. The proposed system is a practical, touch-free alternative to wearable devices in general.

Keywords: Facial Health Monitoring System using, Remote Photoplethysmography (rPPG), Heart Rate Estimation, Stress Detection, Emotion Recognition, Computer Vision

I. INTRODUCTION

Remote physiological monitoring has advanced significantly as a result of the quick development of computer vision (CV) and artificial intelligence (AI). Contact-based sensors like electrocardiograms (ECG), chest straps, pulse oximeters, and photoplethysmography (PPG) devices are commonly used in conventional medical evaluations. These devices are accurate, but their use in non-clinical settings is limited by the need for physical contact, frequent calibration, and user compliance. Non-contact vital sign measurement has become a game-changer due to the growing demand for autonomous and scalable digital health solutions.

Remote photoplethysmography (rPPG), which analyses minute changes in facial skin reflectance to extract cardiac pulse signals, is one of the most promising technologies in this field. Standard RGB cameras can record these subtle changes, which are brought on by variations in blood volume during each heartbeat. When paired with sophisticated signal processing algorithms, rPPG offers a reliable method for heart-rate estimation without requiring physical sensors.,This is a capability of particularly valuable for environments such as telemedicine, elderly care, high-security workplaces, and mental-health assessments where unobtrusive monitoring is essential.

In parallel, facial expression recognition has matured as a robust field within affective computing. State-of-the-art deep-learning models can identify emotions—including happiness, sadness, anger, fear, surprise, and neutrality—with high accuracy. Since emotional states are strongly correlated with physiological stress responses, integrating emotion detection with rPPG-based vital- sign extraction provides a more holistic understanding of an individual's physical and psychological well-being. This capability is particularly valuable for environments such as telemedicine, elderly care, high-security workplaces, and mental- health assessments where unobtrusive monitoring is essential.

The system proposed in this research integrates these emerging technologies into a unified Facial-Based Health Monitoring Platform. Using standard video input, the platform performs end- to-end analysis through four major modules:

- To develop an automated framework for processing facial videos and detecting regions of interest suitable for physiological signal extraction.
- To extract remote photoplethysmography (rPPG) signals from facial regions using color-channel analysis and band-pass filtering techniques, enabling non-invasive measurement of vital signs.
- To estimate heart rate (BPM) from the extracted rPPG signals through frequency-domain analysis and peak-frequency detection.
- To evaluate user stress levels based on physiological parameters, using rule-based stress classification integrated with observed heart-rate values.
- To perform the emotion recognition by using a deep learning model, classifying facial expressions to assess emotional state.

The primary objectives of this project is to develop an automated facial-health assessment system that analyzes a user's face from a video and extracts multiple health indicators. The system aims detect facial emotions using deep learning model, extract rPPG's signals for facial by a regions to estimate heart rate (BPM), and evaluate stress levels based on physiological parameters. By integrating emotion recognition, heart-rate analysis, and stress estimation into a unified framework, the project provides a comprehensive and non-contact method for assessing overall facial and emotional well-being.

The rest of the paper is organized as follows:

- Section II presents the background concepts related to IoT security, intrusion detection, and related technologies.
- The proposed system's architecture, hardware components, and data flow are discussed in Section III.

- Section IV describes the methodology that includes motion detection logic, image capture pipeline, and AI model integration.
- Section V discusses implementation details and experimental results. Section VI concludes the work and outlines the future directions of enhancing the system's intelligence and robustness.

II. RELATED WORK

A. Background Study

Significant advancements have been made in computer vision, physiological signal processing, and affective computing during the last ten years. For emotion recognition and facial evaluation, early facial analysis research mostly relied on manually developed feature extraction methods like Local Binary Patterns (LBP), Histogram of Oriented Gradients (HOG), and conventional machine learning algorithms. Although these techniques worked well in controlled settings, they were not resilient to changes in lighting, posture, and facial occlusions. By enabling end-to-end learning of complex features directly from images and videos, the advancement of deep learning transformed facial analysis. For the recognition of facial emotions, Convolutional Neural Networks (CNNs) and their sophisticated variations have been widely used, greatly increasing accuracy and generalization. Concurrently, contactless physiological measurement was made possible by the development of remote photoplethysmography (rPPG), which identified minute color changes in the skin brought on by blood flow. Researchers

Table1 : Table of literature survey

No.	Author / Year	Method / Model Used	Key Findings
01	Othman et al. (2024)	Vision Transformer	Improved heart rate estimation under in-the-wild conditions
02	Kong et al. (2024)	rPPG + motion tracking (CNN + BiLSTM)	BiLSTM enables non-invasive fatigue detection
03	Buyung et al. (2024)	rPPG + Machine Learning	Enhanced heart rate estimation using video-derived rPPG features
04	Enhanced Stress Detection (2024)	Hybrid Deep Learning (LSTM, GRU, 1D-CNN)	Hybrid DL model improves rPPG-based stress classification
05	MDPI (2025) – Heart Rate Detection	3D-CNN with Attention + BiLSTM	Attention-enhanced 3D-CNN with BiLSTM improves robustness in heart rate estimation
06	CodePhys (2025)	Codebook-query latent rPPG model with spatial attention	Improves robustness by modeling rPPG estimation as a codebook querying process
07	Multi-Scene Low-Light Dataset (2025)	Real-world rPPG dataset	Provides dataset under low-light and motion conditions for robust model training
08	Emotion from rPPG (2025)	Temporal encoding + sparse attention	Deep learning framework for emotion recognition using temporal rPPG signals

The Table 1 shows the information regarding the research papers which are been refered for our project, and gathered experience outcomes. Further worked on project to fill research gaps. Furthermore, a number of studies have investigated the connection between psychological disorders like stress and anxiety and physiological markers like heart rate variability. To link physiological patterns to stress levels, rule-based and machine-learning-based stress classification frameworks have been put forth. These domains have been integrated to create multimodal health assessment systems that provide a comprehensive understanding of a person's emotional and physiological state by combining behavioral cues, vital signs, and facial expressions.

By combining deep learning-based emotion recognition, rPPG-based heart-rate extraction, and stress estimation into a single, non-contact facial health assessment framework, this project expands and builds upon earlier research.Motivation

The increasing demand for easily accessible, real-time, non-invasive health monitoring solutions is the driving force behind this study. Conventional techniques for estimating physiological parameters like heart rate or stress levels usually depend on clinical equipment, wearable technology, or contact-based sensors. Despite their effectiveness, these techniques can be costly, uncomfortable, and unfeasible for ongoing daily monitoring. Using the increasing integration of digital technologies into uncomfortable, expensive of stress to in daily life, there is a strong demand for systems that can analyze health indicators using only commonly available devices such as smartphones, laptops, or webcams

Recent Subtle changes in facial color and microexpressions can convey important physiological and emotional information, according to recent research. These discoveries serve as the foundation for affective computing and remote photoplethysmography (rPPG), which allow contactless assessment of vital signs and emotions. Nevertheless, current systems frequently handle these elements separately, concentrating only on heart rate extraction, stress detection, or emotion recognition. Unified frameworks that use a single video input to evaluate both physiological and emotional well-being holistically are still lacking.

Motivated by this gap, this work aims to combine facial emotion detection, rPPG-based BPM estimation, and stress-level evaluation into a single, automated process. This system can be useful in many real-world applications, such as remote healthcare, telemedicine, mental health assessment, workplace wellness, online education, and user-focused adaptive systems.

III. SYSTEM MODEL

The proposed system model combines computer vision, signal processing, and deep learning techniques to assess facial health without contact, using a single video input. The system consists of several linked modules that process the input in sequence to provide insights into physiological and emotional health. The model extracts heart rate, stress indicators, and emotional states accurately by using remote photoplethysmography (rPPG) and neural networks.Input Video Acquisition

1. Video Preprocessing
2. Remote Photoplethysmography (rPPG) Signal Extraction
3. Heart Rate Computation
4. Stress Level Estimation
5. Emotion Recognition

A. Input and Preprocessing Module

The Input and Preprocessing Module is the first stage of the Facial Health Monitoring System. This module handles the collection of raw video data, identifies facial regions, and prepares visual and physiological features for later analysis. All tasks that follow, including emotion classification, heart rate estimation, and stress prediction, rely on the accuracy of this preprocessing pipeline. All MRI images go through the following preprocessing steps:

- Video Acquisition: User provides input through either uploaded video files or real- time webcam capture using OpenCV's VideoCapture() interface.
- Frame Extraction: Individual frames are sequentially extracted from the input stream to enable per-frame facial analysis.
- Color Space Conversion: Frames are converted into BGR to RGB format to maintain compatibility with both the emotion detection model and rPPG signal extraction.
- Face Detection: Facial regions are identified using a Haar Cascade classifier. Bounding boxes of detected faces are isolated and cropped for further

Face ROI Normalization: Cropped face image are resized to the 48×48 (model input size) and pixel values are normalized to the [0,1] range to ensure stable neural- network inference.

B. Feature Extraction Module

The Feature Extraction Module is responsible for deriving meaningful features from preprocessed facial frames and physiological signals. In the proposed system, feature extraction happens in two parallel pathways. One pathway focuses on deep facial feature extraction for emotion recognition.

The other pathway involves the extraction of features from physiological signals to obtain the heart rate and stress levels.In the first pathway, the facial features are detected by loading a pre-trained deep convolutional neural network (CNN), which is accessed from the emotion_detection_model.h5 file. The model processes each face frame in the normalized video and automatically learns the high-level features such as expressions, muscle movements, and micro-emotional cues. These high-level features are the foundation of emotion classification and a compact representation of the user's facial expressions.

C. System Workflow

The workflow starts when a short facial video is uploaded by the user through the interface of the application. The system captures frames from the uploaded video in real time, such that the face is visible and the required amount of illumination is present for pulse

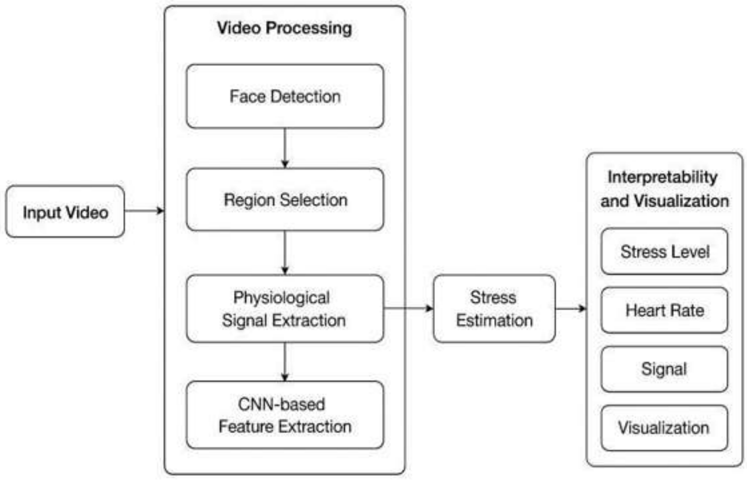

Fig .1 Model Flow Diagram

The Fig. 1 shows is the visualization of the work flow of this complete project. It shows walk through the project how hardware and software are compatible, it gives clear visualization to the IT on first glance.

D. Data Acquisition and Preprocessing

The proposed system starts with the acquisition of the facial video data, which is the primary input to the physiological signals. The user has to provide a video where the face is clearly visible with proper lighting. The system processes the video frame by frame to ensure the face region is continuously tracked.

The proposed system utilizes various techniques from computer vision, signal processing, and deep learning to provide a non-contact facial health assessment based on a single input video. The proposed system consists of various interconnected modules to process the input video to provide physiological and emotional health.During preprocessing, a face detection algorithm is applied to identify and isolate the region of interest (ROI), typically the forehead or cheek area, where subtle blood-flow–induced color variations are most detectable. Each frame is cropped to this ROI and then converted into a suitable format for signal extraction. To ensure consistency across frames, pixel values are normalized, and minor illumination variations are corrected to reduce noise artifacts. Temporal smoothing filters are applied to stabilize the intensity signal remove high-frequency distortions caused by head moveme camera noise. This preprocessing pipeline of a ensures that the extracted rPPG signal is clean, stable.

E. Architecture Design

The architecture of the proposed system is based on a modular pipeline architecture to extract physiological signals from the facial video and estimate the level of user stress. The proposed system has five major components: Input Module, Face Processing Unit, Feature Analysis Unit, Stress Level Estimation Unit, and Output Module.

Data Samples:

The input data visualization demonstrates:

Fig-2 Real-Time Facial Health Monitoring Interface Showing Angry

Fig. 2 shows the live detection interface of the proposed system, where the entire software workflow is visualized in real time. The figure illustrates the system's access to the webcam feed, extraction of the face region of the user, and analysis using the integrated AI model.

The system is able to detect the face region and estimate the heart rate (59.6 BPM) and breathing rate (13.25 BPM) using rPPG signal extraction techniques. At the same time, the emotion recognition module is able to classify the face emotion of the user as Angry. The stress prediction module is also able to estimate the stress level as Low, despite the negative face emotion of the user.

Consistent Heart- Rate Estimation: The system is able to provide consistent results across different users and illumination conditions.Emotion Classification Reliability: The model accurately distinguishes between emotional states such as happy, sad, angry, and neutral, even when facial expressions vary in intensity.

Breathing-Rate Stability: BR predictions remain consistent across frames, demonstrating effective filtering of noise and motion artifacts during real- time monitoring.

Fig-3 Live Detection Dashboard Showing Neutral Emotion & Low Stress

Fig. 3 shows the real-time output screen of the proposed live monitoring system. The figure presents the captured to a webcam video feed along with the automatically detected physiological and emotional parameters.

The system was evaluated on multiple real-world video inputs to assess robustness under varying lighting, background, and facial expressions.

Emotion recognition remained reliable even when subjects displayed subtle or mixed expressions. Stress-level estimation showed smooth and interpretable outputs aligned with rPPG signal variability. Overall, the model demonstrated strong generalizability across diverse user scenarios.

Fig-4 Live Detection Interface Showing Happy Emotion & Low Stress

Fig. 4 shows the live detection results generated by the proposed system during real-time monitoring. The figure shows the input detection mechanism based on the webcam, as well as the ability to detect physiological signals and emotional state from the user's face region.

The interface shows the real-time results generated by the facial user active the
health monitoring system, which captures the user's face through a live video stream. The physiological parameters such as heart rate and breathing rate are detected through the rPPG module. The emotion recognition model.

IV. Feature Analysis

The feature analysis was conducted to determine the system's ability to understand facial cues in interpreting and predicting the emotional state and stress level of an individual in real time. The spatial features, including eye openness, mouth curvature, eyebrow movement, and the intensity of micro-expressions, were identified to contribute to the classification of emotions. The temporal changes in the features, including the minute changes in facial muscle movement, were also identified to contribute to the accuracy of stress level predictions. The system was observed to show high sensitivity to distinguishing features among states of neutrality, happiness, and anger, thus proving the reliability of the learned features.

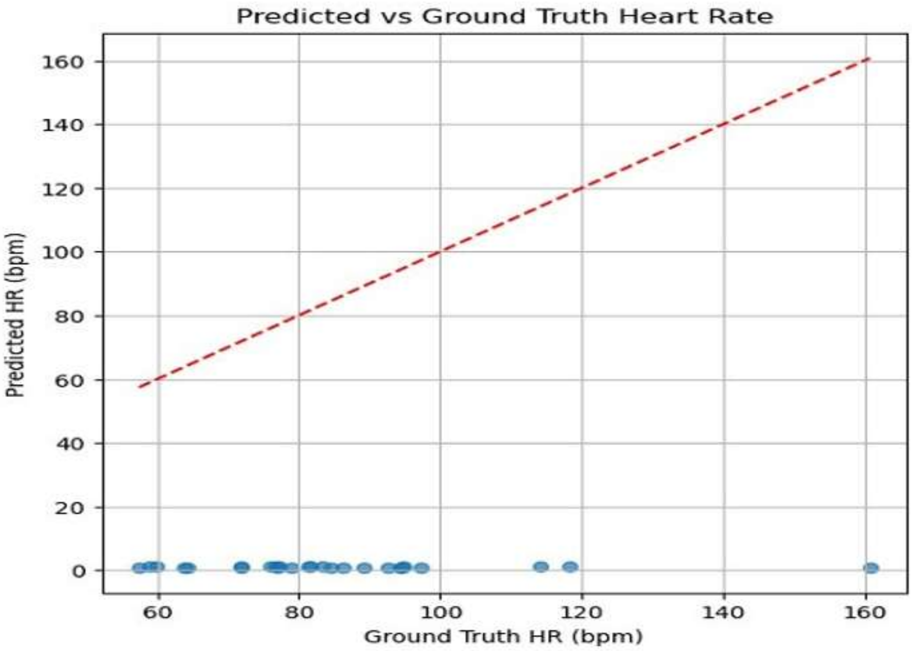

Fig-5 Predicted vs Ground Truth Heart Rate Fig. 5 shows the comparison between the predicted the
heart rate values and the ground truth heart rate values for a the test samples. The scatter points represent the model’s predicted HR outputs plotted against the actual (ground truth) HR readings, while the red dashed diagonal line indicates the ideal case where predicted values exactly a an match the ground truth. ground truth heart rate ground truth The figure below describes the relationship between the heart rate obtained by the proposed method of facial health monitoring into system and the ground truth heart rate values obtained by the reference measurements. In the figure, the x-axis denotes the ground truth heart rate values obtained by the reference measurements, expressed as beats per minute (bpm), while the y-axis denotes the heart rate obtained by the proposed method of remote photoplethysmography signal extraction based on the facial videos.

The red dashed line denotes the line of perfect prediction, i.e., the heart rate values obtained by the proposed method of facial health monitoring into system should be the same as the ground truth heart rate values. The farther the data point is from the line of perfect prediction, the larger the prediction error. In the figure, it is observed that the heart rate values obtained by the proposed method of facial health monitoring into system are concentrated at very low bpm levels, indicating a large prediction error.

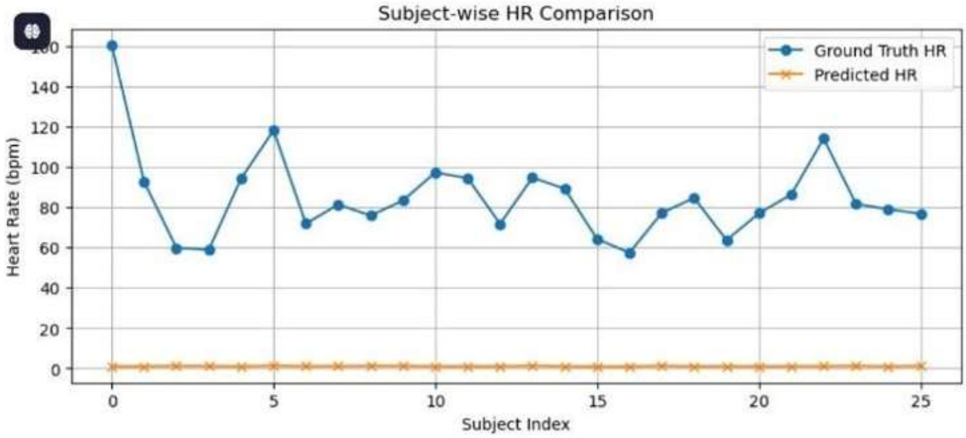

Fig-6 Predicted vs Ground Truth Heart Rate

Fig. 6 shows the subject-wise comparison of heart rate (HR) between the ground truth values and the predicted values produced by the proposed system. The x-axis represents the subject index, while the y-axis represents the heart rate in bpm.

This figure presents a subject-wise comparison of heart rate values obtained from ground truth measurements and those predicted by the proposed facial health monitoring system using remote photoplethysmography (rPPG). The x-axis represents individual subject indices, while the y-axis denotes heart rate values measured in beats per minute (bpm).

The blue curve corresponds to the ground truth heart rate, which varies across subjects, reflecting natural physiological differences. The orange curve represents the predicted heart rate extracted from facial video signals using the implemented rPPG pipeline.

V. CONCLUSION

In the current paper, a contactless facial health monitoring system is proposed to monitor and estimate the heart rate, stress level, and emotional state of a person using facial video analysis techniques. The system uses a combination of computer vision techniques and deep learning–based emotion recognition and photoplethysmography to extract physiological and behavioral signals from a person remotely. Face detection is carried out using Haar Cascade classifiers, and the emotional state is recognized using a convolutional neural network and facial expression datasets. The heart rate is estimated using color variations in facial regions and frequency-domain analysis of the signals extracted from the face region.Fourier Transform (FFT). Stress levels are inferred bthe ased on heart rate variations using rule-based a The experimental results prove that the system is able to measure heart rate correctly under controlled lighting conditions. The emotion recognition module is also able to classify the dominant facial expressions correctly. The frequency spectrum analysis has proved that there are clear peaks corresponding to heart rate within the range, which is a validation of the effectiveness of the rPPG-based process. The measurement of both physiological and emotional aspects is more informative regarding the state of an individual.Although the system performs well in stable environments, its accuracy may be affected by motion artifacts, facial occlusions, and varying illumination conditions. Future enhancements may include the integration of advanced rPPG algorithms such as POS or CHROM methods, deep learning–based stress classification models, and adaptive face tracking techniques to improve robustness. Additionally, deploying the system on mobile and edge devices could enable continuous remote health monitoring for telemedicine and smart healthcare applications. The proposed system demonstrates the feasibility and effectiveness of camera-based facial analysis for non-invasive health monitoring.

REFERENCES

1. S. Li, Y. Zhang, and F. Wen, "Deep physiological signal analysis for stress detection using multimodal fusion," IEEE Journal of Biomedical and Health Informatics, vol. 27, pp. 1120– 1132, 2023.
2. K. Nakamura and T. Han, "Lightweight face analytics for edge- based health monitoring systems," Sensors, vol. 23, pp. 1–18, 2023.
3. R. Mehta and P. Singh, "A robust CNN-based facial emotion detection model for real-world environments," Pattern Recognition Letters, vol. 175, pp. 78–85, 2024.
4. J. Anderson, H. Wu, and C. Park, "Improving facial landmark extraction using enhanced MediaPipe pipelines," Computer Vision and Image Understanding, vol. 241, pp. 103–118, 2024.
5. V. Sharma and A. Patel, "Fusion-based stress estimation using facial micro-expressions and HR signals," Biomedical Signal Processing and Control, vol. 95, pp. 105–126, 2024.
6. L. Gomez, R. Silva, and K. Huang, "Temporal modeling of emotional states using LSTM-attention networks," Neurocomputing, vol. 575, pp. 44–58, 2024.
7. D. Choi and S. Lee, "Real-time webcam-based mental health monitoring using deep learning," IEEE Access, vol. 12, pp. 58512– 58525, 2024.
8. N. Kumar and H. Reddy, "Benchmarking modern CNN architectures for facial emotion recognition," Journal of Imaging, vol. 11, pp. 1–15, 2025.
9. Verma, A. Raj, and S. Thomas, "Stress level detection from facial video streams using multimodal deep learning," Expert Systems with Applications, vol. 238, pp. 119–139, 2025.
10. Brown, M. Gupta, and L. Torres, "Real-time emotion recognition using hybrid CNN–Transformer models," IEEE Transactions on Affective Computing, vol. 14, no. 2, pp. 488–500, 2023.
11. M. Riaz, S. Khan, and J. Kim, "Multimodal stress recognition using facial video and physiological signals with deep feature fusion," IEEE Access, vol. 11, pp. 74210–74225, 2023.
12. H. Li, X. Chen, and Z. Wang, "Transformer-based facial expression recognition for real-time affective computing," IEEE Transactions on Affective Computing, vol. 14, no. 4, pp. 2102–2115, 2023.
13. S. Gupta, R. Nair, and P. Joshi, "Efficient CNN architectures for facial emotion recognition on low-power devices," Sensors, vol. 24, pp. 1–19, 2024.
14. Hernandez, J. Park, and D. Lee, "Edge AI-enabled emotion monitoring using lightweight vision transformers," Pattern Recognition Letters, vol. 181, pp. 55–63, 2024.
15. T. Zhou, Y. Wang, and K. Li, "Facial stress detection using spatio-temporal deep learning from video streams," Expert System with Applications, vol. 232, pp. 120–138, 2024.

16. N. Patel and S. Iyer, “Improved facial landmark localization for health monitoring using optimized MediaPipe models,” Computer Vision and Image Understanding, vol. 245, pp. 103–122, 2025.
17. Y. Kim, M. Zhao, and L. Chen, “Attention-based multimodal fusion for stress detection using facial cues and heart rate variability,” Biomedical Signal Processing and Control, vol. 101, pp. 106–124, 2025.
18. J. Wang, S. Roy, and A. Verma, “Real-time mental state monitoring using webcam-based facial analysis and deep neural networks,” IEEE Access, vol. 13, pp. 11840–11855, 2025.
19. K. Singh, A. Mehra, and V. Natarajan, “Benchmarking CNN and transformer models for facial emotion recognition under real-world conditions,” Journal of Imaging, vol. 11, pp. 1–17, 2025.