iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH
IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 14 Issue: III Month of publication: March 2026

DOI: https://doi.org/10.22214/ijraset.2026.78997

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 14 Issue III Mar 2026- Available at www.ijraset.com

Facial Expression Recognition for Early Detection of Stress and Mental Health Disorders

Bhanu Prakash Palusam[1], Marepalli Sidhu[1], Boini Vanaja[1], Adigamu Harsha Vardhan[1], Kaswa Jayanth[1], Dr. S. Srinivas[2], Dr. B. VenkataRamana[3]

[1]*Department of Data Science, Holy Mary Institute of Technology & Science, Hyderabad – 501303, India*
[2]*Associate Professor, Department of Data Science, Holy Mary Institute of Technology & Science, Hyderabad – 501303, India*
[3]*Associate Professor, Department of Data Science, Holy Mary Institute of Technology & Science, Hyderabad – 501303, India*

Abstract: Stress and mental health issues are increasingly common among office workers and students, often worsening because emotional changes go unrecognized and untracked; to address this, this project proposes a Facial Expression Recognition System for early detection of stress and mental health disorders that delivers intelligent, real-time emotional analysis using modern deep learning and web technologies. The proposed system accepts live webcam streams, single image uploads, and video files, and uses a pre-trained facial emotion recognition model built on PyTorch to classify seven emotions: angry, disgust, fear, happy, sad, surprise, and neutral, while combining MTCNN face detection with a Haar Cascade fallback for robust performance across varied conditions. An advanced stress analysis algorithm monitors emotion patterns over a 30-second rolling window to compute stress levels and detect trends. A session management module maintains user profiles, device tracking, and emotion history. The application is implemented with React 18.2 and Vite on the frontend and Flask with Socket.IO on the backend to ensure a responsive interface and efficient real-time communication. A database is maintained to store user profiles, session details, and emotion history. The system also supports multilingual interaction and concurrent multi-device streaming, making it accessible to users from different regions and deployable in large-scale scenarios. By providing accessible emotion detection and continuous stress monitoring, the system supports mental health assessment, user experience research, educational engagement tracking, and customer satisfaction analysis, enabling data-driven interventions and early support for improved well-being.
Keywords: Deep Learning, Emotion Recognition, Facial Expression Analysis, MTCNN, Stress Detection, Real-Time Processing.

I. INTRODUCTION

Mental health and emotional well-being have emerged as critical concerns in modern society, particularly among office workers and students who face increasing levels of stress, anxiety, and burnout. Despite the growing awareness of mental health issues, early detection and intervention remain challenging due to several factors including lack of accessible monitoring tools, social stigma associated with seeking help, limited awareness of emotional state changes, and delayed recognition of stress accumulation. Traditional mental health assessment methods rely heavily on self-reporting, clinical interviews, or periodic check-ups, which often fail to capture real-time emotional states or identify gradual deterioration in mental well-being. Many individuals are unaware of their stress levels until symptoms become severe, resulting in reduced productivity, impaired decision-making, and serious health consequences.

With the rapid growth of artificial intelligence and computer vision technologies, facial expression recognition has emerged as a powerful non-invasive approach for detecting emotional states and monitoring mental health. Deep learning models, particularly Convolutional Neural Networks (CNNs), can analyse facial expressions with remarkable accuracy, identifying subtle emotional cues that may indicate stress or mental distress. This project introduces a Facial Expression Recognition System for Early Detection of Stress and Mental Health Disorders, designed to provide continuous, accessible, and accurate emotional state monitoring through an intuitive web-based platform.

The proposed system enables users to monitor their emotional states through three primary input modalities: live webcam streaming for continuous real-time monitoring during work or study sessions, single image upload for quick emotional state checks, and video file processing for retrospective analysis of recorded sessions or meetings. By applying state-of-the-art deep learning techniques including a pre-trained FER (Facial Emotion Recognition) model built on PyTorch and MTCNN (Multi-task Cascaded Convolutional Networks) for face detection, the system achieves high accuracy in emotion classification while maintaining real-time performance.

The system classifies seven distinct emotions—angry, disgust, fear, happy, sad, surprise, and neutral—and implements an advanced stress analysis algorithm that examines emotion patterns over a 30-second rolling window to calculate stress levels and identify concerning trends. By delivering timely, accurate, and accessible emotional state monitoring with stress level analysis, this system aims to enable early intervention, improve mental health awareness, reduce stress-related health issues, and promote overall well-being in educational and professional environments.

A. Problem Statement

Individuals, particularly office workers and students, often lack awareness of their emotional states and stress levels, leading to undetected mental health deterioration, burnout, and reduced well-being. Existing mental health monitoring methods suffer from limitations such as dependency on self-reporting, lack of continuous monitoring capabilities, absence of real-time feedback, limited accessibility for non-clinical users, and inability to detect gradual stress accumulation. Thus, there is a critical need for an intelligent, accessible, and real-time system that can automatically detect facial expressions, monitor emotional patterns, calculate stress levels, and provide early warnings for potential mental health concerns.

B. Problem Solution

The proposed solution for the Facial Expression Recognition System is to develop an intelligent, real-time, and user-friendly platform that leverages Deep Learning, Computer Vision, and Web Technologies to provide continuous emotional state monitoring and stress level analysis. The Guardian of a student or the managers in the office can monitor the stress levels of their students and employees. Unlike existing methods, this solution is designed to offer non-invasive, automated, and accessible mental health monitoring while overcoming barriers of clinical dependency, cost, and accessibility.

The system employs a pre-trained Facial Emotion Recognition (FER) model built on PyTorch that classifies seven distinct emotions with high accuracy. Face detection is performed using MTCNN (Multi-task Cascaded Convolutional Networks) with Haar Cascade as a fallback mechanism, ensuring robust face localization across varying lighting conditions, face orientations, and image qualities. This dual-detection approach guarantees consistent performance even in challenging real-world scenarios.

To ensure real-time monitoring and continuous support, the system integrates advanced stress analysis algorithms that track emotion patterns over a 30-second rolling window. The stress calculation employs weighted scoring where each emotion contributes differently based on its association with stress: fear (0.90), angry (0.85), disgust (0.70), sad (0.65), surprise (0.30), neutral (0.20), and happy (0.10). The system calculates the overall stress level and identifies trends (increasing, stable, or decreasing) to provide actionable insights for users.

The user interface is developed using React 18.2 with Vite for fast, responsive interactions, while the backend utilizes Flask with Socket.IO to enable WebSocket-based real-time bidirectional communication. This architecture supports concurrent multi-device streaming, session management with user profiles and device tracking, and comprehensive emotion history storage. The system is scalable, supporting up to 10 simultaneous camera feeds for deployment in offices, classrooms, or research environments. By combining accurate deep learning models with accessible web technologies and intelligent stress monitoring, the proposed solution empowers users to track their emotional well-being, identify stress patterns early, and take proactive measures to maintain mental health.

II. LITERATURE REVIEW

The integration of computer vision and deep learning in emotion recognition has been extensively studied to address challenges related to mental health monitoring, human-computer interaction, and affective computing. Several research works have proposed intelligent systems to improve emotional state detection and reduce dependency on traditional psychological assessment methods. Early research in facial expression recognition focused on geometry-based approaches, where facial landmark positions and distances were used to classify emotions. These methods relied on handcrafted features extracted from facial key points [1]. Although these systems provided basic emotion classification, they lacked robustness and failed to handle variations in lighting, pose, and occlusions. The rigidity of feature-engineering-based systems limited their practical applicability in real-world scenarios.

With the advancement of Deep Learning techniques, researchers explored Convolutional Neural Networks (CNNs) for automated feature extraction and emotion classification. Studies have shown that deep learning models can effectively learn hierarchical representations from facial images, achieving superior performance compared to traditional methods [2]. The FER-2013 dataset and other emotion recognition benchmarks enabled the development and evaluation of CNN-based models for facial expression analysis [3].

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 14 Issue III Mar 2026- Available at www.ijraset.com

Recent studies emphasize real-time emotion detection using pre-trained deep learning models. The FER library and similar frameworks provide accessible tools for facial emotion recognition, enabling researchers and practitioners to deploy emotion detection systems without extensive model training [4]. However, many of these systems focus solely on emotion classification without considering temporal patterns or stress level analysis.

Face detection techniques have also evolved significantly. Multi-task Cascaded Convolutional Networks (MTCNN) have demonstrated high accuracy in detecting faces across varying scales and orientations [5]. MTCNN performs joint face detection and alignment, making it particularly suitable for emotion recognition applications where precise face localization is critical. Research on mental health monitoring using facial expressions has gained attention in recent years. Studies have explored the correlation between facial emotion patterns and stress levels, demonstrating that prolonged exposure to negative emotions correlates with increased stress and mental health concerns [6]. However, most systems lack comprehensive stress analysis algorithms that track emotion patterns over time.

Several web-based emotion recognition platforms have been developed to provide accessible interfaces for users. These platforms utilize modern web technologies and WebSocket communication for real-time interaction [7]. While these systems offer user-friendly interfaces, they often lack advanced features such as stress trend detection, session management, and concurrent multi-device support. From the reviewed literature, it is evident that existing solutions address individual aspects of emotion recognition but lack an integrated approach combining accurate emotion detection, real-time stress monitoring, and accessible web-based deployment. There is a significant research gap in developing a unified system that provides continuous emotional state monitoring with actionable stress level insights. The proposed Facial Expression Recognition System aims to bridge this gap by leveraging deep learning, computer vision, and modern web technologies to deliver real-time, accurate, and user-centric mental health monitoring.

III. METHODOLOGY

The Facial Expression Recognition System is designed to deliver intelligent emotional state monitoring and stress level analysis by integrating Deep Learning, Computer Vision, and Web Technologies into a unified platform. The system architecture consists of multiple interconnected components including input processing, face detection, emotion classification, stress analysis, session management, and real-time communication. The system begins with user interaction through a web-based interface, where users can initiate emotion monitoring through three primary input modalities: live webcam streaming, single image upload, or video file processing. For webcam streaming, frames are captured at regular intervals and transmitted to the backend via WebSocket protocol for real-time processing. Image uploads and video files are processed frame-by-frame to extract emotion data.

Each input frame undergoes pre-processing steps including resizing to a standard dimension (typically 224×224 or 48×48 pixels depending on the model requirements), normalization to scale pixel values to a specific range, and grayscale conversion for certain detection models. Face detection is performed using a dual-approach strategy to ensure robust performance across varying conditions. The primary detection method employs MTCNN (Multi-task Cascaded Convolutional Networks), which performs joint face detection and facial landmark localization through a cascaded architecture consisting of three stages: Proposal Network (P-Net) for initial face candidate generation, Refinement Network (R-Net) for false positive rejection, and Output Network (O-Net) for final face detection and landmark localization. MTCNN is particularly effective in detecting faces across different scales, orientations, and lighting conditions.

If MTCNN fails to detect a face or performs below a confidence threshold, the system automatically falls back to Haar Cascade classifiers, which use handcrafted features based on edge and intensity differences to detect faces. This dual-detection approach ensures consistent face localization even in challenging scenarios. Once a face is detected, the cropped facial region is passed to the emotion classification module. The system employs a pretrained Facial Emotion Recognition (FER) model built on PyTorch, which utilizes Convolutional Neural Networks to classify facial expressions into seven distinct emotion categories: angry, disgust, fear, happy, sad, surprise, and neutral. The CNN architecture consists of multiple convolutional layers with ReLU activation functions, max-pooling layers for dimensionality reduction, and fully connected layers for final classification.

The stress analysis component implements an advanced algorithm that monitors emotion patterns over time to calculate stress levels and detect trends. The system maintains a 30-second rolling window of emotion detections, storing each detected emotion along with its probability and timestamp. Stress calculation employs a weighted scoring mechanism where each emotion is assigned a weight based on its association with stress and negative mental states: fear (0.90), angry (0.85), disgust (0.70), sad (0.65), surprise (0.30), neutral (0.20), and happy (0.10). The overall stress score is calculated as: $\text{Stress} = \Sigma(\text{Emotion\_Probability} \times \text{Emotion\_Weight}) / \text{Total\_Detections}$ where higher weights correspond to emotions typically associated with stress. Trend detection analyzes the stress score trajectory over the rolling window to identify whether stress levels are increasing, stable, or decreasing.

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 14 Issue III Mar 2026- Available at www.ijraset.com

Session management maintains user profiles, device tracking, and emotion history for each connected session. When a user initiates a session, the system generates a unique session ID and associates it with the user's device and profile information. All detected emotions, timestamps, and stress scores are stored in the session history, enabling retrospective analysis and pattern identification. The system supports concurrent multi-device streaming, allowing up to 10 simultaneous camera feeds, which is useful for deployment in classrooms, offices, or research environments. Real-time communication between the frontend and backend is facilitated through WebSocket protocol using Socket.IO. This bidirectional communication channel enables low-latency transmission of video frames from the client to the server and immediate delivery of emotion detection results and stress analysis back to the client. The final output includes the detected emotion, confidence scores for all emotion classes, current stress level, stress trend (increasing, stable, or decreasing), and visualization of emotion patterns over time.

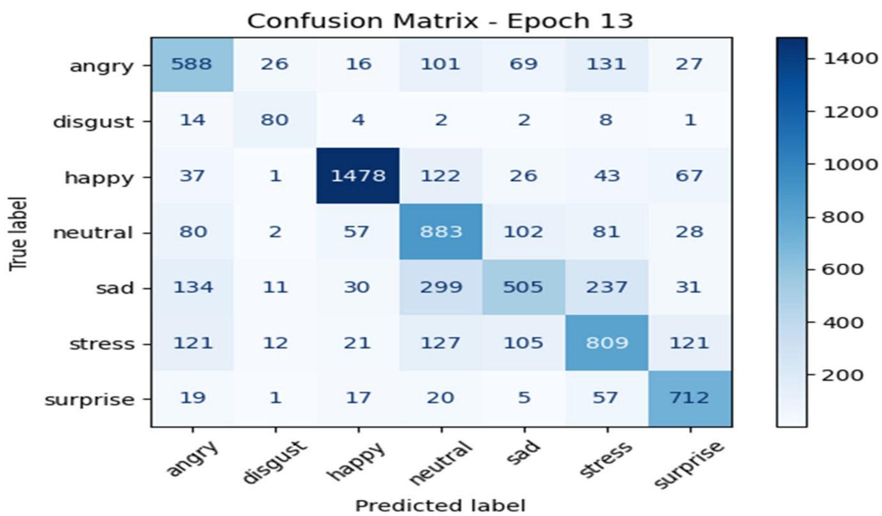


Fig. 1. Confusion Matrix of a Trained model

IV. SYSTEM DESIGN AND IMPLEMENTATION:

A. Architecture Diagram

The architecture of the Emotion Detection System follows a modern three-tier design pattern, separating concerns between the presentation layer, application logic layer, and data processing layer. This architectural approach provides numerous benefits including improved maintainability, scalability, and the ability to independently update or replace components without affecting the entire system. The frontend tier, built using React, handles all user interface rendering and interaction logic. The backend tier, implemented in Flask with SocketIO extensions, manages API endpoints, WebSocket communications, and orchestrates the emotion detection pipeline. The processing tier encompasses the machine learning models and computer vision algorithms that perform the actual face detection and emotion classification tasks.

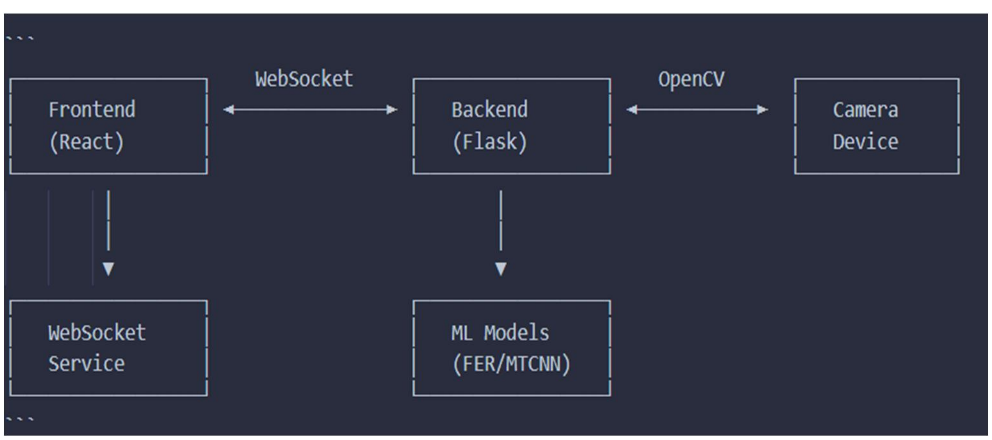


Fig. 2. Facial Expression Recognition System Architecture

B. Frontend / User Interface Layer

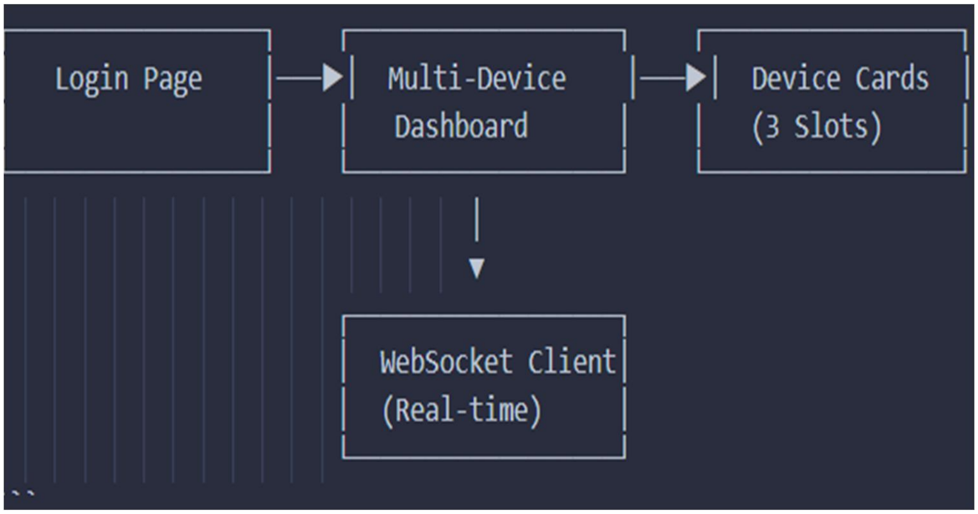


Fig. 3. Architecture of Frontend

The frontend component leverages React 18, a modern JavaScript library for building user interfaces, to create a responsive and dynamic application. The Live-Detection component serves as the primary interface for real-time emotion detection, managing the display of video streams and emotion overlays through canvas-based rendering. This canvas approach provides fine-grained control over video frame display and allows for efficient drawing of bounding boxes and text labels directly onto the video feed. The WebSocket service component handles all real-time communication with the backend server, establishing and maintaining persistent connections, handling connection events, and processing incoming data streams. The frontend architecture implements a service-oriented pattern where different aspects of functionality are encapsulated in dedicated services. The API service centralizes all HTTP requests for non-real-time operations such as uploading images or videos for analysis. Component state management utilizes React hooks, providing a clean and functional approach to managing application state without the complexity of external state management libraries.

C. Backend / API Layer

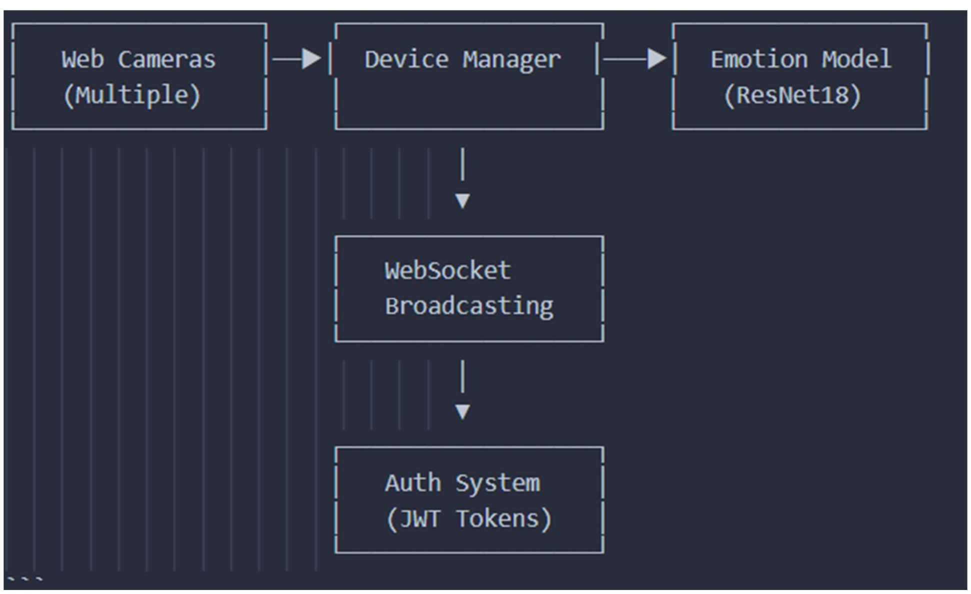


Fig. 4. Architecture of Backend

The backend architecture focuses around Flask, a lightweight Python web framework known for its simplicity and extensibility. Flask-SocketIO extends the basic Flask application with WebSocket support, enabling real-time bidirectional communication essential for streaming video and emotion data. The camera service component encapsulates all logic related to accessing camera hardware, capturing video frames, and managing camera lifecycle. This service is designed with configurability in mind, allowing administrators to specify camera device identifiers, target frame rates, and resolution settings to optimize performance for different deployment scenarios. The emotion prediction service represents the core analytical component of the backend, integrating with machine learning libraries to perform face detection and emotion classification. This service orchestrates a multi-step pipeline that begins with receiving a video frame, applies face detection algorithms to locate facial regions, extracts those regions for analysis, feeds them through emotion classification models, and returns structured results containing detected emotions with confidence scores. Session management is handled through a dedicated service that tracks active connections, allocates resources for each session, and ensures proper cleanup when sessions terminate, preventing memory leaks and resource exhaustion.

D. Machine Learning Integration

The machine learning layer integrates two primary models for face detection and emotion classification. Face detection employs either MTCNN (Multi-task Cascaded Convolutional Networks) or Haar Cascade classifiers, with MTCNN serving as the primary detector due to its superior accuracy in challenging conditions. MTCNN represents a sophisticated approach to face detection that simultaneously handles face localization, facial landmark detection, and face alignment through a cascaded architecture of three convolutional networks. This multi-stage process first generates candidate face regions, refines these candidates, and finally outputs precise face locations along with facial landmarks such as eyes, nose, and mouth positions. The Facial Expression Recognition (FER) library handles emotion classification, utilizing deep convolutional neural networks trained on extensive facial expression datasets. These models have learned to recognize patterns in facial muscle movements, eye positions, mouth shapes, and other facial features that correlate with different emotional states. The FER library supports classification into seven distinct emotion categories: happiness, sadness, anger, surprise, fear, disgust, and neutral expressions. Each classification result includes confidence scores indicating the model's certainty about the detected emotion.

V. RESULTS AND DISCUSSION

The Facial Expression Recognition System was implemented and tested using a combination of live webcam streams, uploaded images, and video files to evaluate its effectiveness in providing accurate emotion detection and stress level analysis. The system demonstrated reliable performance in real-time emotional state monitoring, particularly for commonly occurring emotions such as happy, neutral, sad, and angry. The pretrained FER model successfully classified facial expressions with high confidence under controlled lighting conditions and frontal face orientations. Face detection using MTCNN produced promising results when tested with diverse facial images representing different ages, genders, and facial features. The system was able to correctly detect and localize faces in real-time with minimal latency.

The Haar Cascade fallback mechanism proved effective in scenarios where MTCNN performance degraded due to extreme face angles, occlusions, or poor lighting conditions. This dual-detection strategy significantly improved the system's robustness and reliability across varying real-world conditions. The stress analysis algorithm successfully tracked emotion patterns over the 30-second rolling window and calculated stress levels based on the weighted scoring mechanism. Users who exhibited prolonged negative emotions such as fear, anger, or sadness received higher stress scores, while those displaying predominantly happy or neutral expressions showed lower stress levels.

Real-time communication through WebSocket protocol demonstrated low-latency performance, with emotion detection results being delivered to the frontend within milliseconds of frame transmission. The system maintained stable performance even with concurrent multi-device streaming, supporting up to 10 simultaneous camera feeds without significant degradation in processing speed or accuracy. Session management effectively maintained user profiles, device tracking, and emotion history, allowing users to review their emotional patterns and stress trends over time. The web-based interface built with React and Vite provided a smooth and responsive user experience. Users found the interface intuitive and easy to navigate, with clear visualizations of detected emotions and stress levels.

However, the system's performance was influenced by several factors. Variations in lighting conditions, particularly very dim or very bright environments, affected face detection accuracy. Non-frontal face orientations and partial occlusions (such as hands covering parts of the face or wearing masks) occasionally resulted in incorrect emotion classification or missed detections.

The accuracy of emotion recognition also depended on the quality and resolution of input images, with lower-resolution images producing less reliable results. Overall, the results indicate that the proposed system is effective in providing real-time emotional state monitoring and stress level analysis. While the current implementation demonstrates strong potential for applications in mental health monitoring, user experience research, and educational assessment, future improvements such as enhanced lighting normalization, 3D face pose estimation, and expanded emotion datasets can further improve system accuracy and robustness.

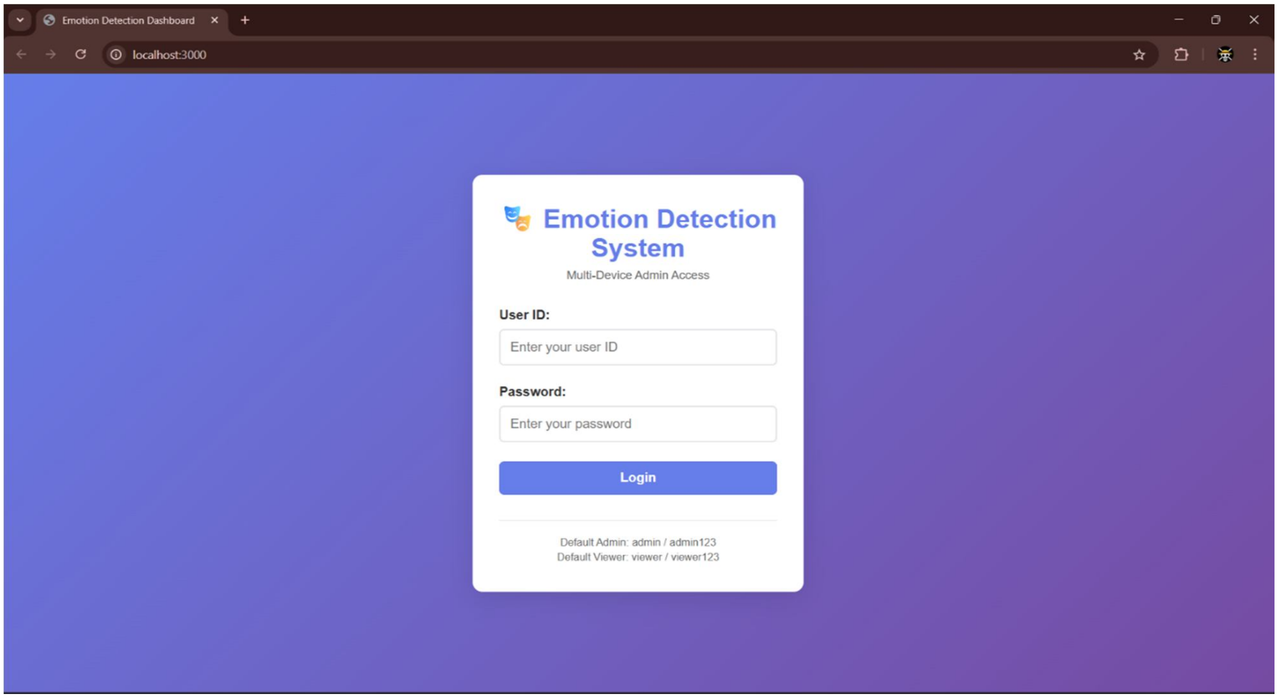


Fig. 5. Login Page

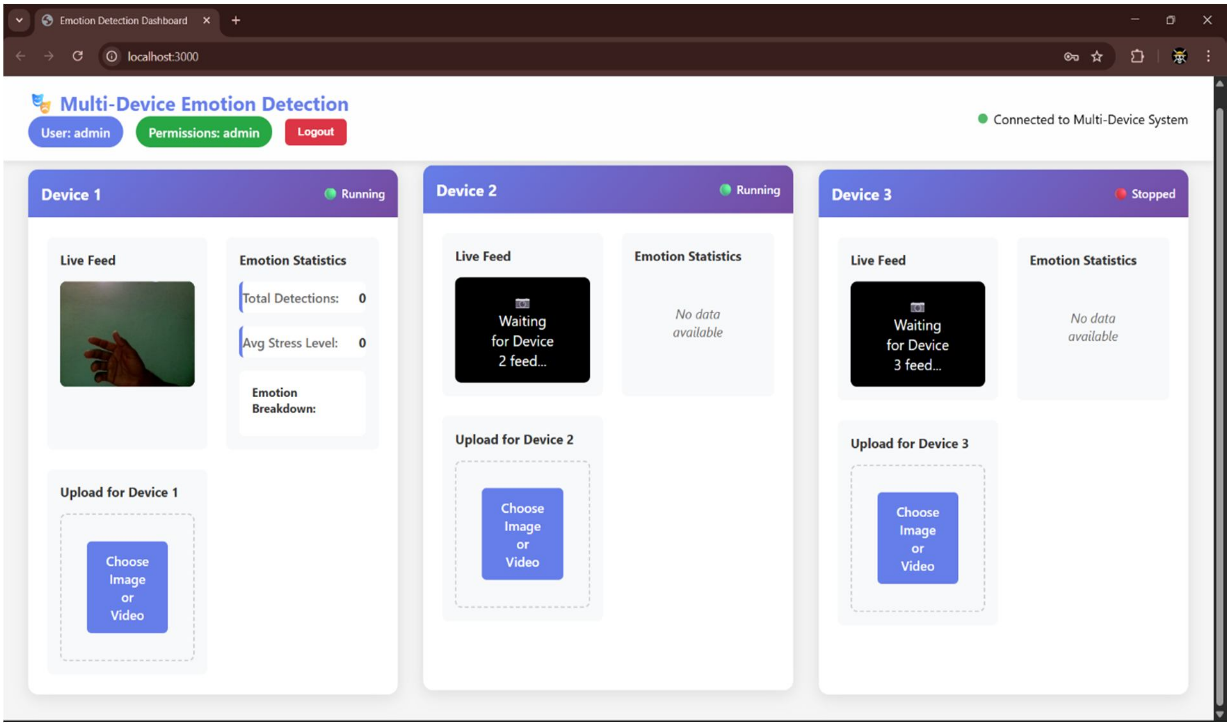


Fig. 6. React frontend displaying the stress Analysis

VI. CONCLUSION

This project successfully presents the design and implementation of a Facial Expression Recognition System for Early Detection of Stress and Mental Health Disorders, aimed at addressing the critical challenges faced by individuals due to limited access to continuous mental health monitoring and early stress detection. By integrating Deep Learning, Computer Vision, Natural Language Processing, and Web Technologies, the system provides an intelligent, accessible, and user-friendly platform for delivering real-time emotional state analysis and stress level monitoring. The developed system effectively detects facial expressions through advanced face detection using MTCNN with Haar Cascade fallback, classifies emotions into seven distinct categories using a pre-trained PyTorch-based FER model, and calculates stress levels through an intelligent weighted scoring algorithm.

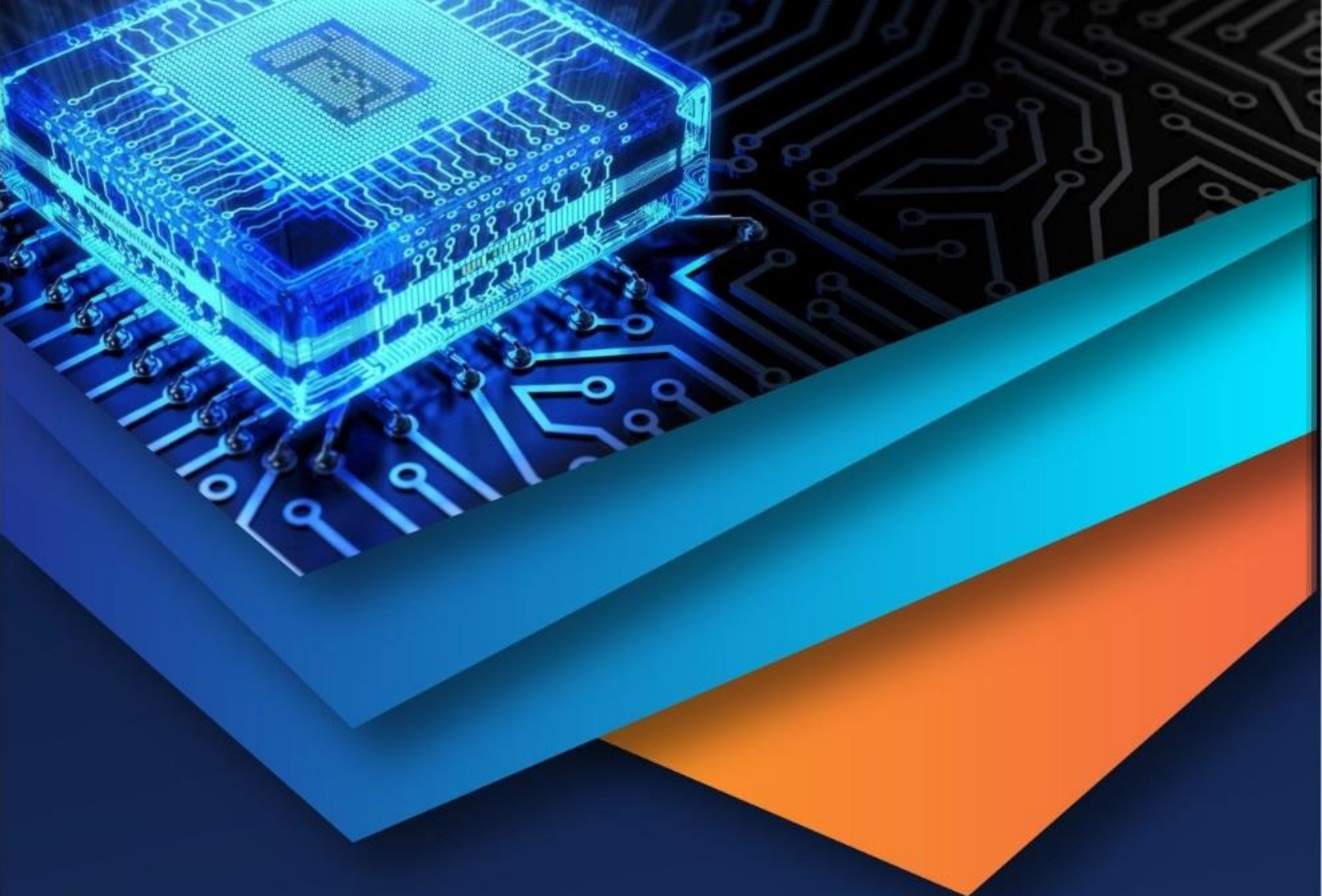
The use of deep learning techniques enables accurate emotion recognition, while the implementation of a 30-second rolling window for stress pattern analysis provides actionable insights into mental health trends. The integration of WebSocket-based real-time communication ensures low-latency performance, supporting continuous monitoring and immediate feedback. Experimental evaluation demonstrates that the system is capable of delivering accurate and timely emotion detection and stress level analysis for common emotional states, thereby reducing dependency on traditional mental health assessment methods and minimizing delays in identifying emotional distress. The web-based interface built with React and Vite ensures accessibility and ease of use, making the system suitable for users with varying levels of technical expertise.

Although the current implementation shows promising results, its effectiveness is influenced by factors such as lighting conditions, face orientation, and image quality. Overall, the Facial Expression Recognition System highlights the potential of artificial intelligence in transforming mental health monitoring and stress detection. The project contributes toward improving emotional well-being awareness, enabling early intervention, reducing stress-related health issues, and promoting proactive mental health management. With further enhancements such as multi-modal emotion detection, advanced 3D face analysis, and expanded deployment platforms, the proposed system can be scaled to support a wider range of applications including workplace wellness programs, educational institutions, healthcare facilities, and remote mental health services, making it a valuable technological solution for modern mental health support.

The Confusion Matrix shows how the model confuses between some of the emotions. Sometimes fear can be confused for sad, the results of emotion prediction can be based on many factors as mentioned above. It purely depends on the quality of an image or the video.

REFERENCES

[1] Z. Zhang, M. Lyons, M. Schuster, and S. Akamatsu, "Comparison between geometry-based and Gabor-wavelets-based facial expression recognition using multi-layer perceptron," in Proc. 3rd IEEE International Conference on Automatic Face and Gesture Recognition, Nara, Japan, 1998, pp. 454-459.
[2] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," Nature, vol. 521, no. 7553, pp. 436-444, 2015.
[3] I. J. Goodfellow, D. Erhan, P. L. Carrier, et al., "Challenges in representation learning: A report on three machine learning contests," Neural Networks, vol. 64, pp. 59-63, 2015.
[4] J. Shenk and M. Schwabacher, "FER: Facial Expression Recognition," GitHub repository, https://github.com/justinshenk/fer, 2020.
[5] K. Zhang, Z. Zhang, Z. Li, and Y. Qiao, "Joint face detection and alignment using multitask cascaded convolutional networks," IEEE Signal Processing Letters, vol. 23, no. 10, pp. 1499-1503, 2016.
[6] B. C. Ko, "A brief review of facial emotion recognition based on visual information," Sensors, vol. 18, no. 2, p. 401, 2018.
[7] M. Pantic and L. J. M. Rothkrantz, "Automatic analysis of facial expressions: The state of the art," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 22, no. 12, pp. 1424-1445, 2000.
[8] R. W. Picard, Affective Computing. Cambridge, MA, USA: MIT Press, 1997.
[9] Z. Zeng, M. Pantic, G. I. Roisman, and T. S. Huang, "A survey of affect recognition methods: Audio, visual, and spontaneous expressions," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 31, no. 1, pp. 39-58, 2009.
[10] S. Li and W. Deng, "Deep facial expression recognition: A survey," IEEE Transactions on Affective Computing, vol. 12, no. 4, pp. 502-524, Oct.-Dec. 2022. doi: 10.1109/TAFFC.2020.2981446



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY