StressLens: Facial Stress and Fatigue Recognition Using Emotion-Weighted Transformer Ensemble

Gauri Shailesh Indani[*1], Dr. H. R. Vyawahare[*2]
[*1]Research Scholar, MTech Computer Science and Engineering, SIPNA College of Engineering and Technology, Amravati.
[*2]Professor, Computer Science and Engineering, SIPNA College of Engineering and Technology, Amravati.

ABSTRACT- Stress and fatigue are significant factors affecting mental health, productivity, and overall quality of life. Traditional stress monitoring methods primarily rely on wearable sensors and physiological measurements, which may be intrusive and unsuitable for continuous monitoring. This paper presents StressLens, a real-time facial stress and fatigue recognition system based on deep learning and emotion-weighted analysis. The proposed framework employs Vision Transformer (ViT) and Bidirectional Encoder Representation from Image Transformers (BEiT) models to extract facial features from webcam-captured images. An emotion-weighting mechanism assigns different importance levels to detected emotions for improved stress estimation, while a majority-voting ensemble strategy combines predictions from both transformer models to enhance classification reliability. The system is implemented using PyTorch, OpenCV, Hugging Face Transformers, and Flask, and is deployed through an interactive web dashboard supporting live monitoring, visualization, and alert generation. Experimental results demonstrate that transformer-based models outperform conventional CNN approaches in stress and fatigue recognition, achieving higher precision and robustness under varying conditions. The proposed framework offers a non-invasive, scalable, and intelligent solution for continuous mental wellness monitoring in healthcare, workplace, and educational environments.

KEYWORDS- Stress Detection, Fatigue Recognition, Vision Transformer, BEiT, Emotion Weighting, Deep Learning, Facial Analysis.

I. INTRODUCTION

Stress and fatigue have become increasingly common challenges in modern society due to demanding work schedules, academic pressure, prolonged screen exposure, and fast-paced lifestyles. These conditions significantly impact an individual's mental and physical well-being, often leading to reduced productivity, poor decision-making, emotional instability, and various health-related complications. Continuous exposure to stressful situations can affect concentration, memory, and overall performance, while fatigue may decrease alertness and increase the likelihood of errors and accidents. As a result, the early detection and monitoring of stress and fatigue have become important for maintaining mental wellness and improving quality of life. Traditional methods for stress assessment primarily rely on self-report questionnaires, psychological evaluations, interviews, and physiological monitoring devices. Although these approaches can provide useful insights, they often suffer from limitations such as subjectivity, inconvenience, and the inability to perform continuous real-time monitoring. Wearable devices that measure physiological signals such as heart rate, skin conductance, and body temperature have gained popularity for stress detection. However, these systems require physical contact with the user, which may cause discomfort during long-term usage and limit their practicality in everyday environments.

Recent advancements in artificial intelligence, computer vision, and deep learning have created new opportunities for non-invasive stress monitoring systems. Human facial expressions contain valuable information about emotional and psychological states. Emotions such as sadness, anger, fear, and frustration are often associated with elevated stress levels, while relaxed facial expressions may indicate lower stress conditions. By analyzing facial features and emotional patterns, it is possible to estimate stress and fatigue levels without requiring specialized sensors or wearable equipment. This approach offers a more convenient and user-friendly alternative for continuous mental wellness monitoring. Deep learning techniques have achieved remarkable success in image classification, facial recognition, and emotion analysis tasks. Convolutional Neural Networks (CNNs) have traditionally been used for extracting facial features and recognizing emotions. Although CNN-based approaches provide satisfactory performance, they may struggle to capture long-range spatial relationships and global contextual information present in facial images. Recently, transformer-based architectures have emerged as powerful alternatives for visual recognition tasks. These models process images by analyzing relationships among image regions, enabling them to learn richer and more meaningful feature representations.

This paper presents StressLens, a real-time facial stress and fatigue recognition system based on an emotion-weighted transformer ensemble. The proposed framework utilizes Vision Transformer (ViT) and Bidirectional Encoder Representation from Image Transformers (BEiT) models to analyze facial expressions captured through a webcam. An emotion-weighting mechanism is introduced to assign different importance levels to detected emotional states, improving the reliability of stress estimation. Furthermore, an ensemble majority-voting strategy combines predictions from multiple transformer models to enhance classification robustness and reduce prediction fluctuations. To facilitate practical deployment, the proposed system is integrated into a Flask-based web dashboard that supports live webcam monitoring, interactive visualizations, user authentication, alert generation, and stress-relief recommendations. The framework aims to provide a scalable, intelligent, and non-invasive solution for continuous stress and fatigue monitoring in healthcare, workplace wellness, educational environments, and other real-world applications. By combining advanced transformer architectures with emotion-aware analysis, StressLens contributes toward the development of effective mental wellness technologies capable of supporting proactive stress management and improved quality of life.

II. LITERATURE REVIEW

Stress detection and monitoring have become important research areas due to the increasing impact of stress on human health, workplace productivity, and overall well-being. Researchers have explored various approaches using wearable sensors, physiological signals, machine learning, deep learning, and multimodal systems to develop reliable stress recognition frameworks. Abd Al-Alim [1] proposed a machine-learning approach for stress detection using wearable sensors in free-living environments. The study utilized physiological signals and demonstrated that machine learning algorithms can effectively classify stress levels when appropriate preprocessing and feature extraction techniques are applied. Similarly, Ghaderi et al. [2] presented a systematic review of wearable-based stress monitoring systems and highlighted the importance of physiological signals such as heart rate variability, electrodermal activity, and body temperature for stress assessment. Bhatia and Goel [3] introduced an explainable deep learning framework for stress detection using physiological sensor data. Their work combined deep neural networks with explainability techniques to improve transparency and decision-making. Bolpagni et al. [4] reviewed personalized stress detection approaches using biosignals from wearable devices and emphasized the need for adaptive models that consider individual variations in stress responses.

Kankal et al. [5] developed an IoT-based stress monitoring system that integrated wearable sensors and machine learning algorithms for real-time stress analysis. Mohammed and Hassan [6] proposed a wearable sensor-based framework utilizing physiological signals and IoT technologies to enable continuous stress monitoring. Jha and Singh [7] implemented a machine-learning-based stress detection system using wearable sensors and demonstrated the effectiveness of traditional classification algorithms for stress prediction. Rehman and Khan [8] proposed a hybrid CNN-based architecture for stress recognition using wrist-based photoplethysmography sensors. Their framework combined handcrafted and deep features to improve classification accuracy. Wataru and Rossi [9] introduced SELF-CARE, a context-aware sensor fusion framework that selectively activates sensors to reduce power consumption while maintaining prediction performance. Nkomo and Potdar [10] further explored context-aware sensor fusion techniques by combining physiological and environmental information to improve stress detection accuracy.

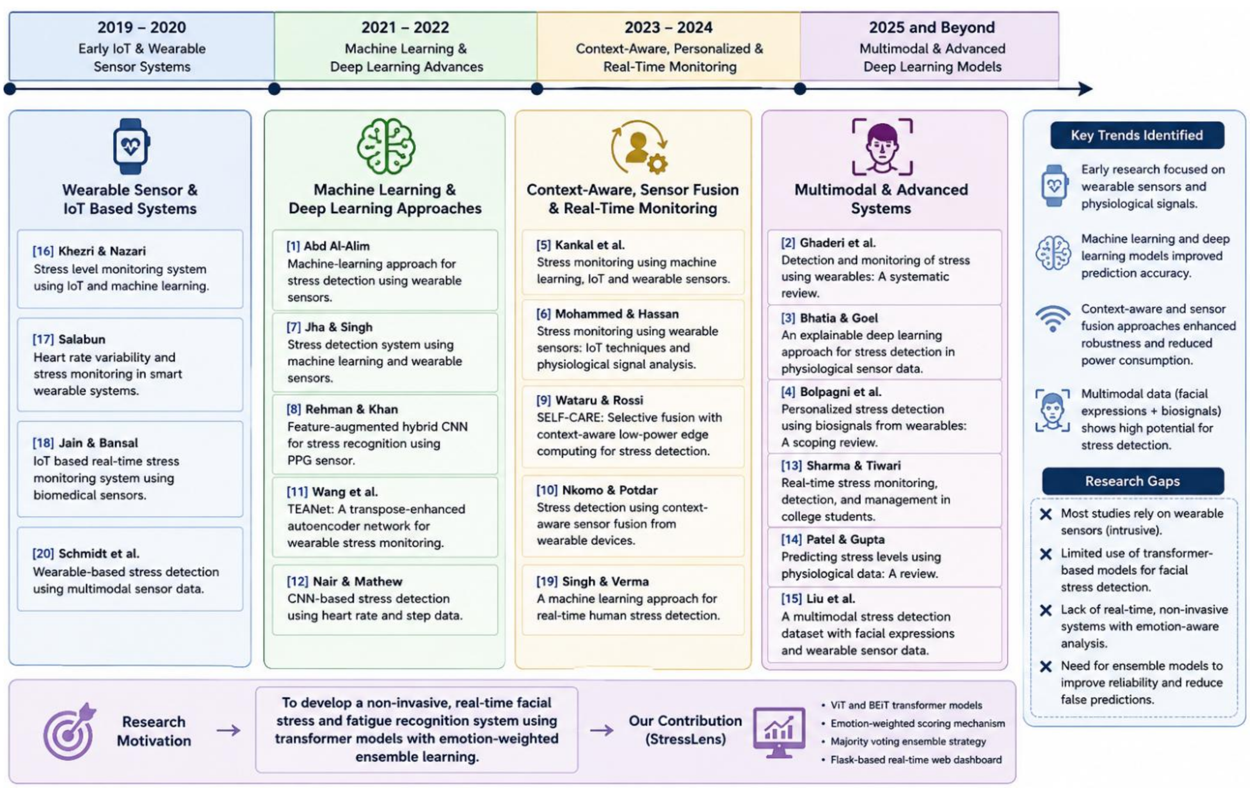


Figure 1: Literature review summary and research gaps leading to the proposed StressLens framework

Wang et al. [11] proposed TEANet, a transpose-enhanced autoencoder network designed for wearable stress monitoring. The model demonstrated improved feature extraction capabilities and achieved promising classification performance. Nair and Mathew [12] developed a CNN-based stress detection framework using heart rate and activity data collected from wearable devices, showing that deep learning models can outperform traditional machine learning approaches in stress recognition tasks. Sharma and Tiwari [13] focused on real-time stress monitoring and management among college students using wearable technologies and machine learning methods. Their work highlighted the importance of early stress detection in educational environments. Patel and Gupta [14] provided a comprehensive review of real-time stress prediction systems and discussed the challenges associated with physiological data acquisition and continuous monitoring.

A significant advancement in non-invasive stress detection was presented by Liu et al. [15], who introduced a multimodal dataset containing facial expressions and wearable sensor data. Their study demonstrated that facial expressions can serve as valuable indicators of emotional and psychological stress. This work encouraged further exploration of computer vision-based stress recognition systems. Khezri and Nazari [16] proposed an IoT and machine learning-based stress monitoring framework capable of real-time stress prediction. Salabun [17] investigated the role of heart rate variability in wearable stress monitoring systems and highlighted its effectiveness as a physiological stress indicator. Jain and Bansal [18] developed an IoT-based biomedical sensor framework for continuous stress monitoring and remote healthcare applications. Singh and Verma [19] presented a machine-learning approach for real-time human stress detection and demonstrated promising classification results using physiological data. Schmidt et al. [20] proposed a multimodal wearable-based stress detection system that combined multiple sensor signals to improve prediction robustness.

III. METHODOLOGY

The proposed methodology aims to develop a real-time and non-invasive facial stress and fatigue recognition framework using deep learning and emotion-weighted analysis. The system combines computer vision, transformer-based architectures, and ensemble learning techniques to monitor stress and fatigue levels through facial expressions captured from a webcam. Unlike traditional wearable sensor-based approaches, the proposed framework eliminates the need for physical sensors and enables continuous monitoring in a practical and user-friendly manner. The overall system architecture consists of five major stages: video acquisition, face detection and preprocessing, transformer-based feature extraction, emotion-weighted ensemble prediction, and real-time visualization. Initially, facial video streams are captured through a webcam in real time.

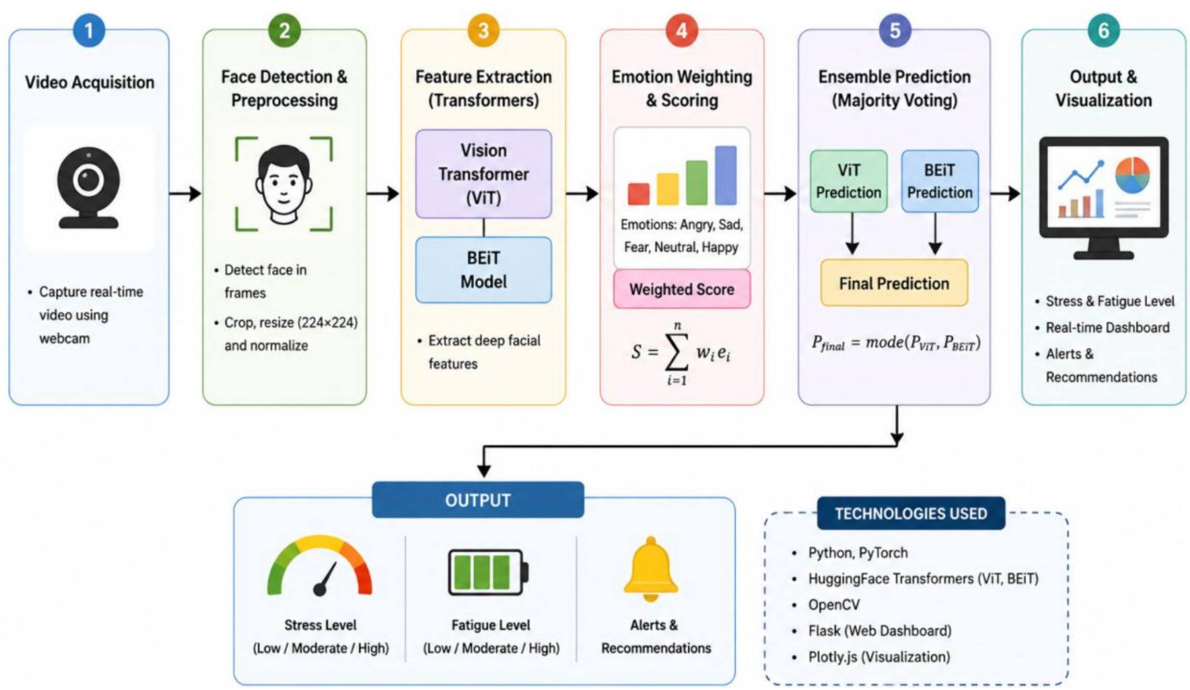


Figure 2: Methodology for Facial Stress and Fatigue Recognition

The video is divided into multiple frames at fixed intervals to ensure efficient processing while maintaining temporal consistency. Each frame is then passed through a face detection module that isolates facial regions from the background. Detected facial images are resized and normalized to match the input requirements of transformer models.

After preprocessing, the extracted facial frames are provided as input to two transformer-based deep learning models: Vision Transformer (ViT) and Bidirectional Encoder Representation from Image Transformers (BEiT). These models are selected due to their superior capability in learning global contextual information and long-range feature dependencies from facial images. Unlike traditional Convolutional Neural Networks (CNNs), transformer architectures analyze relationships between image patches more effectively, resulting in improved facial feature representation and classification accuracy. The proposed framework introduces an emotion-weighting mechanism to improve stress and fatigue estimation. Human facial emotions such as anger, sadness, fear, and neutrality are closely associated with psychological stress levels. Therefore, different emotional states are assigned different weights based on their contribution to stress and fatigue recognition. The weighted emotional features are combined to generate stress and fatigue scores for each frame. The stress score is computed as:

$$S = \sum_{i=1}^{n} w_i \, e_i$$

where (S) represents the final stress score, (w_i) denotes the emotional weight, and (e_i) represents the confidence score of the detected emotion. Emotions associated with negative mental states are assigned higher weights compared to positive emotions. To improve prediction robustness, the framework employs an ensemble majority voting strategy between ViT and BEiT outputs. Predictions generated from multiple frames and both transformer models are aggregated to determine the final stress and fatigue classification. The final prediction is calculated as:

$$P_{final} = \text{mode}\left(P_{ViT} \, , P_{BEiT}\right)$$

P_{final}	:	Final predicted class (stress or fatigue level)
P_{ViT}	:	Prediction from Vision Transformer (ViT) model
P_{BEiT}	:	Prediction from BEiT model
$mode(\cdot)$	:	Function that returns the most frequent (majority) class

This ensemble mechanism reduces prediction instability and improves classification reliability under varying lighting conditions, facial orientations, and emotional variations.

The proposed system is implemented using a Flask-based web dashboard integrated with PyTorch and HuggingFace Transformers for real-time inference. The dashboard supports live webcam analysis, user authentication, alert generation, and interactive visualization using Plotly.js charts. Stress and fatigue levels are continuously monitored and displayed through dynamic graphs and frame-based visual indicators. The framework also provides stress-relief recommendations such as breathing exercises and relaxation techniques when elevated stress levels are detected. The proposed methodology combines advanced transformer architectures, emotion-aware analysis, and real-time deployment to provide an intelligent and scalable solution for facial stress and fatigue recognition in healthcare, educational, and workplace environments.

IV. PROPOSED SYSTEM

The proposed system, **StressLens**, is a real-time facial stress and fatigue recognition framework developed to monitor an individual's mental wellness using computer vision and deep learning techniques. Unlike conventional stress monitoring solutions that depend on wearable sensors and physiological measurements, the proposed system utilizes facial expressions captured through a webcam to estimate stress and fatigue levels in a non-invasive manner. The framework combines transformer-based architectures, emotion-weighted analysis, and ensemble learning to improve prediction accuracy and reliability.

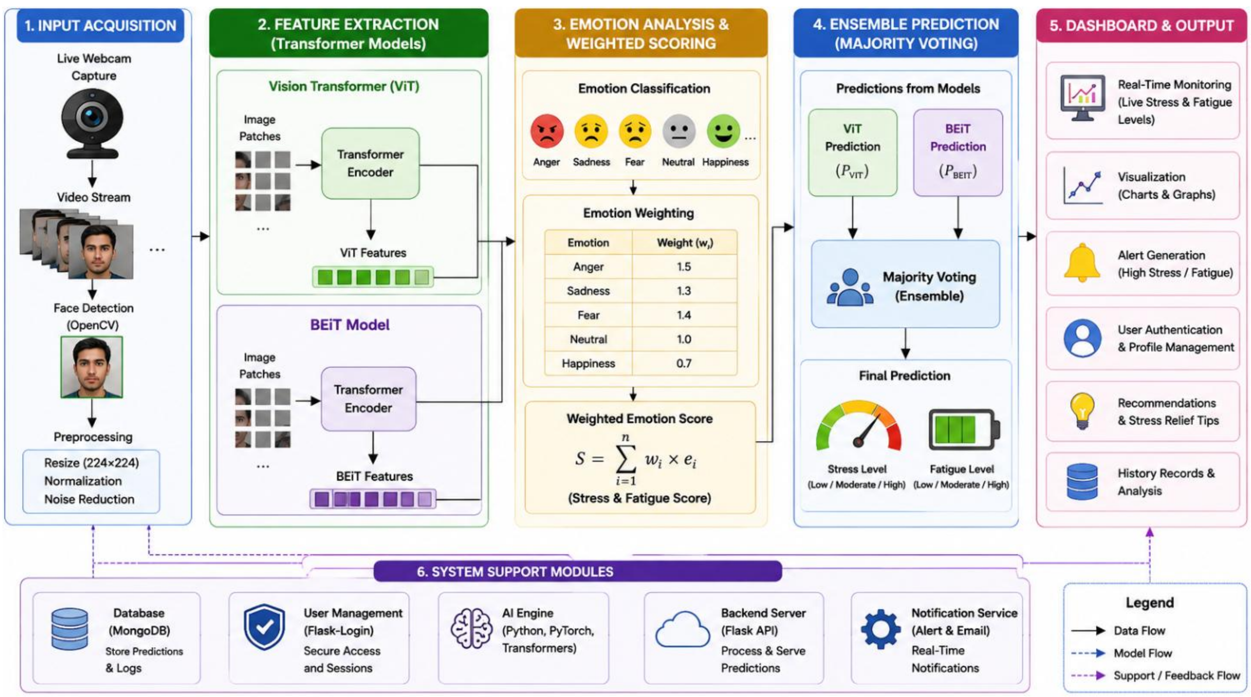


Figure 3: Proposed System Architecture of StressLens for Real-Time Facial Stress and Fatigue Recognition

4.1 Facial Analysis and Feature Extraction

The system begins by acquiring live facial images through a webcam. The captured frames undergo face detection and preprocessing operations to isolate the facial region and improve image quality. The processed facial images are then provided to Vision Transformer (ViT) and Bidirectional Encoder Representation from Image Transformers (BEiT) models. These transformer architectures extract global contextual facial features and learn complex emotional patterns associated with stress and fatigue conditions. The use of transformer models enables more effective feature representation compared to traditional convolutional neural networks.

4.2 Emotion-Weighted Stress and Fatigue Estimation

After feature extraction, the system analyzes facial emotions such as anger, sadness, fear, neutrality, and happiness. Since different emotions contribute differently to stress recognition, an emotion-weighting mechanism is applied to assign varying importance levels to each detected emotional state. Emotions that are commonly associated with psychological stress are assigned higher weights, while positive emotions receive comparatively lower weights. The weighted emotional scores are aggregated to calculate stress and fatigue levels, enabling more accurate and context-aware prediction.

4.3 Ensemble Prediction and Dashboard Integration

To improve robustness and prediction consistency, the outputs generated by the ViT and BEiT models are combined using a majority-voting ensemble strategy. The ensemble mechanism reduces prediction fluctuations and improves classification reliability under varying lighting conditions, facial orientations, and emotional expressions. The final stress and fatigue predictions are displayed through a Flask-based web dashboard that supports real-time monitoring, graphical visualization, user authentication, alert generation, and historical analysis. The dashboard also provides wellness recommendations whenever elevated stress or fatigue levels are detected, making the system suitable for practical deployment in healthcare, workplace, and educational environments.

V. IMPLEMENTATION

The implementation of the proposed StressLens framework focuses on developing a real-time facial stress and fatigue recognition system using transformer-based deep learning models and a web-based monitoring platform. The system is implemented using Python due to its extensive support for machine learning, computer vision, and web application development. Various technologies including PyTorch, OpenCV, Hugging Face Transformers, Flask, HTML, CSS, JavaScript, and Plotly.js are integrated to provide efficient real-time analysis and visualization.

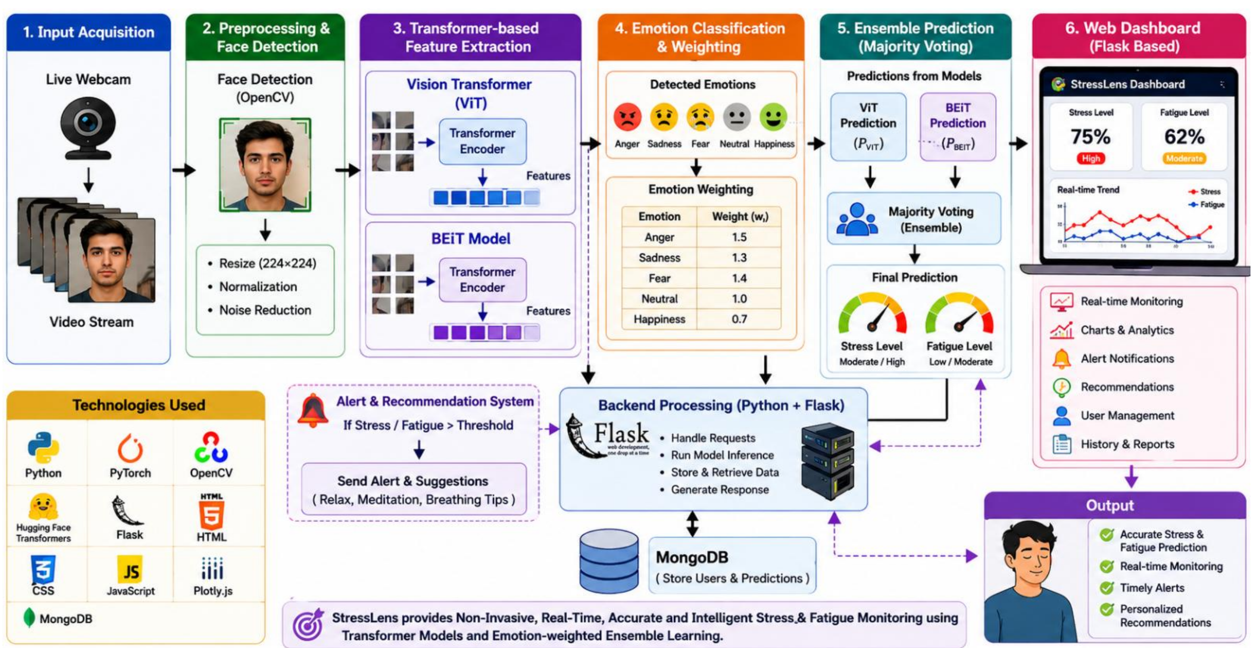


Figure 4: Implementation Overview of the StressLens Framework

5.1 Data Acquisition and Preprocessing

The implementation process begins with real-time facial image acquisition through a webcam. OpenCV is used to continuously capture video frames and perform face detection. The detected facial regions are extracted and separated from the background to focus only on relevant facial information. The extracted images are resized to 224 × 224 pixels and normalized according to the input requirements of the transformer models. Image preprocessing techniques such as scaling, noise reduction, and brightness normalization are applied to improve image quality and maintain consistency during model inference. This preprocessing stage ensures that facial images remain suitable for accurate feature extraction under different environmental conditions.

5.2 Transformer Model Implementation

The feature extraction and classification components are implemented using Vision Transformer (ViT) and Bidirectional Encoder Representation from Image Transformers (BEiT) models. These models are developed using PyTorch and Hugging Face Transformer libraries. The transformer architectures divide facial images into multiple patches and process them through self-attention mechanisms to learn global contextual relationships among facial features. Pre-trained transformer weights are utilized and further fine-tuned for facial stress and fatigue recognition. The extracted features contain valuable emotional and facial information that contributes to stress prediction. Compared to traditional convolutional neural networks, transformer models provide superior representation of long-range facial dependencies and contextual patterns.

5.3 Emotion Weighting and Ensemble Learning

An emotion-weighting mechanism is implemented to improve stress estimation accuracy. The system analyzes detected emotions such as anger, sadness, fear, neutrality, and happiness. Different weights are assigned to each emotion based on its contribution to stress recognition. Negative emotions are assigned higher weights because they are more closely associated with stress conditions. The weighted emotional scores are aggregated to compute final stress and fatigue scores. To improve prediction reliability, outputs generated by the ViT and BEiT models are combined using a majority-voting ensemble strategy. This ensemble approach reduces prediction instability and improves classification consistency across multiple frames.

5.4 Web Dashboard Development

The complete framework is deployed using a Flask-based web application. Flask handles backend processing, model inference requests, authentication, and communication between different system components. The frontend is developed using HTML, CSS, and JavaScript to provide an interactive and user-friendly interface. Plotly.js is integrated to generate dynamic charts for visualizing stress and fatigue levels in real time. The dashboard supports user registration, login, profile management, historical record analysis, and real-time monitoring. Alert notifications are generated whenever stress or fatigue levels exceed predefined thresholds. The system also provides stress-relief recommendations and wellness suggestions to assist users in managing their mental health.

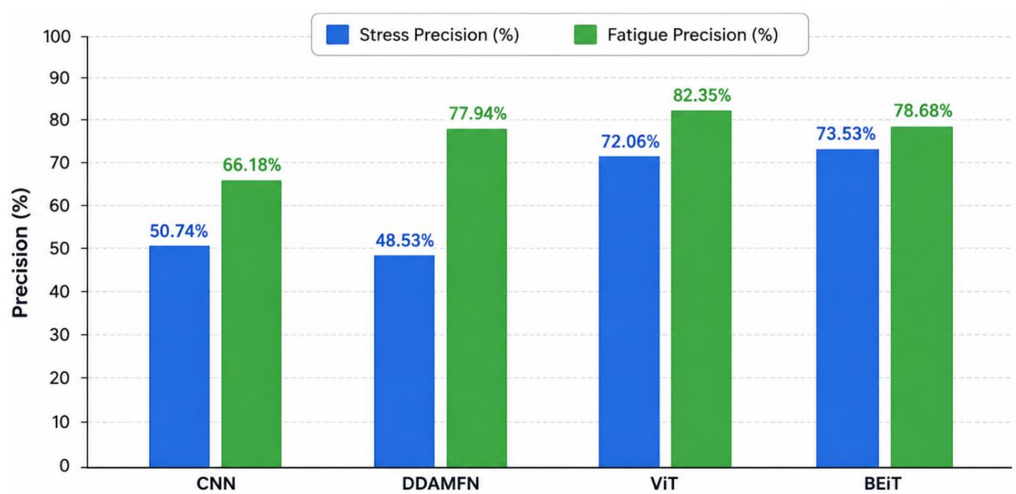
5.5 System Integration and Deployment

All modules are integrated into a unified framework to enable seamless operation. The webcam continuously captures facial images, which are processed by the preprocessing module and forwarded to the transformer models. The generated predictions pass through the emotion-weighting and ensemble-learning modules before being displayed on the dashboard. Prediction results and user information are stored in the database for future analysis and reporting. The deployed system provides low-latency performance, continuous monitoring, and reliable stress detection. The modular architecture allows future enhancements such as mobile deployment, cloud integration, and advanced multimodal stress monitoring capabilities.

The implemented StressLens framework successfully combines computer vision, transformer-based deep learning, emotion-aware analysis, and web technologies to create a practical, scalable, and intelligent solution for real-time facial stress and fatigue recognition.

VI. RESULT

The results show that both transformer models significantly outperform traditional CNN-based approaches. The ViT model achieved the highest fatigue recognition accuracy, while the BEiT model provided the best stress recognition performance. The improved results can be attributed to the ability of transformer architectures to capture long-range spatial relationships and contextual facial information. The emotion-weighting mechanism further enhanced prediction accuracy by assigning higher importance to stress-related emotions such as anger, sadness, and fear. This approach enabled the system to generate more meaningful stress estimates compared to direct emotion classification methods. The weighted emotional analysis reduced false predictions and improved the overall reliability of stress assessment.



Model	Stress Precision (%)	Fatigue Precision (%)
CNN	50.74	66.18
DDAMFN	48.53	77.94
ViT	72.06	82.35
BEiT	73.53	78.68

Figure 5: Performance Comparison of Deep Learning Models

The majority-voting ensemble strategy played a significant role in improving classification stability. By combining predictions from both ViT and BEiT models, the system minimized fluctuations caused by temporary facial expression changes, illumination variations, and partial facial occlusions. The ensemble mechanism produced more consistent stress and fatigue predictions during continuous monitoring sessions.

The Flask-based web dashboard successfully performed real-time monitoring with low latency and smooth visualization. Interactive charts generated through Plotly.js provided clear graphical representations of stress and fatigue trends. The alert module effectively generated notifications whenever stress or fatigue levels exceeded predefined thresholds, while the recommendation module provided personalized wellness suggestions to support stress management. The experimental analysis demonstrates that the proposed StressLens framework provides an efficient, scalable, and non-invasive solution for facial stress and fatigue recognition. The integration of transformer-based feature extraction, emotion-weighted analysis, and ensemble learning significantly improves prediction accuracy and robustness. These results confirm the suitability of the proposed system for deployment in healthcare, workplace wellness monitoring, educational institutions, and intelligent mental health applications.

VII. CONCLUSION

The proposed StressLens framework presents an effective and non-invasive approach for real-time facial stress and fatigue recognition using transformer-based deep learning techniques. The system successfully integrates Vision Transformer (ViT) and Bidirectional Encoder Representation from Image Transformers (BEiT) models with an emotion-weighting mechanism and ensemble majority-voting strategy to improve prediction accuracy and reliability. By utilizing facial expressions captured through a webcam, the framework eliminates the need for wearable sensors and physiological monitoring devices, making it more practical and user-friendly for continuous mental wellness monitoring. Experimental results demonstrate that transformer-based architectures outperform traditional CNN-based approaches in both stress and fatigue recognition tasks. The emotion-weighting mechanism enhances prediction quality by assigning greater importance to stress-related emotions, while the ensemble learning approach improves classification stability under varying environmental conditions. The Flask-based web dashboard provides real-time monitoring, graphical visualization, alert generation, historical analysis, and personalized recommendations, enabling a complete and interactive user experience. The proposed system can be effectively applied in healthcare, workplace wellness programs, educational institutions, and smart mental health applications. Overall, StressLens demonstrates the potential of transformer-based computer vision systems in advancing intelligent stress monitoring technologies and supporting proactive mental health management through accurate and real-time facial analysis.

VIII. REFERENCES

[1] M. Abd Al-Alim, “A machine-learning approach for stress detection using wearable sensors in free-living environments,” Computers in Biology and Medicine, vol. 170, 2024.

[2] G. Ghaderi, M. Heldt, and K. H. Plötz, “Detection and monitoring of stress using wearables: A systematic review,” Frontiers in Computer Science, vol. 6, 2024.

[3] D. S. Bhatia and S. Goel, “An explainable deep learning approach for stress detection in physiological sensor data,” Sensors, vol. 24, no. 16, pp. 1–20, 2024.

[4] M. Bolpagni, S. Pardini, M. Dianti, and S. Gabrielli, “Personalized stress detection using biosignals from wearables: A scoping review,” Sensors, vol. 24, no. 10, 2024.

[5] R. A. Kankal et al., “Stress monitoring using machine learning, IoT and wearable sensors,” Sensors, vol. 23, no. 21, 2023.

[6] S. R. Mohammed and A. G. Hassan, “Stress monitoring using wearable sensors: IoT techniques and physiological signal analysis,” Soft Computing, vol. 27, pp. 1–15, 2023.

[7] A. Jha and P. Singh, “Stress detection system using machine learning and wearable sensors,” International Journal of Advanced Research in Innovative Technology, vol. 11, no. 1, pp. 1–7, 2021.

[8] A. Rehman and M. H. Khan, “Feature-augmented hybrid CNN for stress recognition using a wrist-based photoplethysmography sensor,” arXiv preprint arXiv:2108.03166, 2021.

[9] M. Wataru and L. Rossi, “SELF-CARE: Selective fusion with context-aware low-power edge computing for stress detection,” arXiv preprint arXiv:2205.03974, 2022.

[10] T. C. Nkomo and V. K. Potdar, “Stress detection using context-aware sensor fusion from wearable devices,” arXiv preprint arXiv:2303.08215, 2023.

[11] Y. Wang et al., “TEANet: A transpose-enhanced autoencoder network for wearable stress monitoring,” arXiv preprint arXiv:2503.12657, 2025.

[12] R. S. Nair and J. Mathew, “Stress detection through wearable sensors: A CNN-based approach using heart rate and step data,” International Journal of Machine Learning Research, vol. 55, 2025.

[13] P. Sharma and A. Tiwari, “Real-time stress monitoring, detection, and management in college students: A wearable technology and machine-learning approach,” Journal of Biomedical Informatics Research, vol. 13, pp. 1–12, 2025.

[14] A. Patel and S. Gupta, “Predicting stress levels using physiological data: A real-time stress prediction review,” Journal of Medical Informatics, vol. 45, 2024.

[15] H. Liu et al., “A multimodal stress detection dataset with facial expressions and wearable sensor data,” Scientific Data, vol. 12, 2025.

[16] M. H. Khezri and M. Nazari, “Stress level monitoring system using IoT and machine learning,” International Journal of Advanced Computer Science and Applications, vol. 11, no. 5, pp. 214–219, 2020.

[17] R. Salabun, “Heart rate variability and stress monitoring in smart wearable systems,” IEEE Access, vol. 8, pp. 79550–79560, 2020.

[18] N. K. Jain and M. Bansal, “IoT based real-time stress monitoring system using biomedical sensors,” International Conference on Smart Systems and Inventive Technology, pp. 624–628, 2019.

[19] T. Singh and V. K. Verma, “A machine learning approach for real-time human stress detection,” IEEE International Conference on Advanced Networks and Telecommunications Systems, pp. 45–50, 2021.

[20] P. Schmidt et al., “Wearable-based stress detection using multimodal sensor data,” IEEE Internet of Things Journal, vol. 7, no. 6, pp. 4999–5008, 2020.