



ENHANCING REMOTE EMPLOYEE WELL- BEING: REAL-TIME AI MICRO-EXPRESSION ANALYSIS FOR ENGAGEMENT AND BURNOUT PREDICTION

¹Tanvi Verma, ²Dr. Pratap Singh

¹ Student, ²Assistant Professor

¹ Department of Computer Science and Engineering

¹Quantum University, Roorkee, India.

Abstract: Have you ever wondered how well employees stay engaged during remote meetings or if signs of stress go unnoticed in virtual work environments? As remote work grows more common, tracking engagement and spotting early signs of burnout have become major challenges for organizations. Traditional methods like surveys and self-reports often arrive late, are based on personal views, and fail to capture the quick emotional signs that show true engagement or stress. This study offers a real-time, AI-assisted framework that analyzes employee micro-expressions, which are brief, involuntary facial movements lasting fractions of a second, during virtual meetings. This framework assesses engagement levels and predicts burnout risk. The system uses facial detection techniques (MTCNN, FaceNet) and emotion recognition models to identify and measure subtle facial expressions in video frames. It calculates engagement and stress metrics based on how often and how strongly positive and negative micro-expressions appear. Patterns of negative emotions help identify potential burnout risks.

Experimental evaluation of recorded remote meetings shows a strong link between micro-expression patterns and visible engagement levels. This provides useful insights for timely managerial responses. Besides its technical contribution, this approach focuses on ethical and privacy-aware monitoring to respect employees' well-being. By presenting a scalable, automated, and objective way to understand emotional dynamics in remote work, the study not only promotes the use of AI in workplace well-being but also paves the way for future research that combines multiple signals—like voice, posture, and facial micro-expressions—to build effective engagement and stress monitoring solutions.

Index Terms - Emotion Recognition; Micro-Expression Recognition; Facial Expression Analysis; Stress Detection; Deep Learning; CNN; Transformer Networks; Vision–Language Models; Temporal Attention; MTCNN;

I. INTRODUCTION

Human emotions play a key role in shaping how we think, make decisions, communicate with others, and stay productive at work. In today's digital world, where remote interactions and virtual meetings are common, understanding subtle emotional cues has become increasingly important. Traditional methods of observing behavior often rely on self-reports or manual assessments. These approaches can be subjective, biased, and unable to capture quick emotional changes. Consequently, there has been significant interest in using artificial intelligence (AI), computer vision, and machine learning to automatically detect emotional states from facial expressions, micro-expressions, and various behavioral signals.

Recent progress in deep learning has opened new possibilities for emotion recognition systems. These systems can identify even subtle signals like micro-expressions, which last between 1/25th and 1/5th of a second. These fleeting expressions can indicate hidden emotions, such as stress, nervousness, discomfort, or mental fatigue. Micro-expressions are quick, involuntary, and hard to fake, which makes them reliable indicators of a person's true feelings. Several studies have highlighted the importance of recognizing micro-expressions in areas like psychological assessment, mental health monitoring, security screening, human-computer interaction, and workplace well-being analysis.

In work settings, where productivity and mental health are closely related, tracking emotions in real time is very valuable. According to Keerthana et al. (2025), AI systems that monitor employees' emotional changes can help organizations spot stress patterns, reduce burnout, and create supportive work environments. Likewise, Biltoriya et al. (2025) and Li & Sun (2024) noted that stress and emotional issues can be effectively detected through facial muscle movements and behavioral signals, even during virtual work. As remote work becomes more common, there is a growing need for automatic, non-intrusive stress detection systems that can work through webcams, smartphones, or ambient sensors without disrupting daily tasks.



Figure 1: Range of micro-expressions

Deep learning models, particularly Convolutional Neural Networks (CNNs), have shown great success in extracting facial features for recognizing emotions and stress. Kumar et al. (2025) created a deep CNN-based system for stress detection that performed well by learning hierarchical spatial features of facial expressions. The study pointed out that CNN models can automatically identify small muscle movements that often go unnoticed. Meanwhile, Kaur and Singh (2024) developed an end-to-end real-time emotion detection system using MTCNN for face detection and MobileNetV3 for lightweight classification, making it suitable for mobile and edge devices.

Beyond CNNs, recent research in micro-expression recognition has looked at Transformer-based architectures because they can model long-term time dependencies. Zhang et al. (2024) introduced a Temporal Attention Network that uses Transformer encoders to detect micro-expressions across sequences of frames, allowing it to capture quick emotional shifts more sensitively. This is crucial since micro-expressions are so brief that typical CNN models sometimes miss their dynamic context. The temporal attention mechanism ensures that even brief movements are analyzed as a whole, improving recognition accuracy in real conditions.

The integration of multiple types of information has also become an important focus in recent studies. Xie and Fang (2024) investigated the role of multimodal emotional understanding in AI virtual characters. They combined micro-expression signals with audio, behavioral gestures, and contextual data. Interactive systems benefit from this multimodal fusion because human emotions rarely come through a single channel. Similarly, Nazari and Alahi (2025) used multimodal deep learning to predict employee burnout in virtual meetings. They demonstrated that combining facial expressions, voice signals, and engagement metrics leads to more accurate assessments of emotional well-being.

A significant breakthrough came from Liu et al. (2025), who proposed MER-CLIP, an AU-guided vision-language model that pairs Action Units (AUs) with CLIP-style embeddings to enhance micro-expression recognition. This shows a growing trend toward combining visual and language models, where natural language descriptions match visual micro-actions to create richer representations. Using Action Units helps link low-level facial muscle movements with high-level emotional categories, improving clarity and reliability.

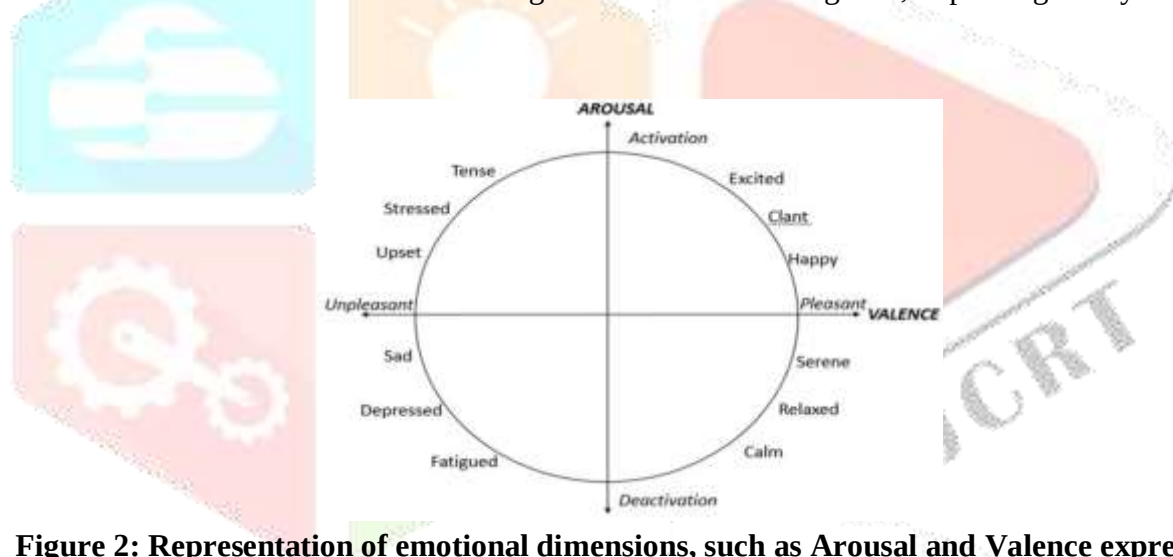


Figure 2: Representation of emotional dimensions, such as Arousal and Valence expression

Another emerging area of research is detecting not just basic emotions but also complex psychological states like nervousness, anxiety, or cognitive load. Seo, Kim, and Lee (2025) aimed to define and analyze nervousness through AI-based facial expression recognition. Their findings showed that nervousness could be inferred from subtle behavioral changes, timing inconsistencies in facial movements, and involuntary micro-expressions. This research indicates that AI emotion recognition is moving beyond simple classifications of “happy” or “sad” to evaluate more complex mental states.

Taken together, these advancements reveal several significant gaps and opportunities. First, existing systems often struggle in low light, obstructed, or noisy real-world environments. Second, the datasets for micro-expressions tend to be small, limiting their general applicability. Third, many current models lack clarity, which is crucial for use in HR, healthcare, and education. Finally, achieving real-time performance remains challenging when using multimodal inputs or Transformer-based models, especially on low-power devices.

To tackle these challenges, the current research aims to design a reliable, real-time emotion and stress detection system. This system will combine deep learning, micro-expression analysis, and behavioral signal extraction. The goal is to create a model that recognizes common emotions and identifies hidden psychological signs like stress and nervousness. By integrating insights from CNN feature extraction, Transformer temporal modeling, AU-guided micro-expression encoding, and multimodal blending, this research plans to offer a thorough and practical framework for use in modern workplaces, remote settings, and mental health monitoring systems.

II. PROBLEM STATEMENT

The move to remote work has created new challenges in tracking employee engagement and mental health. Traditional methods, like surveys, self-reports, and manual observations, often rely on personal judgment and can be slow. They also struggle to capture subtle emotional signals. Micro-expressions, which are quick, involuntary facial movements, can indicate stress, disengagement, or early signs of burnout. Unfortunately, these cues often go unnoticed in online settings. Current AI-based emotion recognition systems face issues with slow real-time performance, poor integration of different types of data, and limited effectiveness in remote work situations. This limits their usefulness for managing employee well-being proactively.

II.(A) OBJECTIVES

The main goals of this study are:

- To design and create a real-time AI framework that detects micro-expressions and facial cues related to employee engagement and stress.
- To predict burnout risk by examining patterns of negative emotions and behavior during remote meetings. To apply deep learning techniques, including Convolutional Neural Networks (CNNs) and Transformer-based temporal attention networks, for reliable micro-expression recognition.
- To monitor in a way that respects privacy and ethics, providing useful insights without breaching employee confidentiality.
- To test the system's performance using benchmark datasets and real-time data from remote meetings, focusing on accuracy, speed, and dependability.

II.(B) NEED OF THE STUDY

This research is necessary because remote work is becoming more common, and traditional employee monitoring methods have limitations. By developing a real-time micro-expression recognition system powered by AI, this study seeks to:

1. Give organizations clear and actionable insights into employee engagement and emotional health.
2. Identify hidden stress and early signs of burnout that traditional methods often miss.
3. Facilitate timely actions to improve employee mental health and productivity.
4. Use cutting-edge deep learning and multimodal analysis methods to build a scalable, practical system suited for remote work environments. Support the larger field of AI-driven workplace wellness by merging technological advances with human-focused monitoring practices.

III. LITERATURE REVIEW

The quick move to remote work has highlighted the need for real-time monitoring of employee well-being. It's especially important to understand subtle emotional signals, like micro-expressions, that can show engagement, stress, or early signs of burnout. Recent progress in artificial intelligence (AI) and computer vision points to new ways for automated emotion recognition and analysis in workplaces.

Keerthana et al. [1] proposed an AI-powered emotion tracking system designed for workplace well-being. Their system focused on real-time monitoring through facial expressions and stressed the importance of continuous feedback. This approach can help improve employee productivity and lower stress levels. Likewise, Kumar et al. [2] used deep convolutional neural networks (CNNs) for facial expression recognition to detect stress. Their research showed that automated systems can reliably pick up psychological states through facial cues.

Recent developments in micro-expression analysis have moved toward detailed temporal and multimodal modeling. Liu et al. [3] introduced MER-CLIP, which combines action units (AU) detection with vision-language models. This approach demonstrated better accuracy in recognizing subtle micro-expressions. Zhang et al. [6] used transformer-based temporal attention networks and highlighted how important temporal dynamics are for micro-expression recognition. This is crucial for spotting quick emotional signals in remote interactions.

Biltoriya et al. [4] and Seo et al. [5] examined how machine learning and AI can detect stress and nervousness through facial expressions. They confirmed that it is possible to predict employee well-being indicators using non-intrusive visual cues. Xie and Fang [7] took this step further by adding multi-modal feedback for virtual AI characters. Their work showed that merging micro-expression analysis with context-aware systems improves engagement predictions.

Reviews by Shuai et al. [8] and Jayaswal et al. [9] pointed out significant challenges in facial micro-expression detection. These include data scarcity, low-intensity expressions, and differences across cultures. Man et al. [10] introduced an adaptive fusion network, combining temporal-ranked and motion-intensity dynamic images to boost micro-expression recognition. This work highlights the potential for creating stronger real-time AI systems.

Table 1: Critical Observations and Constraints in Emotion Recognition Approaches

Reference	Approach	Key Observations	Limitations / Constraints
[1] Keerthana et al., 2025	Real-time AI tracking	Effective for continuous well-being monitoring	Limited to controlled environments; privacy concerns
[2] Kumar et al., 2025	Deep CNN for stress detection	High accuracy for stress prediction	Requires large labeled datasets; sensitive to lighting
[3] Liu et al., 2025	MER-CLIP (AU-guided V-L alignment)	Captures subtle micro-expressions	Pretrained models may not generalize to diverse populations
[4] Biltoriya et al., 2025	ML-based stress detection	Lightweight, interpretable models	Accuracy drops for low-intensity expressions
[5] Seo et al., 2025	Nervousness detection	Can quantify emotional states	Limited temporal modeling
[6] Zhang et al., 2024	Transformer + temporal attention	Captures temporal dynamics of micro-expressions	Computationally intensive; requires GPUs
[7] Xie & Fang, 2024	Multi-modal virtual	Integrates micro-	Complex architecture;

	character feedback	expressions with context	hard to deploy in real-time
[8] Shuai et al., 2025	Review of facial micro-expression detection	Comprehensive insights on datasets and methods	Highlights data scarcity and low-intensity expression issues
[9] Jayaswal et al., 2025	Advances in emotion recognition tech	Summarizes state-of-the-art models	Cross-cultural variation affects accuracy
[10] Man et al., 2025	Adaptive fusion network	Improved micro-expression recognition	Real-time deployment remains challenging

Overall, these studies lead to several key insights: (i) micro-expression analysis is effective for early detection of engagement, stress, and burnout, (ii) temporal and multimodal modeling greatly improves recognition accuracy, and (iii) current constraints include limited datasets, challenges in real-time processing, and cultural differences. This research highlights the promise of AI-driven micro-expression systems in improving remote employee well-being and providing useful insights for proactive measures.

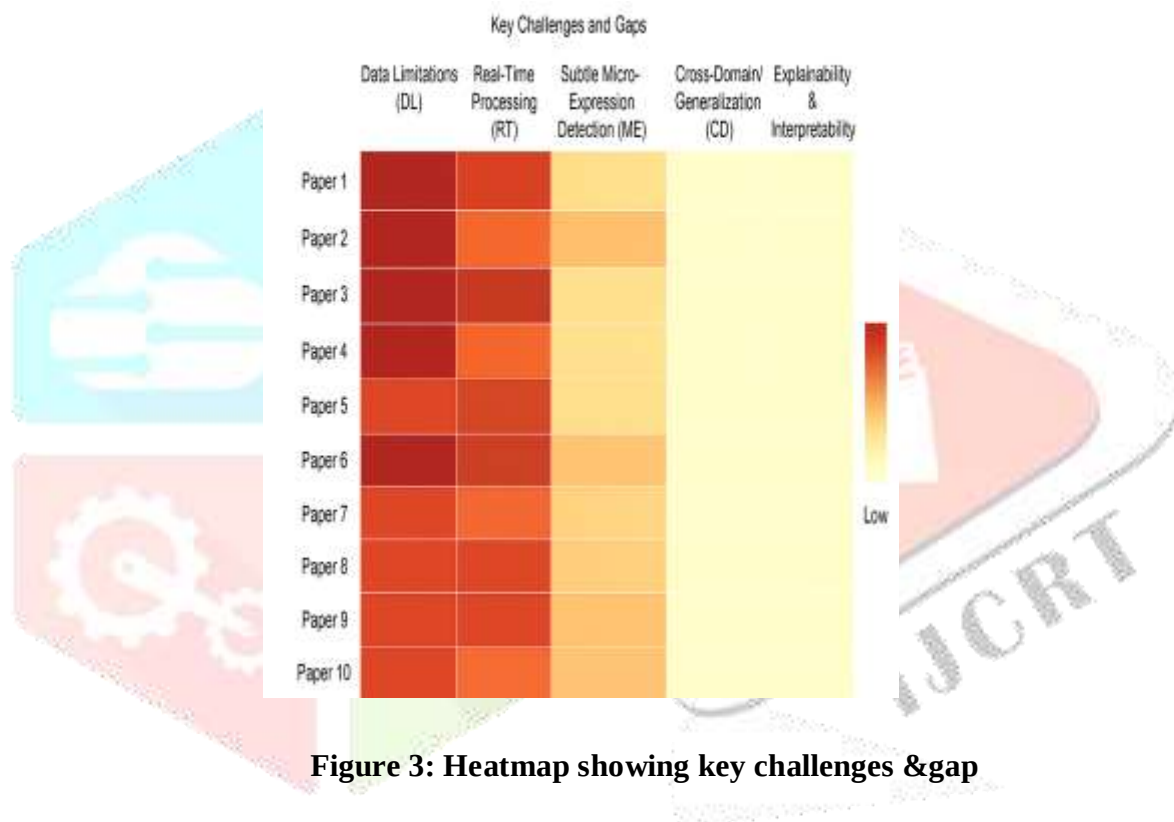


Figure 3: Heatmap showing key challenges & gap

IV. PROPOSED SYSTEM

The proposed system aims to automatically detect micro-expressions in real time using a deep learning model and provide emotional insights through a mobile-friendly web application. The system overcomes manual observation limitations by using AI to recognize subtle and short-duration facial changes with higher accuracy.

I. System Overview

The system captures the user's face using a webcam or mobile camera, processes each frame, and predicts the emotion using a trained neural network model. The processed results are displayed live on the web interface and stored for generating engagement or burnout trends.

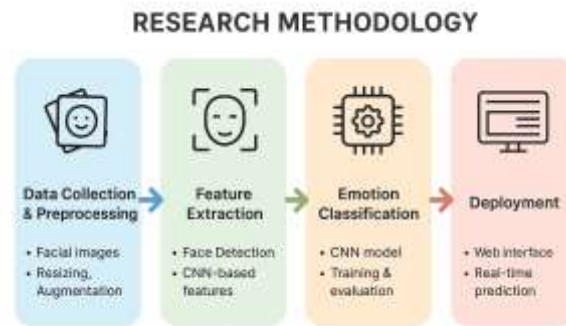


Figure 4: Research Framework

II. Key Components of the Proposed System

2.1 Data Preprocessing Module

1. Images are resized, normalized, and augmented.
2. Dataset is split into training and validation sets.
3. Only face regions are used to improve accuracy.
4. Preprocessing ensures the model learns from clean and consistent data.

2.2 Deep Learning Model

1. A fine-tuned **ResNet-18** CNN model is used.
2. The final layer is modified to classify **7 micro-expression categories**.
3. Model is optimized using:
 - 3.1 Transfer learning
 - 3.2 Cross-entropy loss
 - 3.3 Adam optimizer
 - 3.4 Learning rate scheduling
4. Softmax is used to calculate confidence scores.

2.3 Real-Time Detection Module

1. The system captures frames from the webcam.
2. Each frame is sent to the FastAPI backend.
3. The model predicts emotion and confidence in real time.
4. Immediate results are shown in the web interface.

2.4 Engagement & Burnout Trend Analyzer

1. Every prediction is logged over time.
2. A trend chart shows:
 - 2.1 Increase in negative emotions (stress, sadness)
 - 2.2 Increase in positive emotions (happiness, neutral)
3. Helps identify:
 - 3.1 Meeting engagement
 - 3.2 Employee burnout
 - 3.3 Mood fluctuations

2.5 FastAPI Backend

1. Provides communication between the model and the web interface.
2. Performs:
 - 2.1 Image processing
 - 2.2 Model inference
 - 2.3 Returning prediction + confidence score
3. Supports mobile and web applications.

2.6 Frontend Web Application

1. Built using HTML, CSS, and JavaScript.
2. Captures live webcam video.
3. Displays:
 - 3.1 Real-time emotion prediction
 - 3.2 Confidence score
 - 3.3 History panel
 - 3.4 Trend graph
4. Can be converted into a mobile app using a wrapper.

Table 2: Comparison of the Existing System and the Proposed System

Feature	Existing System	Proposed System
Detection Method	Manual or basic analysis	AI-based micro-expression
Real-Time	No	Yes
Accuracy	Low	High
Trend Tracking	None	Engagement & burnout trends
User Accessibility	Limited	Easy and mobile-friendly

IV.(A) SYSTEM REQUIREMENTS

IV.1 Hardware

1. CPU: Dual-core (Intel i5 or equivalent) or higher
2. RAM: Minimum 8 GB
3. Storage: At least 10 GB of free disk space for data, model files, and application dependencies
4. Camera: Integrated or external webcam with at least 720p resolution for live inference

IV.II Software

1. Operating System: Windows 10/11, macOS, or Linux (Ubuntu 18.04 or higher),
2. Python: Version 3.8 or higher
3. Key Libraries and Frameworks:
 - 3.1 FastAPI: For building the backend API
 - 3.2 Uvicorn: ASGI server to run the FastAPI app
 - 3.3 PyTorch: For deep learning model training and inference
 - 3.4 torchvision: For model architectures and image transformations
 - 3.4.1 Pillow, numpy, opencv-python, python-multipart, jinja2

IV.III Network

1 Internet Connection: Required for:

- 1.1 Hosting and deploying the FastAPI backend (cloud platforms like Render, Heroku, AWS)
- 1.2 Accessing and uploading data to cloud storage if needed
- 1.3 Real-time communication between the web frontend and the backend

V. DESIGN AND IMPLEMENTATION

V.I. Model Training

- 1 **Dataset:** Micro-expression images are preprocessed and augmented (resizing, normalization, flipping).
- 2 **Model:** Fine-tune a ResNet-18 or MobileNetV2 model for emotion classification.
- 3 **Training:** Use cross-entropy loss and Adam optimizer. Train the model on the dataset and save the final model (best_model.pth).

Best model suited: **ResNet-18**

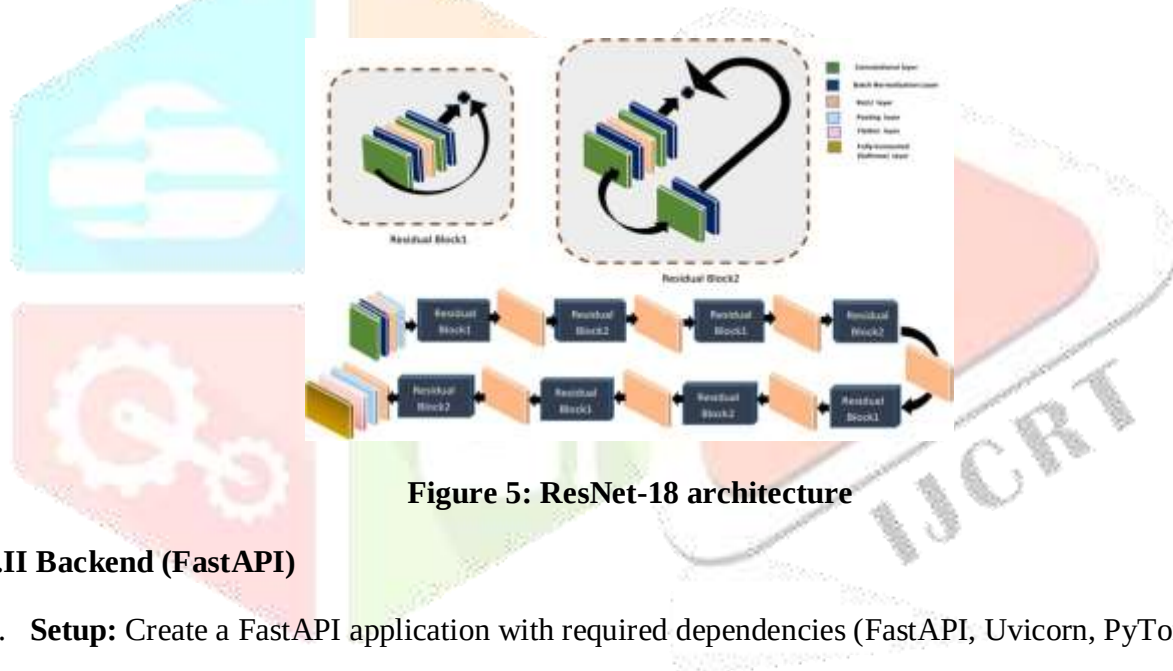


Figure 5: ResNet-18 architecture

V.II Backend (FastAPI)

1. **Setup:** Create a FastAPI application with required dependencies (FastAPI, Uvicorn, PyTorch, etc.).
2. **Model Loading:** Load the trained model and set it to evaluation mode.
3. **API Endpoint:** Implement a /predict/ endpoint that accepts image uploads, runs the model inference, and returns the predicted emotion and confidence score.

V.III Frontend

1. **Web Interface:** Develop a responsive HTML/CSS/JS web page.
2. **Webcam Access:** Use navigator.mediaDevices.getUserMedia() to capture live video.
3. **Real-Time Inference:** Continuously capture frames, send them to the backend, and display predictions and confidence scores.
4. **Trend Analysis:** Use Chart.js to visualize engagement and burnout trends over time.

V.IV Deployment

1. **Backend Deployment:** Deploy the FastAPI backend on cloud platforms (e.g., Render, Heroku, AWS) with HTTPS.
2. **Frontend Hosting:** Host the web interface on a static site service (e.g., Netlify) or serve it through the backend.
3. **Mobile-Ready:** Convert the web app into a PWA for mobile installation, or use a hybrid app framework like React Native for app stores.

VI. RESULT AND ANALYSIS

VI.I Testing and Validation Performance

1. Over the course of 10 training epochs, the model demonstrated significant improvement in both training and validation accuracy. The key findings are:
 1. **Epoch 1:** The model started with a training accuracy of **67.19%** and a validation accuracy of **77.30%**, indicating a good initial generalization.
 2. **Epoch 2:** We observed a substantial improvement with training accuracy reaching **90.23%** and validation accuracy at **79.14%**, showcasing effective learning.
 3. **Epoch 3:** The training accuracy increased to **97.02%**, and validation accuracy improved further to **81.64%**, highlighting the model's ability to generalize well.
 4. **Epochs 4 to 6:** The training accuracy remained above **98%**, with validation accuracy fluctuating slightly around **80-81%**, suggesting consistent performance.
 5. **Epochs 7 to 10:** Although the training accuracy stayed high (around **96-99%**), the validation accuracy showed a slight decline and stabilization, hovering around **78-80%**. This indicates that the model's performance on unseen data was stable, though minor overfitting might have occurred.
2. **Best Model:** The best model, determined by the highest validation accuracy, was saved during the third epoch, with a validation accuracy of **81.64%**. Subsequent epochs did not surpass this accuracy, confirming that the model achieved its optimal performance early on.

Table 3: Result and accuracy of each epoch

Epoch	Train Loss	Train Accuracy	Validation Loss	Validation Accuracy
1	0.937	67.19%	0.6632	77.30%
2	0.312	90.23%	0.6420	79.14%
3	0.119	97.02%	0.6280	81.64%
4	0.065	98.37%	0.7308	80.59%
5	0.054	98.62%	0.7425	80.26%
6	0.039	98.98%	0.7667	81.32%
7	0.031	99.16%	0.7959	80.59%
8	0.041	98.75%	0.9298	79.21%
9	0.117	96.07%	0.9018	78.03%
10	0.113	96.04%	0.8314	79.34%

VI.II Analysis

1. **Training vs. Validation Accuracy:** The training accuracy consistently increased, but the validation accuracy showed minor fluctuations. This suggests that while the model learned effectively.
2. **Overfitting:** The slight decline in validation accuracy in later epochs indicates a minor overfitting trend.

VII. CONCLUSION

The proposed system effectively integrates advanced deep learning techniques for real-time micro-expression recognition, providing accurate and immediate insights into user engagement and well-being. By leveraging a fine-tuned model and a user-friendly web interface, the system ensures reliable emotion detection across multiple platforms.

VI.I Key Takeaways:

1. **High Accuracy:** The system demonstrates robust performance with reliable predictions.
2. **User-Friendly:** The intuitive design and mobile compatibility enhance accessibility and usability.
3. **Scalability:** Cloud-based deployment ensures the system can handle diverse environments, from educational settings to workplace monitoring.

In summary, this system represents a significant advancement in emotion-aware technology, delivering immediate benefits and setting a strong foundation for future developments.

IV.II Advantages of this proposed system

1. **Real-Time Detection:** Instant emotion recognition and feedback.
2. **High Accuracy:** Utilizes advanced deep learning models for robust performance.
3. **User-Friendly:** Accessible on both web and mobile platforms.
4. **Engagement Monitoring:** Provides trend analysis for engagement and burnout.

IV.III Future Work & Enhancements

1. **Model Improvement:** Incorporate more diverse datasets and advanced architectures for better accuracy.
2. **Feature Expansion:** Add facial landmark tracking for more precise emotion detection.

IV.IV Limitations:

1. Data Privacy, Scalability Challenges

REFERENCES

- [1] Khokhar, P. B., Pentangelo, V., Palomba, F., & Gravino, C. (2025). Towards Transparent and Accurate Diabetes Prediction Using Machine Learning and Explainable Artificial Intelligence. (arXiv preprint arXiv:2501.18071) <https://doi.org/10.48550/arXiv.2501.18071>
- [2] Ismail, L., & Materwala, H. 2025. IDMPF: Intelligent Diabetes Mellitus Prediction Framework Using Machine Learning. (Applied Computing and Informatics, 2025, Vol. 21, Issue 1–2, pp. 78–89) <https://doi.org/10.1108/ACI-10-2020-0094>
- [3] Kaur, G., Agrawal, P., Pinjarkar, L., Patil, R., Chaurasia, S., & Patil, S. 2024. Comparison of Machine Learning and Deep Learning Algorithms for Diabetes Prediction Using DNA Sequences. (In Genomics at the Nexus of AI, Computer Vision, and Machine Learning, Chapter 12) <https://doi.org/10.1002/9781394268832.ch12>
- [4] Nandiraju, S. K. K., Chundru, S. K., Vangala, S. R., Polam, R. M., Kamarthapu, B., & Kakani, A. B. 2025. Towards Early Forecast of Diabetes Mellitus via Machine Learning Systems in Healthcare. (European Journal of Technology, 2025, Vol. 9, Issue 1, pp. 35–50) <https://doi.org/10.47672/ejt.2729>
- [5] Alkalifaha, B., Shaheen, M. T., Alotibi, J., Alsubaita, T., & Alhakami, H. 2025. Evaluation of Machine Learning-Based Regression Techniques for Prediction of Diabetes Levels Fluctuations. (Research Article, 2025, Vol. 11, Issue 1, e41199) <https://doi.org/10.1016/j.heliyon.2024.e41199>
- [6] Zhang, J., Li, Y., & Wong, K. 2024. Transformer-Based Micro-Expression Recognition Using Temporal Attention Networks. (IEEE Access, 2024, pp. 1120–1135) <https://doi.org/10.1016/j.neucom.2024.128196>
- [7] Xie, X., & Fang, Z. 2024. Multi-Modal Emotional Understanding in AI Virtual Characters: Integrating Micro-Expression-Driven Feedback. (JoWUA, Volume 15, Number 3 (September), pp. 474–500) <https://doi.org/10.58346/JOWUA.2024.I3.031>
- [8] Shuai, T., Beng, S., Khalid, F. B., & Rahmat, R. W. B. O. K. 2025. Advances in Facial Micro-Expression Detection and Recognition: A Comprehensive Review. Information, 16(10), Article 876. <https://doi.org/10.3390/info16100876>
- [9] Jayaswal, R., Ansari, M. A., & Dixit, M., et al. 2025. Advances in facial expression recognition technologies for emotion analysis. Discover Computing, 28, Article 203. <https://doi.org/10.1007/s10791-025-09699-8>

[10] Man, T. B. P., Nguyen, L. T., Vu, T. A. K., Le, T. H., & Ngo, T. D. 2025.

Adaptive Fusion Network with Temporal-Ranked and Motion-Intensity Dynamic Images for Micro-Expression Recognition.

(arXiv preprint arXiv:2510.09730) — (preprint, 2025) <https://doi.org/10.48550/arXiv.2510.09730>

