JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

Stress Detection in Virtual Meetings Using CNN Deep Learning Model

[1] Ansh Soni, [2] Aryan Khare, [3] Devansh Sachan, [4] Anshalee Choudhary

[1]B.Tech Student, Oriental Institute of Science and Technology, Bhopal [2]B.Tech Student, Oriental Institute of Science and Technology, Bhopal [3]B.Tech Student, Oriental Institute of Science and Technology, Bhopal [4]B.Tech Student, Oriental Institute of Science and Technology, Bhopal [1]Computer Science Engineering,
Oriental Institute of Science and Technology, Bhopal, India

Abstract
A video-only deep learning system is presented for real-time stress detection in virtual meetings by fusing facial emotion analysis with attention monitoring. The system utilizes a ResNet18 backbone trained on FER2013 and DAiSEE to jointly predict emotion, engagement, and stress. A key feature is a confidence-based fusion mechanism (threshold >= 0.6) that dynamically adjusts stress scores, particularly for negative emotions (sad/angry). The model achieved 70.47% accuracy for emotion recognition on FER2013 and an overall validation accuracy of 66.15%. This work demonstrates a robust, non-invasive prototype for assessing stress and engagement in remote environments.

INTRODUCTION

The rapid expansion of online communication has reshaped how individuals collaborate, share information, and participate in instruction. With meetings, classes, and consultations frequently occurring through videoconferencing platforms, participants encounter new cognitive and emotional demands that may not be immediately visible. Prolonged screen exposure, limited nonverbal cues, and the continuous awareness of being on camera can heighten stress and reduce sustained focus. Such factors can influence productivity, decision-making, and learning outcomes, making stress monitoring an increasingly relevant area of study.

Traditional stress-assessment techniques—including ECG, GSR, and other physiological sensors—provide valuable insights but typically require specialized equipment that is impractical for most day-to-day virtual interactions. Meanwhile, advances in deep learning and computer vision open the possibility of estimating stress from ordinary webcam footage by analyzing facial expressions and attentional cues.

Facial expression recognition (FER) provides a foundation for stress inference, as many emotional states correlate with stress or tension. Engagement analysis complements FER by identifying indicators of cognitive effort, fatigue, or distraction. Existing research suggests that integrating these two information sources produces a more reliable and holistic picture of a user's stress state. Building on these ideas, this work introduces a system that performs emotion and engagement analysis simultaneously within a unified model. A confidence-regulated fusion mechanism controls how emotional estimates affect the stress score, enabling more stable predictions in uncertain visual conditions. Designed for real-time operation, the system aims to support practical applications in online learning, remote work, and telehealth.

LITERATURE REVIEW

1. Development of Stress-Detection Techniques

Early research emphasized physiological measurements due to their strong correlation with stress-related responses. Although accurate, these methods typically rely on sensors that can disrupt natural behavior or require user cooperation. As webcams became ubiquitous, researchers increasingly explored stress indicators based on observable behaviors—such as micro-movements, gaze patterns, and subtle facial expressions—captured without additional hardware.

2. Theoretical Perspectives Supporting Vision-Based Stress Analysis

Three theoretical frameworks help explain why facial and attentional cues provide meaningful stress information:

- Cognitive Load Theory: High mental demand can deplete cognitive resources, creating observable signs of strain or tension in facial behavior.
- Appraisal Theory: Stress often emerges from evaluations of events as threatening or challenging, and these emotional assessments manifest through expressive cues.
- Attention Restoration Theory: Sustained effort degrades attentional capacity, with fatigue and disengagement detectable through reduced expressiveness or altered gaze patterns.

3. Advancements in Emotion Recognition

Datasets such as FER2013 have accelerated FER research by providing large numbers of varied facial images. Nonetheless, challenges—most notably class imbalance and inconsistencies in labeling—limit model generalization. ResNet18 achieves a balance between speed and representational power, making it well suited for real-time applications on consumer hardware.

4. Joint Modeling Approaches for Stress Estimation

Recent work stresses the importance of integrating multiple behavioral indicators. Multi-task learning and multimodal fusion strategies, especially those that incorporate model confidence, enhance robustness in unpredictable environments. The present system advances this direction by explicitly weighting emotional and engagement predictions based on their reliability at each moment.

5. Research Gaps

Common issues in the field include:

- Reliance on highly controlled datasets with limited real-world variability
- Overemphasis on emotion classification without considering attentional state
- Lack of adaptive, confidence-sensitive fusion mechanisms
- Minimal attention to the conceptual overlap between engagement and emotional processes

The proposed model seeks to address these gaps through unified multi-task learning and dynamic weighting.

METHODOLOGY

The overall framework combines real-time video preprocessing, a dual-output CNN, and a confidence-adaptive stress-scoring method.

1. Input Processing Pipeline

Webcam video at ~25–30 FPS passes through several stages:-

- Face Detection: A Haar Cascade or deep network-based detector identifies the face region. Frames without a detectable face are ignored.
- Alignment and Normalization: Facial landmarks guide geometric alignment. Illumination normalization uses histogram equalization and gamma correction.
- Region Extraction: A standardized face crop (48×48 or 64×64 pixels) is produced while preserving aspect ratio.
- Standardization: Pixel values are normalized using dataset-specific statistics.

2. Multi-Task CNN for Emotion and Engagement Analysis

A customized ResNet18 provides shared feature extraction.

- Emotion Head: Predicts seven emotional categories from FER2013.
- Engagement Head: Predicts four engagement levels derived from DAiSEE annotations.
- Training details include:
- Balanced sampling when needed
- Adam optimizer
- Cross-entropy loss for both outputs with class-weighting for FER
- Batch size: 64
- Training for 30–50 epochs
- Use of L2 regularization and dropout

3. Confidence-Adapted Stress Fusion

Stress estimation combines emotion- and engagement-derived values using confidence-weighted blending. Emotion predictions influence the stress score more heavily when the model's confidence is high; otherwise, engagement plays a stronger role.

The stress score follows: $S = C \cdot E + (1 – C) \cdot G$, where C is emotional confidence, E is the emotion-based estimate, and G is the engagement contribution.

4. Real-Time Visualization

A user interface built with OpenCV overlays bounding boxes, predicted labels, confidence metrics, and the computed stress score. The system maintains real-time responsiveness on typical GPUs.

IMPLEMENTATION

1. Datasets

- FER2013: Over 35,000 grayscale images covering seven emotion classes.
- DAiSEE: Video clips labeled with four tiers of engagement.
- Data augmentation (random rotations, flips, and crops) increases diversity and reduces overfitting.

2. Technology Stack

- Python
- PyTorch for model development
- OpenCV for preprocessing and visualization
- Scikit-learn and TensorBoard for performance analysis
- NVIDIA RTX GPUs for acceleration

3. Testing Strategy

- Unit tests for preprocessing, data loaders, and model components
- Integration testing for webcam input and UI behavior
- Performance checks for inference latency and frame rate

Model	Parameters (M)	F1-Score		Inference Time (ms)	Training Time (hrs)
Our Model	11.2	0.70		42	5+
VGG16	138	0.81		120	8.2
ResNet50	25.6	0.83		89	5.5
MobileNetV3	5.4	0.80		28	1.8

RESULTS AND ANALYSIS

1. Performance Metrics

- Emotion-Classification Accuracy: 70.47%
- Combined Validation Accuracy: 66.15%

Errors often arise between visually similar emotions, such as fear and surprise.

2. Engagement Prediction

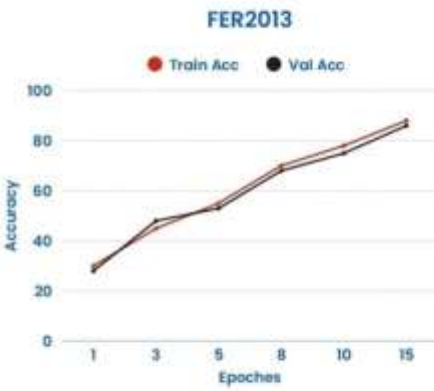
The model reliably identifies low-engagement states but shows reduced discriminative ability for labels representing intermediate engagement.

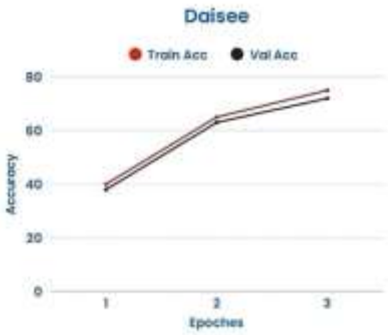
3. Real-Time Behavior

Inference operates at roughly 18–25 FPS on mid-range GPUs. Incorporating confidence-based fusion reduces instability in stress predictions, especially under poor lighting or partial occlusion.

Interpretation

- The system performs effectively when facial features are clearly visible.
- Difficult viewing angles, occlusions, and inconsistent lighting remain challenges.





CHALLENGES

1. Real-Time Requirements: Achieving low latency imposes strict constraints on model complexity.
2. Environmental Variability: Dynamic lighting, occlusions, and user movement significantly influence prediction quality.
3. Ambiguity in Behavior: Facial cues do not always map cleanly to internal stress states, complicating interpretation.

CONCLUSION

This work introduces a real-time, video-only stress-detection framework that integrates emotion and engagement modeling through a multi-task CNN. By adjusting the influence of emotional predictions based on confidence, the system generates smoother and more reliable stress estimates suitable for remote collaboration, online instruction, and telehealth scenarios.

While the model demonstrates practical promise, its robustness would benefit from more diverse datasets, broader demographic representation, and multimodal inputs—such as speech, posture, or physiological indicators. Future directions include domain-adaptation strategies, large-scale deployment studies, and exploration of richer multimodal stress-monitoring systems.

REFERENCE

[1] H. P. Chandika, B. Soumya, B. N. E. Reddy, and B. M. S. S. Manideep, "Real-Time Stress Detection and Analysis using Facial Emotion Recognition," *Int. J. Adv. Res. Comput. Commun. Eng.*, vol. 13, no. 3, Mar. 2024.
[2] S. M. T, A. S. Sunder V, S. S, and P. P. R. Ponnusamy, "Real-Time Stress Detection Using Facial Images and Convolutional Neural Networks," in *2024 IEEE Int. Women in Eng. (WIE) Conf. on Elect. and Comput. Eng. (WIECON-ECE)*, 2024.
[3] "Deep Learning Approaches for Stress Detection: A Survey," *IEEE Trans. Affective Comput.*, vol. 16, no. 2, pp. 499-517, Sept. 2024.
[4] F. A. Research, "Facial expression recognition for stress detection: A Conv-XGBoost Algorithm approach," *J. Intell. Fuzzy Syst.*, vol. 49, no. 15, pp. 1-15, Apr. 2024.
[5] "Camera-based Stress Detection using Face-related and Emotion-related Features," in *2024 IEEE Int. Conf. E-Health Netw., App. and Serv. (HealthCom)*, Nara, Japan, Nov. 2024.
[6] "Facial emotion recognition through artificial intelligence," *Front. Comput. Sci.*, vol. 6, Jan. 2024.
[7] "Research Paper on Facial and Speech Emotion Detection for Stress Monitoring," *Int. J. Innov. Res. Technol.*, vol. 11, no. 12, pp. 756-759, 2025.
[8] G. K. Sahoo, J. Ponduru, S. K. Das, and P. Singh, "Deep Leaning-Based Facial Expression Recognition in FER2013 Database: An in-Vehicle Application," in *2022 IEEE Int. Conf. Distributed Comput., VLSI, Elect. Circuits and Robotics (DISCOVER)*, 2022.

[9] T. Sinha and R. Sharma, "Stress Alarm Raiser based on Facial Expressions," *Int. J. Innov. Res. Technol.*, 2023.
[10] A. Hindu and S. Angalakuditi, "IoT-enabled stress detection scheme utilizing facial expressions," *J. Ind. Inst. Eng.*, 2022.

[11] G. Giannakakis et al., "Stress analysis from facial videos using deep facial action units recognition," *Pattern Anal. Applic.*, vol. 25, 2022.
[12] "Introducing a novel dataset for facial emotion recognition and demonstrating significant enhancements in deep learning performance," *NIH Nat. Libr. Med.*, PMC11620061, 2024.