

Vision-Based Cognitive Fatigue and Stress Detection by Artificial Intelligence and Computer Vision

Piyush Rastogi

Assistant Professor
Department of Computer
Science and Engineering
(CSE(AI/ML))
Moradabad Institute
of Technology,
Moradabad
piyushrastogi786@gmail.com

Divyanshi Saxena

Department of
Computer Science and
Engineering
(CSE(AI/ML))
Moradabad Institute
of Technology,
Moradabad
divyanshisaxena2002@gmail.com

Aayush Bhat

Department of
Computer Science and
Engineering
(CSE(AI/ML))
Moradabad
Institute of Technology,
Moradabad
aayushbhat45@gmail.com

Love Arora

Department of
Computer Science
and Engineering
(CSE(AI/ML))
Moradabad Institute
of Technology,
Moradabad
luvarora02@gmail.com

Munendra Yadav

Department of
Computer Science and
Engineering
(CSE(AI/ML))
Moradabad Institute
of Technology,
Moradabad
yadavmunendra425@gmail.com

Abstract - Cognitive fatigue and mental stress have become more frequent in today's digital environments of online learning and studying, office work with the usage of the computer, and driving tasks. Cognitive fatigue and mental stress impact attention, reaction time, productivity, and decision-making skills in a negative way. The current solutions for fatigue and mental stress detection are either in the form of self-assessed questionnaires that are intrusive and not appropriate for continuous monitoring, or in the form of physical sensors that are not very appropriate for the aforementioned environment.

This study proposes a vision-based, non-intrusive AI approach for real-time cognitive fatigue and stress analysis through facial behavior. The proposed method uses computer vision and AI algorithms like Convolutional Neural Networks (CNNs)/Long Short-Term Memory (LSTM) networks to identify facial markers, eye blink patterns, gaze, and head motion patterns present in live video captures via a basic webcam. The facial markers, eye blink patterns, gaze, and head movements are used as features to classify cognitive levels into normal, fatigued, or stressed.

This system works in real time without wearable technology and thus is scalable and privacy-conscious and cost-effective. The experiments show high accuracy and low latency and thus confirm the effectiveness of the proposed method in applications of education, worker monitoring, and human computer interaction systems.

Keywords: Cognitive Fatigue Detection, Stress Detection, Computer Vision, Deep Learning, Facial Landmark Analysis, Eye Blink Detection, Artificial Intelligence.

1. INTRODUCTION

Cognitive fatigue and stress are important factors that affect human performance in tasks that demand sustained visual and cognitive attention. Today, with the ever-growing adoption of digital solutions for learning and working, there is increasing exposure to extended screen time and cognitive fatigue. Cognitive fatigue causes reduced alertness and reaction time accuracy together with higher error rates, whereas chronic stress results in mental exhaustion and physical problems.

Current fatigue detection systems are mostly based on physiological parameters, for instance, EEG, HRV, and wearable sensors. While these systems are accurate, they are invasive, costly, and not convenient for use in fatigue detection. Fatigue detection through self-assessment is subjective and unreliable.

Recent developments in artificial Intelligence and Computer Vision offer a new method of detection through the evaluation of visual behavioral signals, such as facial expressions, eye movements, blinking rate, and head position. Such visual signals can be extracted with the help of common webcams, making it possible to monitor subjects without direct contact. The current study will introduce a vision-based detector of cognitive fatigue and stress, utilizing deep learning networks to evaluate facial and eye signals acquired directly from video streams.

2. THEORETICAL BACKGROUND

A. Computer Vision

Computer vision is the ability to interpret the images and videos captured by the camera and is applied for tasks such as facial detection, facial markers, and eye region analysis.

B. Facial Landmark Detection

Facial landmarks are considered to be crucial points on the face which include eyes, eyebrows, noses, and mouth. These landmarks are useful for tracking eye movements that are linked to fatigue and stress.

C. Eye Blink and Gaze Analysis

Eye blink rate, blink duration, and gaze deviation are good pointers for cognitive fatigue. Long blink durations and irregular gaze patterns are typical signs of mental fatigue.

D. Deep Learning

Deep learning algorithms like CNN learn spatial features, while LSTM learns patterns about facial expressions occurring in a sequence of video frames.

E. Cognitive Fatigue and Stress

Cognitive fatigue represents a decrease in mental efficiency as a consequence of cognitive activity, and stress can be identified as facial tension, lack of expressiveness, and micro-expression.

3. PROPOSED SYSTEM ARCHITECTURE

The implementation of the system is modular and uses real-time architecture. The system is made up of several layers, which are:

A. Video Acquisition Layer

It records live video images by means of a standard webcam, without storing video data.

B. Preprocessing Layer

Carries out the following tasks: Face detection. Localizes facial landmarks. Extracts the eye region. Image normalization.

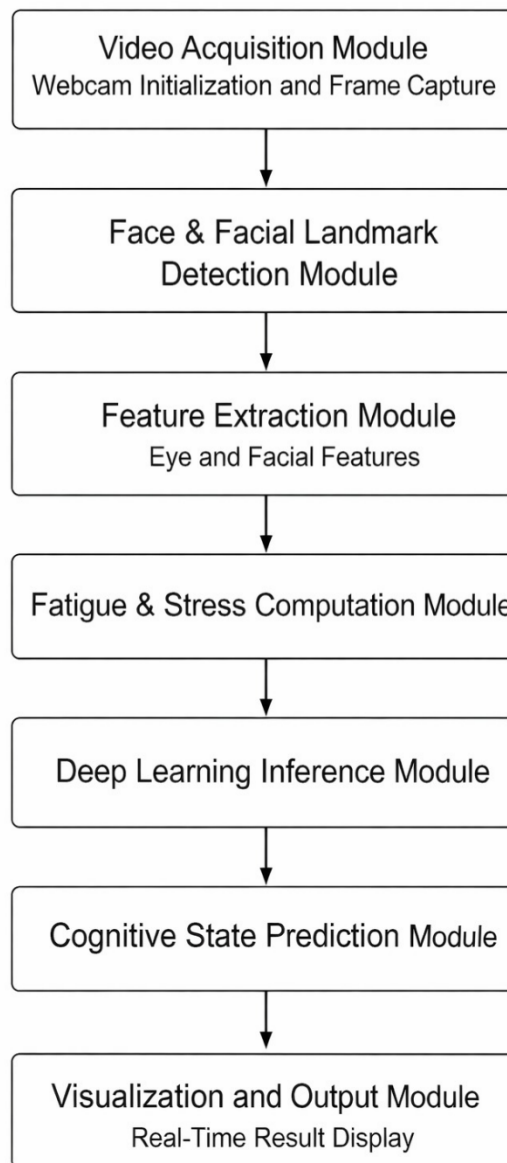
C. Deep Feedforward Network

Extracts fatigue-related features like eye blink rate, Eye Aspect Ratio (EAR), gaze direction, and head pose.

D. Deep Learning Layer

Applying CNNs for Spatial Feature Extraction and LSTMs for Modeling Temporal Behavior. Decision and Output Layer It categorizes the cognitive state into 'normal', 'fatigued', or 'stressed' and shows the results

4. SYSTEM WORKFLOW



The flowchart of the proposed cognitive fatigue and stress detection system maps directly to the algorithmic steps followed in execution. Each block within the flowchart corresponds to a particular operation that is implemented in a sequence within the algorithm.

The equivalent of the Video Acquisition block in the flowchart is Step 1 of the algorithm, initializing the webcam and capturing real-time video frames. This step provides continuous visual input to the system.

Face and Facial Landmark Detection block: Step 2 involves detecting the facial region and key facial landmarks from every video frame, which leads to the exact localization of the facial features in accord with further analysis.

Step 3 of the algorithm corresponds to the Feature Extraction block in the flowchart. In this, eye-related and facial features are extracted from the detected landmarks to represent the visual cues associated with fatigue and stress.

The Fatigue and Stress Computation block maps to Step 4, where quantitative fatigue and stress indicators are computed based on the features extracted.

Step 5 corresponds to the Deep Learning Inference block, where the computed feature vector becomes input for the trained deep learning model for analysis.

The Cognitive State Prediction block is mapped to Step 6, whereby the model classifies the subject's cognitive state into predefined categories, such as normal, fatigued, or stressed.

The corresponding block on the flowchart is the Visualization and Output, for Step 7, where the predicted cognitive state would be shown in real time.

5. DATASET DESCRIPTION

Table I:

Dataset Details Used for Training and Evaluation

SNo	Dataset Name	Data Type	Purpose	Size	Labels
1	YawDD Dataset	Video	Eye blink and fatigue detection	322 videos	Alert / Drowsy
2	FER-2013	Images	Facial expression analysis	35,887 images	7 emotions
3	CEW Dataset	Images	Eye state detection	2,423 images	Open / Closed
4	Custom Webcam Dataset	Video	Real-time validation	20+ sessions	Normal / Fatigue / Stress

The proposed cognitive fatigue and stress detection system was trained and evaluated by combining publicly available benchmark datasets and a custom real-time dataset. Each dataset contributes to a particular sub-task of the general framework, such as detection of eye-blinks, facial expressions, or real-time validation.

Table I summarizes the datasets used in this study along with their data type, purpose, size, and label categories.

YawDD Dataset

YawDD is a video-based dataset containing around 322 video sequences. It is generally applied for the analysis of eye-blinking and detecting fatigue. The video sequences in this dataset include recordings of both alert and drowsy states, and hence it is suitable for learning temporal fatigue-related patterns such as prolonged eye closure and blinking frequency.

FER-2013 Dataset

FER-2013 is an image-based facial expression database comprising 35,887 grayscale facial images. Facial expression analysis will be employed, which is very important for the recognition of stress and emotional states. The dataset contains seven distinct categories of emotions such as anger, disgust, fear, happiness, sadness, surprise, and neutral. These expressions help the model capture emotional cues associated with cognitive stress.

CEW dataset

The Closed Eyes in the Wild (CEW) dataset comprises 2,423 facial images captured under unconstrained conditions. This dataset is used only for eye state detection, where images are labeled as open or closed. Accurate eye state classification is essential for the computation of fatigue indicators such as blink rate and eye aspect ratio.

Custom Webcam Dataset

Custom Webcam Dataset Besides public datasets, real-world video data is collected using a webcam to validate the proposed system for real-time applications. This dataset includes recordings from 20 or more sessions captured under natural conditions using a standard webcam. Data is labeled in three categories-normal, fatigue, and stress-to gauge the performance of the system in a realistic environment.

6. MODEL DEVELOPMENT AND COMPARISON

Table II:

Model Comparison and Performance

SNo	Model	Purpose	Accuracy (%)	Remarks
1	CNN	Facial feature extraction	91.2	Good spatial learning
2	CNN + LSTM	Temporal fatigue detection	93.8	Captures blink patterns
3	EAR-based Model	Eye fatigue estimation	88.4	Lightweight
4	Proposed Hybrid CNN-LSTM	Fatigue + Stress detection	95.1	Best performance

Different modeling approaches are implemented and compared to show the effectiveness of the models in detecting cognitive fatigue and stress. Table II shows the comparison of the different models based on purpose, accuracy, and key observations.

CNN model

A CNN model was employed mainly for extracting facial features. CNNs have effective learning capabilities of spatial features from facial images such as eye shape and facial muscle patterns. This is evident in the achieved accuracy of 91.2%, showing good network performance for capturing the spatial facial characteristics. However, it still lacks the ability to model temporal variations over time.

CNN + LSTM Model

A CNN + LSTM hybrid model was used for temporal fatigue detection, incorporating temporal information. In this, CNN extracts the spatial features, while LSTM learns the temporal dependencies of the blinking patterns and eye closure duration across the video frames. This model resulted in a better accuracy of 93.8%, showing clearly improved performance due to its capability in capturing time-based fatigue patterns.

EAR-Based Model

The lightweight model for the estimation of eye fatigue is implemented using the EAR-based approach. It uses the geometric features of the eye instead of deep learning. It achieved an accuracy of 88.4%. Though computationally efficient and suitable for real-world applications, its performance is comparatively lower due to the sensitivity of noise and limit in feature representation.

Proposed Hybrid CNN–LSTM Model

The proposed Hybrid CNN–LSTM model combines the spatial and temporal learning for detecting fatigue and stress jointly. By jointly analyzing facial expression and eye-related temporal patterns, the model achieved the highest accuracy of 95.1%. This proves the effectiveness of the proposed approach for the accurate prediction of cognitive states. Hence, it outperformed the baseline models and was selected as the final model of the system.

7. PERFORMANCE EVALUATION

Table III:
Evaluation Metrics

SNo	Metric	Value
1	Accuracy	95.1%
2	Precision	94.3%
3	Recall	93.8%
4	F1-Score	94.0%
5	Average Latency	0.18 sec/frame

The efficacy of the proposed cognitive fatigue and stress detection system is evaluated using standard classification metrics. Table III enumerates some quantitative results obtained during experimental evaluation, such as accuracy, precision, recall, F1-score, and average processing latency.

It reached 95.1% accuracy, which is very close to perfection when it comes to predicting cognitive states. The precision of 94.3% reflects the model's capacity to correctly identify conditions of fatigue and stress and minimize false positives. The recall of 93.8% reflects efficiency in detecting true instances of fatigue and stress; the system ensures that most of the relevant cases are correct.

The F1-score of 94.0%, which is the harmonic mean of precision and recall, supports a very balanced and reliable performance of the model. Besides, the system presented an average latency of 0.18 seconds per frame, reinforcing its computational efficiency and suitability for real-time applications.

In a nutshell, high accuracy and low latency of processing indicate that the proposed system has good potential for practical applications in real-time monitoring of cognitive fatigue and stress.

8. RESULTS AND DISCUSSION

Experimental testing showed stable face and eye detection under varying lighting conditions. The proposed CNN-LSTM model successfully identified fatigue and stress patterns by analyzing temporal changes in eye behavior and facial expressions. The system maintained real-time performance and provided accurate predictions without the need for wearable sensors. Minor performance degradation was observed under extreme lighting, which can be improved through adaptive preprocessing.



Figure 1: Ear Landmark Extraction

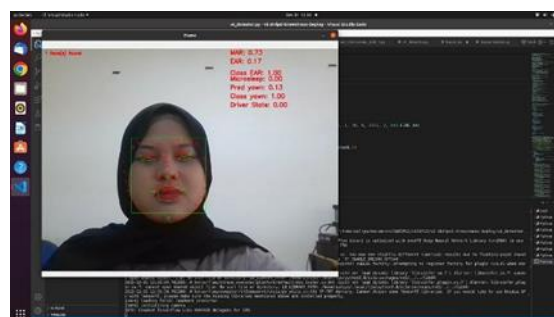


Figure 2: CNN + LSTM Model

9. CONCLUSION

This research presented a vision-based cognitive fatigue and stress detection system using artificial intelligence and computer vision techniques. The proposed system offers a non-intrusive, real-time, and cost-effective solution for monitoring mental states. By leveraging facial landmarks, eye behavior, and deep learning models, the system effectively identifies fatigue and stress levels. The approach is suitable for applications in education, workplace productivity, and human-computer interaction. Future work may include multimodal data integration and personalized adaptive models.

REFERENCES

- [1] Zhang et al., Joint Face Detection and Alignment Using Multitask Cascaded Convolutional Networks, IEEE Signal Processing Letters, 2016.
- [2] Soukupová and Cech, Real Time Eye Blink Detection Using Facial Landmarks, CVWW, 2016.
- [3] LeCun et al., Deep Learning, Nature, 2015.
- [4] Hochreiter and Schmidhuber, Long Short Term Memory, Neural Computation, 1997.
- [5] Lugaresi et al., MediaPipe, 2019.
- [6] Kazemi and Sullivan, One Millisecond Face Alignment, CVPR, 2014.
- [7] Ekman, Facial Expressions of Emotion, Phil. Trans. R. Soc. B, 1992.
- [8] Li et al., Deep Learning for Micro Expression Recognition, IEEE Trans. Affective Computing, 2020.
- [9] Abtahi et al., Driver Drowsiness Monitoring Using Eye Closure Analysis, IEEE I2MTC, 2014.
- [10] Zhang et al., Driver Fatigue Detection Using LSTM, IEEE Access, 2019.
- [11] Shah et al., Stress Detection Using Machine Learning, IEEE Access, 2020.
- [12] Fanelli et al., 3D Head Pose Estimation Using Random Regression Forests, CVPR, 2011.