

Real-Time Driver Emotion Detection Using Convolutional Neural Network and Computer Vision for Road Safety.

Mayuri Salunke

salunkemayuri2705@gmail.com

sanjivani

Research Article

Keywords: Driver Emotion Detection, Convolutional Neural Network, Computer Vision, Road Safety, Face Detection, Real-Time Alerts

Posted Date: October 22nd, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-11409926/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Additional Declarations: The authors declare no competing interests.

Abstract

Most vehicle safety systems monitor the vehicle and the road, but not the emotional state of the driver, although emotions such as anger, stress, and fatigue are generally associated with reduced concentration and slower reactions. This paper presents a preliminary prototype for real-time driver emotion detection that combines Haar-Cascade face detection, a pretrained Convolutional Neural Network (CNN) emotion classifier, and a rule-based alert mechanism that responds only when a potentially unsafe emotion persists over time. The prototype was implemented in Python using OpenCV and the open-source fer library. Preliminary testing was performed on a 115-frame publicly available face video and on a controlled alert-timing test. A face was detected in 114 of the 115 frames, the average processing time was 68.14 ms per frame on a CPU (about 14.7 frames per second), and the alert rule fired at the configured 2-second threshold in the controlled test. The evaluation is preliminary: no ground-truth emotion labels were available, classification accuracy was not measured, the CNN was not trained or fine-tuned by the author, and the footage does not show real driving. The work explores feasibility only and does not claim reliable detection of driver emotion in real driving conditions.

I. INTRODUCTION

Road accidents are strongly influenced by the condition of the driver. Most existing vehicle safety systems focus on monitoring the vehicle and the road environment, but they do not monitor the driver's emotions in real time. If a driver becomes angry, stressed, tired, or sleepy, such a system has no way of detecting it, which may increase the risk of an accident.

Negative emotional and physical states such as anger, stress, fatigue, and drowsiness can reduce a driver's concentration, decision-making ability, and reaction time. A system that observes the driver and gives a timely alert could therefore be a useful addition to existing safety features. Camera-based methods are attractive in this setting because they do not require the driver to wear any sensor.

Recent research has examined driver monitoring mainly from the viewpoint of drowsiness and fatigue, using eye and mouth measurements, deep learning, and physiological signals. Zambrano et al. [1] reported a monitoring system that also includes facial emotion, while most other studies [2]–[8] concentrate on fatigue or drowsiness.

A single emotion prediction from one video frame is also not a reliable basis for an alert, because expressions change quickly and classifiers make occasional errors. An alert mechanism therefore needs to consider how long an unsafe state lasts rather than react to every frame.

This research presents a lightweight prototype in which a face detector and a CNN emotion classifier are followed by a simple persistence-based alert rule. The prototype is evaluated only with preliminary tests. The main objectives of this work are:

- To develop a real-time driver emotion detection prototype using a CNN and Computer Vision.

- To classify facial expressions into emotion categories such as anger, sadness, happiness, and neutral.
- To generate an alert when a potentially unsafe emotion is sustained.
- To study the practical feasibility of such a pipeline on ordinary CPU hardware.

II. RELATED WORK

Research on driver monitoring has developed with the increasing demand for safer vehicles. The studies considered in this work cover emotion-aware monitoring, drowsiness detection from facial features, and physiological sensing.

A. Emotion-Aware Driver Monitoring

Zambrano et al. [1] developed a real-time driver monitoring system using Computer Vision, a MobileNetV2-based CNN [9], and MediaPipe [10]. The system uses facial expressions, eye movements, head pose, and gaze to detect emotion, fatigue, and distraction, and was reported to give good results for happiness, anger, surprise, fatigue, and distraction.

B. Drowsiness and Fatigue Detection from Facial Features

Prasanthi et al. [2] built a drowsiness detection system using OpenCV, Dlib, the Eye Aspect Ratio (EAR), and the Mouth Aspect Ratio (MAR), and raised an alert when measured values crossed a threshold. You et al. [3] used a CNN with an SVM on eye features while considering differences between individual drivers, and reported 94.80% accuracy. Safarov et al. [4] used a CNN and facial landmark tracking to detect blinking and yawning in real time, and reported 95.8% accuracy for eye-based fatigue detection. Dewi et al. [5] showed that EAR-based Computer Vision can be used for real-time drowsiness detection.

Díaz-Santos et al. [6] studied driver identification together with drowsiness detection, and Beles et al. [7] proposed a multi-method approach to drowsiness detection for vehicles with autonomous driving functions.

C. Physiological Signals

Alguindigue et al. [8] used heart rate variability and electrodermal activity with a deep neural network and reported 98% precision and recall for driver fatigue. However, the use of physiological sensors is more intrusive and may be difficult to apply in ordinary vehicles.

The reviewed studies provide the foundation for the present work. All accuracy values above are those reported by the respective authors on their own data and are not directly comparable with the results of this paper.

III. RESEARCH GAP AND OBJECTIVES

The reviewed research shows that camera-based methods can detect drowsiness and fatigue with good reported accuracy. At the same time, most of these systems concentrate on eye and mouth measurements or on physiological sensors. Comparatively less attention has been given to lightweight, camera-only pipelines that recognise a wider range of facial emotions and decide when an alert is justified.

Based on the reviewed studies, this work investigates a pipeline that combines face detection, CNN-based emotion classification, and a persistence-based alert rule. The research objectives are:

1. To develop a real-time processing pipeline for driver emotion detection.
2. To map predicted emotion classes to safe and potentially unsafe driver states.
3. To implement an alert rule that responds only to sustained unsafe states.
4. To evaluate the behaviour of the prototype using controlled test inputs.
5. To identify limitations and possible improvements of the proposed approach.

IV. PROPOSED METHODOLOGY

The proposed methodology treats the alert decision as part of the processing pipeline rather than relying only on a single prediction of the CNN. The processing flow is shown in Fig. 1:

Video Frame → Face Detection → Preprocessing → CNN Emotion Classification → Safe/Unsafe Mapping → Sustained-State Alert Rule

A. Face Detection

Each video frame is first examined by a Haar-Cascade face detector [12] available in OpenCV [13]. When more than one face is found, the largest face is assumed to belong to the driver. If no face is found, the persistence timer of the alert rule is reset.

B. Preprocessing

The detected face region is cropped, converted to grayscale, and resized to 48×48 pixels, which is the input size used by the emotion classifier.

C. CNN Emotion Classification

The cropped face is classified by a pretrained compact CNN distributed with the open-source fer library [14]. The model predicts seven emotion classes (angry, disgust, fear, happy, sad, surprise, neutral) of the FER-2013 dataset [11]. In this work the model is used as provided; it was not trained or fine-tuned by the author.

D. Safe/Unsafe Mapping

The predicted class with the highest probability is mapped to a driver state. Happy, neutral, and surprise are treated as safe, and angry, sad, fear, and disgust are treated as potentially unsafe. This mapping is a design assumption of the prototype and has not been validated against real driving data. The seven classes do not include a fatigue or drowsiness class, so fatigue is not measured directly.

E. Sustained-State Alert Rule

An alert is generated only when the potentially unsafe state continues without interruption for a configurable duration. If a safe state or a missing face occurs, the timer is reset. This reduces the influence of isolated misclassifications on the alert decision.

V. SYSTEM IMPLEMENTATION

A prototype of the proposed pipeline was developed in Python. OpenCV was used for video input, face detection, and drawing the result on each frame, and the pretrained classifier was run through the fer library in its TensorFlow Lite form. The application reads the video frame by frame, applies the pipeline described above, and stores the predicted emotion, the confidence value, the processing time, and the alert status of every frame in a log file.

The implementation keeps the safety logic (mapping and alert rule) separate from the classifier. This allows the mapping or the alert threshold to be changed without modifying the CNN.

VI. RESULTS AND DISCUSSION

The prototype was evaluated using a short test video and a controlled alert-timing test. The results are preliminary and are intended only to show that the pipeline operates as designed.

A. Test Setup

The test video contains 115 frames at 24.7 frames per second (596×336 pixels) and shows one person in front of the camera. It is a short clip distributed with the repository of the fer library. It was used only as an example of a continuous face video and was not recorded inside a vehicle. No ground-truth emotion labels are available for it. All timings were measured on a CPU without GPU acceleration; the exact hardware was not controlled, so the timings are indicative only.

B. Face Detection and Emotion Output

A face was detected in 114 of the 115 frames (99.13%). The reason for the single missed frame (frame 55) was not investigated. For the 114 frames with a detected face, the classifier output “happy” in 103 frames, “sad” in 8 frames, and “angry” in 3 frames (Fig. 2). The mean confidence of the top class was 0.79. Because the video has no ground-truth labels, it is not possible to say how many of these predictions are correct.

Table 1
Summary of the Preliminary Measurements

Measurement	Value
Frames in test video	115
Frames with a detected face	114 (99.13%)
Predicted classes (frames)	happy 103, sad 8, angry 3
Average processing time per frame	68.14 ms (range 54.5–83.2 ms)
Throughput of detection and classification	14.68 frames per second
End-to-end throughput including video reading	14.35 frames per second
Total processing time for 115 frames	8.01 s
C. Alert Rule Testing	

The alert rule was tested in two ways, as shown in Table 2. First, the full test video was processed with a persistence threshold of 0.6 s. No alert was generated, because the frames predicted as potentially unsafe occurred in short groups of at most 5 consecutive frames (about 0.2 s). Second, in a controlled test one frame that had been classified as “sad” was repeatedly supplied to the system with simulated timestamps at 20 frames per second and a threshold of 2.0 s. The alert was generated at $t = 2.00$ s. This second test checks the timing logic only; it does not show that a real driver’s sustained emotion would be detected.

Table 2
Alert Rule Test Cases

Test	Setting	Result
Full test video	Threshold 0.6 s	No alert (unsafe predictions were short and intermittent)
Repeated “sad” frame, simulated timestamps	Threshold 2.0 s	Alert generated at $t = 2.00$ s
D. Runtime Performance		

The average processing time was 68.14 ms per frame, with a standard deviation of 6.23 ms (Fig. 3). This corresponds to about 14.7 frames per second on a CPU, which is lower than the 24.7 frames per second of the source video. In a live system, frames would therefore need to be skipped or processed at a reduced rate, which is acceptable if the alert rule works on durations of seconds, but this was not tested with a live camera.

E. Discussion

The preliminary implementation shows that face detection, CNN-based emotion classification, and a persistence-based alert rule can be combined in a single pipeline running on ordinary CPU hardware. An important observation is that the predictions of a single frame fluctuate. In the test video the potentially unsafe predictions appeared only in short groups, and the persistence rule did not turn them into an alert. This supports the design choice of not alerting on isolated frames, although the video itself does not show whether these predictions were correct.

A second observation is that the person in the test video is looking down at a phone for part of the clip. A possible reason for some “sad” predictions is therefore the head and gaze direction rather than the emotion, which is a known difficulty for expression classifiers. This explanation has not been verified. The proposed pipeline provides a practical starting point, but evaluation on labelled data and on in-vehicle recordings is required before conclusions about its reliability can be drawn.

VII. LIMITATIONS

The present prototype has several limitations.

First, no ground-truth labels were available, so accuracy, precision, recall, and F1-score are not reported. The results show how the pipeline behaves, not how often it is correct.

Second, the test uses one short video of one person, which was not recorded in a vehicle. Lighting changes, glasses, head rotation, and camera position in real driving may affect the results considerably.

Third, the CNN was used as provided by the fer library and was not trained or fine-tuned by the author. The FER-2013 images are not taken in vehicles, and the dataset is known to contain some label noise.

Fourth, the mapping of emotion classes to safe and unsafe states is an assumption of the prototype. The seven classes contain no fatigue or drowsiness class, so the prototype does not measure fatigue, although fatigue is part of the motivation of this work.

Finally, the Haar-Cascade detector may be less robust for rotated faces, the timings were measured in one CPU environment only, and the alert is only displayed on the video frame; no study with drivers was carried out.

VIII. CONCLUSION

This paper presented a preliminary prototype for real-time driver emotion detection that combines Haar-Cascade face detection, a pretrained CNN emotion classifier, and a persistence-based alert rule. Preliminary testing showed that a face was detected in 114 of 115 frames of a test video, that the pipeline processed about 14.7 frames per second on a CPU, and that the alert rule generated an alert at the configured 2-second threshold in a controlled test.

The work demonstrates that such a pipeline can be assembled from open-source components and run without special hardware. However, the current evaluation is preliminary and should not be considered an assessment of how well driver emotions are recognised in real driving.

Future work will focus on evaluating the classifier on the labelled FER-2013 test set and reporting accuracy, precision, recall, and F1-score, recording a small dataset in a vehicle or driving simulator with informed consent, adding eye-based fatigue measurements such as the Eye Aspect Ratio, and comparing the Haar-Cascade detector with more robust face detectors.

Declarations

Ethics statement: This study, titled “Real-Time Driver Emotion Detection Using Convolutional Neural Network and Computer Vision for Road Safety,” presents a preliminary software-based prototype for driver emotion detection. No human participants were recruited, no personal data were collected directly by the author, and no real-world driving experiments were conducted. The preliminary testing used a short, publicly available video clip distributed with the repository of the open-source fer library, along with a controlled software timing test. No images of the person in the video are reproduced in the manuscript. Ethics committee or IRB approval was not sought and was not required for this study.

References

1. Zambrano T, Arias L, Haro E, Santos V, Trujillo-Guerrero M (2026) Driver monitoring system using computer vision for real-time detection of fatigue, distraction and emotion via facial landmarks and deep learning. *Sensors* 26:889
2. Prasanthi BB, Reddy GSM, Vardhan NG, Varshitha I, Rushita N, Sudheer D (2025) Driver drowsiness detection. *Int J Sci Res Eng Manage (IJSREM)* 9:1–4
3. You F, Li X, Gong Y, Wang H, Li H (2019) A real-time driving drowsiness detection algorithm with individual differences consideration. *IEEE Access* 7:179396–179408
4. Safarov F, Akhmedov F, Abdusalomov AB, Nasimov R, Cho YI (2023) Real-time deep learning-based drowsiness detection: Leveraging computer-vision and eye-blink analyses for enhanced road safety. *Sensors* 23:6459
5. Dewi C, Chen RC, Chang CW, Wu SH, Jiang X, Yu H (2022) Eye aspect ratio for real-time drowsiness detection to improve driver safety. *Electronics* 11:3183
6. Díaz-Santos S, Cigala-Álvarez Ó, Gonzalez-Sosa E, Caballero-Gil P, Caballero-Gil C (2024) Driver identification and detection of drowsiness while driving. *Appl Sci* 14:2603
7. Beles H, Vesselenyi T, Rus A, Mitran T, Scurt FB, Tolea BA (2024) Driver drowsiness multi-method detection for vehicles with autonomous driving functions. *Sensors* 24:1541
8. Alguindigue J, Singh A, Narayan A, Samuel S (2024) Biosignals monitoring for driver drowsiness detection using deep neural networks. *IEEE Access* 12:93075–93090

9. Sandler M, Howard A, Zhu M, Zhmoginov A, Chen LC (2018) MobileNetV2: Inverted residuals and linear bottlenecks, in Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR), pp. 4510–4520
10. Lugaresi C et al (2019) MediaPipe: A framework for building perception pipelines. arXiv preprint
11. Goodfellow IJ et al (2015) Challenges in representation learning: A report on three machine learning contests. Neural Netw 64:59–63
12. Viola P, Jones M (2001) Rapid object detection using a boosted cascade of simple features, in Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR), pp. I-511–I-518
13. Bradski G (2000) The OpenCV library. Dr Dobb's J Softw Tools
14. Shenk J fer: Facial expression recognition, open-source Python library. [Online]. Available: <https://github.com/justinshenk/fer>

Figures

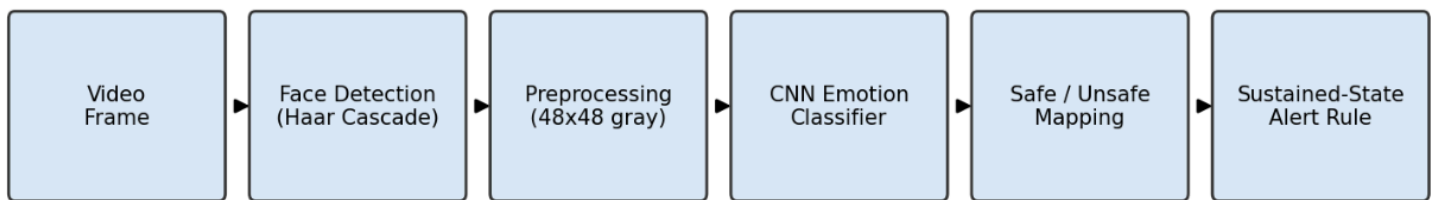


Figure 1

Processing pipeline of the proposed prototype.

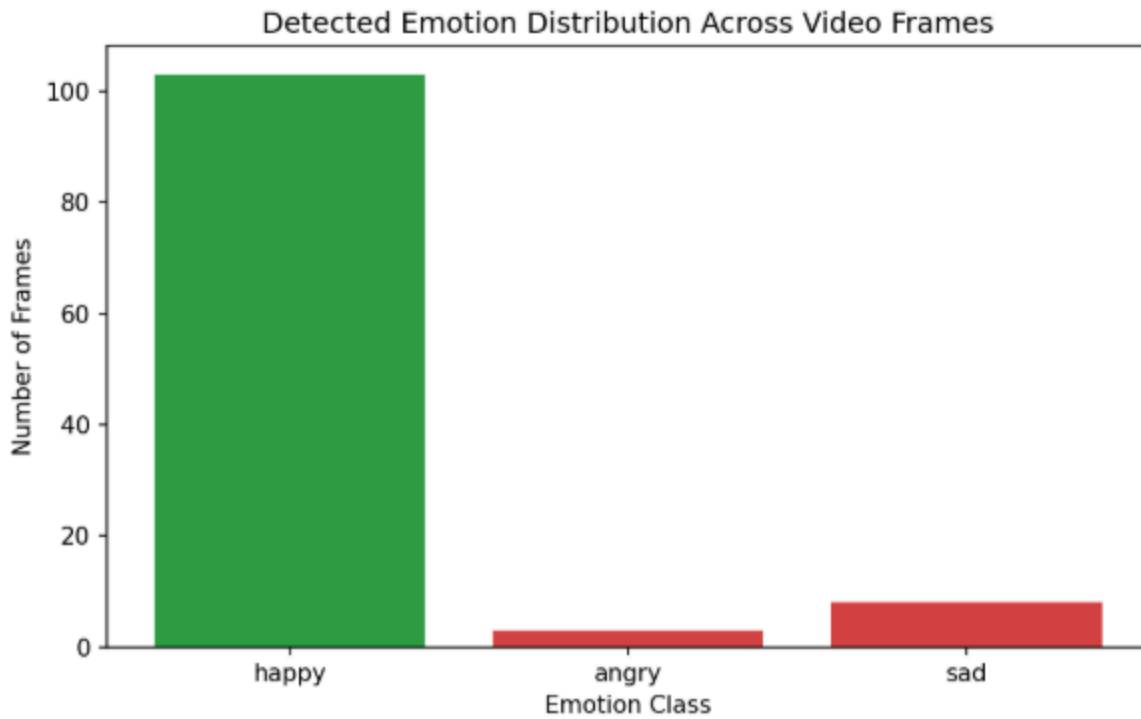


Figure 2

Distribution of predicted emotion classes over the test video.

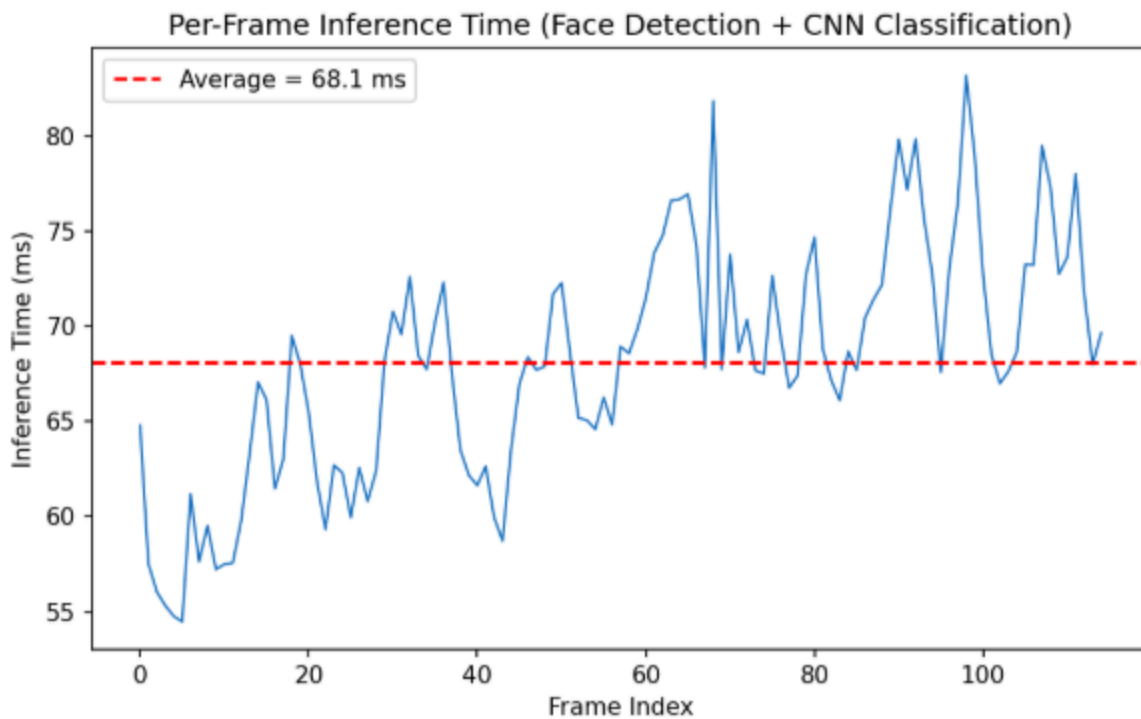


Figure 3

Processing time per frame (face detection and CNN classification).

