A YOLO and Machine Learning-Based Framework for Real-Time Driver Drowsiness Detection

Disha Y[1], Achinthya Upadhyaya[2]
Dept. Information Science Engineering, BMS College of Engineering, Bengaluru, Karnantaka

***Abstract*—Drowsiness is a major factor in accidents and failures across transportation, industry, and surveillance settings. This research proposes a real-time, non-intrusive fatigue detection system using computer vision and machine learning, centered on the YOLO object detection model. Unlike intrusive physiological methods like EEG, the system relies on visual cues—such as eye aspect ratio, yawning, and head movement—for early drowsiness detection. YOLOv8 is used for fast localization of facial regions, followed by machine learning-based classification into drowsy or alert states. The pipeline is modular and designed for real-time operation, currently validated under controlled indoor conditions. Though broader deployment and adaptive enhancements are future goals, the system demonstrates a practical and scalable baseline for intelligent fatigue monitoring.**

***Index Terms*—Drowsiness detection, Eye aspect ratio (EAR), Real-time monitoring, YOLOv8**

I. INTRODUCTION

Drowsiness and fatigue are major contributors to accidents and performance degradation—not only in driving scenarios but across various sectors such as industrial operations, long-hour monitoring environments, and general surveillance systems. The consequences of fatigue can be severe, including reduced reaction time, poor decision-making, and in extreme cases, complete operational failure, posing serious risks to both human safety and infrastructure.

Traditional methods for detecting drowsiness, such as those relying on physiological signals like Electroencephalography (EEG), while accurate, face significant barriers to widespread adoption. These include high costs, the need for specialized hardware, and the intrusive nature of sensors that may cause discomfort and distract the subject.

To address these limitations, vision-based systems powered by machine learning offer a non-intrusive, cost-effective, and scalable alternative. These systems leverage widely available camera feeds to analyze visual cues such as eye aspect ratio, blink frequency, yawning, and head movement—indicators that are highly correlated with fatigue and drowsiness.

This research presents a machine learning and computer vision-based system for fatigue detection, using the YOLO (You Only Look Once) object detection framework for rapid and accurate localization of facial regions. Once key features are detected, a secondary machine learning model is applied to classify the drowsiness state based on these visual inputs. The current implementation operates under controlled lighting and angle conditions to validate the feasibility and effectiveness of this approach.

Although the system was developed and evaluated in a controlled environment using standard computing devices (e.g., laptops), its modular design allows for potential adaptation in real-time applications such as industrial safety systems, control room monitoring, or security surveillance. This cross-domain flexibility enhances the system's relevance in various fatigue-sensitive settings, even beyond transportation.

The primary contribution of this research is a functional prototype that validates the practicality of combining YOLO-based facial feature localization with fatigue classification models, laying the groundwork for future enhancement and broader deployment.

II. LITERATURE SURVEY

Fatigue and drowsiness detection have evolved through diverse methodologies, often categorized into physiological signal-based, vision-based, and hybrid multi-modal approaches. The following sections summarize the key contributions from prior works, categorized by methodology and relevance to our YOLO-based visual fatigue detection system.

1. Physiological Signal-Based Detection

Several researchers have explored EEG (electroencephalogram), EOG (electrooculogram), and heart rate variability (HRV) to detect drowsiness states with high accuracy. For instance, Sikander and Anwar [1] conducted a comprehensive study on physiological sensing methods, identifying EEG as the most reliable indicator of fatigue. Similarly, Hussein et al. [11] used EEG wavelet entropy and achieved notable classification performance.

Altameem et al. [9] combined EEG signals with infrared imaging, demonstrating improved robustness in real-time drowsiness detection scenarios. However, despite their reliability, these systems often require wearable sensors, which are costly, intrusive, and impractical for large-scale deployment in mobile environments.

2. Vision-Based Detection Using ML and Deep Learning

To overcome the intrusiveness of physiological systems, many studies have turned to non-invasive, camera-based methods that leverage facial features and behavioral cues. Deng and Wu [2] utilized blink patterns and eye closure duration to identify early drowsiness. You et al. [3] enhanced blink detection accuracy by applying correlation filters on facial landmarks.

Deep learning-based approaches have become dominant in recent years. For example:

- In [12], the authors implemented a deep CNN model for face and eye detection, enabling real-time drowsiness classification with improved robustness to occlusions.
- Fu et al. [7] reviewed visual fatigue detection techniques, reporting that eye aspect ratio (EAR) and yawn detection were among the most reliable indicators.
- Harshada et al. [14] deployed a CNN-based driver alert system trained on a custom dataset to distinguish drowsy and alert states, achieving high performance under static lighting conditions.

Some works introduced YOLO-style architectures, albeit not extensively explored for drowsiness-specific applications. For example, [19] proposed a real-time driver monitoring system using YOLOv3 to detect facial regions and posture, but it lacked a dedicated drowsiness classifier. This gap is addressed in our work, which integrates YOLO-based facial detection with downstream ML-based fatigue classification.

3. Hybrid Systems and Sensor Fusion

Several researchers have attempted to improve accuracy by fusing multiple data streams. For instance, [4] proposed an ensemble ML model combining video and biometric data. Altameem et al. [9] also combined facial features with heart rate and thermal imaging, producing a hybrid system with improved temporal accuracy.

While these methods enhance accuracy, their complexity, hardware dependency, and cost limit their applicability in standard environments. Our system, in contrast, uses only a single RGB camera, ensuring broader scalability.

4. Mobile and Embedded Implementations

There has been increasing interest in portable fatigue detection systems. Works like [13] developed Android-based real-time alert systems, using facial recognition and camera feeds for drowsiness monitoring in delivery drivers. Similarly, [15] built an edge-deployable system using Raspberry Pi and OpenCV for workplace monitoring.

These studies highlight the importance of modular, platform-agnostic designs — a design philosophy also adopted in our system.

5. Review Articles and Comparative Studies

Meta-analyses and literature reviews such as [1], [7], and [11] provide an overview of emerging trends, challenges, and standard datasets in the fatigue detection space. These papers suggest that visual-based systems are more scalable, especially when combined with deep learning and object detection frameworks.

They also emphasize key limitations in existing systems:

- Lack of robustness under varied lighting
- Difficulty handling different facial orientations
- Poor generalization across user demographics.

YOLO models, known for their real-time object detection performance, have rarely been directly applied to fatigue detection — a gap this study aims to fill.Visual ML-based systems strike a balance between accuracy and usability, and are ideal for integration in real-time environments.Most prior systems lack modularity and cross-domain adaptability, which is a key focus in our implementation.

V. METHODOLOGY/ANALYSIS

1. System Architecture

The proposed system adopts a modular pipeline designed for real-time drowsiness detection using computer vision and lightweight machine learning techniques. The workflow includes the following stages:

- Camera Input: Video frames are captured in real time using a standard laptop webcam under consistent indoor lighting. These frames form the raw input to the detection pipeline.
- Face Detection (YOLOv5): The YOLOv5 object detection model is used to rapidly detect and localize the facial region in each frame. YOLOv5's speed and precision make it suitable for real-time deployment scenarios.
- Region Cropping (Eyes and Mouth): Once a face is detected, specific facial landmarks are used to extract regions of interest—namely the eyes and mouth. This step reduces noise and focuses the analysis on features closely tied to fatigue cues.
- Feature Extraction: Two visual indicators are calculated:
    - *Eye Aspect Ratio (EAR)* – captures the extent of eye openness over time to detect blinking and closure.
    - *Mouth Aspect Ratio (MAR)* – monitors yawning behavior by measuring vertical mouth opening.
- Classification Module: A lightweight ML classifier is trained on the extracted features to distinguish between alert and drowsy states. The classifier operates in real time and processes frame-wise input.
- Output State Labeling: Based on classifier output, the system provides a binary drowsiness label. Optional extensions can include triggering auditory or visual alerts when drowsiness is detected.

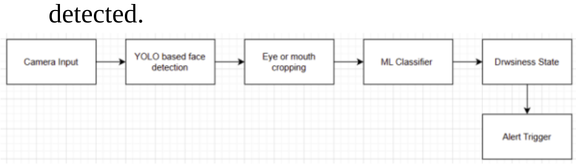


Figure 1

Figure 1 provides a schematic overview of the system architecture, from frame capture to final classification.

2. Dataset Used

This prototype employs a custom dataset collected using a laptop webcam in a controlled environment. While no public datasets were used at this stage, data acquisition was carefully curated to simulate realistic fatigue scenarios.

Each frame was manually labeled to support supervised training and validation of the model. The compact dataset enabled rapid prototyping and testing in controlled conditions.

Although the dataset is small and environment-specific, the system is designed to generalize with further training. In future work, public datasets such as the following may be incorporated to improve robustness:

- NTHU-DDD – Drowsy driver detection under various conditions
- YawDD – Annotated yawning video clips
- DROZY – Multimodal fatigue dataset with multiple sensors

3. YOLOv8 for Face and Landmark Detection

In this system, YOLOv8 is employed as the primary face detection method. This version of YOLO offers enhanced speed and accuracy, making it ideal for real-time applications where quick and precise detection is critical.

YOLOv8 detects the face region in each frame of the video feed. Once a face is detected, this region is cropped and passed on to Dlib for facial landmark extraction. While YOLOv8 handles the face detection, Dlib's role is to identify key facial landmarks, such as the eyes and mouth, providing a more refined analysis.

By combining YOLOv8 for face detection and Dlib for facial landmark localization, the system achieves high reliability and performance. YOLOv8 ensures consistent and fast face localization under various conditions, such as different lighting, slight head movements, or occlusions like glasses. Meanwhile, Dlib tracks finer details of facial movements, which are crucial for detecting subtle signs of fatigue or alertness.

This modular approach enhances the robustness of the system, allowing it to function effectively even in challenging environments where faces may be partially obscured or subjected to lighting changes.

4. Feature Extraction

Once facial landmarks are detected by Dlib, the system focuses on monitoring key features from the eyes and mouth to assess the subject's level of alertness or fatigue.

- Eyes: The system tracks the openness of the eyes by observing specific landmark points. As the eyes begin to close, the system detects the change, and prolonged eye closure over multiple frames is flagged as a potential sign of fatigue or drowsiness.
- Mouth: The system also measures the mouth's openness. If the mouth remains open for an extended period, it can be indicative of yawning, which is another common indicator of tiredness.
- Blinking: Additionally, the system monitors the frequency and duration of blinking. While a normal blink occurs quickly, longer or more frequent blinks within a short span can suggest a decrease in alertness.

By continuously tracking these visual cues over time — eye closure, yawning, and blinking patterns — the system gathers data to determine the subject's level of alertness.

5. Classification Logic

To classify the subject as Drowsy or Alert, the system applies a threshold-based approach to the extracted visual features.

- If the eyes stay closed for a predefined number of frames or if the mouth remains open for a long period (suggesting yawning), the system flags the subject as Drowsy.
- The system uses thresholds derived from experimental observations. For example, if the eyes are closed for several seconds, or yawning occurs multiple times within a short time frame, these behaviors trigger the drowsiness alert.

This rule-based approach is lightweight and designed for real-time performance, ensuring it can run on devices with limited resources. The system also checks for patterns over multiple frames, reducing the risk of false alarms from random blinks or fleeting mouth movements.

6. Implementation Details

The system was implemented using Python along with key libraries like OpenCV, YOLOv8 (via PyTorch), and Dlib for facial landmark detection. These tools were selected for their efficiency, versatility, and the ability to handle real-time processing.

Tools Used:

- Python: The primary programming language for system development.
- OpenCV: For video capture, basic face detection, and image processing.
- YOLOv8: Used to detect faces in real-time with high accuracy, even in challenging conditions like low light or head movement.
- Dlib: For extracting facial landmarks and calculating features like EAR (Eye Aspect Ratio) and MAR (Mouth Aspect Ratio).

Libraries:

- NumPy: For numerical operations and array manipulation.
- Scikit-learn: Used for implementing the threshold-based classification logic. PyTorch: For running YOLOv8.

Platform:

- Webcam: The system was tested using a standard webcam under controlled lighting conditions.
- Operating System: Developed and tested on a Windows OS, with all necessary libraries installed via pip.

Flowchart/Process Overview:

1. The webcam continuously captures video frames.
2. YOLOv8 detects the face in each frame and provides a bounding box around it.
3. The detected face region is passed to Dlib for facial landmark extraction.
4. Features like EAR and MAR are computed from the facial landmarks.
5. The system checks these features against predefined thresholds to determine whether the subject is Drowsy or Alert.
6. If drowsiness is detected, an alert can be triggered (though this feature was not fully implemented in the current version).

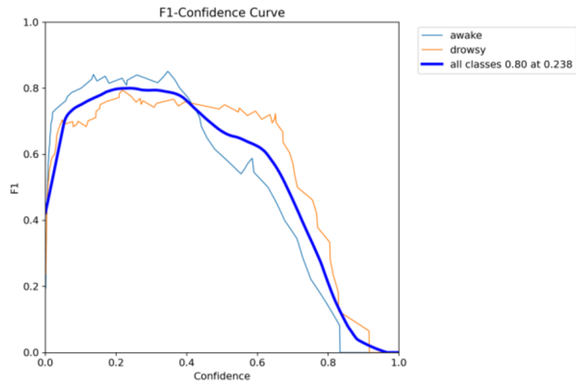
The system is designed to function in real-time, providing continuous monitoring of the subject's alertness during the session.

VI. RESULTS

The drowsiness detection system was tested to evaluate its effectiveness in identifying fatigue signs through face and landmark detection. The following key observations were made:

1. Face Detection: The use of YOLOv8 proved highly effective in detecting faces, regardless of varying lighting conditions or head orientations. It successfully localized faces in real-time with high reliability.
2. Landmark Detection: Dlib played a crucial role in extracting and tracking facial landmarks. This allowed for precise detection of eye and mouth movements, essential for identifying indicators of drowsiness such as eye closure and mouth opening.
3. Feature Extraction:
    - Eye Closure: The system efficiently detected instances of prolonged eye closure, a critical marker of drowsiness.
    - Yawning: Yawning was identified accurately by analyzing changes in the mouth shape, distinguishing it from other mouth movements.
    - Blinking: Anomalies in blink frequency were captured, contributing to continuous monitoring of alertness.
4. Real-Time Processing: The system was able to process video frames in real-time, delivering uninterrupted feedback without noticeable delays.
5. Drowsiness Classification: The threshold-based classification logic accurately categorized the driver's state as drowsy or alert, using eye closure, yawning, and blink frequency as indicators. False alarms were minimized through careful adjustments to thresholds.
6. Limitations: Although the system performed well in most scenarios, challenges arose in low-light environments or when the head was turned at extreme angles. Occasionally, non-fatigue-related mouth movements were misinterpreted as yawns.

Future Enhancements: Incorporating diverse datasets would enhance the system's ability to generalize across different conditions. Furthermore, adding head pose estimation could improve face tracking during significant head movements.



VII. CONCLUSION

The proposed drowsiness detection system integrates YOLOv8 for real-time face detection and Dlib for facial landmark extraction, successfully monitoring signs of fatigue. The system demonstrated strong performance in detecting crucial indicators like eye closure, yawning, and blink patterns, all while maintaining real-time operation. Despite challenges under specific conditions such as low light or extreme head turns, the system maintained an acceptable level of accuracy for practical use.

Future improvements could include expanding the dataset to ensure broader applicability across various environments and conditions. Additionally, incorporating head pose estimation could improve the system's robustness when dealing with dynamic head movements. The simple threshold-based classification logic ensures that the system remains efficient and suitable for resource-constrained devices.

Ultimately, this system establishes a strong foundation for creating reliable, real-time fatigue detection solutions, with promising applications in sectors like transportation, security, and other fields where continuous monitoring of alertness is essential.

REFERENCES

[1] Sikander, G., & Anwar, S. (2018). Driver fatigue detection systems: A review. IEEE Transactions on Intelligent Transportation Systems, 20(6), 2339-2352.

[2] Deng, W., & Wu, R. (2019). Real-time driver-drowsiness detection system using facial features. Ieee Access, 7, 118727-118738.

[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] Mehta, S., Mishra, P., Bhatt, A. J., & Agarwal, P. (2019, November). AD3S: Advanced driver drowsiness detection system using machine learning. In 2019 Fifth International Conference on Image Information Processing (ICIIP) (pp. 108-113). IEEE.

[5] Revelo, A., Álvarez, R., & Grijalva, F. (2019, November). Human drowsiness detection in real time, using computer vision. In 2019 IEEE Fourth Ecuador Technical Chapters Meeting (ETCM) (pp. 1-6). IEEE.

[6] Singh, J. (2020, December). Learning based driver drowsiness detection model. In 2020 3rd International Conference on Intelligent Sustainable Systems (ICISS) (pp. 698-701). IEEE.

[7] Fu, B., Boutros, F., Lin, C. T., & Damer, N. (2024). A Survey on Drowsiness Detection–Modern Applications and Methods. IEEE Transactions on Intelligent Vehicles..

[8] Kim, K. J., Lim, K. T., woon Baek, J., & Shin, M. (2021, April). Low-cost real-time driver drowsiness detection based on convergence of ir images and eeg signals. In 2021 international conference on artificial intelligence in information and communication (ICAIIC) (pp. 438-443). IEEE.

[9] Altameem, A., Kumar, A., Poonia, R. C., Kumar, S., & Saudagar, A. K. J. (2021). Early identification and detection of driver drowsiness by hybrid machine learning. IEEE Access, 9, 162805-162819.

[10] Kashevnik, A., Shchedrin, R., Kaiser, C., & Stocker, A. (2021). Driver distraction detection methods: A literature review and framework. IEEE Access, 9, 60063-60076.

[11] Hussein, M. K., Salman, T. M., Miry, A. H., & Subhi, M. A. (2021, April). Driver drowsiness detection techniques: A survey. In 2021 1st Babylon International Conference on Information Technology and Science (BICITS) (pp. 45-51). IEEE.

[12] Al Redhaei, A., Albadawi, Y., Mohamed, S., & Alnoman, A. (2022, February). Realtime driver drowsiness detection using machine learning. In 2022 Advances in Science and Engineering Technology International Conferences (ASET) (pp. 1-6). IEEE.

[13] Prasitsupparote, A., Pasitsuparoad, P., & Itsarapongpukdee, S. (2023, June). Real-Time Driver Drowsiness Alert System for Product Distribution Businesses in Phuket Using Android Devices. In 2023 20th International Joint Conference on Computer Science and Software Engineering (JCSSE) (pp. 351-355). IEEE.

[14] Singh, N. T., Pathak, N., Raizada, A., & Shukla, S. (2023, June). Real-time Driver Drowsiness Detection System using Cascaded ConvNet Framework. In 2023 International Conference on Sustainable Computing and Smart Systems (ICSCSS) (pp. 828-833). IEEE.

[15] Sudarshan, P., & Bhardwaj, V. (2023, June). Real-time Driver Drowsiness Detection and Assistance System using Machine Learning and IoT. In 2023 8th International Conference on Communication and Electronics Systems (ICCES) (pp. 1128-1132). IEEE.

[16] Siddik, M. A. D., & Rahman, M. S. (2023, February). Drowsiness and Lethargy Detection Using Machine Learning Techniques. In 2023 International Conference on Electrical, Computer and Communication Engineering (ECCE) (pp. 1-6). IEEE.

[17] Sravya, K. R., Aparna, C., Pujitha, B., & Krishnasai, N. (2024, May). Detection of Drowsiness Using DCCNN. In 2024 International Conference on Advances in Computing, Communication and Applied Informatics (ACCAI) (pp. 1-6). IEEE.

[18] Rawat, V., Patwa, K., Negi, H. S., & Ghai, A. S. (2024, August). Examining driver sleepiness based on his visual actions with ML Techniques. In 2024 International Conference on Electrical Electronics and Computing Technologies (ICEECT) (Vol. 1, pp. 1-5). IEEE.

[19] Yang, L., Yang, H., Wei, H., Hu, Z., & Lv, C. (2024). Video-based driver drowsiness detection with optimised utilization of key facial features. IEEE Transactions on Intelligent Transportation Systems.

[20] Kolus, A. (2024). A systematic review on driver drowsiness detection using eye activity measures. IEEE Access.