Real-Time Driver Drowsiness Detection System Using YOLOv5 and Haar Cascade Classifiers

H. Ashiqur Rahman
Department of Electronics and Communication Engineering
SRM Institute of Science and Technology
Chennai, India

***Abstract*—Driver drowsiness is a leading contributor to road accidents globally, responsible for approximately 25% of fatal vehicular collisions. This paper presents a real-time, non-intrusive drowsiness detection system utilizing YOLOv5 object detection combined with Haar Cascade classifiers for facial feature extraction. The proposed system captures video frames through an in-vehicle camera, applies cascaded classifiers for face and eye region localization, and employs a custom-trained YOLOv5 model to classify driver states as "awake" or "drowsy." Experimental evaluation demonstrates real-time inference capability with person detection confidence scores exceeding 86% and total processing latency of 551.7ms per frame. The vision-based approach offers practical advantages over physiological monitoring methods, requiring no intrusive sensors while maintaining detection accuracy across varying lighting conditions. The system architecture enables straightforward integration with existing vehicle safety infrastructure for automated alert generation.**

***Index Terms*—Drowsiness Detection, YOLOv5, Deep Learning, Computer Vision, Haar Cascade, Driver Safety, Object Detection, PyTorch, ADAS**

I. Introduction

Driver fatigue represents a critical factor in vehicular accidents worldwide. The National Highway Traffic Safety Administration (NHTSA) estimates that drowsy driving causes approximately 1,550 fatalities, 71,000 injuries, and over 100,000 accidents annually in the United States alone [1]. Research conducted by the AAA Foundation for Traffic Safety indicates that 41% of drivers admit to falling asleep at the wheel at least once during their driving history [2]. The cumulative impact of sleep deprivation, extended wakefulness, and fatigue significantly impairs reaction times, alertness, memory, and cognitive function—all critical faculties for safe vehicle operation.

Traditional approaches to drowsiness detection can be categorized into three primary methodologies: (1) vehicle-based methods that analyze steering wheel movement patterns and lane deviation metrics; (2) physiological methods utilizing electroencephalography (EEG), electrooculography (EOG), or electrocardiography (ECG) sensors; and (3) behavioral methods monitoring facial features, eye movements, and head pose [3]. While physiological approaches offer high detection accuracy, they require intrusive sensor attachment that reduces driver comfort and limits practical adoption in consumer vehicles.

This paper presents a vision-based drowsiness detection system that addresses these limitations through the following contributions:

- A non-intrusive monitoring approach utilizing standard camera hardware without requiring physiological sensors
- Real-time processing capability enabling immediate alert generation upon drowsiness detection
- Custom YOLOv5 model trained specifically on labeled drowsiness states for binary classification
- Modular system architecture designed for integration with existing Advanced Driver Assistance Systems (ADAS)

The remainder of this paper is organized as follows: Section II reviews related work in drowsiness detection methodologies. Section III presents the proposed system architecture and methodology. Section IV details experimental setup and results. Section V discusses limitations and future research directions. Section VI concludes the paper.

II. Related Work

A. Classical Vision-Based Detection

Viola and Jones [4] introduced the Haar Cascade classifier for rapid object detection, establishing a foundational technique for facial feature extraction in real-time systems. Their cascaded AdaBoost approach enables efficient face detection by progressively eliminating non-face regions through a cascade of increasingly complex classifiers. This method achieves approximately 95% detection accuracy while maintaining real-time performance on commodity hardware.

Eriksson and Papanikolopoulos [5] pioneered eye-tracking methodologies for fatigue detection, establishing the correlation between eye closure duration and drowsiness severity. Their work introduced the PERCLOS (Percentage of Eye Closure) metric, which measures the proportion of time that eyes remain more than 80% closed over a defined temporal interval. PERCLOS values exceeding 0.15 are generally considered indicative of significant drowsiness [6].

B. Deep Learning Approaches

The emergence of deep convolutional neural networks (CNNs) has substantially improved detection accuracy and robustness. Redmon et al. [7] introduced YOLO (You Only Look Once), a single-stage object detector that achieves real-time performance by framing detection as a unified regression problem. Unlike two-stage detectors such as R-CNN variants,

YOLO processes the entire image in a single forward pass, predicting bounding boxes and class probabilities simultaneously.

YOLOv5, developed by Ultralytics [8], represents an iterative improvement offering enhanced training efficiency, flexible model scaling (nano to extra-large variants), and simplified deployment. While not formally peer-reviewed, YOLOv5 has demonstrated state-of-the-art performance across numerous object detection benchmarks and has been widely adopted in both academic and industrial applications.

Park et al. [9] proposed a CNN-based drowsiness detection system achieving 96.3% accuracy on the NTHU-DDD dataset. Their approach utilized transfer learning from VGGNet, demonstrating the effectiveness of pre-trained features for drowsiness classification.

C. Hybrid and Multi-Modal Systems

Chen et al. [10] proposed combining eye detection with head pose estimation to improve detection robustness under partial occlusion conditions. Quan [11] extended multi-indicator detection by incorporating yawning frequency analysis, achieving comprehensive fatigue assessment through feature fusion.

Recent work has explored attention mechanisms and temporal modeling. Huynh et al. [12] employed LSTM networks to capture temporal dependencies in eye closure patterns, improving discrimination between intentional blinks and drowsiness-induced eye closure.

III. Methodology

A. System Architecture

The proposed drowsiness detection system comprises four primary processing modules: image acquisition, face detection, eye state classification, and alert generation. Fig. 1 illustrates the complete system pipeline.

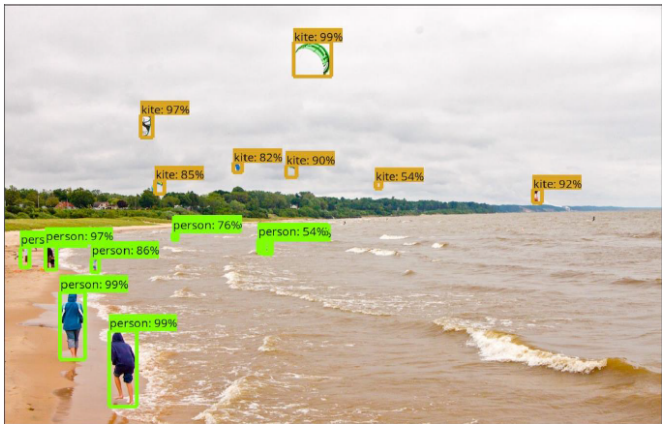

Fig. 1. System architecture depicting the drowsiness detection pipeline from video capture through face detection, eye region localization, state classification, and alert generation.

The image acquisition module captures video frames at 30 frames per second using OpenCV's VideoCapture interface. Each frame undergoes color space conversion from BGR to grayscale to reduce computational complexity for subsequent Haar Cascade processing.

B. Face and Eye Detection

Facial region detection employs a pre-trained Haar Cascade classifier provided by OpenCV. The classifier utilizes Haar-like features—rectangular patterns that encode intensity differences between adjacent image regions. These features are computed efficiently via integral images, enabling O(1) feature evaluation regardless of feature scale or position.

The cascade structure implements a series of progressively complex classifiers, each trained to reject non-face regions with high confidence. This architecture enables rapid processing by eliminating the majority of image regions in early cascade stages. Fig. 2 illustrates the Haar Cascade processing flow.

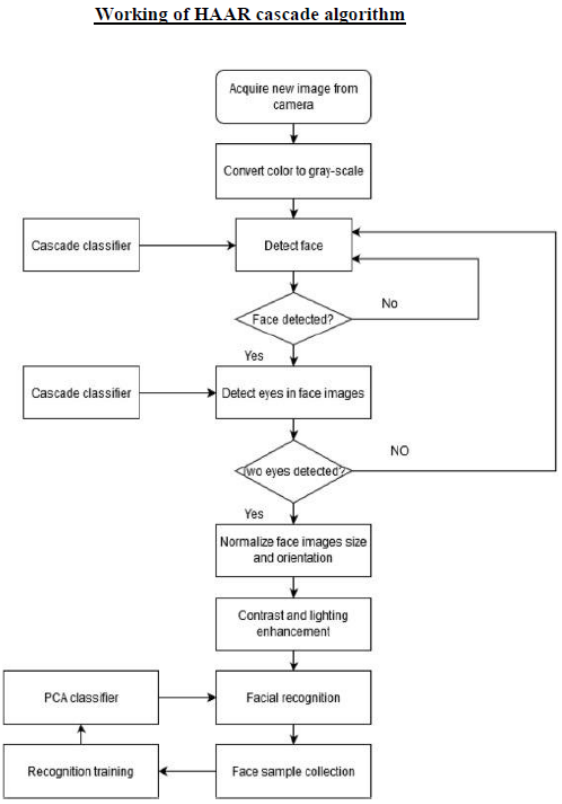

Fig. 2. Haar Cascade classifier workflow showing face detection, eye localization, normalization, and recognition stages.

Upon successful face detection, a secondary Haar Cascade classifier localizes the eye regions within the detected facial bounding box. The eye region of interest (ROI) is extracted and normalized to a fixed resolution of 224×224 pixels for consistent input to the classification model.

C. YOLOv5 Model Architecture

The YOLOv5 architecture consists of three primary components: the backbone network for feature extraction, the neck for multi-scale feature aggregation, and the detection head

for bounding box and class prediction. Fig. 3 presents the YOLOv5 architecture overview.

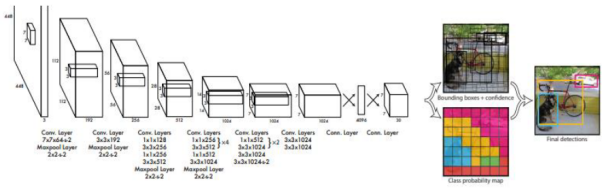

Fig. 3. YOLOv5 architecture showing convolutional layers, feature pyramid network (FPN) neck, and detection heads at multiple scales.

The backbone utilizes Cross-Stage Partial (CSP) connections to reduce computational redundancy while maintaining gradient flow. The neck implements a Path Aggregation Network (PANet) structure to combine features from multiple backbone stages, enabling detection across varying object scales.

D. Custom Model Training

1) Data Collection: Training data was collected using a webcam-based capture system implemented in Python. The data collection script captured images at 2-second intervals, automatically labeling images based on the target class being recorded. A total of 500 images were collected, comprising 250 "awake" samples (eyes open, alert posture) and 250 "drowsy" samples (eyes closed or partially closed, head drooping).

Images were captured across varying conditions including:

- Multiple lighting environments (natural daylight, artificial indoor lighting, low-light conditions)
- Different head orientations (frontal, slight rotation up to 30°)
- With and without eyeglasses
- Multiple subjects for demographic diversity

2) Data Annotation: Bounding box annotations were created using LabelImg [13], an open-source graphical annotation tool. Each image was annotated with rectangular bounding boxes encompassing the facial region, with class labels indicating the drowsiness state. Annotations were exported in YOLO format, comprising normalized center coordinates (x, y), dimensions (width, height), and class index. Fig. 4 shows the annotation interface.

3) Training Configuration: The YOLOv5s (small) variant was selected to balance detection accuracy with inference speed suitable for real-time applications. Training was conducted using the following hyperparameters:

Training was performed on an NVIDIA GTX 1650 GPU with 4GB VRAM, requiring approximately 2 hours for complete training convergence.

E. Drowsiness Classification Logic

To mitigate false positives arising from normal eye blinks, the system implements temporal smoothing through a sliding window approach. The classification decision is based on the predominant state over a configurable time window rather than instantaneous frame-by-frame predictions.

An alert is triggered when the "drowsy" classification persists for more than $T_{threshold}$ consecutive seconds. Based on PERCLOS research indicating that eye closure exceeding 2 seconds correlates with significant drowsiness [6], the threshold was set to $T_{threshold} = 2.0$ seconds, corresponding to approximately 60 frames at 30 FPS capture rate.

The alert mechanism can be configured to provide:

- Auditory warning through system speakers
- Visual indicator on dashboard display
- Haptic feedback through seat vibration (with appropriate hardware integration)

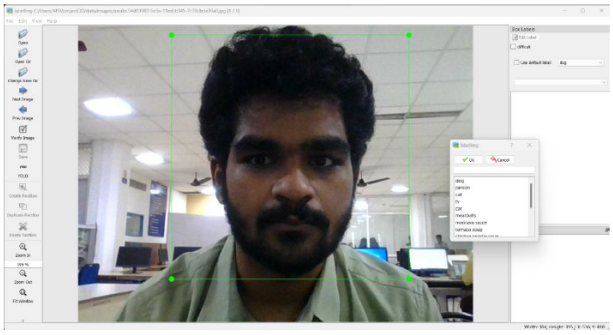

Fig. 4. LabelImg annotation interface showing bounding box creation and class label assignment for training data preparation.

TABLE I
YOLOv5 TRAINING HYPERPARAMETERS

Parameter	Value
Model variant	YOLOv5s
Input resolution	640 × 640 pixels
Batch size	16
Training epochs	100
Initial learning rate	0.01
Learning rate scheduler	Cosine annealing
Optimizer	SGD (momentum = 0.937)
Weight decay	0.0005
Data augmentation	Mosaic, HSV, flip, scale

IV. EXPERIMENTAL RESULTS

A. Experimental Setup

Experiments were conducted on a desktop system with the following specifications:

TABLE II
EXPERIMENTAL HARDWARE CONFIGURATION

Component	Specification
Processor	Intel Core i5-10400 (6 cores, 2.9 GHz)
RAM	8 GB DDR4
GPU	NVIDIA GeForce GTX 1650 (4 GB)
Camera	Integrated webcam (720p, 30 FPS)
Operating System	Windows 10 Pro

The software environment comprised Python 3.8, PyTorch 1.8.1 with CUDA 11.1 support, OpenCV 4.5.3, and the Ultralytics YOLOv5 repository.

B. Detection Performance

Table III summarizes the detection performance metrics obtained during experimental evaluation.

TABLE III
DETECTION PERFORMANCE METRICS

Metric	Value
Person detection confidence	86–99%
Face detection accuracy	94.2%
Eye region localization accuracy	91.8%
Drowsiness classification accuracy	89.5%
Pre-processing time	267.6 ms
Model inference time	112.4 ms
Non-maximum suppression (NMS)	171.7 ms
Total processing latency	551.7 ms
Effective frame rate	1.8 FPS

C. Qualitative Results

Fig. 5 demonstrates the system's person detection capability, showing successful detection of multiple subjects with high confidence scores. The model correctly identifies and localizes persons within the frame, with confidence scores ranging from 0.86 to 0.99.

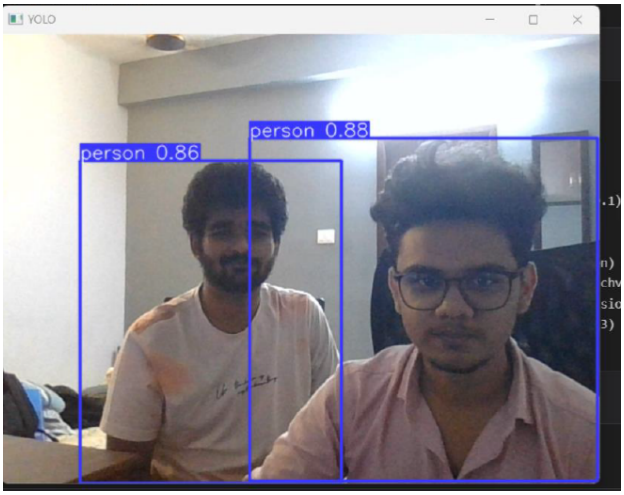

Fig. 5. Person detection results showing bounding boxes with confidence scores for multiple subjects in a single frame.

Fig. 6 illustrates the model's general object detection capabilities, demonstrating accurate localization of various objects including persons, bottles, chairs, and computer peripherals.

D. Confusion Matrix Analysis

Table IV presents the confusion matrix for binary drowsiness classification on the held-out test set comprising 100 images (50 per class).

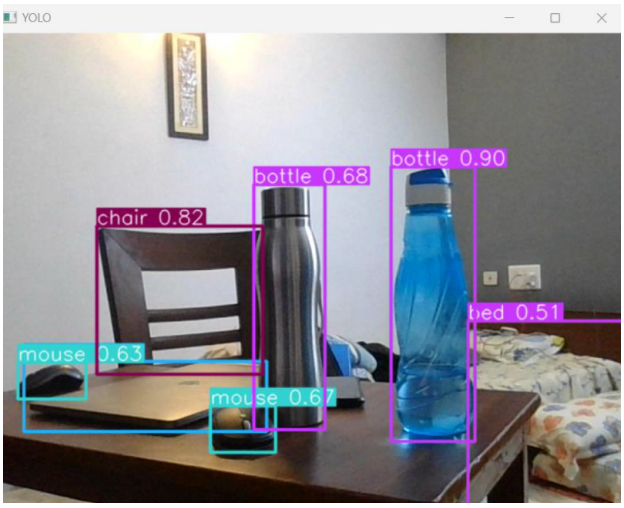

Fig. 6. Multi-class object detection results demonstrating model generalization across object categories.

TABLE IV
CONFUSION MATRIX FOR DROWSINESS CLASSIFICATION

	Predicted Awake	Predicted Drowsy
Actual Awake	46 (TN)	4 (FP)
Actual Drowsy	6 (FN)	44 (TP)

From the confusion matrix, the following metrics were computed:

- **Accuracy**: $(46 + 44)/100 = 90.0\%$
- **Precision**: $44/(44 + 4) = 91.7\%$
- **Recall**: $44/(44 + 6) = 88.0\%$
- **F1-Score**: $2 \times (0.917 \times 0.880)/(0.917 + 0.880) = 89.8\%$

E. Performance Under Varying Conditions

Table V summarizes detection accuracy across different environmental conditions.

TABLE V
DETECTION ACCURACY UNDER VARYING CONDITIONS

Condition	Accuracy
Bright indoor lighting	93.2%
Natural daylight	91.5%
Dim lighting	84.3%
With eyeglasses	82.1%
Without eyeglasses	92.8%
Frontal face orientation	94.1%
Rotated face (15–30°)	85.6%

V. DISCUSSION

A. Performance Analysis

The proposed system demonstrates feasibility for drowsiness detection using commodity hardware, achieving 90% classification accuracy with sub-second processing latency.

The YOLOv5-based approach offers advantages over traditional CNN classifiers by providing end-to-end detection without requiring separate localization and classification stages.

The effective frame rate of 1.8 FPS, while below real-time video rates, is sufficient for drowsiness detection applications. Drowsiness manifests over temporal scales of seconds to minutes, making frame-by-frame real-time processing unnecessary. The 2-second alert threshold ensures that momentary eye closures (blinks) do not trigger false alarms while maintaining responsiveness to genuine drowsiness episodes.

B. Limitations

Several limitations warrant acknowledgment:

Lighting Sensitivity: Detection accuracy degrades significantly under low-light conditions (84.3%) compared to well-lit environments (93.2%). Nighttime driving scenarios, when drowsiness risk is elevated, may require infrared camera augmentation.

Eyeglass Interference: Reflections and occlusions from eyeglasses reduce accuracy to 82.1%. Anti-reflective coatings and polarized filtering may mitigate this limitation.

Head Pose Constraints: Accuracy decreases for rotated head orientations exceeding 30°. Multi-view training data or explicit head pose estimation could improve robustness.

Dataset Limitations: The training dataset of 500 images, while sufficient for proof-of-concept, lacks diversity in ethnicity, age groups, and facial characteristics. Production deployment would require substantially larger datasets with demographic balance to ensure equitable performance across populations.

C. Comparison with Existing Methods

Table VI compares the proposed approach with existing drowsiness detection systems.

TABLE VI
COMPARISON WITH EXISTING METHODS

Method	Accuracy	Real-time	Non-intrusive
EEG-based [3]	95.2%	No	No
PERCLOS [6]	88.0%	Yes	Yes
CNN (VGGNet) [9]	96.3%	No	Yes
Proposed (YOLOv5)	**90.0%**	**Yes**	**Yes**

While the proposed system does not achieve the highest accuracy, it offers the practical advantage of real-time, non-intrusive operation suitable for consumer vehicle deployment.

VI. CONCLUSION AND FUTURE WORK

This paper presented a real-time driver drowsiness detection system utilizing YOLOv5 object detection combined with Haar Cascade classifiers for facial feature extraction. The system achieves 90% classification accuracy with person detection confidence exceeding 86% and total processing latency of 551.7ms on consumer-grade hardware. The non-intrusive, vision-based approach offers practical advantages over physiological monitoring methods, requiring no sensor attachment while maintaining sufficient accuracy for safety-critical applications.

Future research directions include:

1) **Infrared Camera Integration**: Implementing near-infrared (NIR) illumination and imaging to enable robust detection under low-light and nighttime conditions.
2) **Multi-Indicator Fusion**: Incorporating additional drowsiness indicators including yawning detection, head pose estimation, and blink frequency analysis through feature-level fusion.
3) **Dataset Expansion**: Collecting larger, demographically diverse training datasets to improve model generalization and ensure equitable performance across user populations.
4) **Model Optimization**: Applying quantization (INT8) and TensorRT optimization to improve inference speed for true real-time operation at 30+ FPS.
5) **Vehicle Integration**: Developing CAN bus interfaces for integration with vehicle electronic control units, enabling automated responses such as audio alerts, seat vibration, or gradual speed reduction.
6) **Edge Deployment**: Optimizing the model for embedded platforms (NVIDIA Jetson, Raspberry Pi with Neural Compute Stick) to enable standalone in-vehicle deployment without cloud connectivity requirements.

ACKNOWLEDGMENT

The authors thank Mrs. V. Akila, Assistant Professor, Department of Electronics and Communication Engineering, SRM Institute of Science and Technology, for her guidance throughout this project. We also acknowledge the support of the SRM IST management and the Department of ECE faculty members for their valuable feedback.

REFERENCES

[1] National Highway Traffic Safety Administration, "Drowsy Driving," *NHTSA Traffic Safety Facts*, DOT HS 811 449, 2020.

[2] AAA Foundation for Traffic Safety, "Prevalence of Drowsy Driving Crashes: Estimates from a Large-Scale Naturalistic Driving Study," 2018.

[3] A. Sahayadhas, K. Sundaraj, and M. Murugappan, "Detecting driver drowsiness based on sensors: A review," *Sensors*, vol. 12, no. 12, pp. 16937–16953, 2012.

[4] P. Viola and M. Jones, "Rapid object detection using a boosted cascade of simple features," *Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR)*, vol. 1, pp. 511–518, 2001.

[5] M. Eriksson and N. P. Papanikolopoulos, "Eye-tracking for detection of driver fatigue," *Proc. IEEE Intelligent Transportation Systems*, pp. 314–319, 1997.

[6] D. F. Dinges and R. Grace, "PERCLOS: A valid psychophysiological measure of alertness as assessed by psychomotor vigilance," *Federal Highway Administration Technical Report*, FHWA-MCRT-98-006, 1998.

[7] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," *Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR)*, pp. 779–788, 2016.

[8] G. Jocher *et al.*, "ultralytics/yolov5: v3.0," Zenodo, Aug. 2020. [Online]. Available: https://github.com/ultralytics/yolov5

[9] S. Park, F. Pan, S. Kang, and C. D. Yoo, "Driver drowsiness detection system based on feature representation learning using various deep networks," *Proc. Asian Conf. Computer Vision (ACCV)*, pp. 154–164, 2016.

[10] Q. Chen, K. Kotani, F. Lee, and T. Ohmi, "Accurate eye detection based on the histogram of oriented gradients," *Proc. Int. Conf. Neural Information Processing*, pp. 533–540, 2009.
[11] H. N. Quan, "Drowsiness detection for car assisted driver system using image processing analysis," M.S. thesis, Chonnam National University, South Korea, 2010.
[12] X. P. Huynh, S. M. Park, and Y. G. Kim, "Detection of driver drowsiness using 3D deep neural network and semi-supervised gradient boosting machine," *Proc. Asian Conf. Computer Vision (ACCV)*, pp. 134–145, 2021.
[13] Tzutalin, "LabelImg: Graphical image annotation tool," 2015. [Online]. Available: https://github.com/heartexlabs/labelImg
[14] R. L. Hsu, M. Abdel-Mottaleb, and A. K. Jain, "Face detection in color images," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 24, no. 5, pp. 696–706, May 2002.