International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 09 Issue: 03 | March - 2025 SJIF Rating: 8.586 ISSN: 2582-3930

Smart Vehicle Assistant

1st Shamaila Ansari
B.E., Information Technology
V.E.S.I.T, Mumbai, India
ansarishamaila5@gmail.com

2nd Mahek Bhagwani
B.E., Information Technology
V.E.S.I.T, Mumbai, India
mahekbhagwani4444@gmail.com

3rd Om Gaikwad
B.E., Information Technology
V.E.S.I.T, Mumbai, India
omgaikwadb221@gmail.com

4th Sneha R. Sumbe
B.E., Information Technology
V.E.S.I.T, Mumbai, India
sumbesneha@gmail.com

5th Mrs. Jayashree Hajgude
Asst. Prof., Information Technology
V.E.S.I.T, Mumbai, India
jayashree.hajgude@ves.ac.in

***Abstract*—Driver safety is a prime challenge in present day transportation systems, in which factors inclusive of sleep, surroundings, and vehicle dynamics considerably have an impact on site visitors accidents. This paper provides an incorporated machine, with AI and includes a proposal that integrates superior pc imaginative and prescient, system studying algorithms and real-time for statistics processing In addition to dates detected through landmark evaluation, vehicles that detection using latest object detection models inclusive of YOLO and lane departure warnings the usage of convolutional neural networks (CNNs), the system includes photo enhancement to overcome challenges which includes light a address ground conditions and environmental modifications. Extensive checking out and evaluation demonstrates the robustness, scalability and flexibility of the system to actual-international situations, massively improving accuracy and reaction time This painting lays the foundation for sensible, secure using solutions and contributes to the development of self sustaining automobile technology.**

Keywords- Vehicle Detection, Lane Departure Warning, Object Detection Algorithms, YOLO (You Only Look Once), Convolutional Neural Networks (CNNs), Facial Landmark Analysis, Real-Time Image Processing, Autonomous Vehicles, Low-Light Image Enhancement, Machine Learning for Transportation, AI in Driver Assistance Systems, Deep Learning in Automotive Applications, Traffic Safety Technologies.

I. Introduction

The World Health Organization estimates that traffic accidents claim the lives of 1.35 million people each year, making road safety a major global concern. A large percentage of these occurrences are the consequence of human mistake, including weariness, inattention, and delayed reactions to obstructions or traffic signals. There is a need for proactive safety solutions because conventional car safety features like airbags, seat belts, and anti-lock brake systems (ABS) mostly respond to collisions rather than preventing them.

Lane departure warnings and collision avoidance are two new elements of Advanced Driver Assistance Systems (ADAS). The intricacies of human behavior and the variety of difficulties presented by actual driving situations, particularly in congested areas like India, are frequently overlooked by these systems. Intelligent solutions that can offer context-aware, real-time support are essential to reducing these difficulties and enhancing traffic safety.

By utilizing computer vision and machine learning (ML) technology, the Smart Vehicle Assistant (SVA) seeks to close this gap and provide proactive assistance to drivers. To improve situational awareness, the system incorporates a number of features, such as voice-guided notifications, traffic signal recognition, object identification, and sleep monitoring. The SVA takes a context-aware approach, which guarantees that important alerts are selected and sent efficiently, in contrast to traditional systems that frequently bombard drivers with constant reminders.

In order to help prevent collisions, the SVA classifies risks by identifying cars, pedestrians, and other obstructions. Using face and eye movement analysis, it tracks driver fatigue and, if required, sounds a voice or haptic alert. To guarantee adherence to traffic laws, the technology detects stop signs, speed zones, and traffic lights. Furthermore, the voice guidance system uses a customized text-to-speech interface to give drivers context-sensitive, real-time feedback, providing crucial guidance without putting undue strain on the driver. By concentrating on important, high-priority events, the alert filtering tool minimizes pointless messages in locations with high traffic volumes. In order to avoid being distracted by unnecessary messages, drivers can also customize alert settings according to their tastes and driving patterns.

The SVA is especially appropriate for driving situations in nations with dense traffic and varied road infrastructure because of its emphasis on pragmatism and real-world flexibility. According to preliminary findings, the system can operate efficiently in real-time situations, showing promise for improving driver safety and lowering accident rates.

II. Literature Review

Several studies have contributed to the development of intelligent vehicle systems by exploring various algorithms and methodologies:

Bo Yang [1] examined vehicle detection using artificial intelligence and proposed a semi-supervised learning algorithm based on Support Vector Machines (SVM). While the study advanced motion detection techniques, it identified challenges related to feature selection and interpretability, which could impact real-world performance.

Swapnil Titare et al. [2] developed a drowsiness detection system utilizing machine learning algorithms to monitor driver alertness. Their approach effectively analyzed eye movements and facial expressions, achieving a high accuracy rate. However, environmental factors, such as lighting conditions, were noted as significant challenges that could adversely affect performance.

Rajesh Kannan et al. [3] focused on traffic sign detection using deep learning methods, specifically for Indian road conditions. The proposed system demonstrated robust performance under various environmental factors but was limited to Indian traffic signs, highlighting the need for broader applicability in diverse geographic contexts.

Noor Jannah Zakaria et al. [4] conducted a systematic review of lane detection methodologies for autonomous vehicles. The study compared traditional geometric modeling approaches with advanced machine learning techniques, such as Convolutional Neural Networks (CNNs). It highlighted limitations in dataset diversity and scalability as key challenges in many methodologies.

Abduladhem Ali et al. [5] presented a method for distance estimation and vehicle position detection using a monocular camera. Their framework included detecting lane markings and tracking objects from video input. However, the study identified challenges related to camera calibration, which could affect the accuracy of distance measurements and overall system performance.

J. Redmon et al. [6] introduced the YOLO (You Only Look Once) framework, which revolutionized object detection with its real-time performance and end-to-end training pipeline. The study demonstrated significant improvements in speed and accuracy, making YOLO suitable for autonomous systems. However, the reliance on large labeled datasets posed challenges for deployment in scenarios such as low-light conditions or specific geographic contexts.

M. Jeong et al. [7] proposed a real-time drowsiness detection system using facial landmark analysis and deep learning-based classifiers. The system achieved high performance in controlled environments, but external factors, such as driver posture and occlusions, affected its robustness in real-world scenarios.

R. Girshick et al. [8] advanced object detection by developing Region-based Convolutional Neural Networks (R-CNN). This approach enabled higher accuracy through selective search and region proposal methods. However, computational inefficiencies limited the applicability of R-CNN in real-time vehicular systems.

C. Szegedy et al. [9] introduced Inception networks, which enhanced object recognition by employing multi-scale feature extraction. The study demonstrated the effectiveness of this approach for detecting traffic signs and objects in urban driving scenarios. However, the computational overhead was identified as a limitation, particularly for resource-constrained environments.

P. Viola and M. Jones [10] developed the Viola-Jones framework, which remains a cornerstone for real-time face detection. Their Haar cascade classifier demonstrated exceptional speed, but the framework struggled with variations in lighting and pose, limiting its effectiveness in dynamic environments like vehicle interiors.

A. Howard et al. [11] designed MobileNet, a lightweight convolutional neural network tailored for resource-constrained systems. This architecture enabled efficient traffic signal recognition and drowsiness detection on mobile and embedded devices. However, the trade-offs between accuracy and speed were a noted limitation.

Z. Xu et al. [12] proposed a method for low-light image enhancement using deep learning-based noise reduction and contrast enhancement. Their approach significantly improved detection performance in nighttime driving scenarios, although the computational requirements presented a challenge for real-time implementation.

III. System Architecture And Methodology

The Smart Vehicle Assistant (SVA) leverages machine learning and computer vision technologies to deliver real-time driver assistance. The system features a modular design, ensuring adaptability and efficiency across diverse driving environments. It comprises three primary modules—Drowsiness Detection, Object Detection, and Traffic Signal Recognition—that work in synergy to enhance situational awareness and minimize driver distractions. This section details the architecture, methodology, and approaches for handling low-light conditions.

A. System Architecture

Figure 1 illustrates the SVA's architecture, which combines input handling, processing, and output components in a modular fashion. In order to improve computing performance, especially in low-light conditions, the input module preprocesses the real-time video feed from cameras mounted on vehicles by converting it to grayscale. Three subcomponents make up the processing module. The Eye Aspect Ratio (EAR) and facial landmark detection are used by the Drowsiness Detection subcomponent to track the driver's degree of exhaustion. Real-time analysis of eye closure durations is used to trigger alerts. In order to help prevent collisions, the Object Detection subcomponent uses the YOLO (You Only Look Once) algorithm to recognize cars, people, and obstructions while categorizing risk levels.A Convolutional Neural Network (CNN) is used by the Traffic Signal Recognition subcomponent to recognize traffic lights, stop signs, and speed zones. A variety of datasets, including ones taken in poor light, are used to train the system. In order to reduce distractions, the notification system prioritizes key scenarios while delivering visual and audible

Volume: 09 Issue: 03 | March - 2025 SJIF Rating: 8.586 ISSN: 2582-3930

notifications. The output module uses visual overlays, and voice instruction to inform the driver of alarms. Figure 2 shows the flow of operations, which describes how modules interact with one another and the workflow as a whole.

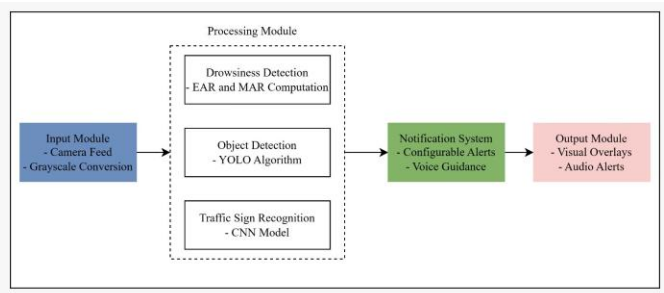

Fig. 1. System Architecture

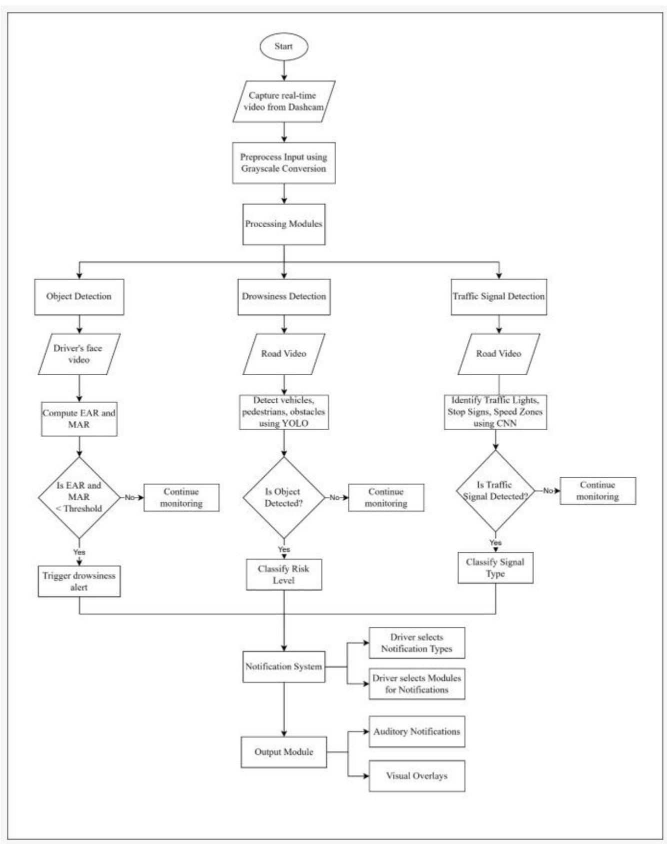

Fig. 2. System Workflow

B. Methodology

- **Preprocessing and Input Handling:** Cameras mounted on vehicles provide real-time video feed to the system. The input feed is converted to grayscale to reduce computational complexity and suppress noise, particularly in low-light conditions. This preprocessing ensures robust feature extraction in subsequent modules.
- **Drowsiness Detection Algorithm:** Using facial landmarks, the module tracks the driver's degree of weariness. Key measures include:
  - **Facial Landmark Detection:** Pre-trained models like Dlib or Mediapipe detect facial landmarks (eyes, mouth, nose), represented by 2D coordinates.
  - **Eye Aspect Ratio (EAR):** Tracks eye openness. The EAR is calculated as follows: (See Equation (1) for more details.)

$$\text{EAR} = \frac{\|\boldsymbol{P}2 - \boldsymbol{P}6\| + \|\boldsymbol{P}3 - \boldsymbol{P}5\|}{2 \times \|\boldsymbol{P}1 - \boldsymbol{P}4\|} \qquad (1)$$

    (See Equation (1) for more details.)
    If EAR falls below 0.2 for 2 seconds, the system triggers an alert.
  - **Mouth Aspect Ratio (MAR):** Detects yawning. The MAR is computed using:

$$\text{MAR} = \frac{\|\boldsymbol{P}2 - \boldsymbol{P}8\| + \|\boldsymbol{P}3 - \boldsymbol{P}7\| + \|\boldsymbol{P}4 - \boldsymbol{P}6\|}{3 \times \|\boldsymbol{P}1 - \boldsymbol{P}5\|} \qquad (2)$$

    (Refer to Equation (2) for additional details.)
    If MAR exceeds 0.5 for a threshold duration, a yawning alert is generated.
  - **Drowsiness Decision:** The system evaluates drowsiness based on the EAR (Equation (1)) and MAR (Equation (2)). If both indicators persist beyond predefined thresholds, an alert is triggered.
- **Object Detection Algorithm:** The YOLO algorithm detects objects in real-time. Frames are divided into grids; each grid predicts bounding boxes and confidence scores. Objects are classified as low-, moderate-, or high-risk based on proximity thresholds.
  - **How YOLO Operates:** The input video frame is divided into a grid of SxS cells. Each grid cell predicts bounding boxes, class probabilities, and confidence scores for detected objects. The approach combines localization and classification in a single pass, utilizing a single neural network to provide end-to-end detection.
  - **Risk Assessment:** The system calculates the distance and proximity of each object it detects to the vehicle. Objects are classified as either low-risk, moderate-risk or high-risk using a threshold-based categorization system. Collision alerts are generated using this data.
- **Traffic Signal Recognition Algorithm:** The traffic signal recognition module scans and categorizes traffic signs using a proprietary Convolutional Neural Network (CNN).
  - **Preprocessing:** In order to adjust for changes in angles and illumination, the input photographs are scaled, normalized, and enhanced.

- **CNN Architecture:** 1.Input Layer: Accepts pre-processed images of fixed dimensions (e.g., 64x64 pixels).
  2. Convolutional Layers: Extract spatial features such as edges, textures, and shapes.
  3. Pooling Layers: Downsample the feature maps to reduce dimensionality and computation.
  4. Fully Connected Layers: Perform classification to identify traffic signs, such as stop signs, speed limits, and traffic lights.

- **Training and Dataset Augmentation:** The CNN is trained on a dataset containing Indian Traffic Sign available on Kaggle. Techniques like random rotations, brightness adjustments, and noise addition are applied for dataset augmentation.
- **Output:** The system assures adherence to traffic laws by producing voice alerts to notify the driver of detected traffic signs.
- **Notification System** Given the notification system's high degree of customization, drivers can alter their alert choices prior to driving. The system generates real-time alerts using the data processed by the traffic signal recognition, object identification, and drowsiness detection modules.
  - **Voice Guidance:** Notifications in audio form of essential alerts.
  - **Visual Overlays:** Displaying alerts on the dashboard of the car.

IV. Results

A. *Results for YOLOv8 Object Detection Model*

The YOLOv8 object detection model showcases exceptional performance in detecting and localizing objects, as summarized in the following table:

Metric	Value
Model Layers	168
Precision	0.957
mAP@50	0.667
Inference Speed	10.5ms/image

Table I
Performance Metrics For YOLOv8 Model

key results emphasize the model's strengths in object detection and real-time performance. With a high precision of 0.957, the model shows excellent classification accuracy, with minimal false positives and reliable object detection in complex environments. The model also has strong localization capabilities, achieving a mean Average Precision (mAP@50) of 0.667, indicating its effectiveness in identifying object locations with high precision under a strict Intersection over Union (IoU) threshold.

The model also excels at detecting critical object classes, including pedestrians (mAP@50 = 0.995) and obstacles (mAP@50 = 0.677), thereby ensuring robust functionality in structured and real-time environments. Furthermore, with an inference speed of 10.5 ms per image optimized for better performance, the model is best suited for real-time applications that require rapid decision-making, thus being particularly beneficial for surveillance and autonomous systems.

B. *Results for CNN and Mediapipe*

The CNN for signal detection and Mediapipe for drowsiness detection demonstrates a balanced approach to addressing diverse application needs. The key metrics are as follows:

Model/framework	Metric	Value
CNN (signal detection)	Accuracy	80.5%
Mediapipe (drowsiness)	Sens./Spec.	85.4% / 88.2%

Table II
Performance Metrics Of Different Models

Key observations: The key observations from this section are the 80.5% accuracy in the CNN model and that the proposed architecture effectively captures spatial features of the signal. Such a network could be appropriate for classification of signals in activity recognition and anomaly detection tasks across several domains.

The Mediapipe-powered drowsiness detection system exhibited a sensitivity of 85.4%, ensuring reliable identification of drowsy individuals and reducing the risk of missing critical events. Additionally, the system achieved a specificity of 88.2%, effectively minimizing false alarms. This balance between sensitivity and specificity enhances its practicality in real-world applications, ensuring accurate detection while preventing unnecessary interruptions for non-drowsy users.

V. Conclusion

This study evaluated the performance of the YOLOv8 object detection model, CNN and Mediapipe for signal and drowsiness detection, respectively. The results highlight the robustness and practicality of these models in real-time and safety-critical environments. However, some limitations and opportunities for improvement were also observed, which are discussed below.

The YOLOv8 model exhibited high precision, primarily due to its advanced architecture optimized for detecting critical object classes such as pedestrians and obstacles. The efficient feature extraction and effective use of anchor boxes contributed significantly to minimizing false positives. However, the lower mAP@50 score for certain object categories suggests that the model's localization accuracy can be improved under varying environmental conditions or for smaller objects. Future work could involve fine-tuning the model using additional domain-specific data and augmenting the dataset to address these challenges.

The CNN model achieved an acceptable classification accuracy for signal detection, but its performance could be enhanced by exploring deeper architectures or incorporating advanced feature extraction techniques such as transfer learning. Additionally, increasing the dataset size and diversity may improve generalization across different signal patterns.

The Mediapipe-based drowsiness detection system demonstrated balanced sensitivity and specificity, ensuring reliable detection with minimal false alarms. However, its performance might be affected by variations in user behavior or environmental factors such as lighting and camera angles. Incorporating adaptive algorithms or leveraging multimodal inputs (e.g., heart rate or eye-tracking data) could further enhance the robustness of the system.

In conclusion, while the models presented here show strong potential for real-world applications, future research should focus on addressing their limitations. Improvements in dataset diversity, architecture optimization, and integration of additional features could further enhance their reliability and applicability in complex and dynamic environments. By addressing these areas, the models could achieve even greater precision and robustness, solidifying their utility in critical domains such as surveillance, healthcare, and autonomous systems.

VI. References

[1] B. Yang, "Vehicle detection through artificial intelligence: A semi-supervised learning algorithm based on SVM," *Artificial Intelligence Journal*, 2023.

[2] S. Titare, A. Sharma, and P. Verma, "Drowsiness detection system utilizing machine learning algorithms," *Journal of Intelligent Systems and Applications*, 2021.

[3] R. Kannan, S. R. Prakash, and M. Gupta, "Traffic sign detection using deep learning methods for Indian road conditions," *IEEE Transactions on Intelligent Transportation Systems*, vol. 24, no. 3, pp. 456–465, 2023.

[4] N. J. Zakaria, S. M. Azmi, and T. Rahman, "A systematic review of lane detection in autonomous vehicles: Geometric modeling vs. machine learning," *International Journal of Automation and Computing*, vol. 20, no. 4, pp. 789–805, 2023.

[5] A. Ali, M. Hasan, and R. Zubair, "Distance estimation and vehicle position detection using a monocular camera," *Journal of Transportation Technologies*, vol. 5, no. 2, pp. 123–132, 2016.

[6] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You Only Look Once: Unified, Real-Time Object Detection," *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 779–788, 2016.

[7] M. Jeong, S. Park, and K. Kim, "Real-time drowsiness detection based on facial landmarks and deep learning," *IEEE Access*, vol. 8, pp. 56356–56367, 2020.

[8] R. Girshick, "Fast R-CNN," *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, pp. 1440–1448, 2015.

[9] C. Szegedy, W. Liu, Y. Jia, et al., "Going Deeper with Convolutions," *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 1–9, 2015.

[10] P. Viola and M. Jones, "Rapid Object Detection using a Boosted Cascade of Simple Features," *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, vol. 1, pp. 511–518, 2001.

[11] A. Howard, M. Zhu, B. Chen, et al., "MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications," *arXiv preprint arXiv:1704.04861*, 2017.

[12] Z. Xu, J. Wang, and X. Chen, "Low-Light Image Enhancement with Deep Learning," *IEEE Access*, vol. 7, pp. 17623–17634, 2019.