Published by :
https://www.ijert.org/
An International Peer-Reviewed Journal

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 04 , April - 2026

Navigating the Gaps: A Comparative Analysis of Vision, Motion, and Hybrid-Based Approaches in Real-Time Driver Monitoring

Ian Carl C. Quiñones, Jimmel Jude P. Tumagcao, Jay Ar P. Esparcia
Department of Computer Engineering
University of Southern Mindanao, Kabacan, Cotabato, Philippines

Abstract - Driving while drowsy remains a leading cause of road accidents worldwide, often resulting from prolonged driving hours and sleep deprivation. This paper presents a comparative analysis of vision-based, motion-based, and hybrid-based approaches for real-time driver monitoring using machine learning techniques. Vision-based systems focus on facial cues such as eye closure, head pose, and gaze tracking, typically implemented through convolutional neural networks and transformer architectures. Motion-based systems infer drowsiness by analyzing how a driver physically operates the vehicle; this includes steering patterns, lane-switching frequency, and acceleration behavior. Hybrid systems integrate multiple data sources to improve detection reliability and reduce false alarms. A thorough review of studies published between 2020 and 2025 was conducted, evaluated across five dimensions: detection accuracy, computational load, real-time viability, environmental robustness, and deployment feasibility. The findings indicate that vision-based models perform well in controlled settings but degrade under low-light conditions. Motion-based models offer stronger privacy since they do not rely on cameras, though their accuracy is susceptible to unpredictable road conditions. Hybrid models emerge as the most balanced approach, albeit at the cost of increased complexity and implementation difficulty. This comparative review highlights existing gaps in adaptability, dataset coverage, and real-world deployment, identifying key directions for future research in driver monitoring systems.

I. INTRODUCTION

Driver drowsiness is a major contributing factor to traffic accidents worldwide because it significantly affects reaction time, situational awareness, and decision-making ability of the driver. Drowsiness-related crashes often happen without prior evasive maneuvers, which results in severe consequences. With the rapid development of intelligent transportation systems and advanced driver-assistance systems (ADAS), real-time driver monitoring has emerged as an important research area aimed at improving road safety and reducing drowsiness-related incidents.

Recent developments in artificial intelligence (AI) and machine learning (ML) have enabled automated driver monitoring systems that can detect drowsiness through physiological and behavioral signs. In general, current methods can be divided into three main categories: vision-based, motion-based, and hybrid-based systems. Vision-based methods use convolutional neural networks (CNNs), recurrent neural networks (RNNs), and other neural networks to analyze facial features such as eye closure rate, blink frequency, head attitude, and gaze direction, as demonstrated in recent transformer- and CNN-based systems [1]–[7]. With the availability of vast annotated datasets and deep learning models that can identify precise visual information, these systems have shown excellent identification accuracy in controlled conditions.

Motion-based systems make use of driving behavior and vehicle motion to detect what state the driver currently is through steering and vehicular dynamics analysis [8]–[11]. Commonly used markers of drowsiness include turn rate, lane changes, acceleration behaviors, and steering wheel angle change. Since motion-based systems are typically discreet and less sensitive to lighting conditions, they are suitable for real-world scenarios. However, their effectiveness may differ depending on the drivers' driving behaviors, traffic patterns, and road conditions.

Hybrid-based systems combine multiple data sources, such as motion signals or other sensor inputs, with visual signals to get around the drawbacks of single-modality systems. These systems typically employ sensor fusion strategies and attention-based architectures to enhance robustness and reduce false alarms, as seen in recent multimodal fusion studies [12]–[15]. Although

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 04 , April - 2026

hybrid-based systems often achieve superior performance, they introduce increased computational complexity, higher deployment costs, and integration challenges in real-time embedded systems.

Despite significant advances, many gaps remain in current research. The use of either controlled data or simulated data to evaluate a model makes its application in actual scenarios difficult. Furthermore, since various models will be using varied data sets, criteria for measuring their effectiveness, and methods of evaluation, comparisons among them might sometimes turn out to be wrong. An exhaustive comparative analysis involving an examination of both the strengths and weaknesses, real-time capabilities, and the challenges associated with the three models is necessary.

The following analysis represents a comparison among the vision-based approach, the motion-based method, and the hybrid method for real-time driver monitoring. Through this analysis, current limitations and future research issues hindering the large-scale deployment of these systems will be identified. In addition, this study will contribute to providing a more detailed understanding of the capabilities and drawbacks of each model, which will help future endeavors to create better real-time driver monitoring models.

II. Methods And Procedures

This study examines and compares vision-based, motion-based, and hybrid-based systems as the three primary methods for real-time driver monitoring. This study analyzes and synthesizes previous machine learning–based research using a narrative review approach rather than conducting experiments. The goal is to identify common performance patterns, strengths, weaknesses, and areas that need further investigation.

A. Literature Search Strategy

The related studies were collected from trusted academic sources such as MDPI, ScienceDirect, and Semantic Scholar. Peer-reviewed journal articles and conference papers published between 2020 and 2025 were included to ensure recency and relevance. However, in order to offer methodological background for motion-based driver monitoring systems, a few foundational research studies that were published prior to 2020 were kept [8], [10]. "Driver drowsiness detection," "real-time driver monitoring," "vision-based detection," "Vehicle patterns," "motion detection," and "hybrid driver monitoring systems" were among the search terms utilized.

Only studies that met the following inclusion criteria were considered:

1. The system has to use machine learning or deep learning methods.
2. The system must demonstrate or assert real-time application.
3. Quantitative performance metrics shall be reported.
4. The methodology must clearly specify the sensing model.

Studies were excluded if they:

- Focus only on physiological sensors that are wearable (e.g., EEG-only systems without vehicle or vision integration).
- Lack quantitative evaluation results,
- Were published before 2020, except for foundational works necessary for methodological background,
- Were to review papers without original methodological contributions.

This filtering process ensured a focused comparison of modern real-time monitoring systems.

B. Categorization Framework

The selected studies were sorted into three categories depending on the type of sensing method they use:

1. *Vision-Based Approaches*
   Face and eye recognition system that uses the camera to detect the facial and eye movements, which include the posture of the head, blink rate, and closed eyes. Such systems use transformer architecture, RNNs, or CNNs.
2. *Motion-Based Approaches*
   These are systems that analyze driver behavior and vehicle dynamics using metrics such as variations in the steering wheel angle, constant lane switching, rate of rotation, acceleration, and the trajectory of the vehicle. Such techniques usually apply traditional machine learning algorithms or small neural network models to time-series data.
3. *Hybrid-Based Approaches*
   Systems that integrate two or more sensing modes (e.g., camera + steering data + inertial measurements). To improve stability and reduce false alarms, these methods often use feature-level or decision-level fusion techniques.

This categorization enables a structured mode-based comparison between different detection paradigms.

C. Evaluation Criteria

To ensure objective comparison, studies were analyzed according to the following evaluation dimensions:

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 04 , April - 2026

1. *Detection Performance*
   Reported accuracy, precision, recall, F1-score, and area under the ROC curve (AUC).
2. *Real-Time Feasibility*
   Model inference speed, computational complexity, and suitability for embedded or edge deployment.
3. *Environmental Robustness*
   Sensitivity to lighting conditions, occlusions, road variability, weather, and driver diversity.
4. *Sensor Intrusiveness and Privacy Considerations*
   Degree of driver monitoring visibility and potential privacy implications.
5. *Deployment Scalability and Cost*
   Hardware requirements, integration complexity, and feasibility for commercial automotive implementation.

These criteria were used to construct a comparative framework highlighting trade-offs among the three modalities.

D. Comparative Analysis Approach

Rather than aggregating results across heterogeneous datasets, this study performs a qualitative and performance-informed comparison. Reported metrics from each study were interpreted within the context of dataset type (simulator-based or real-world), sample size, and experimental conditions.

The analysis focuses on determining:

- Common architectural patterns across modalities,
- Consistent strengths and weaknesses,
- Performance gaps under real-world limitations,
- Limitations related to dataset generalization and deployment readiness.

This structured approach allows for a comprehensive examination of the strengths and weaknesses of each of the three systems while maintaining objectivity regarding the variations in the design of each research project.

III. Comparative Analysis Approaches

In this part, we will present a systematic comparison of vision-based, motion-based, and hybrid-based systems for real-time driver monitoring. The results of selected studies conducted between 2020 and 2025 are presented here. We will specifically address the performance level of these systems, robustness in various conditions, feasibility of real-time operation, and implementation issues.

A. Vision-Based Approaches

Vision-based driver monitoring systems primarily rely on camera inputs to analyze facial and eye features as indicative signs of drowsiness. Common indicators include eye closure duration, blink frequency, eye aspect ratio (EAR), head posture, yawning detection, and gaze tracking. Convolutional neural networks (CNNs) and transformer-based models are examples of deep learning architectures that have replaced conventional computer vision techniques in recent years.

Recent studies utilizing Vision Transformers and Swin Transformer architectures demonstrate strong feature-identifying capabilities [1], [2], especially when it comes to capturing spatial dependencies across facial features. These transformer-based systems often outperform classical CNN-based models in controlled settings, with detection rates exceeding 90% in benchmark datasets. Similarly, CNN-based architectures such as DrowsyDetectNet and other lightweight convolutional models have been optimized for real-time inference while maintaining high classification performance [3]–[7].

The main strengths of vision-based systems include:

- High detection accuracy under controlled lighting conditions
- Direct observation of physiological fatigue indicators
- Strong compatibility with deep learning-based feature extraction

However, several problems still exist. Poor lighting, driving at night, covered faces (using masks or sunglasses), and frequent head rotations affect its performance. In addition, privacy concerns arise due to continuous camera monitoring inside the vehicle cabin. Computational demands may also increase when high-resolution video streams are processed in real time, particularly when transformer-based models are deployed.

Despite these limitations, vision-based approaches remain one of the most mature and widely studied paradigms for driver monitoring due to their rich behavioral feature extraction capability.

B. Motion-Based Approaches

Motion-based systems determine driver drowsiness by analyzing vehicle dynamics and steering behavior rather than directly observing the driver's face. Data including steering wheel angle variations, lane deviation patterns, rotation rate, acceleration changes, and unstable trajectory are mostly used by these systems [8]–[11].

Machine learning techniques that are often used in motion-based models include Support Vector Machines (SVM), Random Forest classifiers, k-Nearest Neighbors (k-NN), and lightweight recurrent models for time-series data

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 04 , April - 2026

[8], [9]. Unlike vision-based systems, motion-based systems are less invasive and do not require camera installation, which lowers privacy concerns and improves usability when driving at night.

The strengths of motion-based systems include the following:

- Insensitivity to lighting conditions
- Lower privacy concerns
- Reduced hardware requirements in vehicles equipped with existing sensors
- Suitability for embedded real-time deployment

However, motion-based systems are highly dependent on environmental factors such as road curvature, traffic jams, vehicle type, and drivers' driving habits. How a person normally drives can make it harder for the model to tell the difference between regular behavior and actual fatigue. For instance, an aggressive steering operator can be detected even without being tired, resulting in a greater likelihood of producing false-positive results. Furthermore, the response rate of this technology is likely to be lower than that of monitoring the driver directly since motion detection is a way of inferring driver fatigue.

On a general basis, motion-based methods provide a practical solution with high deployment rates, yet they tend to exhibit slight inaccuracies when compared to deep learning algorithms using vision technology.

C. Hybrid-Based Approaches

Hybrid systems combine two or more sensing modalities, typically visual and motion-based inputs, to overcome the limitations of single-modality approaches. These systems integrate data at either the feature level or the decision level, and often employ attention mechanisms or multi-stream architectures to improve detection stability and reduce false alarm rates [12]–[15].

Recent studies have explored multimodal systems capable of cross-domain adaptation, demonstrating improved performance across varied real-world scenarios. By fusing visual and vehicular data, hybrid systems are designed so that each modality compensates for the weaknesses of the other [13], [15]. For instance, when camera visibility is compromised due to poor lighting or occlusion, motion-based signals can maintain detection continuity. Conversely, when vehicle dynamics data is noisy or unreliable—such as during irregular road conditions—visual inputs can serve as the primary detection source.

The primary advantages of hybrid approaches include:

- Enhanced robustness under diverse environmental conditions
- Reduced false alarm rates
- Improved generalization across datasets

However, these benefits come with increased system complexity. Hybrid architectures require synchronization of heterogeneous data streams, higher computational overhead, and more intricate system integration. Hardware costs may also increase due to multi-sensor deployment. The cost of the hardware may also go up because many sensors are needed. Compared to systems that only use one type of sensor, it is harder to get hybrid systems to work on time.

Despite these drawbacks, hybrid approaches consistently outperform the other two modalities in terms of overall robustness, particularly in studies evaluating performance under varying environmental conditions.

D. Cross-Modal Comparison

A comparative synthesis across modalities reveals some major trends:

1. *Detection Accuracy*
   Vision-based and hybrid approaches generally yield superior results compared to motion-based approaches, particularly under controlled data sets. The hybrid approach tends to yield consistent results across all scenarios tested.
2. *Real-Time Feasibility*
   The motion-based system consumes fewer computational resources and is easily adaptable within embedded automobile applications. Vision-based models based on transformers can be susceptible to delays.
3. *Environmental Robustness*
   Vision-based systems are sensitive to lighting and occlusion, while motion-based systems are affected by road and traffic variability. Hybrid systems mitigate these weaknesses through modality complementarity
4. *Privacy and Intrusiveness*
   The motion-based system offers the least intrusive method of monitoring. The vision-based system presents substantial privacy concerns because of the presence of in-cabin cameras. The hybrid system lies somewhere between both extremes depending on the combination of sensor types.
5. *Deployment Scalability*

A motion-based system entails lower integration costs for cars that have pre-existing steering and motion

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 04 , April - 2026

sensors. In contrast, the hybrid system demands additional hardware and software integration.

These trends highlight an essential trade-off. While the hybrid approach might prove too complicated to implement broadly, it delivers better resiliency and detection accuracy. At the same time, the motion-based approach allows scalability at the cost of obtaining information regarding the driver's physiological state. The vision-based system strikes a balance by offering high accuracy without significant deployment issues.

Table 1. Cross-Modal Comparison of Driver Monitoring Approaches

Criterion	Vision Based	Motion Based	Hybrid-Based
Detection Accuracy	High (controlled)	Moderate	High (varied environments)
Real-Time Feasibility	Moderate to High	High	Moderate
Environmental Robustness	Low lighting sensitive	Moderate	High
Privacy	Low (in-cabin camera)	High (no camera required)	Moderate
Deployment Scalability	Moderate	High	Low to Moderate
Computational Cost	High (deep learning models)	Low	High
Hardware Requirements	Camera	Existing sensors	Multi-sensor setup

A. DISCUSSION

The comparative analysis reveals that despite substantial progress in real-time driver monitoring, considerable gaps remain in contemporary research. These gaps are primarily associated with data generalization, practical deployment limitations, computational efficiency, and inconsistent evaluation criteria.

A. Dataset Generalization and Real-World Validity

A major limitation across all three modalities is their reliance on controlled or simulator-based data. Vision-based systems are commonly evaluated on datasets consisting of well-lit, front-facing images, while motion-based systems tend to use controlled driving datasets with consistent road curvature, traffic density, and vehicle type. These conditions may inflate performance metrics while failing to capture real-world complexities such as variable lighting, weather, driver positioning, facial occlusions, and unpredictable traffic. Individual variability in driving behavior further limits generalizability across training datasets. Development of large-scale, real-world multimodal datasets is a critical need in this area.

B. Inconsistent Evaluation Metrics and Benchmarking

Another challenge is the lack of standardized benchmarking protocols. Studies report various performance metrics, including accuracy, precision, recall, F1-score, and area under the ROC curve, typically evaluated on heterogeneous datasets with different sample sizes.

Cross-study comparisons are difficult and can be misleading without a unified evaluation framework. For example, a hybrid model reporting 95% accuracy on a limited dataset cannot be directly compared with a motion-based system evaluated on real-world highway data. Additionally, few studies report latency measurements, power consumption, or inference time—factors that are critical for real-time embedded deployment.

C. Computational Complexity and Edge Development

Although multi-modal attention architectures and transformer-based vision models achieve excellent detection accuracy [1], [2], [13], they frequently demand significant computational resources. Multi-stream data fusion and high-resolution video processing raise inference latency and energy consumption, which could restrict deployment on embedded hardware of automotive grade.

Motion-based systems demonstrate stronger compatibility with lightweight classifiers and real-time implementation due to lower data dimensionality [8]–[11]. However, their indirect detection nature may compromise early-stage fatigue detection.

A promising direction is the development of lightweight multimodal architectures that can work on edge hardware. Techniques like model pruning, quantization, and knowledge distillation. And better network designs could help keep detection performance high without going over what automotive-grade hardware can handle.

D. Trade-Off Between Robustness and Scalability

In general, however, it is evident that performance and ease of use have an obvious conflict. Vision systems

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 04 , April - 2026

provide a direct means of identifying fatigue, although such identification is prone to the effects of the environment. Motion systems offer the benefit of scalability and greater privacy, although fatigue cannot be identified at an early stage since the process is indirect. The hybrid system provides the greatest reliability across conditions, although the increased complexity and cost involved make its implementation difficult. [12]–[15].

As far as the manufacturer is concerned, the issue is not only about which system can detect drowsiness in drivers most accurately. The ease of maintenance, difficulty of integration, and economic viability of such an innovation will be equally important. A mere gain in accuracy by one percent is certainly not worth the rise in costs of implementation.

E. Privacy and Ethical Considerations

Vision-based monitoring raises legitimate privacy concerns due to continuous video recording inside the vehicle. User acceptance and regulatory compliance continue to be major obstacles, despite the fact that many systems process data locally without transmitting it externally.

Motion-based systems, which analyze vehicle dynamics instead of facial data, provide better privacy protection. Hybrid systems must carefully manage data handling policies to ensure secure processing and storage.

Developing privacy-aware monitoring architectures—such as on-device inference without cloud storage—may enhance public trust and facilitate adoption.

F. Emerging Research Directions

Based on the identified gaps, multiple promising research recommendations develop:

1. Development of standardized multimodal benchmark datasets.
2. Lightweight multimodal fusion architectures optimized for embedded systems.
3. Cross-driver adaptive learning mechanisms.
4. Domain adaptation techniques for improved generalization.
5. Privacy-preserving edge AI deployment frameworks.

Addressing these challenges will be essential for transitioning driver monitoring systems from controlled experimental settings to reliable real-world automotive applications.

IV. CONCLUSION

This paper presented a structured comparison of vision-based, motion-based, and hybrid-based real-time driver monitoring by compiling the world's leading recent systems that apply machine learning. The comparison outlined the characteristic advantages and disadvantages of the modalities with regard to the applicability and development of the systems.

The results show that vision-based systems are capable of direct physiological monitoring with high detection accuracy in controlled settings, but their reliability is challenged by their sensitivity to lighting variations, occlusions, and privacy issues. Motion-based systems may be non-intrusive and privacy-friendly with high real-time feasibility and lower hardware requirements, but their indirect detection process might compromise sensitivity to early fatigue symptoms. Hybrid systems are more robust and stable across conditions by adopting multimodal fusion strategies, but higher computational requirements and integration costs are scalability issues.

The findings highlight a trade-off between detection accuracy and practical implementation. Although hybrid systems have proven to be the most reliable ones, their implementation still hinges on advancements in building lightweight models and integrating sensors. The variations in the construction of datasets and the way they assess performance metrics are factors that hinder comparative analysis.

Future work should focus on developing standardized benchmarking platforms, collecting large-scale data under realistic road conditions, and building multimodal architectures capable of operating on automotive-grade hardware. Privacy-focused and edge-based deployment strategies must also be prioritized.

By addressing these gaps, more reliable and efficient driver monitoring systems can be developed, it would be possible to develop accurate and efficient driver monitoring systems capable of reducing the occurrence of accidents caused by driver fatigue in intelligent transportation systems.