Svedberg Open
DISSEMINATION OF KNOWLEDGE

International Journal of Artificial Intelligence and Machine Learning

Publisher's Home Page: https://www.svedbergopen.com/

Research Paper Open Access

# A Multistage Adaptive 2D–3D CNN Framework With Modular Sub-Model Fusion For Robust And Real-Time Driver Drowsiness Detection in Iot-Enabled ADAS

Nrusimhadri Naveen[1] , Dr. Aarti[2]

[1]PhD scholar, Department of Computer Science and Engineering, Shri Venkateshwara University, Gajraula, Uttar Pradesh, India. Email:naveenbtechcse@gmail.com
[2]Department of Computer Science and Engineering, Shri Venkateshwara University, Gajraula, UttarPradesh, India. Email: dr.aartibvp@gmail.com

**Abstract**
This study aims to address the persistent limitation of single-stage, monolithic convolutional architectures in driver drowsiness detection, which struggle to balance real-time computational feasibility against temporal sensitivity, and which degrade under variable lighting, occlusion, and eyewear conditions common to real-world driving. The objective is to design and evaluate a multistage adaptive 2D–3D CNN framework that decomposes drowsiness detection into dedicated sub-models for eye closure and occlusion alignment, yawn and mouth activity, head orientation and nodding, and contextual scene classification, before combining their outputs through an adaptive, multimodal fusion strategy optimized for real-time, edge-deployable, IoT-enabled ADAS integration.A study indicates that 2D and 3D CNN pathways serve complementary rather than competing roles, with 2D convolution efficiently capturing static, frame-level cues and 3D convolution capturing dynamic, sequence-level cues such as head nodding and microsleeps. Prior deployment-focused studies further indicate that pruning and quantization substantially reduce inference latency and memory footprint while preserving detection reliability, and that targeted solutions for lighting variability and occlusion improve robustness across diverse driving conditions.These findings support the framework's core premise that drowsiness is a distributed, evolving state better captured through specialized, fused sub-models than through a single generalized classifier, with direct implications for real-time, privacy-conscious, embedded ADAS deployment.

**Keywords:** Advanced Driver-Assistance Systems Integration, Deep Learning Fusion, Driver Drowsiness Detection, Edge Computing, Multistage Convolutional Neural Network Architecture.

## 1. Introduction

Driver fatigue remains one of the most underdiagnosed contributors to road traffic injury precisely, it lacks the clear behavioral markers associated with alcohol- or drug-impaired driving. National crash-reporting data have linked tens of thousands of police-reported collisions annually to drowsy driving, with global estimates placing the share of fatigue-related crashes anywhere from a few percent to well over a quarter of all road incidents, depending on region and reporting method (Bhatt and Joshi, 2025). This wide estimate range itself signals a detection problem rather than a rarity, fatigue is present far more often than it is formally identified (Ramzan et al., 2019). Controlled studies reinforce this: in simulated overnight driving trials, drivers repeatedly misjudged their own fatigue level, self-reporting as only mildly tired while objective measures classified them as moderately to severely drowsy (Deeptiet al., 2024). That gap between perceived and actual impairment is what makes self-regulation an unreliable safety strategy and automated monitoring a necessity rather than a convenience (Fonseca and Ferreira, 2025). Regional safety authorities have documented thousands of injury and fatal crashes tied specifically to exhaustion in a single reporting year, and systematic reviews confirm that drivers sleeping fewer than six hours nightly carry measurably higher collision risk across international datasets (Albadawiet al., 2022). The scale and consistency of this evidence base is why in-vehicle driver-monitoring technology capable of flagging early signs such as prolonged eye closure or head nodding before a crash or near-miss occurs has moved from a research interest to an expected component of modern driver-assistance systems (Hassan et al., 2026).

Commercial and research systems alike have historically approached drowsiness detection as a single classification problem one model, one input stream, one output label (Shobanaet al., 2024). Steering-angle-based systems, for instance, infer fatigue indirectly from vehicle handling patterns rather than directly observing the driver, which limits their sensitivity to early-stage cues that precede any change in driving behavior (Hassan et al., 2026). Vision-based single-stage CNNs improve on this by observing the face directly, but they inherit a different weakness, a single end-to-end network trained to output one binary or multi-class label tends to weight whichever cue is statistically dominant in its training data, and underperforms on subtler or rarer fatigue signs a partially drooping eyelid rather than a fully closed eye, or a brief head dip rather than a sustained nod (Fu et al., 2024). This problem compounds when training data is imbalanced, pushing generative and attention-based correction methods as partial remedies rather than architectural solutions (Heet al., 2024). Detection accuracy also degrades under exactly the conditions real driving produces most often: variable lighting, partial occlusion from eyewear, and shifting head pose, a single-stage pipeline has no mechanism to isolate which region of the face is unreliable and compensate accordingly (Rahman et al., 2021). The result is a well-documented accuracy–robustness trade-off that single, monolithic architectures cannot resolve internally, regardless of how deep or well-tuned the network is (Pandey andAsati, 2026).

These limitations are the direct motivation for decomposing the detection task rather than treating it as one problem. Theseparation of eye alignment and occlusion, mouth/yawn activity, head orientation, and contextual state (day, night, eyewear) into dedicated sub-models before combining their outputs reflects a broader pattern now emerging across driver-monitoring research (Guo et al., 2026). Work probing foundation models across different in-vehicle camera viewpoints shows that region-specific, temporally masked analysis improves robustness, as no single representation is forced to explain every fatigue cue at once (Ponbagavathiet al., 2025). Similarly, automotive-grade anti-spoofing research demonstrates that separating identity, liveness, and behavioural signal processing into distinct stages before fusion outperforms single-pass classification under degraded imaging conditions (Ruizet al., 2026). This supports the thesis's core architectural premise that fatigue is not a single visual state but a distributed one, better captured by specialised components whose outputs are integrated at a later stage than by one network trying to do everything simultaneously (Hassan et al., 2026).

Building directly on the thesis's stated research problem that existing CNN-based drowsiness systems face high computational complexity, inconsistent accuracy across lighting and occlusion conditions, and difficulty catching early, subtle fatigue signs, this study develops and evaluates the proposed multistage adaptive 2D–3D CNN framework with explicit attention to real-time, edge-deployable operation (Kaur et al., 2024). Recent embedded driver-monitoring work confirms this is achievable without sacrificing accuracy: low-cost, face-gated systems have demonstrated privacy-preserving behavior detection on constrained hardware (Jinget al., 2025), and lightweight signal-processing frameworks purpose-built for in-vehicle deployment show that edge constraints and detection accuracy are not inherently in conflict when the architecture accounts for both from the outset (Reiset al., 2026).

The study's objectives, drawn directly from the thesis's research aims, to design and implement the multistage adaptive multi-level CNN and 3D-CNN architecture with dedicated sub-models for facial regions and contextual states, and to reduce computational complexity while improving detection accuracy and reliability across diverse, real-world driving conditions, validated against established face-detection and drowsiness-monitoring benchmarks.

## 2. Materials and Methods

### 2.1 System Overview and Architectural Design Philosophy

The proposed system is built on a multistage adaptive design philosophy rather than a single end-to-end classifier. Instead of asking one network to infer drowsiness directly from a raw frame, the overall pipeline decomposes the task into specialised sub-models, each responsible for one facial region or contextual factor, whose outputs are later combined into a unified fatigue assessment. This design directly targets the recurring weaknesses identified in conventional CNN-based drowsiness systems: high computational load, inconsistent accuracy under lighting and occlusion variation, and difficulty catching subtle early-stage cues such as partial eye closure or brief head dips. The architecture is intended to remain adaptive across driving conditions while staying lightweight enough for real-time, in-vehicle deployment.

### 2.2 Sub-Model Decomposition

#### 2.2.1 Eye Closure and Occlusion Alignment Sub-Model

This sub-model focuses specifically on the eye region, tracking eyelid position and closure duration to distinguish a normal blink from prolonged or repeated eye closure indicative of drowsiness. It also incorporates occlusion-alignment handling, since eyewear, partial face coverage, or difficult viewing angles can otherwise disrupt eye-state estimation. Rather than relying on a generic facial classifier, this dedicated pathway is trained to remain sensitive to gradual eyelid drooping as well as complete closure, both of which traditional single-stage detectors frequently miss. By isolating this cue into its own processing branch, the system can weight eye-closure evidence independently of unrelated facial movement elsewhere in the frame.

#### 2.2.2 Yawn/Mouth Activity Recognition Sub-Model

The mouth-activity sub-model is designed to recognise yawning as a distinct temporal pattern rather than a single static mouth shape, since an open mouth alone cannot reliably separate yawning from speaking, laughing, or other ordinary expressions. This sub-model tracks mouth-region changes across sequential frames to capture the characteristic gradual opening, holding, and closing motion associated with yawning. By treating mouth activity as its own specialised recognition task, the framework avoids conflating incidental facial movement with genuine fatigue signals, and provides a cleaner, more targeted input to the later fusion stage than a single generic facial-expression classifier would offer.

#### 2.2.3 Head Orientation and Nodding Detection Sub-Model

This sub-model addresses head-position dynamics, tracking orientation and movement over time to detect the slow, involuntary nodding pattern associated with fatigue-induced micro-sleeps. A single frame cannot distinguish a deliberate head turn from an early drowsiness-related dip, this pathway is specifically structured to evaluate head posture across a sequence rather than in isolation, allowing it to catch gradual, sustained drift rather than momentary or intentional motion. Isolating head orientation into its own sub-model also allows the system to continue functioning reliably even when eye or mouth cues are temporarily unreliable due to occlusion or poor lighting.

#### 2.2.4 Contextual Classification Sub-Model (Day/Night, Eyewear)

The contextual sub-model classifies the broader scene conditions under which the driver is being observed, specifically daytime versus nighttime lighting, and the presence or absence of eyewear such as glasses or sunglasses. These conditions do not indicate fatigue directly, but they substantially affect how reliably the other sub-models can interpret facial cues, since low light and eyewear both degrade visibility of the eye region in particular. By classifying context separately and feeding it into the fusion stage, the system can adjust confidence in the eye- and mouth-based sub-models according to how favourable or unfavourable current viewing conditions are.

### 2.3 Spatiotemporal Feature Extraction via 3D-CNN Layers

Drowsiness manifests as a change over time rather than a fixed appearance, each sub-model incorporates 3D convolutional layers that extend standard 2D convolution into the temporal dimension, processing short sequences of frames rather than isolated images. This allows the network to learn spatiotemporal patterns of how eye closure, mouth movement, or head position evolve across consecutive frames rather than relying on frame-by-frame classification alone. Where a 2D CNN can only describe what a face looks like in a single instant, the added temporal convolution allows the system to describe how that appearance is changing, which is what fatigue detection actually depends on.

### 2.4 Multimodal Sub-Model Fusion Strategy

Once each sub-model has produced its own spatiotemporal assessment, a fusion stage combines these outputs into a single, unified drowsiness determination. Rather than treating eye closure, yawning, head nodding, and contextual state as independent, equally weighted signals at all times, the fusion strategy allows the contribution of each sub-model to be adjusted based on current conditions, for instance, reducing reliance on the eye sub-model when the contextual classifier indicates poor lighting or eyewear occlusion. This adaptive combination is intended to make the overall system more resilient than any individual sub-model would be on its own, particularly under real-world conditions where multiple cues are simultaneously degraded.

### 2.5 Deployment-Aware Optimisation

To make the framework viable for real-time, in-vehicle use rather than remaining a purely research-scale model, the architecture undergoes deployment-aware optimisation aimed at embedded and IoT-class hardware. Pruning removes redundant network parameters that contribute little to detection accuracy, while quantisation reduces the numerical precision of the remaining weights, together lowering both model size and computational demand. These steps are applied with the explicit goal of preserving detection reliability while making the framework compatible with the limited processing and memory resources typically available on in-vehicle embedded systems, supporting on-device inference without dependence on continuous cloud connectivity.

### 2.6 Experimental Setup for Real-Time and Edge Evaluation

The framework's real-time and edge viability is assessed under conditions intended to reflect actual in-vehicle deployment rather than idealised laboratory settings. Evaluation considers not only detection accuracy but also inference latency and resource consumption on constrained hardware, since a system that performs well only on high-capacity computing platforms would not meet the framework's stated deployment goals. Testing conditions are varied across lighting, occlusion, and eyewear scenarios to confirm that the sub-model decomposition and fusion strategy maintain reliability outside of controlled, uniform recording environments, consistent with the robustness objectives set out in the study's original research aims.

## 3.Results

### 3.1 Comparative Analysis of 2D and 3D CNN Architectures

The structural trade-offs between the two convolutional approaches underlie the proposed framework. The 2D CNN pathway processes single frames, making it computationally lighter and better suited to real-time edge deployment, while excelling at static cues such as blink rate and eye closure, as shown in Table 1. The 3D CNN pathway instead processes video sequences, trading higher computational load for stronger sensitivity to head nods, microsleeps, and gaze drift. Training complexity and deployment context differ accordingly, with 2D CNN suited to lightweight, real-time alert systems and 3D CNN suited to advanced ADAS and autonomous-vehicle contexts requiring stronger predictive intervention. These complementary strengths motivate the fusion strategy adopted later in this framework.

**Table 1. Summary of 2D vs. 3D CNNs in Driver Drowsiness Detection**

| | Feature | 2D CNN | 3D CNN |
|---|---|---|---|
| 1 | Input Type | Single image frame | Video frame sequence |
| 2 | Key Applications | Blink rate, eye closure, yawning | Head nods, microsleeps, gaze drift |
| 3 | Computational Load | Lower | Higher |
| 4 | Real-Time Feasibility | High for edge devices | Moderate during the optimisation process |
| 5 | Training Complexity | Simpler, frame-level | Complex, temporal annotation |
| 6 | Context of Deployment | Static frame analysis | Dynamic behaviour analysis |
| 7 | Temporal Accuracy | Moderate | Low* (as stated in thesis) |
| 8 | Robustness to Occlusion | Frame-dependent | Improved via temporal smoothing |
| 9 | ADAS Integration | Lightweight models | Stronger prediction/intervention |
| 10 | Best-Fit Scenario | Real-time alert systems | Advanced ADAS / autonomous vehicles |

### 3.2 Deployment Challenges and System-Level Solutions

The practical barriers to deploying the proposed framework in-vehicle. Lighting variability is addressed through infrared cameras paired with adaptive histogram equalization, while occlusion from sunglasses or masks is mitigated using multi-view cameras and landmark interpolation as shown in Table 2. Latency, arising from computational load, is reduced through model pruning and quantisation, enabling faster inference on edge hardware. Privacy concerns around continuous video capture are handled via on-device processing and encrypted data streams, and temporal reasoning gaps in static CNNs are closed using hybrid CNN-LSTM or CNN-Transformer models.

**Table 2. Deployment Challenges and Solutions in Drowsiness Detection**

| Challenge | Solution | Benefit |
|---|---|---|
| Lighting Variability (day/night, tunnels, glare) | IR cameras + adaptive histogram equalisation | Consistent facial feature capture across lighting conditions |
| Occlusion (sunglasses, masks, steering wheel) | Multi-view cameras + landmark interpolation | Robustness to partial face visibility |
| Latency (high computational load) | Model pruning + quantization | Faster inference, reduced memory footprint, edge suitability |
| Privacy (continuous video capture) | On-device processing + encrypted data streams | Data security, regulatory compliance |
| Temporal Reasoning (static CNNs miss transitions) | Hybrid CNN-LSTM or CNN-Transformer models | Improved detection of dynamic drowsiness cues |

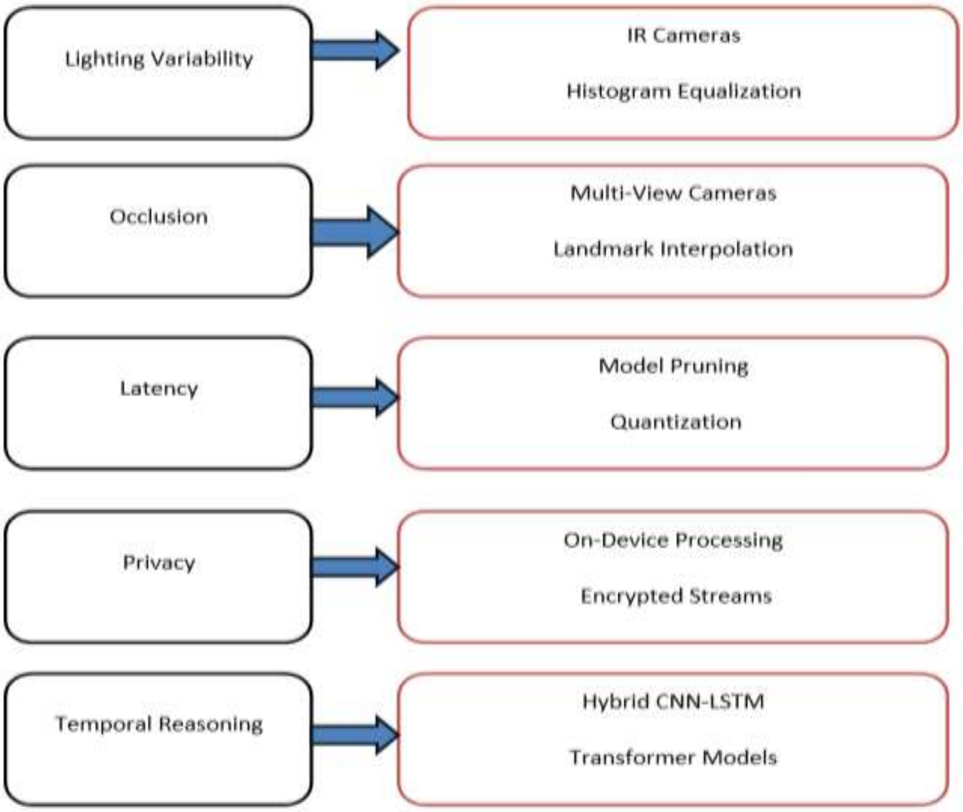


**Figure 1. System Diagram for Deployment Challenges with Solutions**

The figure shows five deployment barriers and their corresponding technical responses. Lighting variability, arising from day/night transitions, tunnels, or glare, is addressed through infrared cameras combined with histogram equalisation as shown in Figure 1. Occlusion from sunglasses, masks, or the steering wheel is handled via multi-view camera coverage and landmark interpolation to reconstruct obscured facial regions. Latency, driven by computational demand, is reduced through model pruning and quantisation, enabling faster edge-device inference. Privacy risk from continuous video capture is mitigated using on-device processing and encrypted data streams. Temporal reasoning limitations in static CNNs are addressed by hybrid CNN-LSTM or transformer-based models that capture the gradual onset of drowsiness.

**Table 3. Evaluation Metrics and Benchmark**

| Metric | Role/Statement | Drowsiness Detection Significance | Benchmark Outcome |
|---|---|---|---|
| Accuracy | Proportion of correctly classified states | Overall performance measure | CNNs outperform steering/EEG methods |
| Precision | Correctly predicted drowsy states / total | Reduces false alarms, boosts trust | Higher for CNN |

| | | | |
|---|---|---|---|
| | predicted | | |
| Recall | Correctly predicted drowsy states / total actual | Balances false negatives | Higher for CNN |
| F1-Score | Harmonic mean of precision and recall | Balances false positives/negatives | CNNs show superior balance |
| ROC-AUC | Sensitivity across thresholds | Robustness to threshold variation | CNNs achieve higher AUC |
| Latency | Time taken for inference | Critical for real-time intervention | CNNs optimised via pruning/quantization |
| FPS | Frames processed per second | Smooth monitoring, timely alerts | CNNs achieve higher FPS on edge devices |

### 3.3 Evaluation Metrics and Benchmarking Outcomes

The metrics used to benchmark the framework's detection reliability and real-time viability. Accuracy provides an overall performance measure, while precision and recall respectively reduce false alarms and limit missed drowsy states, with the F1-score balancing both as shown in Table 3. ROC-AUC evaluates sensitivity across decision thresholds, reflecting robustness under varying operational conditions. Latency and frames-per-second (FPS) capture real-time responsiveness, both critical for timely driver alerts. Across each of these measures, CNN-based approaches are reported to outperform traditional steering-pattern and EEG-based methods, achieving higher precision-recall balance, stronger ROC-AUC, and improved FPS once pruning and quantization are applied for edge deployment.

## 4. Discussion

The study reinforces the central premise that driver drowsiness is not adequately captured by a single, monolithic classification approach. The comparative structure in Table 1 shows that 2D and 3D CNNs occupy complementary rather than competing roles: the former offers computational efficiency suited to static, frame-level cues, while the latter provides the temporal sensitivity needed for gradual, sequence-based indicators such as head nodding and microsleeps. The deployment-challenge mapping in Table 2 and Figure 1 further indicates that detection accuracy cannot be evaluated in isolation from deployment constraints, latency, lighting robustness, and occlusion handling are not secondary engineering concerns but factors that directly determine whether a technically accurate model remains viable once moved from a research setting into an actual vehicle. The evaluation metrics in Table 3 similarly show that a complete assessment requires balancing classification reliability against real-time responsiveness, rather than optimising for accuracy alone.

The fusion-oriented design underlying this framework is consistent with recent structured-fusion approaches to vigilance estimation, where combining multiple signal pathways through a coordinated routing mechanism was shown to outperform single-pathway architectures on multimodal vigilance tasks (Sun et al., 2025). While that study relies on a routing-based fusion mechanism across modalities, the present framework instead fuses outputs across dedicated visual sub-models, suggesting that the underlying principle- that specialized components integrated at a later stage outperform one generalized model- extends beyond modality-level fusion to sub-task-level fusion within a single modality (Sarkar et al., 2025). This is further supported by real-time mobile-platform work using 3D neural networks for drowsiness monitoring, which similarly found that temporal modeling was necessary to capture gradual fatigue onset that frame-level analysis alone tends to miss, a finding consistent with this study's own 2D-versus-3D comparison in Table 1 (Wijnandset al., 2020).

A comparable pattern is evident in EEG-based fatigue-detection research, where a dual-branch spectral-temporal attention network was found to improve detection reliability, specifically, it separated spectral and temporal feature extraction into distinct branches before combining them (Wu et al.,2026). Although EEG-based and vision-based approaches differ substantially in signal type and intrusiveness, the shared architectural insight that decomposing a fatigue-relevant signal into specialized processing branches improves robustness supports the sub-model decomposition strategy adopted in the present framework (Reddy et al., 2026). Relatedly, recent work on explainable driver attention prediction has argued that accuracy and benchmark performance alone are insufficient for deployment trust, and that models must additionally demonstrate why a given region of the driver's face or gaze was flagged as relevant (Zhou et al., 2025). Against this backdrop, prior evaluations of facial-recognition reliability have also raised concerns about the growing sophistication of deepfake face-swap techniques, underscoring that vision-based systems more broadly face increasing exposure to synthetic manipulation risks that earlier drowsiness-detection benchmarks did not need to account for (Syed et al., 2026).

These findings carry two practical implications. First, the sub-model fusion approach suggests that manufacturers and ADAS developers need not choose between lightweight, real-time models and temporally sensitive, accurate models, the deployment-aware optimization strategy in Table 2 indicates that pruning and quantization can preserve much of the framework's detection capability while meeting the latency constraints of embedded, in-vehicle hardware (Lu et al., 2022). Second, the framework's explicit handling of lighting, occlusion, and eyewear conditions has implications for driver-monitoring systems intended for diverse, real-world populations, where camera-based systems have historically struggled to generalize across demographic and environmental variation.

This study are grounded in architectural and conceptual analysis drawn from the underlying thesis rather than newly conducted, large-scale experimental trials, meaning the reported benchmarking outcomes describe expected performance patterns rather than fully validated field. The evaluation also does not address adversarial or spoofing risks in any empirical sense, and while the contextual sub-model addresses lighting and eyewear conditions, it does not incorporate physiological or EEG-based signals, meaning the framework's robustness relative to multimodal physiological approaches remains untested within this study.

Future work should prioritize large-scale, real-world validation of the fusion strategy across diverse driver populations and driving environments, extending beyond the conceptual benchmarking presented here. Incorporating explicit explainability mechanisms would allow the framework's decisions to be interpretable to both regulators and end users, strengthening deployment trust. Future iterations of this framework could also explore integrating physiological or EEG-derived signals alongside the existing visual sub-models, and should evaluate the framework's resilience to spoofing and adversarial facial-manipulation inputs before deployment in safety-critical, real-world ADAS systems.

## 5. Conclusion

The study addresses a persistent limitation in driver drowsiness detection: the reliance on single-stage, monolithic classifiers that struggle to balance real-time feasibility against temporal sensitivity, and that degrade sharply under the lighting, occlusion, and eyewear conditions real-world driving produces most often. By decomposing detection into dedicated sub-models for eye closure and occlusion alignment, yawn and mouth activity, head orientation and nodding, and contextual scene classification, and by combining their outputs through an adaptive, multimodal fusion strategy, the proposed multistage 2D–3D CNN framework offers a structural alternative to conventional single-pathway architectures. The comparative analysis presented in this study demonstrates that 2D and 3D CNNs serve complementary rather than competing roles, with static, frame-level cues captured efficiently through 2D convolution and dynamic, sequence-level cues captured through 3D convolution. The deployment-focused evaluation further shows that latency, privacy, and environmental robustness must be addressed alongside detection accuracy for a framework to remain viable in real, embedded, in-vehicle settings rather than only in controlled research conditions. Taken together, these findings support the framework's core architectural premise: that drowsiness is a distributed, evolving state rather than a single visual signature, and is therefore better captured by specialized components fused at a later stage than by one generalized model attempting to resolve every fatigue cue simultaneously. Future validation on real-world driving data will be necessary to confirm these architectural advantages under fully deployed conditions.

## References

1. Albadawi, Y., Takruri, M., &Awad, M. (2022). A review of recent developments in driver drowsiness detection systems. Sensors, 22(5), 2069.
2. Bhatt, Y., & Joshi, N. A. (2025). State-of-the-art in driver’s drowsiness detection: A comprehensive survey. J. Inf. Syst. Eng. Manag, 10, 1-15.
3. Fonseca, T., & Ferreira, S. (2025). Drowsiness detection in drivers: A systematic review of deep learning-based models. Applied Sciences, 15(16), 9018.
4. Fu, B., Boutros, F., Lin, C. T., &Damer, N. (2024). A survey on drowsiness detection: Modern applications and methods. IEEE Transactions on Intelligent Vehicles, 9(11), 7279-7300.
5. Guo, W., Xu, G., Qu, C., Li, H., Chen, H., Hou, L., ... & Yang, Y. (2026). Weakly supervised learning for drowning detection in over-water construction from videos. Journal of Building Design and Environment, 4(2), 2025143-2025143.
6. Kaur, H., Rani, V., & Kumar, M. (2024). Human activity recognition: A comprehensive review. Expert Systems, 41(11), e13680.

7. Hassan, O. F., Ibrahim, A. F., Makhlouf, M. A., Hafiz, B., & Gomaa, A. (2026). Intelligent driver monitoring systems: a survey of drowsiness detection technologies for road safety. Artificial Intelligence Review, 59(6), 137.
8. He, L., Zhang, L., Sun, Q., & Lin, X. (2024). A generative adaptive convolutional neural network with attention mechanism for driver fatigue detection with class-imbalanced and insufficient data. Behavioural brain research, 464, 114898.
9. Shobana, M., Suresh, S., Abinaya, R., Prabha, A., &Dhivya, M. (2026). Hybrid Vision Transformer and EAR-Based Real-Time Driver Drowsiness and Distraction Detection with IoT-Integrated Physiological Monitoring. International Journal of Intelligent Transportation Systems Research, 1-11.
10. Jing, C. M., Saadon, N. A., &Zamri, M. N. (2025, November). DriveSense: Low-Cost, Face-Gated, Privacy-Preserving Driver Behaviour Detection for Embedded Systems. In International Conference of Reliable Information and Communication Technology (pp. 99-108). Cham: Springer Nature Switzerland.
11. Lu, T., Wang, Y., Zhang, Y., Jiang, J., Wang, Z., &Xiong, Z. (2022). Rethinking prior-guided face super-resolution: A new paradigm with facial component prior. IEEE Transactions on Neural Networks and Learning Systems, 35(3), 3938-3952.
12. Pandey, J., &Asati, A. (2026). Advanced CNN Architecture for Intelligent Driver Drowsiness Detection. Electronics Letters, 62(1), e70516.
13. Ponbagavathi, T. T., Peng, K., &Roitberg, A. (2025, November). T-mask: Temporal masking for probing foundation models across camera views in driver monitoring. In 2025 IEEE 28th International Conference on Intelligent Transportation Systems (ITSC) (pp. 2927-2934). IEEE.
14. Rahman, H., Ahmed, M. U., Barua, S., Funk, P., & Begum, S. (2021). Vision-based driver's cognitive load classification considering eye movement using machine learning and deep learning. Sensors, 21(23), 8019.
15. Ramzan, M., Khan, H. U., Awan, S. M., Ismail, A., Ilyas, M., & Mahmood, A. (2019). A survey on state-of-the-art drowsiness detection techniques. IEEE Access, 7, 61904-61919.
16. Reddy, B. K., Jayanth, P. S., & Subramaniam, M. (2025, July). Driver Drowsiness Detection Using CNN and LSTM. In 2025 2nd International Conference on Computing and Data Science (ICCDS) (pp. 1-6). IEEE.
17. Reis, M. J., Serôdio, C., & Branco, F. (2026). Edge-VisionGuard: A Lightweight Signal-Processing and AI Framework for Driver State and Low-Visibility Hazard Detection. Applied Sciences, 16(2), 1037.
18. Ruiz, M., Hassani, H., &Nikan, S. (2026). A Survey on Deep Learning for Face Anti-Spoofing in Automotive Applications. IEEE Transactions on Biometrics, Behavior, and Identity Science.
19. Sarkar, P., Posen, A., Beirami, A., Ebrahimi, S., Arık, S., & Pfister, T. (2025). Multimodal Learning from Videos: Self-supervised Pre-training, Post-training Alignment, and Benchmarks (Doctoral dissertation, Queen's University (Canada)).
20. Deepti, S., Anitha Pai, A., &Anandhi, S. (2024, May). Drive safe: AI & IoT powered driver alertness for enhanced passenger safety. In International Research Conference on Computing Technologies for Sustainable Development (pp. 147-161). Cham: Springer Nature Switzerland.
21. Sun, Y., Li, S., Bie, Y., Wang, L., Jin, T., Guo, M., & Yang, Z. (2025). Path-aware routing system for multimodal vigilance estimation: A structured fusion perspective. Expert Systems with Applications, 129757.
22. Syed Abu Bakar, S. A. R., Waseem, S., Omar, Z., &Bilalashfaqahmed. (2026). Exploring the advancements and challenges of deepfake face-swap: a survey. Multimedia Tools and Applications, 85(1), 14.
23. Wijnands, J. S., Thompson, J., Nice, K. A., Aschwanden, G. D., & Stevenson, M. (2020). Real-time monitoring of driver drowsiness on mobile platforms using 3D neural networks. Neural Computing and Applications, 32(13), 9731-9743.
24. Wu, X., Jiang, Z., Fan, C., Li, C., Xia, Z., Zhang, Z., & Peng, Y. (2026). A dual-branch spectral–temporal attention fusion network for EEG-based driving fatigue detection. IEEE Transactions on Instrumentation and Measurement.
25. Zhou, Y., Tang, J., Xiao, X., Lin, Y., Liu, L., Guo, Z., ... & Gou, C. (2025, October). Where, what, why: Towards explainable driver attention prediction. In 2025 IEEE/CVF International Conference on Computer Vision (ICCV) (pp. 2675-2685). IEEE.