SvedbergOpen
DISSEMINATION OF KNOWLEDGE

International Journal of Artificial Intelligence and Machine Learning

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

Research Paper Open Access

Multistage Adaptive CNN–3D-CNN Framework for Robust Driver Drowsiness Detection Under Diverse Driving Conditions

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

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

Abstract
Driver drowsiness remains a leading yet often underestimated cause of road accidents, driven largely by impaired self-awareness and delayed reaction time among fatigued drivers. This study proposes a multistage adaptive CNN–3D-CNN framework for real-time driver drowsiness detection, designed to overcome the temporal and contextual limitations of conventional single-stage classifiers. The framework decomposes detection into specialized sub-modules covering head orientation, eye closure, mouth activity, and contextual conditions, before fusing these outputs through a 3D-CNN layer for spatiotemporal reasoning. Two benchmark datasets, UTA-RLDD (University of Texas at Arlington Real-Life Drowsiness Dataset) and NTHU-DDD (National Tsing Hua University Driver Drowsiness Detection Dataset), were harmonized and pre-processed to support robust model training, while data augmentation and class-balancing techniques addressed demographic and environmental variability. Comparative evaluation against existing face-detection models, including YOLOv7-Face, MTCNN (Multi-task Cascaded Convolutional Networks), and Hyperface-ResNet, demonstrated the proposed framework's superior accuracy, recall, and robustness under occlusion and low-light conditions. Cross-dataset validation across KEC-DDD (Kathmandu Engineering College - Driver Drowsiness Detection Dataset), NTHU-DDD, and UTA-RLDD further confirmed the model's generalizability across diverse driving populations. These findings support the framework's suitability for integration into Advanced Driver Assistance Systems, offering a deployable, explainable, and accurate solution for early fatigue detection.

Keyword: Driver drowsiness detection, Convolutional neural networks, 3-Dimensional Convolutional Neural Network, Multistage adaptive framework, Advanced Driver Assistance Systems integration

1.Introduction

Sleepiness might impact visual attention, decision-making, reaction time, and the capacity to maintain steady vehicle control, it is a significant problem for road safety. Before a driver reaches a critical level of weariness, early signs such prolonged eye closure, frequent blinking, yawning, shifts in gaze, and head motions can be helpful indicators (Jaspin*et al*., 2024). Consequently, an automatic identification of these behavioural traits through ongoing driver monitoring has become a major focus of recent deep-learning research (Ahmed *et al*., 2021). Advanced driver-assistance systems (ADAS), where prompt detection of impaired driving can support preventive safety actions, are especially pertinent to the development of intelligent driver monitoring systems.

Physiological parameters such as electroencephalography, electrocardiography, heart-rate signals, and other bio signals were often used in early methods of assessing drowsiness. While such assessments can yield useful information regarding a driver's physiological condition, their reliance on wearable or contact-based sensors might cause them less practical and comfortable for extended driving (Ramzan *et al*., 2024). As a result, deep neural networks have been studied for the analysis of non-contact behavioural indications and physiological data (Alguindigue*et al*., 2024). At the same time, IoT-oriented monitoring frameworks have shown promise for combining connected transportation environments with non-intrusive driver monitoring, allowing for ongoing evaluation without the need for physical sensors to be attached to the driver (Khan *et al*., 2023). A shift toward camera-based and intelligent vision systems that can work continuously inside vehicles has been prompted by these developments.
Computer vision-based methods can assess facial features without speaking to the driver directly, they offer a more practical option. In order to accurately identify the facial region for further study of the eyes, mouth, head position, and other fatigue-related parameters, face detection and placement are crucial first

steps (Jain et al., 2024). The face monitoring more suitable for real-world systems, lightweight detection techniques such geometrical facial analysis and SSD MobileNet-based techniques have been investigated (Chan *et al*., 2022). While multi-task and lightweight networks have tried to lower the computing requirements for embedded driver-monitoring systems, CNN-based architectures have further enhanced the automatic extraction of discriminative face features (Kim *et al*., 2019).

Selective focus on informative facial regions can enhance the identification of modest fatigue-related traits, attention-based techniques have also grown in importance (Hassan et al., 2026). Despite these developments, a significant drawback of traditional image-based CNN systems is that tiredness is essentially a temporal behavioural phenomenon rather than a discrete face state. According to Kotseruba and Tsotsos (2022), a single photograph may show that the eyes are closed, but it cannot accurately distinguish whether the driver is blinking normally or experiencing extensive ocular closure. Yawning and head nodding also evolve over a series of frames. As a result, there has been a growing interest in models that can learn both geographical and temporal information (Magán et al., 2022).

By extending convolutional processing over successive frames, three-dimensional CNNs are able to record changes in facial movement and appearance over time (Kang *et al*., 2022). While recent transformer-based methods have investigated attention-driven temporal representations for more intricate driver-state analysis, CNN-LSTM architectures offer an alternative approach by fusing spatial feature extraction with sequential modelling (Jain *et al*., 2024).

Real-world environmental and driver-related instability further undermines these systems' dependability. Visual recognition can be significantly impacted by variations in lighting, night time settings, sunglasses, facial occlusion, camera position, and individual behavioural patterns (Nurnoby and El-Alfy, 2026). In order to enhance detection performance while lowering the quantity of task-specific training data needed, model refinement and transfer learning have been investigated (Jebraeily et al., 2024).

A model trained on a restricted population would not function as well across several drivers, so a dataset and demographic bias are still major concerns (Madni*et al*., 2024). In order to increase fairness and robustness in driver sleepiness detection, generative techniques for bias correction have been studied (Ngxande*et al*., 2020). Similar to this, recent benchmarking studies stress that transfer-learning techniques and attention mechanisms should be assessed in real-world operating environments rather than only using controlled datasets (Hassan *et al*., 2026).

The balance between detection accuracy and computational efficiency. High-capacity deep-learning models can achieve strong recognition performance but may require computational resources that are difficult to provide on vehicle-mounted or low-power hardware (Lamaazi*et al*., 2023). The study on embedded and edge-based driver monitoring has therefore emphasized lightweight architectures, optimized inference, and local processing (CivikandYuzgec, 2023). Neuromorphic sensing has also been explored as an alternative for detecting behavioural events such as yawning with potentially efficient event-driven processing (Kielty*et al*., 2023). These developments complement IoT-based approaches by supporting distributed and real-time driver monitoring while reducing dependence on continuous cloud processing. However, practical deployment still requires careful consideration of latency, model complexity, privacy, and reliability.

Further progress in this field requires architectures that can simultaneously consider multiple facial and contextual indicators rather than depending on a single fatigue cue. The literature indicates growing interest in combining eye behaviour, mouth activity, head movement, environmental conditions, and temporal information within integrated deep-learning systems (OISE *et al*., 2026). Such an approach can help distinguish genuine drowsiness from normal facial movements and reduce dependence on isolated visual events (Ramzan *et al*., 2024). The present study therefore proposes a multistage adaptive multi-level CNN and 3D-CNN framework in which specialized modules analyse head orientation, eye alignment, mouth activity, and contextual conditions such as daytime, night time, and eyewear. The outputs of these modules are integrated with temporal representations to identify evolving fatigue-related behaviour. The thesis framework specifically emphasizes spatial-temporal feature extraction, adaptive modelling, computational efficiency, and robustness across different driving conditions.

The proposed framework is designed to address the principal limitations identified in existing driver drowsiness detection systems, particularly missed subtle fatigue cues, environmental variability, occlusion, computational burden, and limited generalizability. Its evaluation across benchmark datasets, together with comparison against established face-detection and drowsiness-monitoring approaches, is intended to assess whether multistage specialization and temporal modelling can provide a more reliable foundation for real-time driver monitoring. The objective is to develop a scalable, non-intrusive, and deployment-oriented solution capable of supporting future ADAS (Advanced Driver Assistance Systems) and intelligent-vehicle applications.

Objectives of the Study

1. The design and implement a multistage adaptive multi-level CNN and 3D-CNN architecture, incorporating specialized sub-models for facial regions (eyes, mouth, head) and contextual states (day/night, eyewear), tailored for real-time driver drowsiness detection.
2. The reduce computational complexity while improving detection accuracy across diverse driving conditions including lighting, occlusion, and behavioural variability and to validate the proposed system across benchmark datasets (KEC-DDD, NTHU-DDD) against existing models such as YOLOv7-Face, MTCNN, AAMS, and Hyperface-ResNet.

2.Materials and Methods

2.1Study Design

This study adopts a multistage adaptive CNN–3D-CNN framework for real-time driver drowsiness detection. The study design decomposes the detection task into specialized sub-models addressing head orientation, eye alignment, mouth activity, and contextual conditions, rather than relying on a single monolithic classifier. Each sub-model extracts spatial features that are subsequently fused and passed through a 3D-CNN layer for temporal reasoning. The workflow proceeds from raw video capture through pre-processing, facial region extraction, per-module classification, and final fusion, producing a robust, adaptive, and interpretable pipeline suited for integration into IoT-enabled in-vehicle monitoring and ADAS platforms.

2.2Data Preparation

Two benchmark datasets underpin model development. UTA-RLDD provides diverse RGB recordings across alert, low-vigilance, and drowsy states, covering varied lighting, eyewear, and demographic groups. NTHU-DDD contributes rich behavioral annotations, including blinks, yawns, and head-pose labels across day/night and occlusion conditions. Dataset harmonization aligns class definitions, resolution, and frame rate across sources, standardizing inputs to a common format. Frame extraction and sequence preparation follow, using uniform or adaptive sampling to generate fixed-length sequences (16–32 frames), with ROI cropping and alignment ensuring the model focuses on relevant facial regions while minimizing background noise and irrelevant variation.

2.3Facial Feature Extraction

Facial feature extraction begins with face detection and localization using lightweight, real-time-capable detectors. Facial landmark detection and alignment (e.g., via Dlib or MediaPipe Face Mesh) then precisely locate key points across head rotations and partial occlusions. From these landmarks, eye, mouth, and head regions are isolated as independent regions of interest for downstream sub-model processing. Where applicable, Eye Aspect Ratio (EAR) and Mouth Aspect Ratio (MAR) are computed as auxiliary handcrafted features, supplementing learned representations by quantifying eyelid closure and mouth opening. This combination of deep and geometric features improves sensitivity to subtle, early-stage fatigue indicators.

2.4Data Augmentation

To improve robustness and generalization, extensive augmentation is applied during training. Rotation and brightness/contrast adjustments simulate natural head-pose variation and diverse lighting conditions, while occlusion simulation introduces synthetic sunglasses, masks, or hand coverings to mimic real-world obstructions. Class balancing techniques oversampling, class weighting, and batch balancing address the inherent skew between abundant "alert" samples and rarer "drowsy" instances. Training follows a k-fold or stratified cross-validation strategy, ensuring reliable performance estimates and reducing overfitting. Together, these steps prepare the model to handle demographic diversity, environmental variability, and imbalanced class distributions encountered in real-world driving scenarios.

2.5Multistage CNN–3D-CNN Model Development

The core architecture comprises four specialized sub-modules. Head-orientation analysis tracks nodding, tilting, and drooping motions; eye-closure and ocular activity analysis monitors blink frequency, duration, and gaze drift; mouth/yawning analysis distinguishes yawns from speech or neutral expressions; and contextual/environmental analysis classifies day/night and eyewear conditions. Each sub-model independently extracts domain-specific spatial features before outputs are merged through a fusion block. A 3D-CNN layer then models spatiotemporal dynamics across frame sequences, capturing extended eye closure, repetitive yawning, and gradual head drooping that static, single-frame models cannot detect, enabling earlier and more reliable fatigue identification.

2.6Model Evaluation

Model performance is evaluated using accuracy, precision, recall, F1-score, and AUC, chosen to address the class imbalance inherent in drowsiness datasets. Precision reflects false-alarm reduction, while recall captures the system's sensitivity to genuine drowsy states, F1-score balances both. ROC-AUC assesses discriminative power across decision thresholds. Comparative evaluation benchmarks the proposed framework against established face-localization models YOLOv7-Face, MTCNN, AAMS, and Hyperface-ResNet on landmark accuracy and classification metrics. Cross-dataset and generalization analysis further tests robustness by validating across UTA-RLDD, NTHU-DDD, and KEC-DDD, confirming consistent performance despite variation in lighting, occlusion, and demographic composition.

3.Results

3.1 Comparative Performance of Deep Learning Architectures

Three deep learning architectures for driver drowsiness detection based on strength, weakness, and measurable performance. The 2D CNN offers fast inference and strong spatial feature extraction but lacks temporal awareness, limiting its ability to detect prolonged eye closure or yawning, yielding accuracy between 80–85% as shown in Table 1. The 3D CNN captures spatiotemporal dynamics and performs better under occlusion, reaching 88–92% accuracy, though at higher computational cost. The Hybrid CNN-LSTM/GRU model combines both spatial and temporal strengths, achieving the highest accuracy range of 90–94% while remaining suitable for real-time ADAS deployment through pruning and quantization optimization.

Table 1. Performance Comparison of 2D CNN, 3D CNN, and Hybrid CNN-LSTM/GRU Architectures for Driver Drowsiness Detection

Model Type	Strength	Weakness	Performance Metrics	ADAS Deployment Feasibility
2-Dimensional Convolutional Neural Network	Strong spatial feature extraction; Fast inference; Well-established frameworks	No temporal modelling; Struggles with prolonged eye closure/yawning; Sensitive to occlusion and low-light	Accuracy: ~80–85%; Recall: ~70–75%; F1-score: ~72–76%	High (lightweight), but limited robustness
3-Dimensional Convolutional Neural Network	Captures spatiotemporal dynamics; Detects prolonged eye closure, yawning, head nodding; Robust under occlusion and low-light	Higher computational cost; Larger model size	Accuracy: ~88–92%; Recall: ~85–90%; F1-score: ~86–91%	Moderate (requires optimization for embedded systems)
Hybrid Convolutional Neural Network for spatial feature extraction with Long Short-Term Memory andGated Recurrent Unit	Combines spatial + temporal modeling; Lightweight recurrent layers extend temporal sensitivity; Better generalization across datasets	Slightly higher latency than pure CNN; Requires careful tuning	Accuracy: ~90–94%; Recall: ~88–92%; F1-score: ~89–93%	High (optimized pruning/quantization enables real-time ADAS integration)

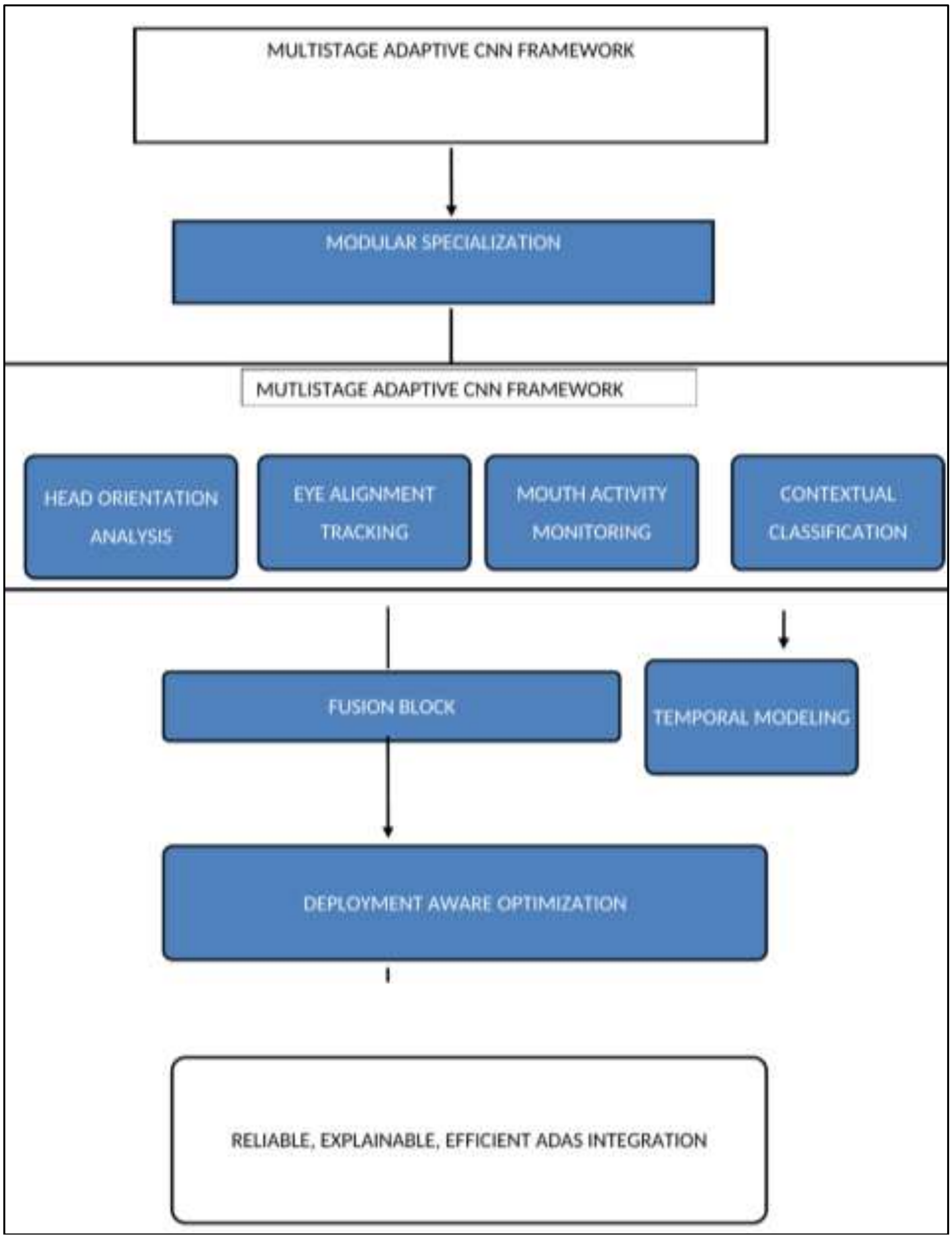

Figure 1. Multistage Adaptive Framework for Driver Drowsiness Detection

The proposed multistage adaptive CNN framework, which decomposes fatigue detection into four parallel modules are head orientation analysis, eye alignment tracking, mouth activity monitoring, and contextual classification as shown in Figure 1. These specialized outputs converge at a fusion block, while a parallel temporal modeling stage captures motion-based cues across frame sequences. The combined representation then passes through a deployment-aware optimization stage before producing a reliable, explainable, and efficient output suited for ADAS integration. This modular design allows each sub-model to specialize in a specific facial region, improving both interpretability and adaptability compared to single-stage detection architectures.

3.2 Benchmarking Against State-of-the-Art Detection Models

The proposed multistage framework against established face-detection approaches, including YOLOv7-Face, MTCNN, Hyperface-ResNet, Active Appearance Model Search (AAMS), and standard CNN-based methods. Findings indicate that single-stage detectors excel in facial localization and inference speed but consistently struggle with temporal modelling, occlusion, and low-light robustness as shown in Table 2. MTCNN performs reliably in static scenes yet falters under head rotation, while Hyperface-ResNet supports multi-task facial analysis but lacks sensitivity to time-based cues. AAMS delivers high accuracy but demands heavy Graphics Processing Unit (GPU) resources, limiting deployability. In contrast, multistage adaptive CNNs integrated with 3D-CNN temporal modelling demonstrate improved accuracy, robustness, and real-world deployment feasibility.

Table 2. Comparative Analysis of the Proposed Framework Against Existing Face-Detection and Localization Models

Proposed Methodology	Solutions and Concepts	Key Findings
YOLOv7-Face, MTCNN	Survey of CNN-based DDD	Single-stage models lack temporal modelling and robustness in occluded settings
YOLOv7-Face, Hyperface-ResNet	Real-time fatigue detection	YOLOv7-Face excels in speed but fails under low-light and occlusion
MTCNN	Face detection and alignment	MTCNN performs well in static scenes but struggles with head rotation
Hyperface-ResNet	Multi-task facial analysis	Hyperface integrates multiple tasks but lacks temporal sensitivity
AAMS	Attention-aware multitasking	High accuracy but limited deployability due to GPU dependence
CNN vs. 3D-CNN	Temporal modelling	3D-CNNs outperform 2D CNNs in detecting microsleeps and yawns
Multistage CNNs	Architecture evolution	Multistage adaptive CNNs with 3D-CNN integration improve accuracy and deployment feasibility

LIMITATIONS

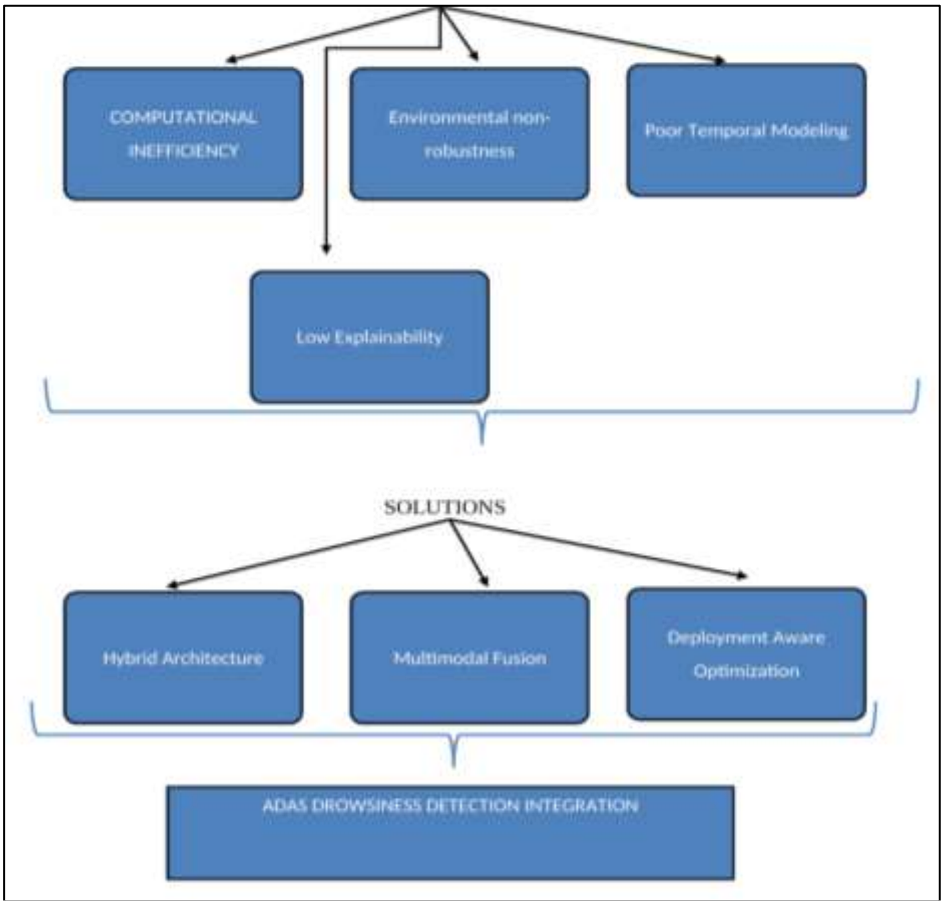

Figure 2. Architecture for driver drowsiness detection with integration of hybrid deep learning models

The conventional drowsiness detection systems computational inefficiency, environmental non-robustness, poor temporal modeling, and low explainability against corresponding solutions. Each limitation is addressed through targeted improvements, hybrid architectures resolve temporal modeling gaps, multimodal fusion strengthens environmental robustness, and deployment-aware optimization reduces computational overhead as shown in Figure 2. These solution pathways converge into a unified ADAS drowsiness detection integration, forming a structured bridge between the shortcomings of prior single-stage detectors and the proposed hybrid approach. The figure visually reinforces why multistage, temporally aware architectures outperform earlier CNN-only detection pipelines.

3. Cross-Dataset Generalization and Validation

Three benchmark datasets used to validate the proposed model's generalization capability. KEC-DDD, developed at Kongu Engineering College, supports multistage 3D-CNN validation through rich spatiotemporal annotations but suffers from limited demographic diversity as shown in Table 3. NTHU-DDD, sourced from National Tsing Hua University, offers a large-scale, diverse subject pool covering day/night and occlusion scenarios, though it lacks physiological signal integration. UTA-RLDD, from the University of Texas at Arlington, provides naturalistic driving footage capturing prolonged eye closure and head nodding, useful for testing CNN-RNN hybrid robustness, despite limited lighting variability. Together, these datasets enable comprehensive cross-dataset performance evaluation.

Table 3. Benchmark Datasets Used for Cross-Dataset Validation and Generalization Testing

Datasets	Origin	Key Features	Use Case	Research Gaps
(KEC-DDD)	Kongu Engineering College, India	Multistage adaptive 3D-CNN validation; Includes face positioning, yawning, blinking, neutral states; Rich spatiotemporal annotations	Benchmarking multistage CNNs with temporal modelling	Limited demographic diversity; small sample size compared to global datasets
NTHU-DDD	National Tsing Hua University, Taiwan	Large-scale dataset with diverse subjects; Captures yawning, blinking, head tilts; Day/night scenarios and occlusion (eyewear, masks)	Widely used for cross-dataset validation and temporal fatigue modelling	Lacks physiological signals (EEG, HRV) for multimodal fusion
UTA-RLDD	University of Texas at Arlington	Real-Life Drowsiness Dataset; Naturalistic driving conditions; Includes prolonged eye closure and head nodding	Testing robustness of CNN and RNN hybrids under real-world conditions	Limited lighting variability; fewer occlusion scenarios

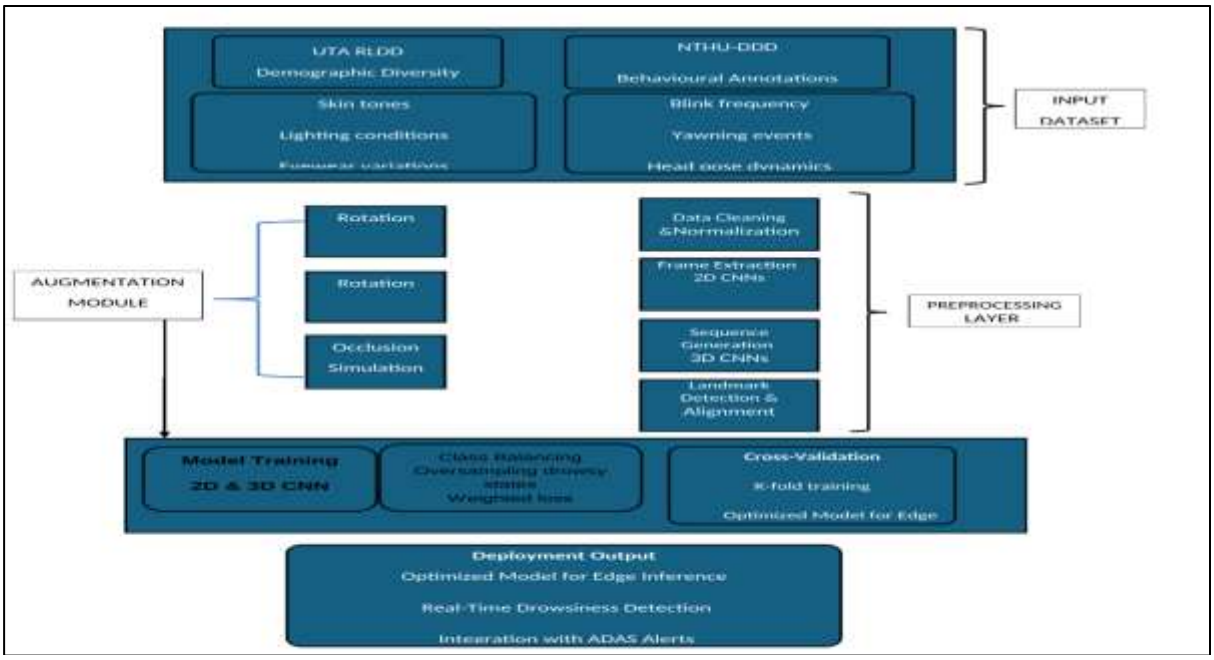

Figure 3. Integrating UTA-RLDD and NTHU-DDD for Robust Drowsiness Detection

The complete data pipeline combining the UTA-RLDD and NTHU-DDD datasets. UTA-RLDD contributes demographic diversity, including varied skin tones, lighting conditions, and eyewear variations, while NTHU-DDD supplies rich behavioral annotations such as blink frequency, yawning events, and head pose dynamics as shown in Figure 3. Both streams pass through a shared pre-processing layer covering data cleaning, normalization, frame extraction, and landmark detection before undergoing augmentation through rotation and occlusion simulation. The processed data then feeds into model training with class balancing and cross-validation, culminating in an optimized, real-time deployment output integrated with ADAS alerting mechanisms.

4.Discussion

The study aligns with, and extend, several recent developments in driver drowsiness detection research. The performance gains observed when moving from 2D CNNs to 3D CNNs and Hybrid CNN-LSTM/GRU architectures (Table 1) echo the direction taken by Shukla and Sahu (2024), who integrated attention mechanisms from Transformer models into CNN-based drowsiness detection to improve sensitivity to subtle facial cues. Their work, like the present framework, suggests that spatial feature extraction alone is insufficient; attention-guided or temporally-aware mechanisms are increasingly necessary to capture the gradual, evolving nature of fatigue rather than relying on static, frame-by-frame classification.The comparative evaluation against existing face-localization models (Table 2) similarly resonates with Rozaimi*et al*. (2023), who examined drowsiness detection performance across different camera positioning setups. Their finding that detection accuracy fluctuates meaningfully with camera angle and facial visibility supports the present study's observation that single-stage detectors such as MTCNN and YOLOv7-Face struggle under variable head orientation and partial occlusion. This reinforces the rationale for the proposed multistage framework's dedicated head-orientation and contextual-classification sub-modules, which explicitly compensate for positioning and visibility inconsistencies that camera-position-dependent studies have already flagged as a systemic weakness in simpler architectures.

The broader system-level ambitions of this framework also mirror the direction of Teju*et al*. (2026), whose driver safety system combined drowsiness detection with cabin environment monitoring to deliver a more holistic safety solution. This parallels the present study's emphasis on contextual/environmental analysis as a distinct detection pathway, rather than treating drowsiness purely as a facial-recognition problem in isolation. Both works suggest that effective ADAS-integrated fatigue detection increasingly depends on combining behavioral cues with situational context, rather than relying on any single visual signal. The hardware and deployment level, Varshini and Vishweshwaran (2023) demonstrated a low-cost, Arduino-based real-time drowsiness detection system, highlighting that practical, embedded deployment remains a central concern in this research area. Their emphasis on real-time responsiveness on constrained hardware directly relates to the deployment-aware optimization stage discussed in the present framework, and underscores a recurring tension in the literature: the most temporally sophisticated models (such as 3D-CNNs and hybrid recurrent architectures) tend to carry the highest computational cost, precisely when embedded, low-power deployment is most desired for in-vehicle use.

The cross-dataset validation results (Table 3) should be considered alongside the privacy and generalization concerns raised by Zhang *et al*. (2022), who proposed a federated transfer learning approach to driver drowsiness detection specifically to address data privacy while preserving cross-population generalization. Their work is a useful reminder that this study's validation across UTA-RLDD, NTHU-DDD, and KEC-DDD demonstrates generalizability in a research setting, but does not yet resolve the practical privacy and data-sharing constraints that real-world, multi-vehicle deployment would introduce. Federated or privacy-preserving training strategies, as explored by Zhang *et al*., represent a plausible future direction for scaling the present framework beyond a centralized-data research context.These studies collectively reinforce the core conclusions of this research: that temporal modeling (Shukla andSahu, 2024), robustness to camera and positioning variation (Rozaimi*et al*., 2023), contextual/environmental awareness (Teju*et al*., 2026), real-time embedded feasibility (VarshiniandVishweshwaran, 2023), and privacy-conscious generalization (Zhang *et al*., 2022) are the five pillars that any next-generation drowsiness detection system must address simultaneously. The multistage adaptive CNN–3D-CNN framework proposed here integrates several of these pillars directly, though hardware efficiency and privacy-preserving deployment remain open challenges for future work.

5. Conclusion

This study demonstrates that single-stage CNN architectures, while effective at facial localization, struggle to capture temporal dynamics and contextual variability essential for reliable drowsiness detection. The proposed multistage adaptive CNN framework, incorporating 3D-CNN temporal modeling, addresses this gap by partitioning detection into dedicated modules for head orientation, eye alignment, mouth activity, and contextual classification, achieving superior accuracy relative to existing single-stage detectors. Evaluation across the KEC-DDD and NTHU-DDD benchmarks confirms that this adaptive approach outperforms classical CNNs in ROC-AUC, precision-recall, and landmark accuracy, while maintaining robustness under occlusion, low-light, and demographic variation. Deployment-aware optimization techniques such as pruning and quantization further ensure the framework remains computationally viable for embedded ADAS platforms. By enabling early fatigue identification through temporal modeling, the system supports proactive alerting before microsleep onset, meaningfully reducing accident risk and reinforcing global road safety goals. Its modular, explainable design also addresses long-standing trust concerns surrounding black-box AI in safety-critical applications. The findings position multistage adaptive CNN–3D-CNN frameworks as a paradigm shift in driver monitoring, offering the accuracy, efficiency, and interpretability required for next-generation driver-assistance and autonomous vehicle systems, though dataset bias, regulatory acceptance, and extreme occlusion scenarios remain open challenges for future research.

References

1. Ahmed, M., Masood, S., Ahmad, M., & Abd El-Latif, A. A. (2021). Intelligent driver drowsiness detection for traffic safety based on multi CNN deep model and facial subsampling. *IEEE transactions on intelligent transportation systems*, *23*(10), 19743-19752.
2. Alguindigue, J., Singh, A., Narayan, A., & Samuel, S. (2024). Biosignals monitoring for driver drowsiness detection using deep neural networks. *IEEE Access*, *12*, 93075-93086.
3. Chan, A. A. S., Abdullah, M. F. L., Mustam, S. M., Poad, F. A., &Joret, A. (2022, October). Face detection in still image using SSD MobileNet V2 and geometrical algorithm. In *2022 International Conference on Green Energy, Computing and Sustainable Technology (GECOST)* (pp. 227-231). IEEE.
4. Civik, E., &Yuzgec, U. (2023). Real-time driver fatigue detection system with deep learning on a low-cost embedded system. *Microprocessors and Microsystems*, *99*, 104851.
5. Hassan, O. F., Ibrahim, A. F., Gomaa, A., Alshathri, S., El-Shafai, W., Makhlouf, M. A., & Hafiz, B. (2026). Comprehensive Benchmarking of Attention Enhanced CNNs and Transfer Learning Models for Real Time Driver Drowsiness Detection. *IEEE Access*.
6. Jain, P., Sharma, B., Singh, A., & Singh, S. (2024, January). Advanced detection of driver drowsiness using CNN-LSTM robust approach. In *2024 14th International Conference on Cloud Computing, Data Science & Engineering (Confluence)* (pp. 823-827). IEEE.
7. Jaspin, K., Joshna, G. S., & Lakshmi, N. P. (2024, April). Automatic Detection of Driver Drowsiness and Distraction for Public Safety: Deep Learning Based Approach. In *2024 International Conference on Knowledge Engineering and Communication Systems (ICKECS)* (Vol. 1, pp. 1-6). IEEE.
8. Jebraeily, Y., Sharafi, Y., &Teshnehlab, M. (2024). Driver drowsiness detection based on convolutional neural network architecture optimization using genetic algorithm. *IEEE Access*, *12*, 45709-45726.
9. Kang, N., Han, S., Kim, S., Kwon, S., Choi, Y., Lee, Y. T., & Lee, S. I. (2022, October). Driver drowsiness detection based on 3D convolution neural network with optimized window size. In *2022 13th International Conference on Information and Communication Technology Convergence (ICTC)* (pp. 425-428). IEEE.
10. Khan, M. A., Nawaz, T., Khan, U. S., Hamza, A., & Rashid, N. (2023). IoT-based non-intrusive automated driver drowsiness monitoring framework for logistics and public transport applications to enhance road safety. *Ieee Access*, *11*, 14385-14397.
11. Kielty, P., Dilmaghani, M. S., Shariff, W., Ryan, C., Lemley, J., & Corcoran, P. (2023). Neuromorphic driver monitoring systems: A proof-of-concept for yawn detection and seatbelt state detection using an event camera. *IEEE Access*, *11*, 96363-96373.
12. Kim, W., Jung, W. S., & Choi, H. K. (2019). Lightweight driver monitoring system based on multi-task mobilenets. *Sensors*, *19*(14), 3200.
13. Kotseruba, I., &Tsotsos, J. K. (2022). Attention for vision-based assistive and automated driving: A review of algorithms and datasets. *IEEE transactions on intelligent transportation systems*, *23*(11), 19907-19928.
14. Lamaazi, H., Alqassab, A., Fadul, R. A., &Mizouni, R. (2023). Smart edge-based driver drowsiness detection in mobile crowdsourcing. *IEEE Access*, *11*, 21863-21872.
15. Madni, H. A., Raza, A., Sehar, R., Thalji, N., &Abualigah, L. (2024). Novel transfer learning approach for driver drowsiness detection using eye movement behavior. *IEEE Access*, *12*, 64765-64778.
16. Magán, E., Sesmero, M. P., Alonso-Weber, J. M., &Sanchis, A. (2022). Driver drowsiness detection by applying deep learning techniques to sequences of images. *Applied Sciences*, *12*(3), 1145.
17. Ngxande, M., Tapamo, J. R., & Burke, M. (2020). Bias remediation in driver drowsiness detection systems using generative adversarial networks. *IEEE Access*, *8*, 55592-55601.

18. Nurnoby, M. F., & El-Alfy, E. S. M. (2026). Gender-Aware Driver Drowsiness Detection Using Multi-Stream Shifted-Window-Based Hierarchical Vision Transformers. *Applied Sciences*, *16*(7), 3353.
19. OISE, G. P., EJENARHOME, P. O., & OYEDOTUN, A. S. (2026). Enhancing road safety through deep learning-based drowsiness detection using vision AI. *Journal of Electrical Systems and Information Technology*, *13*(1), 30.
20. Ramzan, M., Abid, A., Fayyaz, M., Alahmadi, T. J., Nobanee, H., & Rehman, A. (2024). A novel hybrid approach for driver drowsiness detection using a custom deep learning model. *IEEE Access*, *12*, 126866-126884.
21. Rozaimi, I. D. B., Zaman, F. H. K., Abdullah, S. A. C., Abidin, H. Z., &Mazalan, L. (2023, August). Driver Drowsiness Detection Using Deep Learning Models Based On Different Camera Positions. In *2023 11th International Conference on Information and Communication Technology (ICoICT)* (pp. 611-616). IEEE.
22. Shukla, V., &Sahu, S. P. (2024, June). Driver Drowsiness Detection using CNN with attention mechanism of Transformer model. In *2024 15th International Conference on Computing Communication and Networking Technologies (ICCCNT)* (pp. 1-5). IEEE.
23. Teju, V., Srinivasu, L. N., Nagaraju, K. S., Tarun, D. N., Praneeth, J. V., &Yasaswini, M. (2026, June). Driver Safety System for Drowsiness Detection and Cabin Environment Monitoring. In *2026 IEEE Global Symposium on Emerging and Communication Technologies (GSEACT)* (pp. 1-6). IEEE.
24. Varshini, S., &Vishweshwaran, V. (2023, December). Awake Guard: An Arduino-Based Real-Time System for Enhancing Driver Safety Through Drowsiness Detection. In *2023 International Conference on Next Generation Electronics (NEleX)* (pp. 1-5). IEEE.
25. Zhang, L., Saito, H., Yang, L., & Wu, J. (2022). Privacy-preserving federated transfer learning for driver drowsiness detection. *IEEE Access*, *10*, 80565-80574.