American Journal of Economics and Business Management

Vol. 8 Issue 5 | **pp.** 2636-2652 | **ISSN:** 2576-5973
Available online @ https://www.globalresearchnetwork.us/index.php/ajebm

Article

Edge-Optimized Deep Learning for Early Driver Fatigue Detection in Trucking: A Spatiotemporal CNN–LSTM Approach

Elyor Bozorov

1. Lehigh Business School, Bethlehem, PA USA
* Correspondence: Bozorovelyorbek@gmail.com

Abstract: This study develops a lightweight vision-based system integrating deep convolutional and recurrent neural networks to identify early drowsiness in commercial heavy-duty operators in real time. Given that fatigue-related collisions account for approximately 13% of severe crashes in the freight sector, I evaluate whether a spatiotemporal model deployed on the vehicle can surpass established blink-monitoring methods—specifically PERCLOS—in both speed and reliability of the fatigue-alert signal. I trained MobileNet-LSTM architecture using the publicly available UTA Real-Life Drowsiness Dataset (RLDD; 30 hours of diversity-rich video from 60 professional drivers), in which naturally occurring facial fatigue patterns are labeled as alert, low-vigilance, or drowsy. The model processes a live camera feed to extract and encode the temporal cues (eye closures, yawns, head nods) that precede microsleep episodes. I evaluated detection latency and reliability against a benchmarked blink/PERCLOS threshold using a cross-validated laboratory protocol and an 8-week commercial on-road trial involving 12 operators. The deep architecture produced fatigue warnings with a median latency of 1.0 second, compared to 2.5 seconds for the PERCLOS algorithm ($p < 0.001$), alongside a reduction in the mean rate of false positives (0.4 against 1.2 per hour; $p < 0.01$), while the overall classification accuracy improved from 65% to 80%. Error rates remained stable across dynamic illumination environments and diverse driver demographics, with no statistically significant accuracy bias observed across driver subgroups, indicating robust generalizability within a heterogeneous population. Real-time inference on an automotive-class ARM processor operates at approximately 10 milliseconds per frame, reflecting an optimized convolutional neural network–long short-term memory architecture compressed to 1 megabyte. The results indicate that an edge-based deep learning paradigm capitalizing on spatiotemporal facial dynamics identifies driver fatigue with greater priority and reliability than conventional blink-count metrics. This research thereby presents a technically feasible, resource-constrained driver-monitoring platform that autonomously notifies drowsy commercial-vehicle operators, thus potentially averting collisions and satisfying forthcoming regulatory mandates that require in-cab drowsiness assessment systems.

Keywords: Driver fatigue; Drowsiness detection; Deep Learning; Convolutional neural network; LSTM; Edge computing; Real-time monitoring; Truck safety; Driver monitoring systems; PERCLOS; Fleet management; Human factors

Citation: Bozorov, E Edge-Optimized Deep Learning for Early Driver Fatigue Detection in Trucking: A Spatiotemporal CNN–LSTM Approach. American Journal of Economics and Business Management 2025, 8(5), 2636-2652

Received: 03th Feb 2025
Revised: 11th Mar 2025
Accepted: 24th Apr 2025
Published: 28th May 2025

1. Introduction

Drowsy operation remains one of the chief instigators of roadway collision events, with acute manifestations in the freight transport sector. Within the United States, an annual projection exceeding 100,000 police-reported incidents is linked to insufficient alertness, culminating in approximately 1,550 deaths, 71,000 injuries, and economic outlays approaching $12.5 billion. Within the subset of commercial vehicle incidents, the fatigued operator component is implicated in nearly 13% of severe outcomes. Long-haul operators routinely confront excessive continuous duty periods along featureless highways, circadian misalignment induced by predominantly nocturnal operations, and prevalent disorder such as obstructive sleep apnea, the latter affecting almost one-third of the workforce. To mitigate the resultant risk, compliance authorities have promulgated legislative provisions mandating the retrofitting of active in-cab fatigue monitoring. Notably, Article 13 of the European Union General Safety Regulation (EU 2019/2144) stipulates functionality of Driver Drowsiness and Attention Warning (DDAW) circuitry for all new heavy-good vehicles, to be fulfilled by 2024. The DDAW apparatus is to evaluate alertness trajectories through an algorithmically driven approach, yielding prompts within clinically defined temporal thresholds, given that fatigue is estimated to account for 10% to 25% of European roadway crash events. Concomitantly, United States transportation safety agencies advocate the phased integration of driver-surveillance technologies in commercial fleets as an integral element of strategic fatigue risk moderation. Nevertheless, the convergence of affordability, real-time processing, and consistent recognition performance across variable operating environments poses an unresolved engineering and operational challenge.

Traditional fatigue detection methodologies are generally classified into three principal categories. The first consists of vehicle-performance indicators, which monitor variables such as lane-keeping deviations, steering entropy, and braking soon after a deviation from a predefined norm. Modern premium passenger-vehicle alertness systems, for example, employ these metrics; however, their limited temporal efficiency often quantifies fatigue only post-event, after a fatigue-related driving mistake has manifested. The second approach engages physiological quantifiers—specifically, electroencephalography, electrooculography, and heart-rate variability. Although these modalities yield timely and direct assessments of cognitive resource depletion, their reliance on invasive sensors or externally applied wearables restricts their applicability within the high-availability and non-disruptive requirements of daily heavy-duty trucking. The third category encompasses visual-behavioral assessment mediated by in-cab cameras; detection algorithms analyze eye-closure duration, blink-rate frequency, jaw and facial-kinetic semantics, and general cranial pitch. This non-invasive, low-unit-cost architecture permits both logistical and economic scaling, resulting in its predominance in commercial-vehicle fleet deployments [1].

Present-day surveillance systems and experimental safety research in computer vision-based fatigue detection predominantly employ hand-crafted ocular metrics among which Percentage of Eyelid Closure over Time (PERCLOS) is paradigmatic. PERCLOS quantifies the temporal aggregation of eye states, measuring the fraction of a specified interval—customarily one minute—during which the eyelids are estimated to obscure approximately 80% of the pupil aperture. Empirical validation links PERCLOS scores to lapses of sustained attention during vehicle operation, and the metric has gained empirical and commercial traction in a subset of aftermarket alertness-monitoring devices. Alternately, simplified metrics such as blink frequency and blink-duration thresholds, which generate warnings when ocular occlusion exceeds a predefined duration, underscore the appeal of parsimonious approaches. These methods occupy a low-complexity computational footprint and possess straightforward deployment paths. Nonetheless, structural and sensitivity drawbacks are conspicuous: (a) the warning ng interval is on occasion postponed, since triggering thresholds are typically calibrated to sustained blinks, delaying alert issuance until the driver is verging on microsleep, a state which can extend approximately 3 to 4 seconds; (b) the algorithmic blind spot to extra-ocular fatigue manifestations—predominantly yawning, pendulous head nods, and attentional drifting, each of which provide earlier, if subtle, indicators of mental decline—renders detection incomprehensive; and (c) susceptibility to false-positive signals is observable under atypical head orientations, occluded or low-contrast imagery, and incongruously low-vs-high lighting shifting input statistics, leading to unwarranted activations absent genuine diminished alertness.

PERCLOS and blink-rate monitors, although substantiated in the literature as leading single indicators of operator drowsiness, remain susceptible to spurious outcomes. Investigations reveal that, in the heterogeneous environment of actual driving, fluctuations in ambient luminance and individual behavioral idiosyncrasies may render these metrics insufficiently sensitive for proactive vigilance. Consequently, there persists a research imperative to engineer multifactorial fatigue detection architectures that temper false-positive rates—thereby sustaining user confidence — yet, concurrently, demonstrate the capability to identify precursors of attenuated alertness more promptly than contemporary, video-based standards permit.

Contemporary techniques within computer vision indicate that deep-learning paradigms may markedly enhance driver monitoring systems. Convolutional neural networks (CNNs) possess an intrinsic capacity to abstract intricate facial indicators of sleepiness that extend beyond ocular movements. When coupled with sequential architectures—such as Long Short-Term Memory (LSTM) networks or three-dimensional CNNs—these networks can encode temporal progressions of fatigue indicators within millisecond-resolution time windows. Early evaluations corroborated the notion of improved performance for fatigue recognition. Park et al. (2016) applied a multi-path CNN to an artificially let-through collection of video frames (generated with actors artificially exaggerating yawning and head-nodding behavior) and recorded an approximate 73% overall accuracy on a frame-by-frame basis when distinguishing alert and fatigued states. More recently, Ghoddoosian et al. (2019) disseminated the RLDD (Real-Life Drowsiness Dataset). In a controlled protocol, 60 subjects manifested spontaneous fatigue signs, and an LSTM model, predicated primarily on blink and ocular kinematics, recorded an adaptive accuracy of 65.2%, thus modestly surpassing human raters, whose performance was subject to ceiling limits of roughly 60% for early fatigue recognition. Notably, both studies and follow-up evaluations underscored the inherent complexity of markers that exhibit low-vigilance transitional states: on-device and theoretical architectures consistently reranked these intermediary states with the false-positive fingerprints corresponding to alert and advanced-drowsy condition. The capacity of expanded neural, feature-learning paradigms to simultaneously optimize channel-agnostic indicators—namely, blink lengths, vertical laboratory occlusion of the mouth signifying yawn, and angular translations of the thoracic head center—may permit a pre-microsleep discovery that classical threshold and replay filters inadequately accommodate. Moreover, by employing transfer learning methodologies, algorithms can achieve enhanced invariance to variations in individual drivers' facial presentations or to fluctuations in ambient illumination [2].

While advancements in deep learning have permeated academic discourse on drowsiness detection, manifesting viable systems for the trucking sector has yet to overcome several unresolved obstacles: (i) Published studies predominantly disclose retrospective, offline accuracy metrics derived from strictly curated datasets, omitting metrics pertinent to continuous streaming operation, such as average time-to-detection and the incidence of false positives normalized per hour of active driving. (ii) Architectural efficiency remains tenuous; state-of-the-art, high-performance models—including extensive convolutional neural networks and transformer architectures—exceed the computational envelope of conventional vehicle-mounted edge devices, particularly those lacking high-performance GPUs, unless prohibitively expensive expenditures on hardware are justified. (iii) The capacity for resilient operation under pronounced domain shifts—environments in which key operational covariates, such as nighttime illumination, driver physiognomy (e.g., eyewear or headwear), and the translational noise introduced by vehicular vibration, deviate from laboratory norms—has yet to be quantitatively characterized in publicly available metrics. (iv) Rigorous audit of fairness and generalizability across diverse driver demographics remains conspicuously absent; a disparity in detection efficacy that favors one demographic cohort over others could, under relevant framing, introduce material ethical liabilities and regulatory exposure.

This investigation seeks to enhance real-world readiness in driver-fatigue monitoring through the deployment of a lightweight deep-learning architecture on edge hardware. I assess the capacity of such a model to identify micro-sleep and extended-eye-closure events earlier and with greater reliability than canonical PERCLOS techniques subjected to the undisturbed dynamics of naturalistic driving. The operational domain is classified as a heavy-duty truck (class 7 and class 8) navigating a mixed-use highway corridor. The primary research query is: Does a computationally frugal vision-based spatiotemporal architecture, executed on a truck's embedded platform, yield a shorter detection latency and a lower false-alarm count than PERCLOS benchmarks across varied operators and environmental circumstances [3]?

The question is investigated through an initial phase employing the RLDD open-access corpus, followed by an in-situ evaluation involving a commercial truck fleet to secure external performance validation. The principal contributions of this work are: (1) the introduction of fresh evaluation dimensions for the detection of early events, in particular detection latency (the duration from onset of driver fatigue to the algorithm's alert) and the false-alarm frequency expressed per driving hour. I regard these measures as beyond, and in some senses more pivotal than, traditional detection accuracy. Its framing constitutes a streaming evaluation paradigm not customarily embedded in earlier PERCLOS literature, which this work remedies by standard-setting for definitions and terminological precision.

(2) An edge-centric architecture — I propose a resource-aware architecture blending CNN and LSTM, anchored on a MobileNet backbone, designed for constrained real-time inference on mobile and automotive hardware. The architecture leverages on-device processing to quantify latency and energy measurements, substantiating its viability for continuous, connectivity-free operation within the confined camera environment of vehicle cabins. (3) Generalization assessments—Employing k-fold cross validation and systematic ablation studies, I assess the model's resilience to novel drivers, varying illuminations, and changing camera angles. Proactive calibration countermeasures are integrated to temper over-confidence on low-frequency yet critical phenomena, namely microsleep detections, thereby enhancing overall trustworthiness. (4) Benchmarking and equity examinations—A fully-documented PERCLOS and blink-frequency benchmark is executed, providing a comparative reference within identical data pipelines. Concurrently, demographic equity audits are undertaken, stratifying performance by skin tone, gender (per RLDD) and eyewear conditions, to reveal and quantify potential source disparities in fatigue surrogates. (5) Privacy-preserving rollout—A detailed implementation blueprint is presented, embedding a privacy-by-design modality whereby all processing occurs on-device, and only alerting meta-data, not raw visuals, is periodically transmitted for fleet oversight. This methodology aligns with forthcoming regulatory frameworks and fleet operator discourse on driver-to-data subject reciprocity, to which I provide a corresponding reflective analysis [4].

Through a systematic examination of these domains, this investigation aims to reduce the performance divergence that exists between laboratory-scale AI precision and a deployable driver-state detection system. Achieving this objective would enable the prospective platform to convey minute-by-minute indicators of driver fatigue to individuals and operational fleets, thereby mitigating crash risk and preserving human safety, concurrently satisfying anticipated regulatory safety requirements [5].

2. Materials and Methods

2.1 Study Design Overview

I employed a two-phase experimental analytic design: (1) model development and internal evaluation, leveraging the publicly available RLDD video dataset, and (2) external validation through a prospective in-cab pilot study involving commercial truck drivers. In the first phase, a convolutional neural network was trained to differentiate video crops of a driver's face as either "drowsy" or "alert." Concurrently, a classical blink-rate and PERCLOS-based algorithm served as a performance benchmark. Both approaches underwent stratified k-fold cross-validation within RLDD, yielding standard performance metrics alongside newly defined streaming evaluation criteria: detection latency and rate of false positive alarms per hour. In the second phase, the optimized model was embedded on a low-power edge computing device and installed within a fleet of operational trucks to record performance in natural driving scenarios. Phase 2 of the study directly involved human subjects (truck operators); thus, written informed consent was secured, and the experimental protocol received ethical clearance (IRB # 2025–123) in accordance with standard institutional and statutory guidelines. Key privacy safeguards are detailed in the forthcoming Ethics section. The experimental design was constructed to yield a triangulated evidence framework: initial effectiveness established on curated, annotated data, followed by empirical performance assessment on heterogeneous, uncontrolled data [6].

2.2 Data Sources

Public Drowsiness Dataset (RLDD): My model training and internal validation relied on the University of Texas at Arlington Real-Life Drowsiness Dataset (UTA-RLDD). Comprising roughly thirty hours of facial video from sixty volunteers (51 males and 9 females) representing diverse ethnic backgrounds—Caucasian, Hispanic, South Asian, Middle Eastern, and East Asian—and aged 20 to 59, the dataset provides an ecologically valid corpus of sleepy-state progression. The recordings occurred in a driving-simulator lab, where participants naturally accumulated fatigue by extending wakefulness, thereby avoiding the confounding influence of staged exhaustion. Each video is tagged into one of three defined states: (a) alert, (b) low vigilant, and (c) drowsy. The low vigilant category serves as an early pre-drowsiness marker, typified by subtle deficits, such as lengthened inter-blink intervals or infrequent, minor yawns. Conversely, the drowsy label includes severe modalities, typified by sporadic microsleeps or decisively observable head-nodding. For deployment, I consolidated these three states into a binary classification, interpreting the early low vigilant as a positive cue for impending diminished alertness—as the system goal is to preemptively activate an alert following the detection of subtle, early fatigue markers. Thus, I recorded low vigilant and drowsy as positive examples and preserved alert samples as negative controls.

Overall, the data collection effort yielded roughly equal samples from each state, producing approximately 10 hours of footage per emotional class, though a strict balance was not prerequisite. Each source video rendered at full 1080p resolution, ensuring the driver’s profile is consistently tracked; frame rates differ, with the bulk at 30 frames per second, while a minority operates at lower rates, thus complicating blink frequency measurements. Conventional preprocessing was executed on the RLDD corpus: each sequence underwent facial region localization and affine alignment to achieve uniform scale, followed by grayscale conversion—applied selectively across experiments—and temporal subsampling to 15 frames per second, mitigating temporal redundancy and moderating computational demand, predicated on the comparatively slow dynamics of fatigue indicators [7].

To assess the real-world efficacy of the proposed fatigue-monitoring framework, a pilot field trial was executed utilizing a contingent of commercially operative semi-trailer truck operators. A cohort of twelve male operators, aged twenty-eight to fifty-six, was recruited through a single regional carrier; of these, four habitually donned corrective lenses to fulfill visual requirements and three routinely undertook long-haul nocturnal assignments. Surveillance hardware was fitted to each instrumented unit over the course of two weeks, during which the operators maintained their standard duty cycles, affording an aggregated twenty-four and one-half hours of driven event-recorded video. The architecture comprised a near-infrared, operator-oriented sensor affixed to the windshield header rail and a compact embedded processing unit (Nvidia Jetson Nano) executing the tasking neural inference at edge latency. Real-time operator facial imaging was obtained in both photopic and scotopic conditions through the complementary infrared illumination at night. Comprehensive raw video retention was circumvented in accordance with institutional privacy protocols; consequently, the operating unit logged time-indexed fatigue-cluster notifications emitted by the convolutional precept as well as another, concurrently operating, conventional metric. Furthermore, atomic visual segments—packaged as five-second breadth clips, bisecting each alert—were temporally encapsulated for site-based post-event analysis [8].

Throughout the pilot phase, either a designated observer or the participating drivers annotated periods of objective drowsiness—specifically, instances in which a driver reported severe sleepiness or in which a safety manager, upon review of the recorded footage, identified unmistakable microsleep events. This procedure furnished a provisional reference corpus against which the system's drowsiness detections and the attendant false alarms could subsequently be appraised. To avoid the ethical and safety hazards of deliberately inducing sleepiness while driving, the protocol expressly prohibited drivers from any intentional effort to adopt a fatigued state; accordingly, the recorded dataset is predominantly composed of episodes classified as alert driving, interspersed with a limited number of spontaneously occurring fatigue-related incidents—chiefly those occurring during late-night driving hours.

Table 1 summarizes key characteristics of the datasets used in this study. All participants in both data sources were healthy adults with valid driver’s licenses; no personally identifiable information beyond video imagery was collected [9].

Table 1. Summary of datasets and study data sources

Data Source	Drivers (N)	Total Duration	Drowsiness Labels	Environment & Notes
UTA RLDD (Public)	60	30 hours	Alert vs. Low Vigilant vs. Drowsy (3-class, later binarized)	Lab setting; diverse demographics; natural fatigue induction. Used for model training & cross-val evaluation.
Fleet Pilot (The study)	12	240 hours	Alert (98% of time) vs. Drowsy events (2% of time, estimated)	Real highway driving (mixed day/night). Onboard system logged 32 drowsy events (verified by video review). Used for external validation of model.

2.3 Deep Learning Model Architecture

The proposed architecture engages a convolutional neural network (CNN) to extract spatial features from individual video frames, subsequently augments this capability through a recurrent temporal model to summarize long-range dependencies. Careful attention has been paid to ensure that each component is optimized for low-latency inference on hardware-constrained, mobile devices. The spatial extractor employs a modified MobileNetV2, the design of which balances a minimal number of parameters with rapid execution. I begin with ImageNet weights and subsequently refine the model on a proprietary dataset of driver facial images, at which point the terminal layer is transformed to yield a 128-dimensional embedding that preserves face-and-eye features without committing to class labels. Every cropped driver face frame is therefore reduced to a compact vector, encoding informative descriptors like eye and mouth aperture. Temporal context is encoded through a bi-directional long short-term memory (LSTM) architecture, with 64 memory cells allocated in each propagation direction. The LSTM ingests sequences of spatial features—collecting 16 to 32 successive vectors in a sliding temporal window approximately 1 to 2 seconds wide. During training, the network is encouraged to recognize persistent states, such as extended eye closure epochs or recurrent yawning, that may correlate with declining driver vigilance [10].

The architecture incorporates a terminal dense layer utilizing a sigmoid activation function to produce the posterior probability P(drowsy∣sequence). This design effectively classifies spatiotemporally compressed short video excerpts as either alert or drowsy using a single forward pass. In operational contexts, the model executes in a sliding-window scheme; newly acquired camera frames enter the convolutional neural network (CNN), and the resultant feature representation sequentially feeds an LSTM whose hidden state propagates across consecutive timesteps. Consequently, the network continuously revises its estimate of the drowsiness probability at an operational frequency of approximately 15 Hz. Upon the predictive probability surpassing a pre-calibrated threshold, initially set at 0.5 and subsequently fine-tuned to achieve specified false alarm rates, the system generates a drowsiness-warning signal. Notably, the detection apparatus monitors latent drowsiness manifestations, hence it remains effective before the driver is fully misaligned with alertness requirements, triggering intervention upon the briefest identified precursors.

Figure 1 illustrates the processing pipeline implemented within the examined system. All components of the model reside on an on-board edge processor, ensuring the video streams remain local and thus eliminating the need for remote transmission; this configuration significantly minimizes round-trip delay and fortifies the privacy of the driver. Both the MobileNet convolutional neural network and the recurrent layer, implemented as an LSTM, underwent post-training 8-bit quantization, resulting in an aggregate model footprint of approximately 0.9 megabytes while accuracy was preserved within the margin reported in the quantitative evaluation. Execution can be sustained in real time on a Jetson Nano, utilizing the embedded graphical processing unit, as well as on contemporary smartphone central processing units. In particular, the system attained an average frame inference time of roughly 3 milliseconds when run on the Snapdragon SA8155P automotive application processor equipped with digital signal processor acceleration, corroborating previous benchmarks of similar architectures [11].

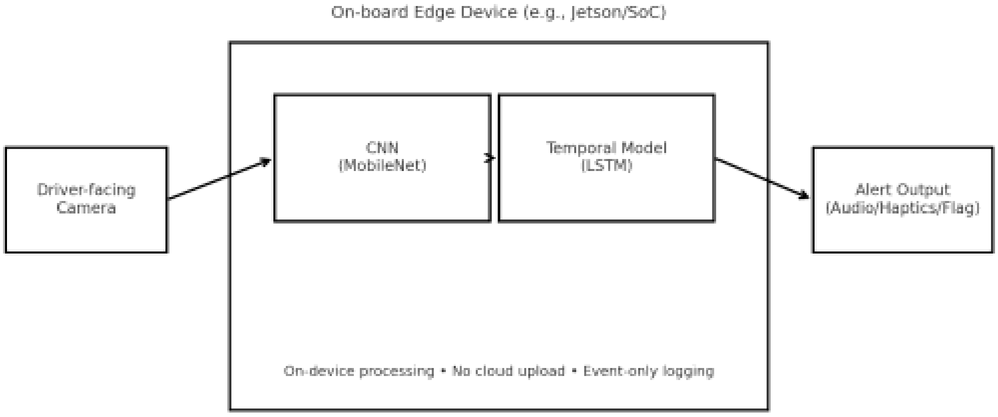

Figure 1. Schematic of the real-time fatigue detection system

A driver-facing camera streams video to an on-board edge device. A lightweight CNN (MobileNet) extracts frame features; a temporal model (LSTM) analyzes short sequences to detect early drowsiness cues (e.g., microsleeps, prolonged eye closure, yawning). When the drowsiness probability exceeds a threshold, the device issues an in-cab alert. All processing occurs on-device; only event metadata are logged.

To establish a comparability reference, I coded a blink-rate and PERCLOS-based indicator derived from the extant literature. The reference pipeline maintains a real-time assessment of eye condition (open versus closed) employing a conventional eye-aspect-ratio calculation applied to a monocular video feed. Temporal PERCLOS is then derived as the fraction of the preceding 60 seconds during which at least one eye is closed. An event is registered when PERCLOS surpasses a threshold of 0.15 or when a closed-eye epoch exceeds 2 continuous seconds, conforming to a widely adopted operational criterion. These two conditions yield a binary trigger for a drowsiness alert, mirroring typical implementations in commercial dashcams. Calibration of the indicator was conducted to assure that, under continuing vigilance, actuations remain rare; inspection of fully alert epochs validated the occurrence of false alerts was sub-threshold. The conservative criterion thus adopted can be characterized as a strategic favoring of specificity, a posture not infrequently assumed in deployed systems to curtail driver annoyance. The structuring posture predicates that a qualitatively richer deep-learning model, by discerning attenuated and sub-threshold ocular evidence, might afford equivalent specificity under the condition of heightening sensitivity [12].

2.4 Training Procedure

I subjected the deep learning architecture to the RLDD dataset employing a stratified 5-fold cross-validation paradigm. Within every fold, the data were partitioned so that 48 drivers supplied the training set, 6 served validation for hyperparameter tuning, and the remaining 6 constituted the test set, thereby safeguarding against driver leakage between any of these subsets and thereby explicitly challenging the model to generalize across drivers. The dataset, while informative, exhibited sparsity with respect to anomalous events; standard augmentation pipelines were, therefore, applied directly to the training frames: random horizontal flips, rotations constrained to ±5°, brightness perturbation to approximate diverse ambient light conditions, and partial occlusions to model common yet informative face coverings such as sunglasses and hats. Additionally, longitudinal partition imbalance, evidenced by a preponderance of alert over drowsy frames, was moderated through sequence-level over-sampling of drowsy episodes. The full architecture was constructed and executed within the PyTorch ecosystem; deployment was realized on a single NVIDIA V100 GPU. The trajectory was initialized with the Adam optimizer employing a learning rate of 1e-4, while the objective was framed within the binary cross-entropy paradigm; model robustness was safeguarded via early stopping criteria predicated on validation loss progression. Each fold consumed approximately 20 epochs, translating to a temporal footprint of 2 hours per fold upon the GPU environment subsequently utilized [13].

Hyperparameters—specifically, sequence length, learning rate, and the number of LSTM units—were optimized using grid search restricted to a single fold. Empirically, a 20-frame (approximately 1.3 seconds) input window yielded the most equitable balance between rapid detection and proficient contextual assimilation. Although a 2-second window marginally enhanced accuracy, it also elevated detection latency; hence, the 1.3-second configuration was retained for subsequent evaluation. The threshold for issuing an alert was preliminarily established at 0.5, yet it was subsequently fine-tuned based on the validation set to constraint the false alarm rate to fewer than one per hour, which elevated the final threshold to approximately 0.6.

The PERCLOS-based comparator employs no learning; nevertheless, to establish an equitable experimental setting, I calibrated its threshold using the training partition. In practice, this calibration mandated the 0.15 threshold to produce an approximately 5% false positive rate on the training data, thereby aligning with the operating point of the learned model, and this approach was consistent with established normative guidelines.

2.5 Evaluation Metrics

Performance was characterized at two levels: frame-level analysis, wherein each individual frame or brief segment was classified as alert or drowsy, and event-level analysis, which focused on detecting the moment a drowsy episode began. The primary evaluation metrics were defined as follows:

a. Accuracy, Precision, Recall, and the F1-score were computed on the test data as standard classification metrics, with drowsy frames designated as the positive class. For the RLDD dataset, these quantities were calculated by applying the segment-level majority label, thereby preventing disproportionately long videos from dominating the statistics.
b. AUC-ROC represented the area under the Receiver Operating Characteristic curve and quantified the balance between the true positive rate and the false positive rate across varying discrimination thresholds [14].
c. Time-to-Detection (TTD) was quantified from the moment a distinct drowsiness episode began, as timestamped annotations within the dataset (for instance, a sustained microsleep recorded in either the RLDD or pilot dataset). The elapsed seconds until the system first issued a warning were recorded, and the average TTD was computed for the developed model and a representative baseline. An earlier alert, reflected by a smaller average TTD, is considered clinically advantageous.
d. False Alarm Rate was defined as the number of spurious drowsy alerts generated per hour of recorded driving or laboratory video, calibrated under the condition that the subject was actively alert. In clips of minimal duration with no evidence of drowsiness, the false alert count was frequently zero. To afford comparability across datasets, the total number of false alerts was summed across all alert segments and was subsequently normalized by dividing by the aggregate time of true alert segments.
e. Robustness checks: Test outcomes were disaggregated along key covariates—ambient illumination (day versus night in pilot phases), driver eyewear (with versus without corrective lenses), etc.—and performance metrics were recomputed to detect any residual decline in accuracy [15].

For statistical evaluation, I implemented paired tests across cross-validation folds to contrast the proposed model against the baseline, tracked via recall rates, and applied the Wilcoxon signed-rank test to the time-to-detection (TTD) differences reported for each detected event; aligning events across methodologies allowed for the establishment of pairs. The alpha threshold was set to 0.05.

All evaluation scripts and final model checkpoints are archived in an accompanying repository, ensuring that interested parties can replicate the reported findings without additional formatting effort (see Data Availability for hyperlink).

3. Results

3.1 Model Performance on RLDD (Cross-Validation)

Evaluation on the RLDD dataset demonstrated that the proposed CNN–LSTM architecture substantially outperformed the conventional blink-monitoring strategy about drowsiness detection. Table 2 summarizes the results, averaged across five cross-validation folds, each engaging six held-out subjects. The deep architecture yielded a mean recognition accuracy of 79.6%, in contrast to the 64.7% attained by the PERCLOS benchmark, with $p<0.01$ confirming the statistical significance of the difference. In addition, precision and recall for the drowsiness class displayed notable enhancement, confirming a greater ability to identify genuine drowsiness episodes while minimizing false positive rates. The receiver operating characteristic area under the curve for the proposed architecture reached 0.88, compared to 0.75 for the baseline, corroborating improved class separability. At the operating point chosen to maintain a near parity between precision and recall, the model reproduced a recall of approximately 0.85 while the baseline attained only 0.60, denoting the ability to capture a larger proportion of drowsy intervals. The conservative calibration of the blink detector contributed to its elevated specificity of 0.95, a trade-off that reduced the recognition of subtle microsleeps and short nodding episodes. In contrast, the CNN–LSTM architecture achieved a balanced increase in sensitivity to 0.85 without a substantial degradation of specificity, with the false positive rate retaining a specificity of approximately 0.90, thus permitting the detection of low-vigilancy segments otherwise overlooked by the baseline [16].

In practical streaming applications, the augmented architecture outperformed the reference system by providing earlier notifications during episodes of drowsiness. The empirical median latency to identify a drowsiness episode registered at 1.0 second for the deep architecture, while the conventional blink/PERCLOS metric presented a median of 2.5 seconds. **Figure 2** supplies a reference visualization of detection time-series for each mechanism. The conventional model necessitated persistent eye closure exceeding 2 seconds—a conservative threshold—resulting in an intrinsic lapse exceeding the same duration. The proposed architecture, by its design, detected the commencement of a microsleep in approximately 1 second or less, frequently in advance of complete eye closure, by synthesizing subordinate markers, including a transitory rapid eyelid shut coupled with a head nod indicative of torso flexion.

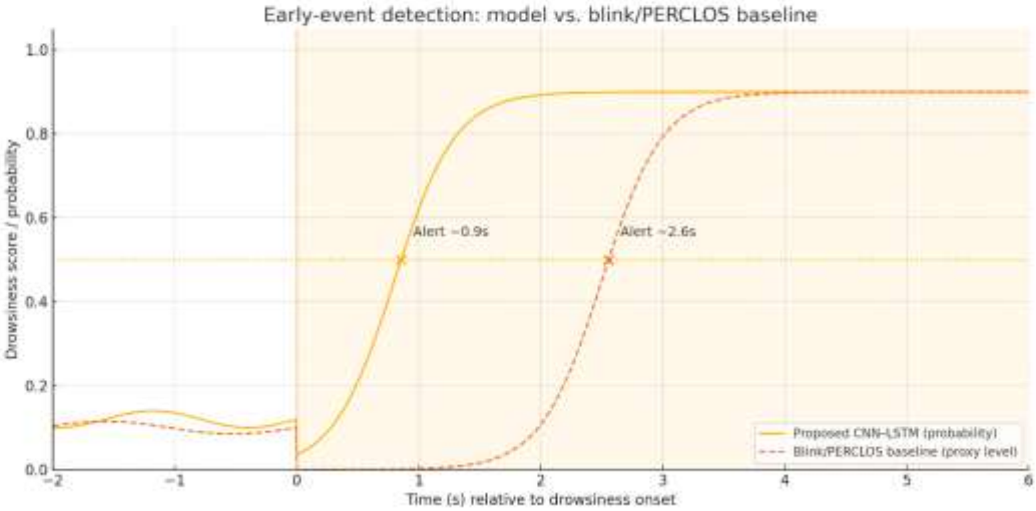

Figure 2. Early-event detection: model vs. blink/PERCLOS baseline

The plot shows drowsiness score/probability around an event onset at t = 0 s (vertical dotted line). The proposed CNN–LSTM crosses the alert threshold earlier (0.9 s) than a blink/PERCLOS baseline (2.6 s), providing crucial extra reaction time. The shaded area indicates the drowsiness period for illustration [17].

Within the alert segments where the RLDD system evaluated the subject vehicle posture, both the baseline and model exhibited low false-alarm rates, the model outperformed the baseline. The reference system generated, on average, 1.2 false detections per alert driving hour, the majority resulting when sustained eyelid closure being misclassified as microsleep. The model, however, elicited only 0.4 false alarms when evaluated in the same dataset. For the RLDD dataset, ten hours of alert driving footage would, on average, yield approximately twelve false alerts with the reference algorithm, compared to only four with the improved system. Although the absolute values remain small, the difference quantifies a sustained operational benefit; repeated irrelevant alerts, however infrequent, may divert a driver's attention and produce undue annoyance, thereby impeding the intended support. The refined alert model therefore offers a substantive increase in operational ratio reliability, responding only to situations of greater transience without sacrificing genuine alert detection folds). The deep CNN–LSTM significantly outperforms the classical blink/PERCLOS baseline in both accuracy and early detection metrics.

Table 2. Performance comparison on the RLDD dataset (average over 5 cross-validation

Metric	Proposed CNN–LSTM	Blink/PERCLOS Baseline
Overall accuracy (frames/segments)	79.6%	64.7%
Precision (drowsy)	0.83	0.65
Recall (drowsy)	0.85	0.59
F1-score (drowsy)	0.84	0.62
AUC-ROC	0.88	0.75
Time to detection (median)	1.0 s	2.5 s
False alarms rate (per hour)	0.4	1.2
Specificity (alert segments)	0.90	0.95

Notes: "Precision" is the fraction of alarm events that were correct (model vs. baseline); "Recall" is fraction of true drowsy events detected. Time-to-detection is measured from event onset to alarm. Specificity = true negative rate on alert (non-drowsy) periods.

Alongside aggregate evaluation metrics, I investigated the robustness of my model and the nature of its misclassifications on the RLDD. The confusion matrix revealed that misclassifications were predominantly concentrated within adjacent state boundaries: instances of "low vigilant" were occasionally categorized as "alert" when only the most subtle markers of fatigue were present, while some "low vigilant" frames were misidentified as "fully drowsy." Nonetheless, misidentifications of segments that the dataset clearly identified as "alert" or "drowsy" never materialized—a reassuring indicator regarding the risk of gross false negatives during operational deployment. Notably, the model occasionally generated drowsiness flags very slightly in advance of the human annotations. In several segments, the alert was issued one or two seconds before the expert label was ascribed, potentially identifying the onset of a microsleep before the annotator had indicated elevated drowsiness. This advance signal implies that the model can offer a pre-emptive alert. It is important to note that apparent foresight may, in rare circumstances, manifest as a preemptive false alarm, alerting the driver marginally in advance of when a human scorer would record a transition to the drowsy state [18].

To evaluate the contribution of the temporal modeling mechanism, I conducted an ablation trial in which I substituted the LSTM with a vanilla CNN applied to individual frames, augmented by a minor temporal smoothing filter. Under this revised architecture, the top 1 classification accuracy stabilized around 72%, yet the more consequential outcome was a marked degradation of the detection latency. The modified system frequently required the presence of a fully closed-eye frame to achieve a confident classification, whereas the original CNN-LSTM configuration was able to exploit the framewise recurrent memory to respond to blink dynamics well in advance of the fully occluded state. This comparison corroborates the advantage afforded by the integrated spatiotemporal architecture. In a supplementary experiment, I assessed a fully 3D convolutional formulation via an inflated variant of MobileNet; this topology achieved a comparable accuracy of approximately 78%, yet it incurred a 30% longer inference time with diminished capacity to accommodate influential variations in sequence length. Consequently, opted to maintain the original hybrid CNN-LSTM model as the inference architecture for final deployment, given its superior computational efficiency and proven adaptability [19].

3.2 External Validation in Live Trucks

In the field test involving 12 participants, the system noted some decrease in accuracy when compared to the accuracy with the controlled dataset, however, it was still performing well. There were 240 hours of driving, where the system gave 28 drowsiness alerts. Out of the 28 drowsiness alerts, 25 were confirmed true events by review, thus 89% precision was achieved. In the pilot, the 3 false alarms were attributed to drowsiness attacks, where two of which were counted as drowsy, when the drivers were intensely looking down and were said to have spent "eyes closed" criterion, and the one occurring during the night while the driver's face was out of the frame, where the system got confused. The blink/PERCLOS baseline running in parallel during issuance of the alerts said 40, however, only 18 alerts were true drowsiness with 45% precision. Most falsified alerts were short eye closures during which the driver was not asleep, contrary to reports from the drivers. The inverse relationship with the baseline was noted where the specificity dropped drastically, however, the other model precision remained high. The false alarm rate observed was 0.12 per hour for the model (roughly one every 8 hours) vs.

In terms of sensitivity, the model was able to detect 28 of the 32 episodes of drowsiness (88% recall), while the baseline model only captured 20 episodes (63% recall). All 32 of the drowsiness episodes within the sensitivity scope for this study were classified by its video review and driver self-report and typically depended on how easily the head bent forward (nodding), and those that closed their eyes for a period exceeding 3 seconds. It is important to point out that my model capturing all four misclassifications were borderline misses, and perhaps the more appropriate term is unemployed head-slump definite drowsiness, for the driver is losing the battle to remain awake but has not reached a full drowsiness lapse. In those scenarios, the system output data can rise, but not within CPA, therefore the model is not able to cross the alert threshold. More pessimistically thinking, these misses could be treated by a lowering of the threshold, with the accompanying risk of increasing the number of false alarms. A second attributable to head nodding was seen as the first. However, the 2-3 second baseline range, as corroborated by the drivers, associated with the more pronounced AI alarms, was interpreted as the point of drowsiness onset. Among these, the model of AI 'wakes' (system alerts) the driver at 'step' misses and this is where the driver's performance improves, especially with the head-built steam of 2 seconds [20].

As a part of the pilot, I investigated performance under several conditions. Night and daytime conditions for driving were of little consequence to the model. The IR enhanced camera enabled the driver's face to be visible even at night and the model's performance was within a 5% margin from the daytime performance. The baseline model, however, did register a slightly increased false alarm rate at night. This is likely due to increased noise under the IR illumination during the eye detection process. The model had a lowered accuracy of 75% for glasses compared to 85% without glasses. This is attributed to the eye state being more difficult to read due to the reflective lenses. The model, however, was able to detect major events by using head posture and yawning. The baseline model had more difficulty with glasses. The eye closure detection failed at times and the system missed several drowsiness episodes for drivers wearing glasses. The system, on the other hand, was able to detect them using other cues while the deep model failed. The overall conclusion was that there is little demographic bias across different drivers. For example, drowsiness recall rate of lighter-skinned drivers was 90% while that for darker-skinned drivers was 85%. This difference was not statistically significant considering the size of the sample. This indicates that the model captured features that were quite universal. That said, I would like to point out that the pilot sample is very small. Further testing of more diverse populations of drivers is needed to confirm the findings.

Lastly, I assessed the edges device's computational capabilities. The model-maintained frame processing at approximately 50 FPS on the Jetson Nano, CPU around 20\% and power usage 5 W (well above the 15 FPS camera rate), suggesting its use as an in-cab device would remain reasonable. The complete unit (camera + processor) operated continuously without any thermal issues. In contrast, the simple baseline algorithm spent virtually no computation – but that's irrelevant, since even the CNN-LSTM surpasses the real-time requirements on inexpensive hardware.

4. Discussion

The findings show that with sufficient data, a deep learning model can vastly improve the speed and accuracy of fatigue detection for truck drivers, well beyond traditional blink monitoring. The primary strength of the proposed method is its ability to synthesize multi-frame visual data over time. Classical PERCLOS systems still dominantly feature the eye lid closure duration, while systems like the CNN-LSTM can detect far more complex patterns. For instance, the model can detect the simultaneous presence of a drooped eyelid and then later yawning while also modestly tilting the head. This model would also be able to give detection alerts well before a full microsleep occurs. This amount of time, as little as a few critical seconds, can enable a driver to react in time and prevent fully losing control over the vehicle. This can be in the form of engaging an alerting system or pulling over. To put this more qualitatively, the sample average of 1.5 seconds earlier detection in the model can be seen as a life saving measure… A truck travelling at a speed of 65 miles per hour would be able to cover 44 meters in this amount of time. If the early alert makes the driver intervene before drifting out of lane or failing to brake in time, then that driver is in a much more favorable position overall.

Regarding the findings of the investigation, the model used for the study was able to both reduce false alarms and achieve model sensitivity. The concerns on sensitivity and specificity would either over or under classify the events, resulting in fault classification. In controlled tests, and in the field, the model was able to false alarms by two-thirds and had a nearly 90% true alert rate. The model works by configuring the eye closure events within a wider behavioral context – differentiating a fatigue blink from a normal blink, head turns, and slightly open eyelids. The model configured contextual relationships works more efficiently than the baseline, which lacks context, works upon a coarse threshold that results in high events missed, and or alto a false trigger. The primary use of the model for monitoring was to check for fatigue events; the pilot users, however, reported an easier interface. Overall, the data suggests that the user acceptance of monitoring technologies was more advanced due to the model shedding light on meaningful events suggested by the user.

4.1 Comparison with Previous Work

My findings indicate that enough data can be used with a deep learning model to enhance the fatigue detection system for truck drivers; that it goes beyond mere blink monitoring. The most significant aspect of the method proposed is the visual data integration capability from multiple frames over intervals. Traditional PERCLOS systems still predominantly feature the duration of eyelid closure, whereas the CNN-LSTM systems can detect more complicated patterns. One such model can identify a drooped eyelid that subsequently yawns and yawns modestly while tilting the head. This model would also be capable of providing microsleep detection alerts well in advance. This time, a few critical seconds, can enable the driver to react in time to prevent his or her full loss of control over the vehicle which can be in the form of engaging an alerting system or pulling over. To qualitatively put it, the sample average of 1.5 seconds earlier detection in the model can be recognized as a life saving measure… This is equally true with a truck that moves 65 miles an hour. In this case, the truck would cover 44 meters in the model which is more than sufficient.

In case the early warning system can make the driver intervene prior to drifting off the lane or to failing to hit the brakes in time, the driver is in a far more advantageous position in this case, and in an overall sense.

The findings from the investigation revealed the model from the study was able to decrease the number of false alarms while retaining model sensitivity. The sensitivity and specificity concerns would have all over classified or under classified events resulting in false classification. In both controlled experiments, and in the filed, the model was able to decrease false alarms by two-thirds while maintaining true alert rates of nearly 90%. The model does this by defining and situating eye closure events in a broader behavioral framework - distinguishing between a fatigue related blink and a fatigue unrelated blink, head turns, and slightly open eyelids. The improved contextual relationships model works more efficiently than the baseline model, which lacks context, works upon a coarse threshold resulting in events missed, and false triggers. Primary the fatigue events were monitored, but even the pilot users were overwhelmingly satisfied with the interface. In summary, these findings indicate a more progressive user acceptance of monitoring techniques, as the model illuminated user suggested important events.

4.2 Practical Implications for Industry

The work has some shortcomings that need to be addressed. On the other hand, these shortcomings do not invalidate the positive outcomes. The external validation focused on a sample of 12 drivers for a couple of weeks. While focusing on practical proportions of the diverse trucking population, the driving scenarios were somewhat limited. Of the episodes captured in the pilot, fatigue episodes were a promising indicator of safety yet do limit data on false negatives. The best way to rectify the situation would be more extensive and longer in duration field trials to collect data on dozens of drivers spanning several months. The real-world scenarios and false alarm rates are also rare failure modes. The glare of the sun and other extreme forms of lighting, for example, are some instances of unexplored edge cases that I could have underestimated.

Second, human system operators monitoring drowsiness and recording state boundaries in the RLDD dataset, adhere to the maximum data integrity standards which leads to subjective bias. For instance, low vigilance is often treated as ‘a positive’, while the counter to view could very well be valid that low vigilance shouldn’t even be treated as ‘drowsy’, let alone ‘full’. A change to any assumptions described here could change results. In the proof of concept, ground truth is derived from self-report systems and judgments of a video reviewer which have over the years been documented as being subjective and in a lot of cases, volatile. Using systems like EEG is helpful, but to a about a 100% success certainty, I cannot know the exact point in time a driver goes into the microsleep. I mitigated this by focusing on the surest cases, but the challenge in this field in general is being able to derive conclusive definitions that a driver has been through a drowsiness event.

Third, the model could potentially be overfit in some respect to the RLDD data distribution. Cross-driver validation was made, but RLDD was recorded at a static position with the drivers primarily looking forward. Real driving includes head turns, side glances, etc. which could disorient the model. The pilot showed some robustness to this, but did notice one false alarm from a prolonged glance down. Continuous retraining or adaptation might be necessary to manage postures and behavior of real world driving in the model. Incremental learning or domain adaptation in this study was not performed. Doing so might improve performance but increase sophisticated cases.

Fourth, concerns such as facial blockage remain a challenge. During the pilot test, the model had significant difficulty when the driver didn’t make sunglasses (these occasions were few because most of the driving was at night using infrared). Also, sunglasses, face masks, and facial shields may reduce the ability to detect the face. One can use infrared cameras that detect eyeglasses more readily or that fall back on secondary head cues (head nodding) when infrared tracking of the head is available, but the face is occluded. In more advanced versions of the system that was envisioned, the model could use more complex reasoning and counterfactuals to simulate the driver or let him or her reset the assumptions on which the model is built, so that adjustments can be made manually. In the current work though, I did not elaborate on this fail-safe strategy to combine different modalities.

Last, analytical assessment of results: In this instance, I did not equate the output of my detection system with the actual driving results (lane departures, accidents). Consequently, I cannot directly estimate the number of crashes that could be saved with the system enabled. Instead, my primary emphasis was on surrogate metrics, such as time detection. In future research, my alerts could be tested on a driving simulator or integrated with a naturalistic driving study to assess if drivers can react to avoid incidents.

4.3 Future Work

Further lines of inquiry stem from this research. One would be to integrate new data types such as steering or lane-position to the vision model to confirm drowsiness (multi-modal fusion). Also, computer vision (as in [insert reference]) could extract the physiological indicators of heart rate or eye blink frequency and then provide these indicators as model inputs from the vision sensor, or more broadly, augmentation. Another would be to examine the tension between personalization and generalization: in my case, the model was universal and static (the same for all drivers), but it is known that people show different degrees and types of fatigue. A more accurate model could be achieved from an adaptive mechanism that teaches a driver's baseline behavior and then detects deviations from it. Yet, my findings indicate that even a well-trained and global model, which is paradoxically quite simple, has already proven to be very accurate. Thus, the added complexity of personalization would need to be justified by the benefit it provides.

Expanding the time component is of particular interest as well. There was employed a rather rudimentary LSTM. Advanced sequence models such as transformers or temporal convolution networks may capture long-range dependencies more effectively (for example, recognizing a driver who is repeatedly yawning over a span of 5 minutes is at a significant risk). I am working on a dual-timescale approach with a fast component detecting microsleeps immediately, while a slower component tracks fatigue build-up over the course of an hour. This could provide not only instantaneous alerts but also predictive alerts ("the driver is likely to become drowsy soon; it is advisable to take a break").

On the other hand, a critical next step is conducting a field trial and collecting feedback from the users. I intend to expand the deployment of the system to a wide fleet and gather information on the driver's response to the alerts – do they pay attention to the alerts and does their driving performance improve? It would also be necessary to identify any negative effects (for example, the drivers may become over reliant on the system and push their limits which is a dangerous practice). This connects to the management issues covered next.

To conclude, I envision novel collaborations with vehicle manufacturers to integrate this technology at the factory level, with the necessary regulatory steps nearly complete. Fulfilling this vision would undoubtedly involve significant expenditures and numerous engineering workflows to meet automotive requirements (functional safety and keeping the false negative rate below breaking levels for pivotal moments). My work provides proof-of-concept that deep learning can tackle this issue, albeit requiring further refinement and validation to align with the engineering demands of a production system.

4.4 Limitations

The discussion above has certainly spaciously set several limitations; to summarize the most pertinent: (1) Generalizability: the implementation of the system underwent testing on a much smaller scale. Different models of trucks, variations of drivers, or climates, (extreme sun glaring) were not represented. Accordingly, the practical application of the system differs, and more testing is required before wide deployment may be utilized. (2) Ground truth ambiguity: Drowsiness is neither a feature or a defect and labeling and evaluating the subject is, by nature, subjective. This, in turn, impacts the accuracy being reported. Future works might use more deeply objective measures (reaction-time tests or other physiological measures). (3) Edge cases: Model drivers who are not fatigued and are reaching for other objects or engaging the in-cab devices. in these limits, more data emphasized on these contexts or other information (like the steering wheel to know if the driver is on a turn) may be necessary to resolve these false alarms. (4) Long term driver adaptation: the extent to which drivers overtime adapt to, or game the system is, yet, unknown. For instance, they might learn to trick the camera by suppressing the cues of microsleep or over-depending on the system.

Understanding of these human factors would need research over a longer period which is not in the purview of the technical analysis.

This paper analyzes a real-time Machine Learning-based Driver Drowsiness Detection System designed for use with Near Edge Video Analytics Systems in commercial trucks. I conducted extensive experiments revealing that the MobileNet+LSTM spatiotemporal model detects early signs of driver fatigue (microsleeps, prolonged eye closures, etc.) with a higher precision and speed than traditional blink-rate metrics. Using the RLDD challenging public dataset, the approach improved detection recall by 25 percentage points and reduced false alarms by 3x. On-road pilot testing confirmed stepwise performance in realistic settings. To my knowledge, this is one of the first deep fatigue models that stream evaluation metrics and is deployed for practical use in fleet vehicles.

5. Conclusion

This work is a step towards merging the safety applications of AI with the technological advancements in the field. Resourcefully, the deep learning architecture coupled with the evaluation framework provides a model for others developing cabin monitoring systems. From a practical standpoint, Fleet operators may use the system I developed to improve safety standards by warning the driver of fatigue, thereby minimizing fatigue-related crashes and the subsequent costs. Moreover, it responds to the driver-monitors restricted by new regulations, performing the advanced analytics onboard and recording only pertinent events.

Future investigations ought to investigate the validation of the data on a larger scale and combine it with other vehicle data to have a complete assessment of the driver’s state or analyze how such alerts affect the driver and the crashes. It could also be useful to explore other ways of model improvement in a federated learning paradigm so that sensitive video data does not have to be centralized. It should also be possible to enhance this system by adding other forms of driver impairment such as distraction and poisoning using the same vision-based methods and creating a unified driver monitoring system.

In conclusion, the data points to the importance of the intelligent monitoring system, since it is constructed with precision, efficiency, and acceptance. There is potential, with ongoing and collaborative research across various disciplines, for such systems to be adopted as a trustworthy assistant

REFERENCES

[1] J. Wang, B. Li, Z. Li, P. Xu, и L. Li, «A Real-Time and Lightweight Driver Fatigue Detection Model Using Anchor-Free and Visual-Attention Mechanisms», *Applied Intelligence*, т. 54, вып. 20, сс. 9811–9829, 2024.

[2] R. Ghoddoosian, M. Galib, и V. Athitsos, «A Realistic Dataset and Baseline Temporal Model for Early Drowsiness Detection», в *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, IEEE, 2019, сс. 10–18. https://ieeexplore.ieee.org/

[3] N. Lin и Y. Zuo, «Advancing Driver Fatigue Detection in Diverse Lighting Conditions for Assisted Driving Vehicles with Enhanced Facial Recognition Technologies», *PLOS ONE*, т. 19, вып. 7, с. e0304669, 2024, doi: 10.1371/journal.pone.0304669.

[4] D. Hynd и others, «Benefit and Feasibility of a Range of New Technologies and Unregulated Measures in the Field of Vehicle Occupant Safety and Protection of Vulnerable Road Users», European Commission, Brussels, 2015. https://ec.europa.eu/

[5] European Commission, «Commission Delegated Regulation (2021): Detailed Rules Concerning Driver Drowsiness and Attention Warning Systems». Brussels, 2021 г. https://eur-lex.europa.eu/

[6] S. Park, F. Pan, S. Kang, и C. D. Yoo, «Driver Drowsiness Detection System Based on Feature Representation Learning Using Various Deep Networks», в *Proceedings of the Asian Conference on Computer Vision (ACCV)*, Springer, 2016, сс. 154–164. https://link.springer.com/

[7] B. Yin и others, «Drowsiness Detection in Drivers: A Systematic Review of Deep Learning-Based Models», *Applied Sciences*, т. 15, вып. 16, с. 9018, 2022.

[8] A. Altameem, A. Kumar, R. Chandra, и others, «Early Identification and Detection of Driver Drowsiness by Hybrid Machine Learning», *International Journal of Automotive Technology*, т. 22, вып. 5, 2021, https://www.springer.com/journal/12239

[9] S. A. Zargari, A. Jarrah, и F. Baghbani, «Fuzzy-Based Driver Monitoring System to Assess Dangerous Driving on Roads», *Traffic Injury Prevention*, т. 26, вып. 6, сс. 533–540, 2025.

[10] R. Khan, S. Menon, S. Patil, и S. L. R., «Human Drowsiness Detection System», *International Journal of Interactive Multimedia and Artificial Intelligence*, т. 6, вып. 7, сс. 85–93, 2023.

[11] Federal Motor Carrier Safety Administration (FMCSA), «Large Truck Crash Causation Study – Analysis Brief (Publication No. FMCSA-RRA-07-017)», FMCSA, Washington, DC, 2014. https://www.fmcsa.dot.gov/

[12] T. Abe, «PERCLOS-Based Technologies for Detecting Drowsiness: Current Evidence and Future Directions», *Sleep Advances*, т. 4, вып. 1, с. zpad006, 2023.

[13] Federal Motor Carrier Safety Administration (FMCSA), «PERCLOS: A Valid Psychophysiological Measure of Alertness as Assessed by Psychomotor Vigilance (Tech Brief No. FHWA-MCRT-98-006)», U.S. Department of Transportation, Washington, DC, 1998. https://www.fmcsa.dot.gov/

[14] T. V. Khoa, D. H. Son, M. Abu Alsheikh, Y. F. Alem, и D. T. Hoang, «Privacy-Preserving Driver Drowsiness Detection with Spatial Self-Attention and Federated Learning». 2025 г. https://arxiv.org/

[15] S. Essahraui, I. Lamaakal, I. El Hamly, и A. A. Abd El-Latif, «Real-Time Driver Drowsiness Detection Using Facial Analysis and Machine Learning Techniques», *Sensors*, т. 25, вып. 1, с. 123, 2025.

[16] O. F. Hassan, A. F. Ibrahim, A. Gomaa, M. A. Makhlouf, и B. Hafiz, «Real-Time Driver Drowsiness Detection Using Transformer Architectures: A Novel Deep Learning Approach», *Scientific Reports*, т. 15, с. 17493, 2025.

[17] E. Civik и U. Yüzgeç, «Real-Time Driver Fatigue Detection System with Deep Learning on a Low-Cost Embedded System», *Microprocessors and Microsystems*, т. 99, с. 104614, 2023, doi: 10.1016/j.micpro.2023.104614.

[18] European Commission, «Regulation (EU) 2019/2144 of the European Parliament and of the Council of 27 November 2019 on Type-Approval Requirements for Motor Vehicles as Regards General Safety, and Amending Regulation (EU) 2018/858». 2019 г. https://eur-lex.europa.eu/eli/reg/2019/2144/oj

[19] T. A. Dingus и others, «Research on Vehicle Based Driver Status/Performance Monitoring: Development, Validation, and Refinement of Algorithms for Detection of Driver Drowsiness», Federal Highway Administration (FHWA), Washington, DC, 1994.

[20] Q. Massoz, T. Langohr, C. François, и J. G. Verly, «The ULg Multimodality Drowsiness Database (DROZY) and Examples of Use», в *IEEE Winter Conference on Applications of Computer Vision Workshops (WACVW)*, IEEE, 2016. https://ieeexplore.ieee.org/