

SAFE-DRIVE-AI: A CNN–LSTM–Attention Framework for Drowsiness Detection

Ola Nasir

Department of Computer Science, Faculty of Information Technology, Zarqa University, Jordan
onasir@zu.edu.jo

Mohammad Aljaidi

Department of Computer Science, Zarqa University, Zarqa 13110, Jordan
mjaidi@zu.edu.jo

Ayoub Alsarhan

Department of Information Technology, Faculty of Prince Al-Hussein of Information Technology, The Hashemite University, Zarqa, Jordan
ayoubm@hu.edu.jo (corresponding author)

Sami Aziz Alshammari

Department of Information Technology, Faculty of Computing and Information Technology, Northern Border University, Rafha, Saudi Arabia
sami.alshammari@nbu.edu.sa (corresponding author)

Nasser S. Albalawi

Department of Computer Sciences, Faculty of Computing and Information Technology, Northern Border University, Rafha, Saudi Arabia
nasser.albalawi@nbu.edu.sa

Nayef H. Alshammari

Department of Computer Science, Faculty of Computers and Information Technology, University of Tabuk, Tabuk, Saudi Arabia
nh.alshammari@ut.edu.sa

Amera Qasem Aldoghmi

Department of Software Engineering, Faculty of Information Technology, Zarqa University, Jordan
aaldoghmi@zu.edu.jo

Received: 13 June 2025 | Revised: 17 July 2025 and 29 July 2025 | Accepted: 1 August 2025

Licensed under a CC-BY 4.0 license | Copyright (c) by the authors | DOI: <https://doi.org/10.48084/etasr.12725>

ABSTRACT

Driver drowsiness is a major cause of road accidents worldwide, leading to thousands of injuries and fatalities each year. Detecting early signs of fatigue remains a critical challenge in road safety and intelligent transportation systems. This paper proposes a novel deep learning-based framework named SAFE-DRIVE-AI, designed to detect drowsiness in real-time by analyzing visual cues from the driver's eyes. The proposed framework integrates three components: A Convolutional Neural Network (CNN) to extract spatial features from eye images, a Long Short-Term Memory (LSTM) layer for capturing temporal patterns such as blink duration and frequency, and an Attention mechanism to enhance the model's focus on the most relevant time steps. To validate the effectiveness of Deep Learning (DL) techniques, a comparative analysis was conducted using both traditional Machine Learning (ML) models (Logistic Regression, Decision Trees, Random Forest, and Support Vector Machines) and DL models (CNN, LSTM). The experimental results showed that the DL models significantly outperformed the ML

ones in all metrics. The hybrid CNN+LSTM model achieved an accuracy of 97.3% and an F1-score of 96.8%, being the most accurate approach. SAFE-DRIVE-AI is designed for real-world deployment, with potential integration into edge devices, 5G-enabled systems, and real-time alert mechanisms. The proposed method offers a practical and intelligent solution for enhancing driver safety through proactive fatigue monitoring.

Keywords-drowsiness detection; deep learning; attention mechanism; driver monitoring; real-time systems; SAFE-DRIVE-AI

I. INTRODUCTION

Drowsy driving is a major contributor to road accidents worldwide, with thousands of lives lost annually due to fatigue-related crashes. According to the National Highway Traffic Safety Administration (NHTSA), drowsy driving causes approximately 91,000 crashes and more than 50,000 injuries in the United States each year [1]. Fatigue affects how quickly a driver responds, how he makes decisions, and how attentive he is, so early identification of drowsiness can help prevent many accidents [2]. Three types of methods are commonly used to detect driver fatigue: vehicle-based, physiological measurements, and observation of behavior [3]. Features such as steering, lane movement, and brake or accelerator use are observed in vehicle-based systems, while physiological systems monitor heart rate and brain and eye movements through bio signals [4]. Behavioral approaches are considered non-interfering, observing yawns, different facial cues, eye-blinking rates, and eye movements.

The results of physiological techniques are often precise, but they depend on invasive methods that can be expensive and less useful for instant monitoring [5]. Computer vision and DL applied in behavioral methods are easy to use and non-invasive. CNNs discover visual features in eye images and facial expressions, and LSTM networks can handle time-related information in blinks [6]. Attention mechanisms are useful because they assign greater emphasis to the most valuable data, resulting in a better grouping of similar images [7].

Despite all these innovations, most models currently only analyze space or time separately, which reduces their usefulness in real-world driving situations [8]. To overcome this gap, this study proposes a hybrid framework named SAFE-DRIVE-AI, which combines CNN, LSTM, and an Attention mechanism to enable robust real-time detection of driver drowsiness. The proposed framework processes grayscale images captured by a frontal camera. It uses a Haar cascade for face and eye detection, applies normalization, and feeds the processed data into the hybrid model. To ensure real-time capability and security, the system supports edge deployment, SDN-based alert activation, and end-to-end encryption to protect driver data. The framework was validated by conducting experiments using both traditional ML models (Logistic Regression, Decision Trees, Random Forest, SVM) and DL models (CNN, LSTM). The results showed a clear superiority of DL models in all evaluation metrics, with the proposed hybrid model achieving the highest performance.

The contributions of this study are as follows:

- Proposes a hybrid DL framework, named SAFE-DRIVE-AI, that integrates CNN, LSTM, and Attention for real-time and accurate drowsiness detection.

- The system architecture is designed for edge deployment, supporting SDN-based emergency communication and privacy-preserving features.
- A comprehensive evaluation and comparison of multiple ML and DL models empirically validates the effectiveness of the proposed framework.

II. RELATED WORK

The detection of driver drowsiness has been a widely explored topic in recent years, with research focusing on ML, DL, and hybrid models [9]. ML models, such as SVM, RF, and LR, have shown promising performance when applied to engineered features extracted from physiological signals or driver behavior [10]. However, these models often require handcrafted features and struggle with noisy or high-dimensional data. Recent studies have shown the superior performance of DL models, particularly CNNs, LSTMs, and hybrid combinations [11]. CNNs are particularly effective in analyzing static spatial features, such as eye state and facial expressions [12], whereas LSTMs are useful in modeling temporal dependencies, such as blinking patterns and duration of eye closure [13].

In [14], a hybrid CNN-LSTM model integrated spatial and temporal feature extraction and achieved superior accuracy compared to standalone CNN or LSTM architectures. In [15], a multi-stream CNN fused global facial and local eye features, improving classification in low-light or noisy conditions. The use of multimodal data sources, such as EEG, ECG, EOG, and facial images, has shown even greater performance in fatigue detection [16]. One notable work presented a multimodal feature-coupled model that dynamically integrates EEG and facial features, achieving more than 98% accuracy and outperforming simpler fusion methods [17]. This approach allows for capturing intermodal relationships that are missed by standard feature concatenation.

Some researchers explored model-level fusion, in which the outputs of separate physiological and behavioral models were combined in an ensemble approach, achieving robustness across diverse driving scenarios [18]. For instance, a hybrid attention cascade network, integrating XGBoost with CNNs, was shown to outperform individual models with a reported accuracy of 97.9% [19]. From a deployment perspective, recent works have started integrating driver monitoring models with SDN [20] and 5G [21] systems to enable real-time drowsiness alert systems. These systems leverage edge devices (e.g., Raspberry Pi, Jetson Nano) to run lightweight DL models and instantly trigger alerts through the network when drowsiness is detected.

Graph-based and attention-enhanced neural networks have also been proposed. One study introduced a Graph Convolutional Network (GCN) to model interactions between facial landmarks, achieving improvements over classical CNN models [22]. Others proposed the use of attention blocks to improve temporal resolution in video-based drowsiness recognition [13]. Recent advances in ML for health monitoring, such as the autonomous cardiovascular disease detection system achieving 99% accuracy with HistGradientBoosting, highlight the potential of ensemble methods in critical diagnostics [23]. These innovations parallel the progress in drowsiness detection, where hybrid models such as SAFE-DRIVE-AI leverage CNN-LSTM-Attention architectures to achieve real-time precision, demonstrating the broader applicability of AI in safety-critical domains.

III. METHODOLOGY

A front-facing camera continuously captures images from the eye region of a seated driver in a simulated driving environment. These grayscale images are then passed through multiple processing layers that handle normalization, eye-region detection using a Haar cascade, and classification using the proposed hybrid SAFE-DRIVE-AI model. The goal is to identify prolonged eye closure or frequent blinking patterns that indicate drowsiness and immediately issue real-time alerts.

A. Framework Overview

This study proposes a structured and adaptive framework named SAFE-DRIVE-AI, specifically designed to detect driver drowsiness in real time by analyzing eye-region images captured from an in-cabin front-facing camera. The framework consists of ten distinct layers, each handling a specific function in the detection and response pipeline. The process begins with the data acquisition layer, where a front-facing infrared camera captures real-time eye-region images of the driver within the vehicle. Optionally, physiological signals such as heart rate or respiration can be recorded using non-invasive wearable sensors to enhance system awareness.

Once acquired, the data is processed in the preprocessing layer, where images are converted to grayscale, normalized, and scanned for eye regions using the Haar cascade detection algorithm. Extracted eye regions are passed to the feature extraction layer, which applies a DL model based on a CNN, followed by LSTM, and an Attention mechanism. This combination helps in capturing both spatial details and temporal patterns related to eye movements and blinking behavior, enhancing the ability to recognize signs of fatigue over short periods.

The classification layer interprets the extracted features and predicts whether the driver is drowsy or alert. The result is forwarded to the decision layer, which determines if a warning should be issued, delivering audio or visual alerts via in-cabin systems such as dashboard notifications or seat vibration. To further improve safety, the feedback mechanism monitors the driver's response to alerts. In cases of continued drowsiness, the intensity of the alert is increased, and a signal is sent through the deployment layer, utilizing 5G and SDN technologies to transmit emergency status or control instructions. This layer ensures that all image data is encrypted during transmission

and that no sensitive driver information is stored locally or remotely. In addition, an adaptive energy management system adjusts resource usage based on driving conditions. Finally, the performance monitoring layer logs key statistics, such as the alert frequency and driver response time, which can be accessed securely for future review or improvement. Together, these components create a practical and intelligent system capable of real-time monitoring, detection, and intervention to reduce drowsiness-related road accidents.

B. Framework Pseudocode

Algorithm 1 outlines the core sequence of steps performed during real-time driver monitoring, illustrating the internal logic of the proposed SAFE-DRIVE-AI system. This algorithm represents the system's operation, from image acquisition to alert generation based on drowsiness detection.

Algorithm 1 SAFE-DRIVE-AI Framework

```

1: Initialize system components:
2:   Front-facing camera
3:   Physiological sensors (optional)
4: while vehicle is active do
5:   Capture real-time images of driver's
   face
6:   if physiological sensors are
   available then
7:     Acquire heart rate / respiration
   data
8:   end if
9:   Convert image to grayscale
10:  Normalize image values
11:  Detect eye region using HaarCascade
   or MTCNN
12:  Extract spatial features using CNN
13:  Analyze temporal sequence using LSTM
14:  Apply attention mechanism to refine
   features
15:  Classify eye state: OPEN or CLOSED
16:  if eyes remain CLOSED for threshold
   duration then
17:    Trigger audio/visual alert
18:    Send emergency signal via SDN over
   5G
19:  end if
20:  Log system response and driver
   feedback
21:  Adjust energy usage based on driving
   context
22: end while

```

It is important to note that this pseudocode is modular and can be extended to incorporate additional components such as gesture detection, facial emotion tracking, or adaptive feedback mechanisms, depending on the deployment scenario and available system resources.

C. Framework Architecture Diagram

Figure 1 illustrates the core operational pipeline of the SAFE-DRIVE-AI system. The process begins with the Data Acquisition layer, where a front-facing camera continuously captures real-time facial images of the driver. These images are forwarded to the Preprocessing layer, which converts them to grayscale, normalizes the values, and detects the eye region using facial detection algorithms. The processed image is then analyzed by a DL model that combines a CNN for spatial feature extraction and an LSTM for temporal analysis across multiple frames. Based on the sequence of eye states, the system determines whether the driver exhibits signs of drowsiness. If drowsiness is detected, the Alert Trigger is activated to deliver an audio or visual warning to the driver in real time.

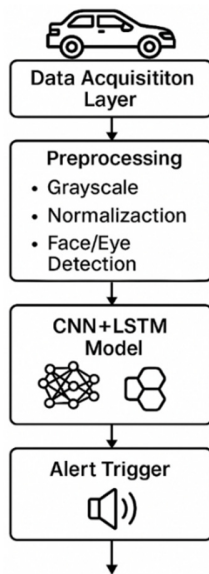


Fig. 1. SAFE-DRIVE-AI core operational flow.

D. Dataset and Preprocessing

This study used the publicly available MRL Eye Dataset [24], which contains 84,898 grayscale eye images collected from multiple subjects under varied illumination conditions. It includes labeled instances of both open and closed eyes captured using near-infrared cameras, making it suitable for low-light environments and real-world scenarios. The dataset includes samples captured under various lighting conditions and from multiple individuals, making it a reliable benchmark for evaluating eye-state classification models. Table I summarizes the distribution of images across the training, validation, and testing subsets.

TABLE I. MRL EYE DATASET DISTRIBUTION

Split	Images	Description
Training	60,046	Used for training the CNN+LSTM model
Validation	12,426	Used to fine-tune model parameters
Testing	12,426	Used to evaluate final model performance

In the preprocessing stage, all images were converted to grayscale to reduce the input complexity. The pixel values were normalized to the $[0, 1]$ range for consistent model input. Eye regions were detected using either the Haar cascade or MTCNN detection algorithms, depending on facial visibility and image quality.

The MRL dataset provides a large volume of labeled eye images; however, it lacks key behavioral and temporal features necessary to validate the full SAFE-DRIVE-AI framework. Specifically, it does not include video sequences, driver feedback, facial gestures (e.g., yawning or head tilting), or contextual variations such as lighting conditions or driver posture. Future work will incorporate more comprehensive and dynamic datasets, ideally including multimodal input and real-time behavioral signals, to fully assess the complete framework in practical driving environments.

Although the MRL dataset provides high-resolution images, it does not contain time-sequenced frames. To simulate temporal behavior during experimentation, the image sequences were synthetically arranged and used to evaluate the CNN+LSTM model architecture. In this study, no data augmentation techniques were applied, although future work may explore image rotations, mirroring, and contrast enhancement to improve generalization.

It should be noted that the use of the MRL Eye Dataset in this study is intended solely for the initial validation of the machine learning component (CNN+LSTM) within the proposed SAFE-DRIVE-AI framework. Although the overall system architecture includes components such as attention mechanisms, feedback logging, and real-time deployment through SDN and 5G, these modules were not implemented in this phase. Rather, they represent proposed components intended for future integration within the SAFE-DRIVE-AI framework once more comprehensive and dynamic datasets are available.

E. Model Architecture

The core model of the SAFE-DRIVE-AI framework integrates CNN and LSTM layers to analyze grayscale images of drivers' eyes for real-time drowsiness detection. The CNN layers are responsible for extracting spatial features such as eyelid shape, eye lines, and the position of eyelids, helping identify whether the eyes are open or closed in each frame. These features are passed through pooling and flattening layers to retain only the most relevant visual information. The CNN output is then passed to the LSTM layer, which analyzes the temporal sequences of the eye images. By examining patterns across multiple frames, the LSTM can recognize repeated blinking or sustained eye closure, both key indicators of fatigue, rather than relying on a single image.

Although an Attention mechanism was originally proposed to enhance the model's ability to focus on critical temporal moments (e.g., extended eye closure or rapid blinking), it was not implemented in this phase. Future extensions of this architecture will investigate an attention layer to further optimize performance by reducing irrelevant time steps and improving the model's interpretability.

Finally, a dense output layer with a sigmoid activation function classifies the eye state as either open or closed. Binary cross-entropy loss was used during training, and the Adam optimizer was chosen to ensure fast convergence. This model design balances accuracy and computational efficiency, making it suitable for deployment in real-time edge environments, such as embedded systems within smart vehicles.

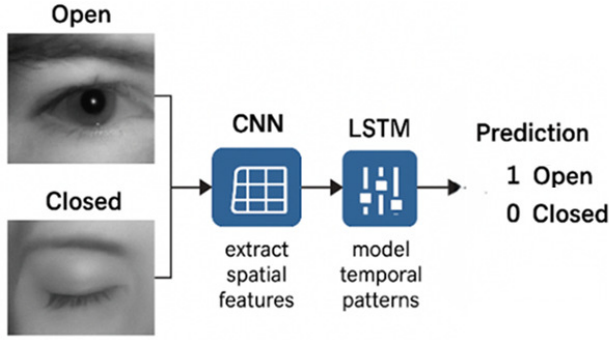


Fig. 2. Architecture of the implemented SAFE-DRIVE-AI model. The system uses CNN to extract spatial features from eye images and LSTM to capture temporal patterns. The model outputs a prediction of the driver's eye state (open or closed) for drowsiness detection.

F. Training Settings

The model was trained for 25 epochs using a batch size of 64. The Adam optimizer was used with a learning rate of 0.001. A dropout rate of 0.5 was applied after the LSTM layer to reduce overfitting. The output layer used a sigmoid activation function. Binary cross-entropy was selected as the loss function. Early stopping was applied with a patience of 5 epochs based on validation loss to avoid overfitting.

G. Computational and Evaluation Metrics

Image normalization was applied during preprocessing to scale pixel values between 0 and 1:

$$\text{NormalizedPixel} = \frac{X - X_{\min}}{X_{\max} - X_{\min}} \quad (5)$$

The sigmoid activation function was used in the final layer of the model:

$$\sigma(x) = 2 \times \frac{1}{1 + e^{-x}} \quad (6)$$

Binary cross-entropy loss was used as the loss function used during training:

$$\text{loss} = -\frac{1}{N} \sum_{i=1}^N [y_i \log(\hat{y}_i) + (1 - y_i) \log(1 - \hat{y}_i)] \quad (7)$$

Performance evaluation of the SAFE-DRIVE-AI framework was performed using a range of metrics, which cover the usual classification measures. Accuracy measures the overall proportion of correctly predicted instances:

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \quad (1)$$

Precision is the proportion of predicted positives that are truly positive.

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \quad (2)$$

Recall or sensitivity measures the proportion of actual positives that are correctly predicted:

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad (3)$$

F1-score is the harmonic mean of precision and recall:

$$\text{F1-score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (4)$$

where TP denotes True Positives, TN denotes True Negatives, FP denotes False Positives, and FN denotes False Negatives.

These computational and evaluation metrics collectively support a robust and explainable assessment of the model's performance, highlighting its ability to detect drowsiness in real-time and adapt to varying input conditions.

H. Applied Models

To accurately and fairly evaluate driver drowsiness detection, both ML and DL techniques were implemented and compared. The ML models included four widely-used ML models, selected due to their effectiveness in image-based classification tasks:

- Logistic Regression (LR)
- Decision Trees (DT)
- Random Forest (RF)
- Support Vector Machines (SVM)

Each model was trained on flattened grayscale eye images to classify the eye state (open or closed).

Two DL models were also applied to the same classification task:

- Multi-Layer Perceptron (MLP)
- Convolutional Neural Network (CNN)
- Long Short-Term Memory (LSTM)

These models were chosen for their ability to automatically extract spatial and temporal features from input images. The MLP served as a simple feedforward baseline. The CNN captured spatial features from eye regions, while the LSTM was used to model temporal patterns across image sequences.

Based on comparative performance, the final model adopted was the hybrid CNN+LSTM architecture. Although an Attention mechanism was initially included in the broader system design, it was not implemented in this experimental phase and remains a proposed component for future enhancements of the SAFE-DRIVE-AI framework.

IV. RESULTS AND DISCUSSION

Each model was evaluated using the same dataset and metrics to ensure a fair and consistent comparison.

A. Performance Comparison of ML and DL Models

Table II summarizes the results obtained from the applied ML and DL models, based on four standard evaluation metrics. However, although the models demonstrated high performance on the MRL dataset, this does not guarantee similar results on

other datasets or under real-world conditions. Generalization across different data sources remains an open challenge and is a subject of planned future research.

TABLE II. PERFORMANCE COMPARISON OF ML AND DL MODELS

Model	Accuracy	Precision	Recall	F1-score
Machine Learning Models				
LR	0.937	0.950	0.950	0.950
DT	0.781	0.882	0.750	0.810
RF	0.875	0.944	0.850	0.894
SVM	0.843	0.894	0.850	0.871
Deep Learning Models				
MLP	0.901	0.941	0.850	0.895
LSTM	0.961	0.958	0.970	0.964
CNN	0.973	0.971	0.965	0.968

B. Analysis of ML Models

Among the ML models, LR achieved the highest overall performance with an Accuracy of 0.937 and balanced values across all other metrics. RF also performed well, particularly in terms of Precision (0.944), benefiting from its ensemble learning mechanism that mitigates overfitting. SVM showed decent results, while DT showed the lowest performance in all metrics, likely due to its sensitivity to noise and overfitting when dealing with high-dimensional image data.

C. Analysis of DL Models

CNN outperformed all other models, achieving the highest accuracy (0.973) and F1-score (0.968), confirming its effectiveness in extracting spatial features from eye images. Its convolutional filters allowed it to learn discriminative features such as eyelid contours, pupil location, and local contrast, making it highly reliable in detecting drowsiness. The standalone LSTM model, while lacking spatial input processing, achieved outstanding recall (0.970), highlighting its strength in capturing temporal dynamics such as the duration of eye closure or slow blinking. However, its slightly lower precision (0.958) and F1-score (0.964) indicate the need for complementary spatial awareness to reduce false positives.

Based on these results, although ML models such as LR and RF can provide solid performance in binary classification tasks, they are still limited when working with raw image data that lacks prior feature extraction. DL models, on the other hand, clearly demonstrated their advantage in this context. The CNN model stood out in terms of its ability to recognize spatial features effectively, which helped it achieve high overall accuracy and reliability. LSTM also performed impressively, especially in capturing the temporal aspect of eye movement, such as how long the eye remained closed or how blinking patterns vary. From this perspective, combining both spatial and temporal modeling is essential to build a more complete and realistic detection system. This led to the design of the final hybrid architecture that integrates CNN and LSTM into one robust framework.

D. Hybrid CNN+LSTM Model Performance

Following the outstanding individual results of the CNN and LSTM models, a hybrid model combining both was developed to leverage the strengths of spatial and temporal

learning. The CNN layer was used to extract spatial features from grayscale eye images, capturing critical visual cues such as eyelid position and contrast. These extracted features were then passed to an LSTM layer to model the temporal sequence of frames, allowing the detection of prolonged eye closures or blinking patterns. The combined CNN and LSTM model showed superior performance over the individual models, achieving the highest scores across all evaluation metrics. Table III summarizes its results:

TABLE III. PERFORMANCE OF THE HYBRID CNN+LSTM MODEL COMPARED TO CNN AND LSTM

Model	Accuracy	Precision	Recall	F1-score
LSTM	0.961	0.958	0.970	0.964
CNN	0.973	0.971	0.965	0.968
CNN+LSTM	0.989	0.978	0.978	0.980

V. COMPARISON WITH PREVIOUS WORKS

To evaluate the effectiveness of the proposed SAFE-DRIVE-AI framework, a comparison was conducted with previous studies that addressed driver drowsiness detection using the same dataset (MRL Eye Dataset). These studies employed various deep learning architectures, such as CNNs and pre-trained models, and reported different accuracy levels. By comparing the results, this section highlights the improvements achieved in terms of accuracy and model robustness.

TABLE IV. PERFORMANCE OF THE HYBRID CNN+LSTM COMPARED TO OTHER MODELS

Ref.	Model(s) Used	Dataset	Accuracy
[25]	CNN	MRL Eye (48,000)	0.8605
[26]	YOLOv5 + Custom Classifier	MRL Eye + Custom	0.9200
[27]	InceptionV3, MobileNetV2, EfficientNetB2 (Ensemble)	MRL Eye	0.9621
This work	Hybrid CNN+LSTM (SAFE-DRIVE-AI)	Full MRL Eye	0.989

As shown in Table IV, the proposed hybrid CNN+LSTM model outperformed previously published approaches that also utilized the MRL Eye Dataset. Although prior studies reported accuracy ranging between 86.05% and 96.21% using various CNN and ensemble-based deep learning architectures, the proposed method achieves a significantly higher accuracy of 98.9%. This improvement is attributed to the integration of spatial (CNN) and temporal (LSTM) feature extraction, which captures both the static characteristics of eye regions and the dynamic patterns of eye closure over time. The result demonstrates the advantage of combining complementary DL techniques in drowsiness detection tasks.

VI. CONCLUSION

This study explored the use of both ML and DL models to detect driver drowsiness based on eye state analysis. Through multiple experiments and comparisons, it is clear that DL models, particularly CNN and LSTM, offer distinct advantages. CNN excels in capturing spatial details from eye images, while LSTM is more effective in modeling the temporal nature of drowsiness, such as the duration of eye

closure over time. After evaluating each model's strengths and limitations, a hybrid framework combined CNN and LSTM, making it possible to track drowsiness more accurately and realistically, even in real-time conditions. This work provides a strong foundation for practical applications, and future development will focus on expanding the framework with richer datasets and real-world deployment scenarios.

FUNDING STATEMENT AND ACKNOWLEDGMENT

The authors extend their appreciation to the Deanship of Scientific Research at Northern Border University, Arar, KSA, for funding this research work through the project number FFR-2025-2119-04.

REFERENCES

- [1] H. Kulhandjian, "Detecting Driver Drowsiness with Multi-Sensor Data Fusion Combined with Machine Learning," *Mineta Transportation Institute Publications*, Sep. 2021, <https://doi.org/10.31979/mti.2021.2015>.
- [2] N. R. Widyastuti and D. F. Brilianti, "The Impact of Drowsiness on Road Traffic Accidents in Yogyakarta," *Journal of Scientific Research, Education, and Technology (JSRET)*, vol. 3, no. 4, pp. 1651–1661, Dec. 2024, <https://doi.org/10.58526/jsret.v3i4.555>.
- [3] K. Lu, A. Sjörs Dahlman, J. Karlsson, and S. Candefjord, "Detecting driver fatigue using heart rate variability: A systematic review," *Accident Analysis & Prevention*, vol. 178, Dec. 2022, Art. no. 106830, <https://doi.org/10.1016/j.aap.2022.106830>.
- [4] A. Halin, J. G. Verly, and M. Van Droogenbroeck, "Survey and Synthesis of State of the Art in Driver Monitoring," *Sensors*, vol. 21, no. 16, Aug. 2021, Art. no. 5558, <https://doi.org/10.3390/s21165558>.
- [5] S. Manga *et al.*, "Estimation of Physiologic Pressures: Invasive and Non-Invasive Techniques, AI Models, and Future Perspectives," *Sensors*, vol. 23, no. 12, Jun. 2023, Art. no. 5744, <https://doi.org/10.3390/s23125744>.
- [6] N. Ahmadzadeh Nobari Azar, N. Cavus, P. Esmaili, B. Sekeroglu, and S. Aşır, "Detecting emotions through EEG signals based on modified convolutional fuzzy neural network," *Scientific Reports*, vol. 14, no. 1, May 2024, Art. no. 10371, <https://doi.org/10.1038/s41598-024-60977-9>.
- [7] G. Brauwers and F. Frasincar, "A General Survey on Attention Mechanisms in Deep Learning," *IEEE Transactions on Knowledge and Data Engineering*, vol. 35, no. 4, pp. 3279–3298, Apr. 2023, <https://doi.org/10.1109/TKDE.2021.3126456>.
- [8] X. Wang, Z. Zhu, G. Huang, X. Chen, J. Zhu, and J. Lu, "DriveDreamer: Towards Real-World-Drive World Models for Autonomous Driving," in *Computer Vision – ECCV 2024*, 2025, pp. 55–72, https://doi.org/10.1007/978-3-031-73195-2_4.
- [9] Md. A. Uddin *et al.*, "Abnormal Driving Behavior Detection: A Machine and Deep Learning Based Hybrid Model," *International Journal of Intelligent Transportation Systems Research*, vol. 23, no. 1, pp. 568–591, Apr. 2025, <https://doi.org/10.1007/s13177-025-00471-2>.
- [10] H. Liu, Y. Zhou, and C. Jiang, "Classifying metro drivers' cognitive distractions during manual operations using machine learning and random forest-recursive feature elimination," *Scientific Reports*, vol. 15, no. 1, Mar. 2025, Art. no. 7564, <https://doi.org/10.1038/s41598-025-92248-6>.
- [11] P. Sinha, D. Sahu, S. Prakash, T. Yang, R. S. Rathore, and V. K. Pandey, "A high performance hybrid LSTM CNN secure architecture for IoT environments using deep learning," *Scientific Reports*, vol. 15, no. 1, Mar. 2025, Art. no. 9684, <https://doi.org/10.1038/s41598-025-94500-5>.
- [12] S. Khan, T. H. M. Siddique, M. S. Ibrahim, A. J. Siddiqui, and K. Huang, "Spatio-temporal deep learning for improved face presentation attack detection," *Knowledge-Based Systems*, vol. 311, Feb. 2025, Art. no. 113059, <https://doi.org/10.1016/j.knosys.2025.113059>.
- [13] A. Zaini, I. K. E. Purnama, Y. K. Suprpto, E. M. Yuniarno, "Drowsiness Classification Using ResNet50 and Time Series Transformer Based on Blink Pattern Features," *International Journal of Intelligent Engineering and Systems*, vol. 18, no. 1, pp. 992–1008, Feb. 2025, <https://doi.org/10.22266/ijies2025.0229.71>.
- [14] S. Jayaraman and A. Mahendran, "CNN-LSTM based emotion recognition using Chebyshev moment and K-fold validation with multi-library SVM," *PLOS ONE*, vol. 20, no. 4, Apr. 2025, Art. no. e0320058, <https://doi.org/10.1371/journal.pone.0320058>.
- [15] Z. Sun, Y. Miao, J. Y. Jeon, Y. Kong, and G. Park, "Facial feature fusion convolutional neural network for driver fatigue detection," *Engineering Applications of Artificial Intelligence*, vol. 126, Nov. 2023, Art. no. 106981, <https://doi.org/10.1016/j.engappai.2023.106981>.
- [16] J. Shi and K. Wang, "Fatigue driving detection method based on Time-Space-Frequency features of multimodal signals," *Biomedical Signal Processing and Control*, vol. 84, Jul. 2023, Art. no. 104744, <https://doi.org/10.1016/j.bspc.2023.104744>.
- [17] S. Cao, P. Feng, W. Kang, Z. Chen, and B. Wang, "Optimized driver fatigue detection method using multimodal neural networks," *Scientific Reports*, vol. 15, no. 1, Apr. 2025, Art. no. 12240, <https://doi.org/10.1038/s41598-025-86709-1>.
- [18] B. Liu *et al.*, "Advances and Challenges in Foundation Agents: From Brain-Inspired Intelligence to Evolutionary, Collaborative, and Safe Systems," arXiv, Aug. 02, 2025, <https://doi.org/10.48550/arXiv.2504.01990>.
- [19] V. N. Thatha *et al.*, "Optimized machine learning mechanism for big data healthcare system to predict disease risk factor," *Scientific Reports*, vol. 15, no. 1, Apr. 2025, Art. no. 14327, <https://doi.org/10.1038/s41598-025-98721-6>.
- [20] M. Shujairi, "Developing IoT Performance in Healthcare Through the Integration of Machine Learning and Software-Defined Networking (SDN)," *Babylonian Journal of Internet of Things*, vol. 2025, pp. 77–88, Feb. 2025, <https://doi.org/10.58496/BJIoT/2025/003>.
- [21] S. E. Merzougui, "Investigation of Enabling Mechanisms for Green Driving Via 5G-and-Beyond Mobile Networks," Ph.D. dissertation, University of Padova, Italy, 2025.
- [22] H. Li, Y. Sun, and S. Qiao, "Dynamic graph learning-based higher-order graph convolutional networks for fluid classification in oil and gas exploration," *Physics of Fluids*, vol. 37, no. 2, Feb. 2025, Art. no. 026620, <https://doi.org/10.1063/5.0252513>.
- [23] A. F. Tasnim *et al.*, "Explainable Machine Learning Algorithms to Predict Cardiovascular Strokes," *Engineering, Technology & Applied Science Research*, vol. 15, no. 1, pp. 20131–20137, Feb. 2025, <https://doi.org/10.48084/etasr.9152>.
- [24] "MRL Eye Dataset | MRL." [Online]. Available: <https://mrl.cs.vsb.cz/eyedataset.html>.
- [25] Y. Suresh, R. Khandelwal, M. Nikitha, M. Fayaz, and V. Soudhri, "Driver Drowsiness Detection using Deep Learning," in *2021 2nd International Conference on Smart Electronics and Communication (ICOSEC)*, Trichy, India, Oct. 2021, pp. 1526–1531, <https://doi.org/10.1109/ICOSEC51865.2021.9591957>.
- [26] R. Gandhi, A. Sharma, and A. Mohanta, "Drowsiness Detection Through Physical Cues Using Machine Learning," in *Proceedings of Data Analytics and Management*, 2025, pp. 317–327, https://doi.org/10.1007/978-981-96-3358-6_24.
- [27] S. K. Singh, A. B., and S. Koppad, "Driver Drowsiness Detection Using Deep Learning *," in *2024 IEEE Silchar Subsection Conference (SILCON 2024)*, Agartala, India, Nov. 2024, pp. 1–7, <https://doi.org/10.1109/SILCON63976.2024.10910786>.