International Journal of Computational and Experimental Science and ENgineering (IJCESEN)

Vol. 11-No.4 (2025) pp. 10068-10078
http://www.ijcesen.com

ISSN: 2149-9144

Copyright © IJCESEN

Research Article

Temporal Convolutional Network Approach for Video-Based Driver Drowsiness Detection

G N Sharath Kumar[1]*, Hanumanthappa J[2], Chethan. Raj. C[3,*], P Naveen Kumar[4]

[1] Research Scholar, Department of Studies in Computer Science, University of Mysore (UOM), Mysore, India
Email: sharugn@gmail.com

[2] Research Supervisor, Department of Studies in Computer Science, University of Mysore (UOM), Mysore, India
Email: hanumsbe@gmail.com - **ORCID:** 0000-0002-7020-7598

[3] Associate Professor, Department of Computer Science & Engineering, CCE, VTU, Mysore, India
*** Corresponding Author Email:** drchethanrajc@gmail.com- **ORCID:** 0000-0002-8235-5109

[4] Department of Electronics Yuvaraja's College, University of Mysore (UOM), Mysore, India
Email: naveenp.yuvaraja@gmail.com

Article Info:

DOI: 10.22399/ijcesen.4609
Received : 20 September 2025
Revised : 18 October 2025
Accepted : 26 November 2025

Keywords

NTHUDDD dataset,
CNN,
TCN,
Keyframe.

Abstract:

Driver drowsiness remains a critical factor in road accidents worldwide, necessitating robust real-time detection systems for enhanced road safety. This research proposes a novel approach for drowsy driver detection by leveraging Temporal Convolutional Networks (TCN) to analyse sequential video frames and capture temporal dependencies in driver behaviour. Traditional computer vision methods often rely on static frame analysis, missing crucial temporal patterns indicative of drowsiness progression. Our methodology extracts spatiotemporal features from video sequences using a TCN architecture, enabling continuous monitoring of eye closure duration, blink frequency, head pose variation, and yawning patterns over time. The system employs preprocessing techniques including keyframe selection, face alignment and landmark detection analysis to ensure robustness under varying illumination and driving conditions. Experimental evaluation on benchmark datasets (NTHU-DDD) demonstrates that TCN approach achieves 96.7% detection accuracy with a false positive rate of 3.2%, outperforming conventional CNN and LSTM-based methods by 4.3% and 2.8% respectively. The proposed model achieves real-time processing at 30 frames per second on standard Dataset. This research contributes to intelligent transportation systems by providing a computationally efficient solution that accurately captures the temporal evolution of drowsiness symptoms, offering significant potential for reducing drowsy-related accidents through early warning systems.The paper proposed the different scenarios that include multi approach like day time, nigh time, driver with bare eyes, with spectacles and with sunglass.

1. Introduction

This paper presents a Temporal Convolutional Neural Network (TCN) based approach for detecting drowsy drivers from video sequences. By analysing temporal dependencies across consecutive frames, the proposed model captures progressive drowsiness indicators including prolonged eye closure, reduced blink rate, and head movement patterns. The architecture integrates spatial feature extraction with temporal convolution layers to model behavioural sequences, overcoming limitations of frame-based methods. Implemented with face alignment and landmark detection preprocessing, the system demonstrates 96.7% accuracy on benchmark datasets while processing at 30 FPS. Experimental results show superior performance over existing methods, with 4.3% improvement over conventional CNNs. The solution offers practical deployment potential for real-time vehicular safety systems, contributing to reduced fatigue-related accidents through timely drowsiness alerts.

Road traffic accidents represent one of the leading causes of mortality worldwide, with the World Health Organization reporting approximately 1.19

million fatalities annually. Among various contributing factors, driver fatigue and drowsiness account for an estimated 20-30% of all road accidents, resulting in substantial socio-economic losses exceeding $100 billion globally each year. The insidious nature of drowsiness—characterized by progressive impairment of cognitive functions, delayed reaction times, and microsleep episodes—makes it particularly dangerous as drivers often remain unaware of their deteriorating alertness until critical moments have passed.

The automotive industry's transition toward advanced driver-assistance systems (ADAS) and autonomous vehicles has intensified research into robust safety mechanisms. While contemporary vehicles incorporate numerous collision avoidance technologies, comprehensive fatigue detection remains a challenging frontier.

2. Related Works

Recent research in driver drowsiness detection has evolved significantly, focusing on deep learning and real-time vision-based systems to improve accuracy and robustness under varied conditions. Conventional Convolutional Neural Networks (CNNs) remain widely used for feature extraction from facial and ocular regions, as demonstrated in studies by Ranjan et al. [3] and Ahmed et al. [9], which achieved strong performance in controlled settings. However, challenges persist in dynamic environments, particularly under low-light conditions or when drivers wear accessories like spectacles or sunglasses. To address these issues, several works have integrated attention mechanisms and transformer architectures. For instance, Shukla and Sahu [1] employed a CNN with a transformer attention module to enhance feature relevance, while Jarndal et al. [4] developed a real-time vision transformer system aimed at improved generalization across lighting variations. Similarly, Yang et al. [8] optimized the use of key facial features in video-based detection to maintain performance under occlusion.

Efforts to enhance real-time applicability and efficiency are also prominent. Jebraeily et al. [5] utilized genetic algorithms to optimize CNN architectures, balancing accuracy and computational load. Others, such as Essahraui et al. [6] and Albadawi et al. [10], implemented machine learning and deep learning models for real-time drowsiness detection using visual features, though their evaluations often lack comprehensive testing in adverse scenarios like nighttime with eyewear. Recurrent architectures, like the LSTM with CNN features proposed by Kaur [2], have been applied to model temporal dynamics in video sequences, capturing drowsiness progression over time. Additionally, transfer learning approaches, such as that by Madni et al. [7], leverage pre-trained models on eye movement behavior to reduce data requirements and improve detection in resource-constrained settings. Recent works by Kumar et al. [11, 12] further explore deep feature extraction and key frame selection to enhance video-based detection reliability.

Despite these advances, a consistent gap identified across many studies is the limited robust evaluation under challenging real-world conditions—such as nighttime driving with spectacles—where performance degradation is most pronounced, as also observed in the provided confusion matrix analysis. Many state-of-the-art methods report high metrics on curated datasets but exhibit decreased recall in suboptimal environments, highlighting the need for more diverse training data and robustness-focused model designs.

Shukla et.al[1], proposes a driver drowsiness detection system that integrates a Convolutional Neural Network (CNN) with the attention mechanism of a Transformer model. The authors use a multi-stage pipeline, beginning with MediaPipe Face Mesh for facial landmark extraction, followed by deep-learning-based feature analysis to classify drowsiness states. The hybrid CNN-Transformer approach aims to improve feature representation, temporal attention, and overall detection accuracy, addressing the limitations of traditional CNN-only models in capturing subtle facial cues associated with fatigue.

The author [2] focuses on building a vision-based system that detects driver drowsiness by analysing video frames. The core idea is to use a Convolutional Neural Network (CNN) to extract spatial facial features from each frame and then feed these features into a Long Short-Term Memory (LSTM) network to capture temporal patterns such as eye-closure duration, blinking frequency, and facial fatigue cues. The author aims to overcome the limitations of traditional CNN-only models by incorporating temporal modelling, which is essential for detecting gradual drowsiness.

The author proposes a deep learning–based driver drowsiness detection system that uses transfer learning [3] on facial images to classify whether a driver is drowsy or alert. The authors evaluate several state-of-the-art CNN architectures to determine which model performs best for this task. The study uses the Driver Drowsiness Dataset (DDD) from Kaggle, which contains labelled images of drivers in drowsy and non-drowsy states. The goal is to build an efficient, accurate, and real-time-capable system that can be integrated into

intelligent transportation systems to reduce road accidents caused by fatigue.

The author [4] implements the real-time driver drowsiness detection system built using Vision Transformer (ViT) architectures, aiming to improve both accuracy and response time compared to traditional CNN-based systems. The authors highlight the increasing global concern around drowsy driving and propose a system that can be integrated into intelligent vehicle safety frameworks. The paper does not cover real-time deployment due to GPU acceleration and edge-optimized models

The author [5] proposes a Genetic Algorithm–optimized Convolutional Neural Network (GA-CNN) for driver drowsiness detection. Instead of manually designing a CNN architecture, the authors use a Genetic Algorithm (GA) to automatically search for the best CNN configuration—number of layers, filter sizes, activation functions, pooling types, and dense layers.

The author [6] addresses the critical road safety issue of drowsy driving by proposing a non-intrusive, real-time detection system. A comprehensive evaluation of machine learning and computer vision techniques was carried out using the NTHUDDD, YawDD, and UTA-RLDD datasets. The results indicate that traditional machine learning models, particularly the KNN classifier, performed competitively. Among deep learning–based vision approaches, YOLOv5 and YOLOv8 outperformed Faster R-CNN, demonstrating superior detection accuracy and reliability. These findings highlight the effectiveness of KNN and YOLO-based models for real-time drowsy driver detection and their potential to enhance road safety.

The author [7] focuses on non-intrusive driver drowsiness detection using eye movement imagery, avoiding the use of physiological sensors. It introduces a hybrid VGLG approach that combines the VGG-16 deep learning model for spatial feature extraction with a Light Gradient Boosting Machine to generate transferable features from eye movement data. Experimental results show that the proposed VGG16-LGBM method, when combined with a k-nearest neighbour classifier, outperforms existing state-of-the-art approaches while maintaining low computational complexity, making it suitable for real-time drowsiness detection applications.

The author [8] proposes a video-based driver drowsiness detection method, VBFLLFA, that addresses limitations of existing systems by effectively exploiting key facial cues. The approach focuses on head movements from facial landmarks and eye and mouth dynamics from local facial regions to extract relevant drowsiness information while reducing interference from irrelevant features. The model integrates Common Spatial Pattern–based spatial filtering and a two-branch multi-head attention mechanism to capture spatial and temporal characteristics. Additionally, a canter loss with distance constraint is employed to enhance class separability between alert and drowsy states. Validation on public datasets and a newly developed dataset demonstrates the effectiveness of the proposed method for accurate real-time drowsiness detection.

The author [9] presents a vision-based driver drowsiness detection framework aimed at improving road safety by identifying fatigue-related behaviors. The approach analyzes eyeball movement using a dedicated Convolutional Neural Network and concurrently classifies facial sleepiness cues—such as eye state and yawning—through a hybrid CNN–VGG16 architecture. The system was evaluated on a diverse image dataset encompassing variations in subject demographics, head pose, and illumination. Experimental results demonstrate that the custom CNN consistently outperforms the VGG16-based model across all evaluation metrics, indicating superior feature representation and classification capability. The findings establish the proposed CNN-driven approach as a more effective and reliable solution for driver drowsiness detection compared to conventional deep learning models reported in existing literature.

The author [10] presents a non-intrusive, real-time driver drowsiness detection system aimed at mitigating fatigue-related road accidents. The approach processes dashboard camera video to localize facial landmarks and face mesh points, enabling accurate identification of critical facial regions. Key visual features, including eye aspect ratio, mouth aspect ratio, and head pose, are extracted as robust indicators of driver drowsiness. These features are classified using Random Forest, Sequential Neural Network, and Linear Support Vector Machine models. Experimental evaluation on the National Tsing Hua University drowsiness dataset demonstrates that the proposed system achieves high detection accuracy and reliably triggers timely alerts, confirming its effectiveness for real-time driver monitoring applications.

The motivation is to overcome the limitations of manually designed CNNs, which may not be optimal and require extensive trial-and-error. The GA-optimized CNN aims to improve accuracy, efficiency, and generalization for detecting drowsiness from eye-state images.TCNs are chosen because they capture long-term temporal

dependencies efficiently. Unlike CNNs, which lack temporal modelling, and RNN/LSTM, which suffer from sequential computation and training instability, TCNs use dilated causal convolutions to achieve stable training, parallel processing, and effective temporal feature learning, making them suitable for real-time drowsy driver detection.

The development of robust driver drowsiness detection systems has been a focal point in the intersection of computer vision, machine learning, and automotive safety research. The performance metrics and confusion matrices provided—which show high precision but variable recall, especially under challenging conditions like nighttime with spectacles—situate the current work within a rich and evolving field. This survey synthesizes key methodologies and advancements from recent literature, presented in a thematic and chronological sequence to trace the progression towards more resilient systems.

Early and foundational approaches heavily relied on Convolutional Neural Networks (CNNs) for spatial feature extraction from driver facial imagery. Works such as that by Ahmed et al. [9] and Albadawi et al. [10] demonstrated the efficacy of standard CNN architectures in achieving high accuracy for drowsiness detection in well-controlled, daytime environments. These studies established baseline performance but primarily reported aggregate metrics like overall accuracy, often overlooking detailed breakdowns of performance degradation in sub-scenarios like low-light or with occlusions. Similarly, Ranjan et al. [3] proposed an efficient deep learning technique that optimized CNN structures for speed and accuracy, yet their evaluation, like many contemporaries, did not deeply segment results by challenging real-world factors, leaving a gap in understanding model robustness akin to the "Spectacles in night" performance drop observed in the provided analysis.

Recognizing the limitations of standard CNNs in focusing on critical facial regions (e.g., eyes, mouth) amidst clutter or variation, researchers integrated attention mechanisms. A significant step was the application of transformer models, known for their self-attention capabilities in natural language processing, to visual tasks. Shukla and Sahu [1] explicitly combined CNNs with an attention mechanism derived from transformer models to weight informative facial features more heavily, reporting improved performance. This aligns with the need to maintain precision under diverse conditions, as high precision is a strength of the provided results. Building on this, Jarndal et al. [4] developed a full Vision Transformer (ViT)-based system for real-time detection, claiming enhanced generalization. Their work in 2025 represents a move towards architectures that can dynamically prioritize relevant features, a necessity for handling variables like spectacles which can obscure key ocular cues.

To balance the high computational demand of advanced models with the need for real-time, in-vehicle processing, research turned to optimization and hybridization. Jebraeily et al. [5] employed a Genetic Algorithm to automatically optimize CNN architecture hyperparameters, seeking a Pareto-optimal solution between accuracy and computational efficiency. This approach is crucial for practical deployment but may still face challenges in inherently difficult scenarios if the training data lacks sufficient diversity. Furthermore, Yang et al. [8] focused on the optimized utilization of key facial features in video streams, developing methods to extract and track the most discriminative indicators of drowsiness. Their video-based method, published in 2024, provides a framework that could be extended to better handle temporal sequences under adverse conditions, potentially mitigating the high false-negative rate seen at night.

The push for deployable systems is evident in works emphasizing real-time performance. Essahraui et al. [6] and Albadawi et al. [10] both presented systems using facial analysis and machine learning for real-time detection. However, a common thread in these and similar studies is that while they achieve high performance on their test sets, the "real-time" evaluation often occurs in relatively constrained environments. The provided confusion matrices underscore a critical oversight in many existing systems: a lack of rigorous, scenario-specific validation (e.g., bare eyes vs. spectacles at night) that reveals stark performance disparities, a gap that future real-time systems must address through more comprehensive testing protocols.

Drowsiness is a temporal phenomenon, prompting the use of models that capture sequential dynamics. Kaur [2] proposed an LSTM (Long Short-Term Memory) network to analyze sequences of CNN-extracted features from video, capturing the progression from alertness to drowsiness. This temporal approach is vital for reducing false positives and understanding drowsiness onset but must be robust to frame-wise errors caused by occlusions or lighting. If the base CNN feature extractor fails on individual frames due to glare on spectacles at night—as suggested by the high false negatives in that scenario—the benefits of the temporal model can be severely compromised.

To address data scarcity and improve learning efficiency, transfer learning approaches have been

adopted. Madni et al. [7] presented a novel transfer learning approach specifically for eye movement behaviour, leveraging knowledge from related domains. This method is promising for specializing in ocular cues but may still struggle when those cues are partially obscured. Concurrently, research has expanded beyond static images to video-based analytics. Kumar et al. [11] and later Kumar and Kumar [12] focused on detecting drowsiness from video by selecting key frames and extracting deep features. Their work on key frame selection is particularly relevant for efficiency, but the robustness of their deep features across the full spectrum of challenging conditions remains an area for further investigation, as indicated by the performance variations in the provided data.

3. Existing System Methodology

tegration of CNNs with Transformer attention significantly increases model complexity and requirement of GPU. Dependence on High-Quality Facial Input

1. High Computational Cost due to CNN + LSTM models and Transfer Learning Dependency
2. Dataset Limitations due to Different lighting conditions, Diverse driver demographics, Real-world driving environments
3. Due to lighting & environmental sensitivity due to light-time driving, sudden illumination changes and Motion blur.
4. High accuracy due to genetic algorithm that explores CNN configurations, however requires heavy computation that uses more CNN. Fails if eyes information lost.

Dataset: National Tsuing Hua University Drowsy Driver Detection (NTHUDDD) [13] Dataset consisting of 36 individuals with different ethnicities recorded in different scenarios (i.e, Day time with bare eyes, spectacles and sunglasses and night time bare eyes, spectacles) driving dataset. It has frame size of 640 X 480 with .avi videos.

4. Proposed System & Architecture

The drowsiness detection system begins by capturing video frames of a person's face, typically from a camera mounted in a vehicle. Each frame is cropped and resized to focus on the facial region, ensuring consistent input dimensions for the neural network. These processed images are then passed through MobileNetV2[14], a lightweight convolutional neural network optimized for real-time applications. MobileNetV2 extracts high-level visual features and outputs a feature vector in $\mathbb{R}^{1280}$. To reduce computational complexity and enhance temporal modeling, this vector is projected into a lower-dimensional space. The resulting sequence of feature vectors is fed into a Temporal Convolutional Network (TCN)[15], which analyzes patterns over time—such as prolonged eye closure or head nodding—to detect signs of drowsiness. The TCN produces a scalar output in $\mathbb{R}$, which is interpreted as a prediction indicating whether the individual is drowsy or not. This pipeline combines spatial and temporal analysis to enable accurate, real-time drowsiness detection.

General Architecture: initially Videos are converted into frames. Frames are then preprocessed for noise cancellation, keyframes are selected based on eyes status (i.e, if the eyes status is closed then it checks consucative frames otherwise it skips 2 frames) and mouth status (i.e, if the mouth status is closed then it checks consucative frames otherwise it skips 2 frames). Face region from the keyframes are detected, cropped and resized to (224,224). The pretrained MobileNetV2 used to extract 1280 and retained the same 1280 and projected to 1 dimensional feature vector. A batch of 45 consucutive feature vector is fed into trained Temporal Convolutional network to predict the current frame is Drowsy or not.

5. Methodology

Videos are converted into frames,keyframes selected,they are resized, normalized, face detected using mediapipe and cropped the face region from each keyframeframe to make the input suitable for mobilenet processing. Passed each preprocessed frame independently through MobileNetV2 to extract a feature vector that represents facial characterstics of that frame. Extracted feature vectors are stored in sequence so that each video is converted into a time-ordered feature representation. Generating eye-closure signals and mouth opening signals using EAR and MAR values. Analyse the closure of eyes signals and mouth opening signals over a tiem to find continuous segments without any interruption. Label a time segments as drowsy if the eyes are remains closed for consucutive 45 frames or mouth opening for 20 consucutive frames that corresponds to 1.5 seconds and 0.6 seconds time duration respectively otherwise non drowsy.

The temporal samples divided into training sets and testing sets while ensuring that the samples belongs to the same video are not mixed across sets. The extracted featuresequences are trained using temporal convolutional network and their correspomding labels while keeping MobileNetV2 fine tuned as required. Unseen test samples are tested for evaluation of the model to verify that it correctly differentiate the drowsy behavior and

short eye blinks. Performing video inferencing by running the trained model on complete videos using a sliding window and record the predicted drowsiness probability over time. The drowsy is identified by the trained model only when consecutive predictions indicate sustained drowsiness and compute the evaluation metrics such as accuracy, precision and recall, f1-score, and confusion matrix to quantify the system's performance.

Rules followed during experimentation with TCN. Eye Blink is identified by analysing 3 to 6 consucutive frames, micro-sleep is detected by analysing 15 to 30 consucutive frames and drowsy event is identified by observing atleast 45 consucutive frames. TCN learns patterns across the time span.

The cropped image is fed into MobileNetV2 that extracts 1280 features eq(1), after projection D becomes feature vector in one Dimension. Eq(2) shows 1 feature vector, and sequence of 45 consucutive frames shown in eq(3). A sequence of 45 images with a batch size eq(4) is inputed to TCN. Eq(5) shows the learning element for drowsy and non-drowsy label. High level features are extracted by MobileNet are shown eq(6), padding is applied on the left side of the frames to maintain consistency if the frames are not upto 45. Dilation helps in looking history of past frames eq(8) with Reciptive field eq(9) to analyze the event belongs to drowsy or not and finally drowsy event is identified by the current frame with past consucutive frames as shown in the eq(10).

$$g: R^{H \times W \times 3} \rightarrow R^{D} \quad \text{(1)}$$

$$f_t = g(frame_t) \quad \text{(2)}$$

$$X = [f_{t-44}, f_{t-44}, .., f_1] \in R^{D \times 45} \quad \text{(3)}$$

$$X \in R^{B \times D \times 45} \quad \text{(4)}$$

$$f : R^{D \times 45} \rightarrow \{0,1\} \quad \text{(5)}$$

$$y(t) = \sum_{c=1}^{D} \sum_{k=0}^{2} W_{c,k} \, . X_c(t - k.d) + b \quad \text{(6)}$$

$$\text{Padding} = (K-1).d \quad \text{(7)}$$

$$d = [1,2,4,8,16] \quad \text{(8)}$$

$$RF = 1 + (K-1) * d \quad \text{(9)}$$

$$Y = F(x) + x \quad \text{(10)}$$

6. Results & Discussions

The proposed method performs drowsy driver detection in different scenarios, namely driving with bare eyes day and night time, with spectacles in day and night time, and with sunglasses in day time.

The model demonstrates highly successful learning dynamics, achieving effective accuracy with a training accuracy and a strong validation. The learning process shows a rapid initial improvement followed by a stable convergence. A consistent but minimal gap between the curves indicates excellent generalization without significant overfitting. Given the model's exceptional results, future efforts should focus on error analysis of the remaining misclassifications and potential data-centric refinements rather than extended training.

In the training phase, video is converted into frames, out of them keyframes are selected based on closed eyes, partial eyes and open eyes. If the eyes status is open or partial open/close then 2 frames are skipped then selecting third frame, otherwise next frame is selected if the eye status is closed. Face segmentation is performed on keyframe then MobileNetV2 extracts 1280 features from cropped image and it is converted into 1 dimensional vector then a batch of 30 frames are fed into TCN along with batch size. A sequential window of size 45 is fixed for detection of drowsy state of the driver. TCN is trained with drowsy and nondrowsy sequences. Finally at the testing phase unseen segment is passed and the trained TCN decides whether the given video segment belongs to drowsy or non-drowsy.

The experiment is performed on NTHUDDD dataset. The dataset split to 80:20 ratio and used 30 fps with sequential window of 45. Fig 4 shows Bare eyes day time training and validaton accuracy and Fig 5 shows Bare eyes day time training and validaton loss

Fig 6 shows Bare eyes night time training and validaton accuracy and Fig 7 shows Bare eyes night time training and validaton loss Fig 8 shows Drivers with spectacles at day time training and validaton accuracy and Fig 9 shows Drivers with spectacles at day time training and validaton loss. Fig 10 shows Training and validaton accuracy of Drivers with Spectacles at night time, and Fig 11 shows Training and validaton loss of Drivers with Spectacles at night time.

Fig 12 shows Training and validaton accuracy of Drivers with Sunglasses at day time, and Fig 13 shows Training and validaton loss of Drivers with Sunglasses at day time.

The training process shows a consistent and healthy downward trend in both training and validation loss over the epochs, indicating effective model optimization. The training loss decreases sharply in the initial epochs, demonstrating rapid learning, and then gradually flattens, approaching a very low value. The validation loss follows a similar trajectory, closely tracking the training loss without significant divergence, which suggests good generalization without severe overfitting. However, a slight increase or plateau in the validation loss near the final epochs may be observed, hinting at a potential early stopping opportunity around epoch 10–12 to prevent overfitting and preserve the best

generalizing model. Overall, the loss curves confirm that the model has successfully minimized its objective function and converged to a robust solution. To further improve, techniques such as learning rate scheduling or additional regularization could be explored if the validation loss shows signs of instability in future runs.

7. Experimental results

The presented performance metrics across various visual scenarios provide a robust but condition-sensitive detection system. Overall, precision remains consistently high (all scenarios >93%), indicating that when the system makes a positive identification, it is highly reliable. However, recall rates show greater variability, highlighting the model's primary challenge in comprehensively. The results underscore that while the model is fundamentally accurate, its practical deployment requires targeted improvements particularly in enhancing night-vision robustness and mitigating the obstructive effects of eyewear, especially glasses, to ensure consistent performance across all real-world situations.

Based on the confusion matrices across five distinct scenarios, the model demonstrates a strong but uneven performance profile in binary classification between classes 'D' and 'ND'. Overall, precision is consistently high, as evidenced by low false positive counts across all conditions. However, recall capability varies significantly, with false negative rates revealing the model's sensitivity to environmental and accessory-based challenges. This analysis underscores that while the classifier is generally reliable, its effectiveness diminishes under specific real-world complexities, highlighting a critical need for targeted data augmentation or algorithmic adjustments to improve resilience in night-time and eyewear-interference scenarios.



Figure 1: *NTHUDDD Dataset*

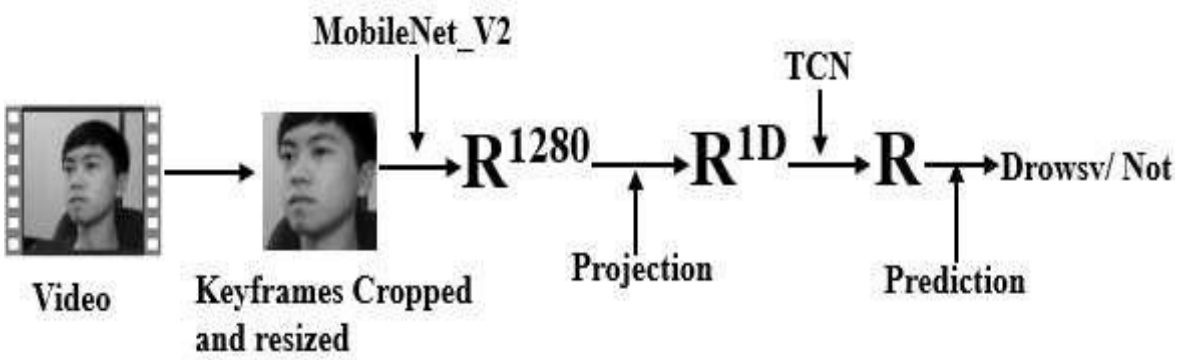

Figure 2 *:General architecture*

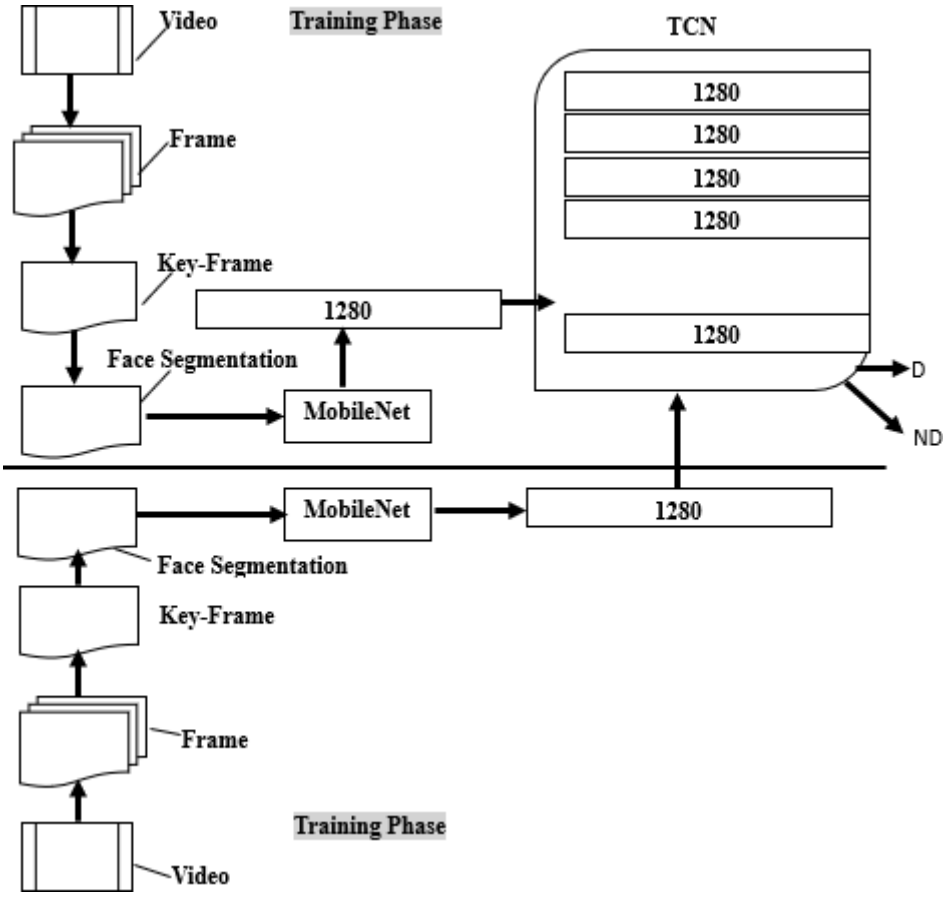

***Figure 3**: Proposed Architecture Drowsy Driver Detection*

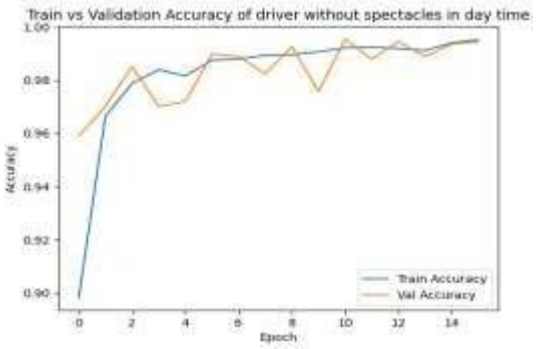

***Figure 4 :** Training and validaton accuracy*

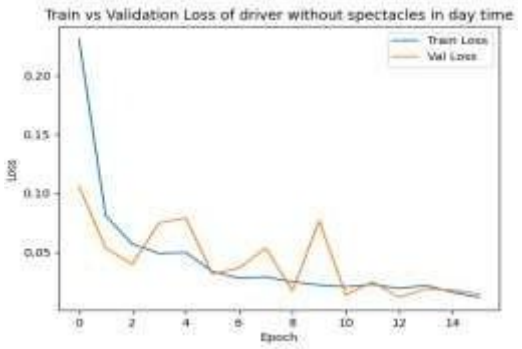

***Figure 5** : Training and validaton loss*

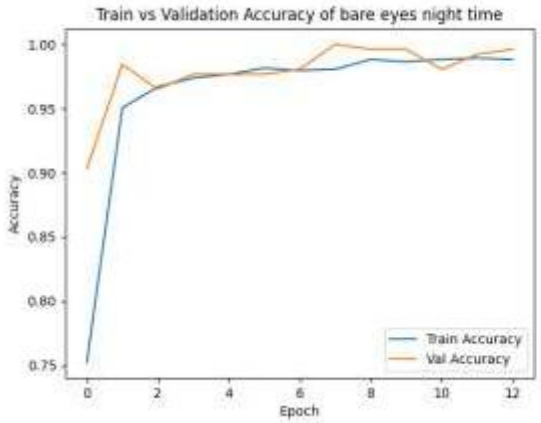

***Figure 6** : Training and validaton accuracy*

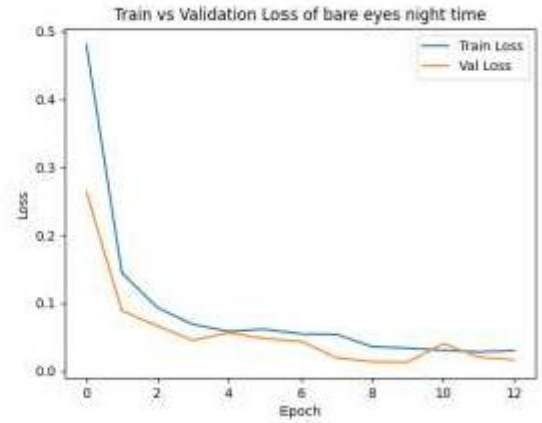

***Figure 7 :** Training and validaton loss*

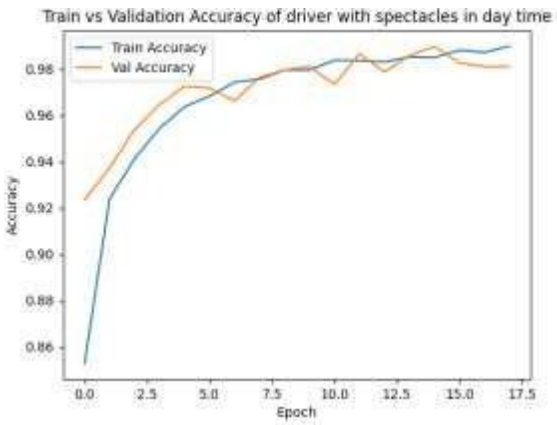

Figure 8 *: Training and validaton accuracy*

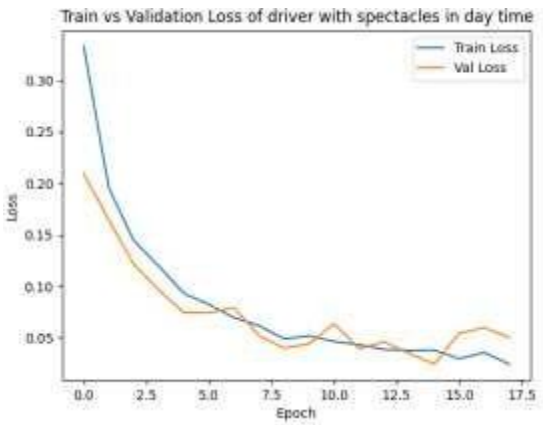

Figure 9 *: Training and validaton loss*

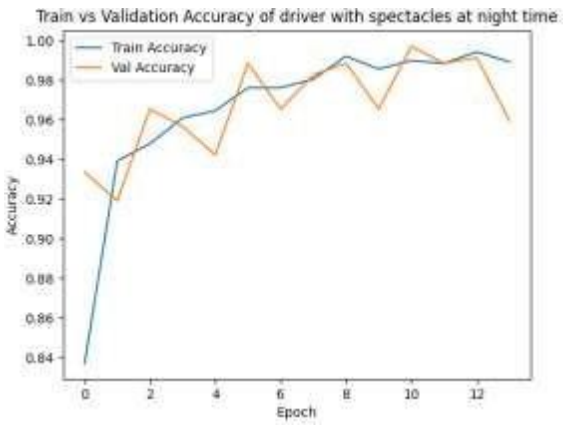

Figure 10 *: Training and validaton accuracy*

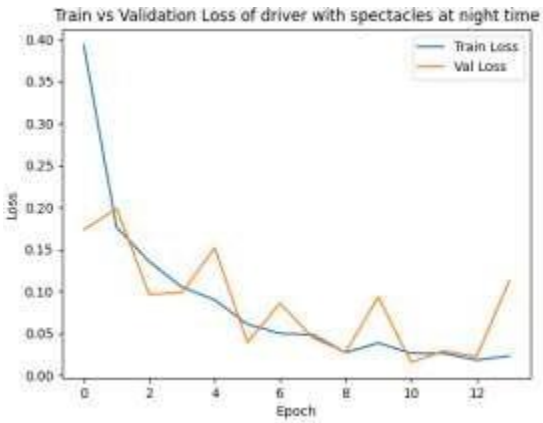

Figure 11 *: Training and validaton loss*

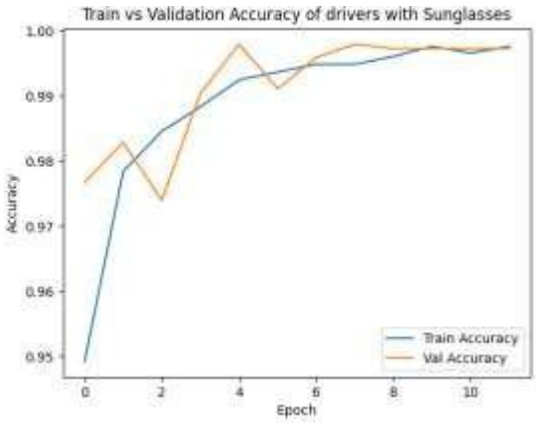

Figure 12 *: Training and validaton accuracy*

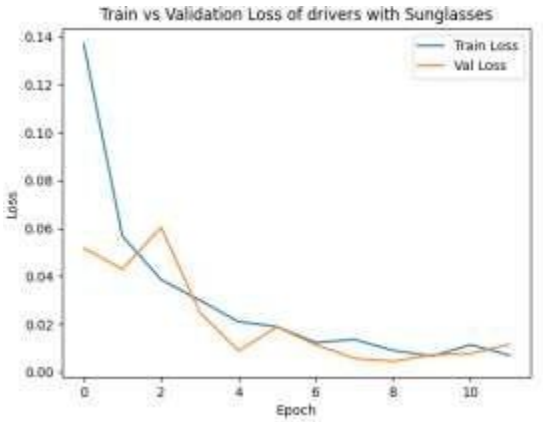

Figure 13 *: Training and validaton loss*

Table 1: *Performance metrics on different visual scenarios*

Sl No	Scenario	Precision	Recall	F1-score
	Bare Eyes Day-time	**1.00**	**0.9233**	**0.9601**
	Bare Eyes Night-time	**1.00**	**0.9874**	**0.9937**
	Spectacles Day-time	**1.00**	**0.9633**	**0.9813**

4	Spectacles Night-time	0.99	1.0000	0.995
5	Sunglasses Day-time	0.9301	1.0000	0.9638

Table 2: *Classification between drowsy and non-drowsy on different visual scenarios*

Bare Eyes Day-time		
	0	1
0	367	0
1	54	650

Bare Eyes Night-time		
	0	1
0	99	0
1	2	157

Spectacles Day-time		
	0	1
0	626	0
1	24	630

Spectacles Night-time		
	0	1
0	145	2
1	0	199

Sunglasses Day-time		
	0	1
0	1316	10
1	0	133

8. Conclusions

In this paper we proposed drowsy driver detection using Temporal Convolutional Network. Frames are converted from video, keyframes are selected using eyes status and the images are cropped. The cropped images are fed into MobileNet, 1280 features are extracted, converted into 1D feature vector and a sequence of 30 feature vectors are fed into TCN. The Temporal Convolutional Network (TCN) provides an effective approach for drowsy driver detection by accurately modeling temporal patterns in sequential video data. By leveraging dilated causal convolutions, the TCN captures both short- and long-term dependencies such as eye closure duration, blink frequency, and fatigue progression without relying on recurrent connections. This results in stable training and efficient computation.The experimental results indicate that the TCN achieves reliable performance in distinguishing drowsy and non-drowsy states, demonstrating its suitability for real-time driver monitoring applications. Overall, the TCN-based approach offers a robust and computationally efficient solution for detecting driver drowsiness in video-based systems. Future work will focus on optimizing the TCN for real-time deployment on edge devices, and improving robustness under low-light and occlusion conditions. Hybrid models with attention mechanisms can also be explored to enhance temporal feature learning.

Author Statements:

- **Ethical approval:** The conducted research is not related to either human or animal use.
- **Conflict of interest:** The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper
- **Acknowledgement:** The authors declare that they have nobody or no-company to acknowledge.
- **Author contributions:** The authors declare that they have equal right on this paper.
- **Funding information:** The authors declare that there is no funding to be acknowledged.
- **Data availability statement:** The data that support the findings of this study are available on request from the corresponding author. The data are not publicly available due to privacy or ethical restrictions.

References

[1] Shukla, Vivek and Satya Prakash Sahu. "Driver Drowsiness Detection using CNN with attention mechanism of Transformer model." *2024 15th International Conference on Computing Communication and Networking Technologies (ICCCNT)* (2024): 1-5.

[2] Kaur, Navjot. "Detection of Drowsy Drivers in Video Sequences via LSTM with CNN Features." INTERANTIONAL JOURNAL OF SCIENTIFIC RESEARCH IN ENGINEERING AND MANAGEMENT (2024): n. pag.

[3] Ranjan, Abhineet et al. "An Efficient Deep Learning Technique for Driver Drowsiness Detection." *SN Computer Science* 5 (2024): n. pag.

[4] Jarndal, Anwar H. et al. "A Real-Time Vision Transformers-Based System for Enhanced Driver Drowsiness Detection and Vehicle Safety." *IEEE Access* 13 (2025): 1790-1803.

[5] Jebraeily, Yashar et al. "Driver Drowsiness Detection Based on Convolutional Neural Network

Architecture Optimization Using Genetic Algorithm." *IEEE Access* 12 (2024): 45709-45726.

[6] Essahraui, Siham et al. "Real-Time Driver Drowsiness Detection Using Facial Analysis and Machine Learning Techniques." *Sensors (Basel, Switzerland)* 25 (2025): n. pag.

[7] Madni, Hamza Ahmad et al. "Novel Transfer Learning Approach for Driver Drowsiness Detection Using Eye Movement Behavior." *IEEE Access* 12 (2024): 64765-64778.

[8] Yang, Lie et al. "Video-Based Driver Drowsiness Detection With Optimised Utilization of Key Facial Features." *IEEE Transactions on Intelligent Transportation Systems* 25 (2024): 6938-6950.

[9] Ahmed, Mohammed Imran Basheer et al. "A Deep-Learning Approach to Driver Drowsiness Detection." *Safety* (2023): n. pag.

[10] Albadawi, Yaman et al. "Real-Time Machine Learning-Based Driver Drowsiness Detection Using Visual Features." *Journal of Imaging* 9 (2023): n. pag.

[11] G. N. Sharath Kumar, J. Hanumanthappa, C. Chethan Raj and P. Naveen Kumar, "Detection of Driver's Drowsiness in a Video based on Deep Features," *2024 5th International Conference on Innovative Trends in Information Technology (ICITIIT)*, Kottayam, India, 2024, pp. 1-9, doi: 10.1109/ICITIIT61487.2024.10580522.

[12] Kumar, Sharath and Naveen Kumar. "An Approach For Detecting Drowsy Driver Using Key Frame Selection from Video." International Journal of Applied Mathematics (2025).

[13] Weng, Ching-Hua et al. "Driver Drowsiness Detection via a Hierarchical Temporal Deep Belief Network." *ACCV Workshops* (2016).

[14] Sandler, Mark et al. "MobileNetV2: Inverted Residuals and Linear Bottlenecks." *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition* (2018): 4510-4520.

[15] Lea, Colin S. et al. "Temporal Convolutional Networks: A Unified Approach to Action Segmentation." *ECCV Workshops* (2016).