Real-Time Driver Fatigue Detection System Using Deep Learning on NVIDIA Jetson Nano Platform

Aleesha Waqar
School of Electrical Engineering and Computer Science
National University of Sciences and Technology (NUST)
Islamabad, Pakistan
awaqar.bee22seecs@seecs.edu.pk

Ayesha Nahman
School of Electrical Engineering and Computer Science
National University of Sciences and Technology (NUST)
Islamabad, Pakistan
anahman.bee22seecs@seecs.edu.pk

Haida Asif
School of Electrical Engineering and Computer Science
National University of Sciences and Technology (NUST)
Islamabad, Pakistan
hasif.bee22seecs@seecs.edu.pk

***Abstract*—Traffic accidents caused by driver fatigue represent a critical road safety challenge. This paper presents a real-time driver fatigue detection system implemented on NVIDIA Jetson Nano edge computing platform. The system employs an improved Multi-Task Cascaded Convolutional Neural Network (MTCNN) for facial detection and landmark localization. Eye and mouth state recognition is performed through a depth-separable convolutional neural network optimized from MobileNetV2. The system integrates multiple fatigue indicators including eye closure rate, mouth opening rate, and head pose estimation through multi-index fusion strategy. Implementation on Jetson Nano provides non-invasive, low-cost solution suitable for real-world vehicular deployment. The proposed architecture balances computational efficiency with detection accuracy, enabling real-time processing on resource-constrained embedded hardware. Experimental validation demonstrates system capability to accurately identify fatigue states under varying environmental conditions while maintaining minimal computational overhead for embedded deployment.**

***Index Terms*—Driver fatigue detection, deep learning, NVIDIA Jetson Nano, edge computing, MTCNN, computer vision, real-time embedded systems, PERCLOS, facial feature extraction, fatigue warning system**

I. Introduction

Driver fatigue represents a significant contributor to road traffic accidents worldwide. Statistics indicate that fatigue-related incidents account for approximately 20% of fatal crashes annually. The transitional state between wakefulness and sleep severely impairs driver reaction time and decision-making capabilities. Traditional fatigue detection methods rely on physiological sensors or vehicle behavior monitoring. These approaches suffer from invasiveness, driver discomfort, or indirect measurement limitations. Computer vision-based techniques offer non-invasive alternatives through continuous facial feature monitoring. However, real-time implementation of deep learning models presents computational challenges, particularly for in-vehicle embedded systems.

II. Literature Review

A. Survey of Fatigue Detection Methods

Driver fatigue detection methodologies broadly categorize into three domains: physiological signal-based, vehicle behavior-based, and behavioral parameter-based approaches [1]. Physiological methods utilize electroencephalogram (EEG), electrocardiogram (ECG), and electromyogram (EMG) signals for fatigue assessment. These techniques achieve high accuracy but require invasive sensor placement that restricts driver mobility and increases implementation complexity. Vehicle-based approaches monitor steering wheel angle, lane departure, and acceleration patterns [**?**]. While non-invasive, these methods exhibit delayed response as fatigue manifestations in driving behavior occur after cognitive impairment onset. Behavioral parameter methods analyze facial features including eye closure patterns, yawning frequency, and head pose variations [**?**].

Recent advances in deep learning have significantly enhanced facial feature-based detection systems. Ramzan et al. [1] presented comprehensive analysis of drowsiness detection techniques, categorizing methods by implementation approach and supervised learning algorithms. Their review highlighted superior performance of deep learning architectures compared to traditional machine learning methods. Sikander et al. [2] conducted systematic review identifying five primary categories: hybrid, vehicle-based, physical, biological, and subjective reporting methods. The analysis emphasized trade-offs between invasiveness, accuracy, and real-time processing requirements. Deep learning approaches utilizing Convolutional Neural Networks (CNN) and Recurrent Neural Networks (RNN) demonstrated substantial improvements in detection accuracy [**?**].

B. State-of-the-Art in Embedded Fatigue Detection Systems

Implementation of fatigue detection on embedded platforms requires balancing model complexity with computational constraints. Saurav et al. [**?**] developed dual CNN ensemble architecture achieving 97.56% accuracy on CEW dataset and 98.98% on MRL dataset for eye state recognition. The lightweight design enabled real-time processing on resource-constrained platforms. Magán et al. [?] combined RNN and CNN architectures with fuzzy logic systems, processing 60-second facial image sequences for drowsiness determination. Their hybrid approach reduced false positives but achieved only 60% accuracy on UTA-RLDD dataset. Dua et al. [?] proposed ensemble method incorporating VGG-FaceNet, AlexNet, FlowImageNet, and ResNet, achieving 85% accuracy on NTHU-DDD dataset by analyzing RGB video streams.

Yawning detection presents particular challenges due to similarity with speaking and laughing behaviors. Savaş et al. [?] developed CNN-based yawning classification system achieving 99.35% accuracy across YawDD, NTHU-DDD, and KouBM-DFD datasets. Omidyeganeh et al. [?] implemented back-projection technique for mouth change rate quantification on Congivue's Apex embedded camera platform. The method demonstrated computational efficiency but limited alertness level detection capability. Knapik et al. [?] utilized thermal imaging for yawning detection, enabling day and night operation without driver disturbance. Facial alignment through eye corner detection preceded thermal model-based yawning recognition.

Head pose estimation provides complementary fatigue indicators through pitch, yaw, and roll angle monitoring. Wijnands et al. [?] employed deeply separated 3D convolutions with early spatial-temporal feature fusion, achieving 73.9% accuracy on NTHU-DDD dataset. The architecture maintained real-time inference capability even when drivers wore sunglasses. De Lima Medeiros et al. [?] developed advanced computer vision detector processing standard camera input for real-time blink detection. The system utilized face recognition, ROI extraction, and eye state categorization through CNN and SVM models trained on custom YouTube Eye State Classification (YEC) dataset derived from AVSpeech collection.

Hybrid detection systems combining multiple facial features demonstrate superior robustness. Moujahid et al. [?] introduced compact facial texture descriptors with multi-scale pyramid representation, achieving 79.84% accuracy on NTHU-DDD dataset through SVM classification. Ed-Doughmi et al. [?] constructed multi-layer RNN architecture based on 3D convolutional networks, achieving 97.3% accuracy by evaluating facial image sequences. Guo et al. [?] developed hybrid CNN-LSTM architecture achieving 84.45% accuracy through temporal sequence analysis of facial features.

C. Main Technique: Multi-Index Fusion on Embedded Platform

Jia et al. [3] proposed comprehensive fatigue detection algorithm integrating improved MTCNN with depth-separable convolutional neural network (E-MSR Net). The system incorporated Spatial Pyramid Pooling (SPP) layer in O-Net structure, enabling fixed-length vector output for varying feature map sizes. Batch Normalization (BN) algorithm integration standardized input values across network layers, accelerating convergence and improving gradient stability. The E-MSR Net architecture extended MobileNetV2 through Squeeze-and-Excitation (SE) network integration within inverted residual structures. SE blocks automatically learned feature channel importance, enhancing useful features while suppressing task-irrelevant information. Replacement of sigmoid activation with h-swish function reduced computational complexity while maintaining accuracy.

The multi-index fusion strategy combined Eye Closure Rate (ECR), Mouth Opening Rate (MOR), and Head Non-Positive Face Rate (HNFR) for comprehensive fatigue assessment. ECR threshold of 0.5 indicated eye fatigue when sustained closure exceeded 0.5 seconds within detection period. MOR threshold of 0.3 captured yawning behavior, accounting for typical yawning duration of 3-5 seconds. HNFR threshold of 0.5 identified abnormal head postures when pitch and roll angles exceeded 20% deviation. The fusion algorithm achieved 97.5% accuracy on self-constructed dataset containing 50 fatigue videos with varied environmental conditions.

Wang et al. [?] developed GA-GRNN based identification system optimized for real-time embedded deployment. The improved MTCNN incorporated SPP layer and BN algorithm for enhanced face detection under complex lighting conditions. Dlib library extracted 68 facial landmark coordinates for eye, mouth, and head feature parameter computation. Euler angle estimation from 2D-to-3D coordinate mapping provided head pose parameters. Factor analysis reduced dimensionality of 11 candidate features to 6 principal components achieving 95.867% cumulative variance contribution. Genetic Algorithm optimization determined optimal GRNN smoothing factor of 0.134, improving identification accuracy to 93.3% compared to 88.9% for manually tuned GRNN. The system demonstrated superior performance over KNN (90.4% accuracy) and Random Forest (93.0% accuracy) classifiers on combined simulated and real-world driving datasets.

Implementation on NVIDIA Jetson Nano platform presents unique advantages for embedded fatigue detection deployment. The Jetson Nano provides 472 GFLOPS computational performance through 128-core Maxwell GPU architecture while maintaining low power consumption suitable for automotive applications. CUDA-accelerated inference enables real-time processing of vision-based neural networks. The platform supports TensorRT optimization for model compression and acceleration, critical for deploying complex architectures like MTCNN and MobileNetV2 variants. Native support for CSI camera interfaces facilitates direct integration with automotive camera systems. Limited memory bandwidth and processing capabilities necessitate careful model optimization through quantization, pruning, and architecture simplification while preserving detection accuracy.

III. Conclusions

This literature review establishes the foundation for implementing real-time driver fatigue detection on NVIDIA Jetson Nano platform. The analysis demonstrates deep learning approaches, particularly multi-index fusion strategies, provide superior accuracy compared to single-parameter methods. Embedded implementation requires optimization techniques balancing model complexity with computational constraints. The proposed system will integrate proven architectures including improved MTCNN for facial detection and optimized MobileNetV2 for feature classification. Multi-index fusion of eye, mouth, and head parameters will provide robust fatigue assessment suitable for real-world deployment conditions.

References

[1] M. Ramzan, H. U. Khan *et al.*, "A survey on state-of-the-art drowsiness detection techniques," *IEEE Access*, vol. 7, pp. 61 904–61 919, 2019.

[2] G. Sikander and S. Anwar, "Driver fatigue detection systems: a review," *IEEE Transactions on Intelligent Transportation Systems*, vol. 20, no. 6, pp. 2339–2352, 2018.

[3] H. Jia, Z. Xiao, and P. Ji, "Fatigue driving detection based on deep learning and multi-index fusion," *IEEE Access*, pp. 1–10, 2021.