STATISTICS, OPTIMIZATION AND INFORMATION COMPUTING
Stat., Optim. Inf. Comput., Vol. **151**, January 2026, pp 274–294.
Published online in International Academic Press (www.IAPress.org)

Bayesian-Optimized CLAHE for Enhanced Drowsiness Detection in Low-Light Conditions Using Time-Distributed MobileNetV2-GRU Architecture

Farrikh Alzami[1,*], Muhammad Naufal[1], Ruri Suko Basuki[1], Sri Winarno[1], Harun Al Azies[1], Syaheerah Lebai Lutfi[2,3], Rivaldo Mersis Brilianto[4]

[1] *Faculty of Computer Science, Universitas Dian Nuswantoro, Semarang, 50131, Indonesia*
[2] *Sultan Qaboos University, Sultanate of Oman*
[3] *School of Computer Sciences, Universiti Sains Malaysia, Penang, 11800, Malaysia*
[4] *School of Mechanical Engineering, Pusan National University, 2, Busandaehak-ro 63beon-gil, Geumjeong-gu, 46241, Busan, Republic of Korea*

Abstract Driver drowsiness remains a critical factor in road traffic accidents, particularly under low-light conditions where conventional computer vision approaches struggle with poor image quality. This study presents a novel approach combining Bayesian-optimized Contrast Limited Adaptive Histogram Equalization (CLAHE) with a Time-Distributed MobileNetV2-GRU architecture for robust drowsiness detection in challenging lighting conditions. Using the NITYMED dataset containing 128 video sequences, this paper systematically compares three preprocessing strategies: original frames, fixed-parameter CLAHE (clip_limit=2.0), and Bayesian-optimized CLAHE. The methodology employs Bayesian Optimization to adaptively determine optimal CLAHE parameters based on Perceptual Image Quality Evaluator (PIQE) scores, transforming preprocessing into a task-aware component. Statistical analysis using Wilcoxon Signed-Rank Test demonstrates that the Bayesian-optimized approach significantly outperforms baseline methods, achieving mean accuracy of 93.77% $\pm$ 0.0521, F1-score of 93.77% $\pm$ 0.0522, and AUC of 97.85% $\pm$ 0.0145 across 10-fold cross-validation, with peak performance reaching 98.11% accuracy under optimal configuration (p-values < 0.05 for accuracy and F1-score comparisons). The integration of lightweight MobileNetV2 with GRU enables efficient temporal modeling while maintaining computational efficiency with only 62,449 trainable parameters. Results indicate that adaptive preprocessing significantly improves feature visibility and model convergence, demonstrating practical viability for deployment in Advanced Driver Assistance Systems (ADAS) when implemented with periodic optimization strategies.

Keywords Drowsiness detection, Bayesian optimization, CLAHE, Low-light conditions, Deep learning, MobileNetV2, GRU

DOI: 10.19139/soic-2310-5070-3024

1. Introduction

Road traffic accidents represent a significant global public health challenge, with driver drowsiness being identified as a contributing factor in approximately 20% of all crashes and up to 50% of fatal accidents [1, 2]. The severity of this problem becomes worse under low-light driving conditions, where conventional computer vision systems face substantial challenges in accurately detecting drowsiness indicators due to poor illumination, reduced contrast, and increased noise levels [3, 4]. Traditional drowsiness detection systems, while effective under optimal lighting conditions, often exhibit degraded performance during nighttime driving scenarios when the risk of fatigue-related accidents is statistically highest [5, 6].

*Correspondence to: Farrikh Alzami (Email: alzami@dsn.dinus.ac.id). Faculty of Computer Science, Universitas Dian Nuswantoro, Semarang, 50131, Indonesia.

ISSN 2310-5070 (online) ISSN 2311-004X (print)
Copyright © 2026 International Academic Press

The fundamental challenge in low-light drowsiness detection lies in the tradeoff between image enhancement and computational efficiency. Current preprocessing techniques either rely on fixed parameters that may not generalize across diverse lighting conditions, or employ computationally expensive adaptive methods unsuitable for real-time applications [7, 8]. Furthermore, existing deep learning approaches often struggle with temporal dependencies in video sequences under suboptimal lighting, leading to inconsistent detection performance [9, 10].

This study addresses the aforementioned challenges through three primary objectives: **Objective 1,** Develop an adaptive image preprocessing pipeline that automatically optimizes CLAHE parameters using Bayesian Optimization, guided by perceptual quality metrics to enhance feature visibility in low-light conditions. **Objective 2,** Design a lightweight yet effective deep learning architecture combining Time-Distributed MobileNetV2 for spatial feature extraction with GRU for temporal modeling, ensuring both accuracy and computational efficiency. Then, **Objective 3,** Provide comprehensive statistical validation demonstrating the significance of adaptive preprocessing over conventional fixed-parameter approaches, with specific focus on deployment feasibility in real-world ADAS applications.

The primary contributions of this work include: (1) **Methodological Innovation**, Introduction of Bayesian-optimized CLAHE as an adaptive preprocessing strategy that transforms traditional fixed-parameter enhancement into a task-aware, quality-driven process; (2) **Architectural Efficiency**, Integration of lightweight MobileNetV2 with GRU in a Time-Distributed framework, enabling effective temporal modeling while maintaining computational constraints suitable for edge deployment; (3) **Empirical Validation**, Comprehensive statistical analysis using robust non-parametric tests demonstrating significant performance improvements in low-light drowsiness detection scenarios.

2. Literature Review

2.1. Drowsiness Detection in Low-Light Conditions

Recent advances in computer vision-based drowsiness detection have primarily focused on optimal lighting conditions, with limited attention to challenging low-light scenarios. Pandey and Muppalaneni [6] conducted a comprehensive survey of visual and non-visual features for drowsiness detection, highlighting the persistent challenges in nighttime driving scenarios. Similarly, Hidalgo-Gadea et al. [5] demonstrated that personality traits and IQ factors influence microsleep prediction accuracy, yet their approach did not address illumination variability.

Recent work by Soman et al. [10] proposed an IoT-assisted CNN-LSTM framework for drowsiness monitoring using multiple datasets including YawDD, NITYMED, FL3D, and custom datasets, achieving promising results under controlled conditions. However, their methodology lacks adaptive preprocessing strategies for handling low-light environments. Alzami et al. [11] specifically addressed low-light drowsiness detection using Time Distributed MobileNetV2 with Auto-CLAHE on the DROZY dataset, achieving 93.62% accuracy with improvements in eye region analysis. While this work shares similarities with our approach, it relies on fixed automated CLAHE parameters rather than optimization-driven adaptation.

2.2. Image Enhancement for Low-Light Computer Vision

The application of histogram equalization techniques in low-light image processing has gained significant attention in recent years. Tian et al. [7] provided a comprehensive survey of deep learning-based low-light image enhancement, emphasizing the importance of adaptive preprocessing in computer vision pipelines. Kim [8] presented a comparative review of low-light enhancement prospects, highlighting CLAHE as a particularly effective technique for local contrast improvement.

Guo et al. [9] conducted an extensive survey on image enhancement for low-light images, demonstrating that adaptive histogram equalization methods consistently outperform global enhancement techniques. However, existing approaches typically employ heuristic parameter selection rather than optimization-based strategies. Borra et al. [12] proposed multi-objective improved Cat Swarm Optimization for CLAHE parameter selection, yet their approach lacks integration with deep learning frameworks and perceptual quality metrics.

2.3. Bayesian Optimization in Computer Vision

The application of Bayesian Optimization to computer vision preprocessing has emerged as a promising research direction. Bian and Priyadarshi [13] provided a comprehensive classification of machine learning optimization techniques, emphasizing the effectiveness of Bayesian approaches for hyperparameter tuning in complex systems. Salhi and Woodward [14] proposed SMBOX, a scalable method for sequential model-based parameter optimization, demonstrating superior performance over traditional grid search approaches.

Mei et al. [15] extended Bayesian Optimization through Gaussian Cox Process models for spatio-temporal data, showing particular relevance to video-based applications. Zhang et al. [16] developed sequential Gaussian processes for online learning of nonstationary functions, providing theoretical foundations for adaptive parameter optimization in dynamic environments. However, the specific application of Bayesian Optimization to CLAHE parameter tuning for drowsiness detection remains largely unexplored.

2.4. Lightweight Deep Learning Architectures

The development of efficient deep learning models for real-time applications has become increasingly critical. Hu et al. [17] demonstrated lightweight deep learning for real-time road distress detection on mobile devices, emphasizing the importance of computational efficiency in deployment scenarios. Rajak et al. [18] proposed efficient lightweight models specifically for drowsiness detection, yet their approach lacks temporal modeling capabilities.

MobileNetV2 has proven particularly effective for mobile and edge computing applications due to its depthwise separable convolutions and inverted residual structure. When combined with recurrent networks like GRU, these architectures can effectively model temporal dependencies while maintaining computational constraints. However, the specific combination of Time-Distributed MobileNetV2 with GRU for drowsiness detection under low-light conditions represents a novel architectural contribution.

2.5. Research Gaps and Motivation

Based on the comprehensive literature review, several critical gaps emerge: (1) **Adaptive Preprocessing Gap**, Existing low-light enhancement methods predominantly rely on fixed parameters or heuristic adaptation, lacking optimization-driven approaches guided by perceptual quality metrics; (2) **Integration Gap**, Limited integration between adaptive preprocessing and deep learning architectures, with most approaches treating enhancement and classification as separate pipeline components; (3) **Statistical Validation Gap**, Insufficient rigorous statistical analysis of preprocessing impact on model performance, particularly in comparative studies between adaptive and fixed-parameter approaches; (4) **Real-time Deployment Gap**, Lack of comprehensive evaluation considering both accuracy and computational efficiency requirements for practical ADAS deployment.

This study addresses these gaps by introducing a novel framework that seamlessly integrates Bayesian-optimized CLAHE with lightweight deep learning architectures, providing both theoretical innovation and practical deployment viability.

3. Methodology

This study followed a structured pipeline designed to enhance video quality under low-light conditions and improve the accuracy of driver drowsiness detection. The overall workflow is summarized in Figure 1, which outlines the primary stages of the methodology. The process begins with dataset preparation and segmentation into frame sequences, followed by preprocessing steps including face detection, cropping, resizing, and normalization. CLAHE is applied, where the clipLimit parameter is optimized through Bayesian Optimization using PIQE as the objective function. The enhanced sequences are then fed into a deep learning architecture that combines MobileNetV2 for spatial feature extraction with Gated Recurrent Units (GRUs) for temporal modeling. The model is trained and evaluated using cross-validation and multiple performance metrics.

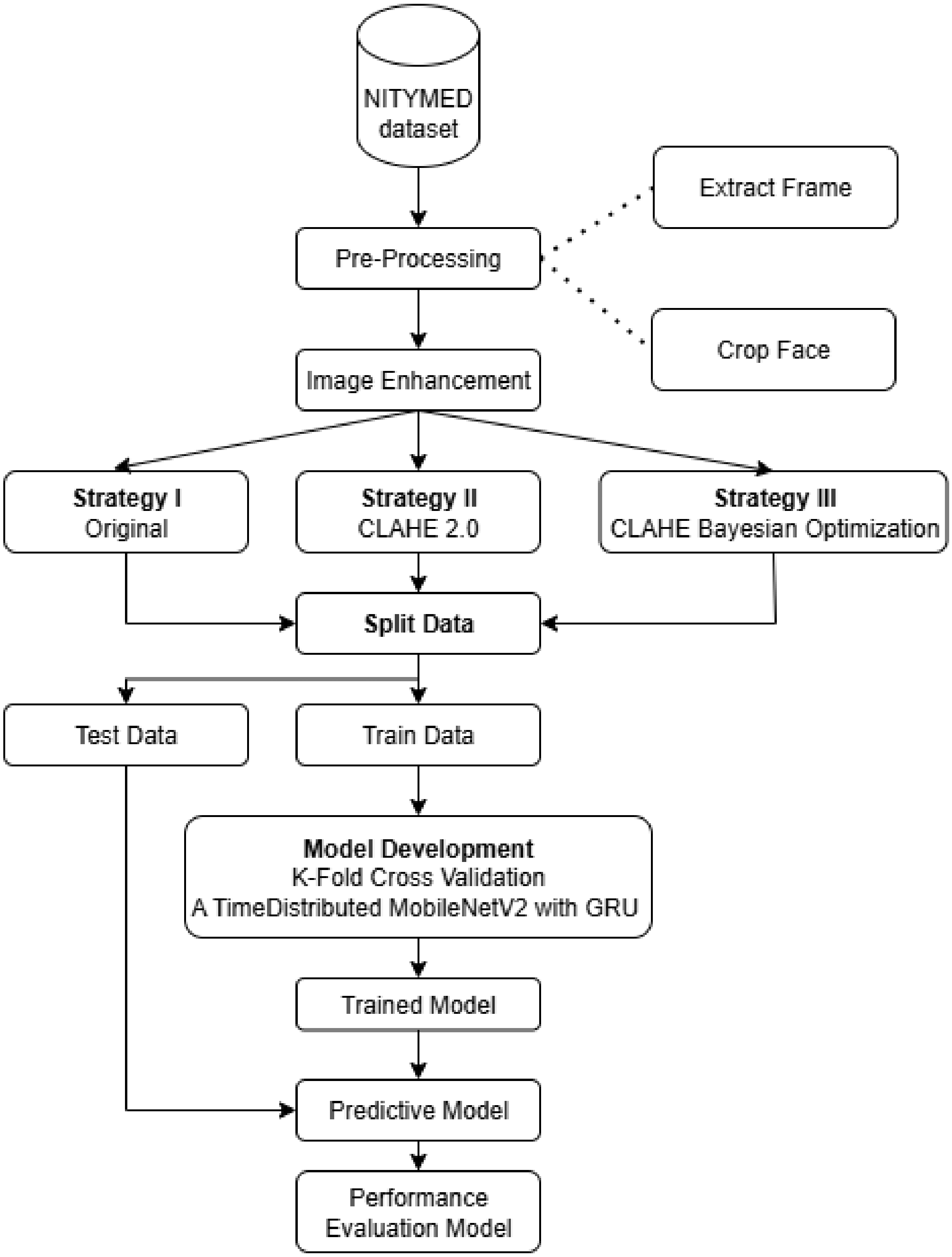


Figure 1. Flowchart of the proposed methodology for drowsiness detection under low-light conditions

3.1. Dataset Description

This study utilized the NITYMED dataset [21], which consists of 128 video recordings specifically designed for drowsiness detection under low-light conditions. The dataset contains two primary categories: yawning episodes (106 recordings) and microsleep incidents (23 recordings), providing a realistic representation of driver fatigue behaviors. The videos were recorded at 1920×1080 resolution with varying durations ranging from 15-25 seconds for yawning episodes to approximately 2 minutes for microsleep sequences.

The dataset exhibits balanced gender distribution with 11 male and 10 female participants captured in Patras, Greece, displaying drivers in real cars moving under nighttime conditions. Participants demonstrate diverse phenotypic features including variations in hair color, facial hair, and eyewear. Each video was segmented into sequences of 20 consecutive frames, resulting in 528 sequences for experimental purposes. The dataset was stratified and split using an 80:20 ratio for training and testing, with 10-fold cross-validation employed to ensure robust performance evaluation and minimize overfitting risks.

3.2. Preprocessing Pipeline

3.2.1. Frame Extraction and Face Detection To ensure uniformity and effective feature representation, preprocessing steps were applied to all video sequences. Faces were detected and localized using YOLOv8, and the detected regions were cropped to focus only on the driver's face [19]. Each frame was then resized to 128×128 pixels and normalized to the range [0,1]. A sliding window approach was employed to capture temporal dependencies in driver behavior, where each sequence consisted of 20 consecutive frames sampled at two frames per second. This approach ensured that both short-duration actions, such as yawning, and longer-duration behaviors, such as microsleep, were adequately represented in the input data.

3.2.2. Bayesian-Optimized CLAHE Implementation The core innovation of this study lies in the adaptive CLAHE optimization strategy. Traditional CLAHE implementations rely on fixed clip_limit parameters, typically set to 2.0 based on empirical observations. Our approach employs Bayesian Optimization to determine optimal clip_limit values for each frame, guided by Perceptual Image Quality Evaluator (PIQE) scores.

The CLAHE enhancement is applied to the L channel in LAB color space according to:

$$I_{enhanced}(x, y) = CDF^{-1}_{local}(\min(CDF_{local}(I(x, y)), clip\_limit)) \tag{1}$$

where CDF_{local} represents the cumulative distribution function of the local tile, and the clip_limit parameter controls contrast amplification.

The PIQE objective function is formulated as:

$$PIQE(I) = \alpha \cdot Block\_Variance + \beta \cdot Activity\_Mask + \gamma \cdot Noise\_Estimation \tag{2}$$

where lower PIQE scores indicate better perceptual quality.

While PIQE evaluates overall image quality rather than task-specific facial features, it provides a practical no-reference metric suitable for optimization without requiring ground-truth annotations. The rationale for employing PIQE lies in the hypothesis that improved perceptual quality in low-light conditions enhances the discriminability of subtle facial features critical for drowsiness detection, such as eyelid closure degree and mouth opening angle. The empirical correlation between PIQE improvement and classification accuracy gains (demonstrated in Section 4.1) validates this optimization strategy. Alternative no-reference metrics such as NIQE or BRISQUE operate on similar principles but would introduce additional computational overhead without demonstrated superiority for this specific application.

The relationship between CLAHE optimization and drowsiness detection performance can be understood through the following mechanism: Low-light conditions compress pixel intensity distributions into narrow ranges, obscuring subtle facial features essential for drowsiness recognition. CLAHE's adaptive histogram equalization expands local contrast, but excessive clip_limit values (e.g., 2.0) amplify noise and create over-enhancement artifacts that degrade CNN feature extraction. Bayesian Optimization navigates this tradeoff by identifying clip_limit values (approximately 1.42 in our experiments) that maximize perceptual quality (minimize PIQE) while preserving feature discriminability.

As seen in Algorithm 1, to identify the optimal clip_limit value, Bayesian Optimization was applied within the range [0.1, 2.0]. The optimization was initialized with three random exploration points (init_points = 3) and continued with ten guided iterations (n_iter = 10). The Perception-based Image Quality Evaluator (PIQE), a no-reference metric that quantifies perceptual image quality based on blur, noise, and artifact presence, was used as

Algorithm 1 Bayesian-Optimized CLAHE using Gaussian Process with Upper Confidence Bound acquisition

Require: Original frame I, initial exploration points = 3, iterations = 10
Ensure: Enhanced frame $I_{enhanced}$, optimal clip_limit

```
Initialize Gaussian Process surrogate model
for each exploration point in range [0.5 - 2.0] do
    Apply CLAHE with current clip_limit to L channel in LAB space
    Compute PIQE score as objective function
    Update surrogate model
end for
for iteration = 1 to 10 do
    Select next clip_limit using acquisition function
    Apply CLAHE and evaluate PIQE score
    Update Gaussian Process model
end for
return frame with optimal clip_limit and enhancement result
```

the objective function. This procedure ensures that CLAHE is only applied if it leads to perceptual improvements compared to the original frame.

Bayesian Optimization was implemented using the `bayesian-optimization` Python library with a Gaussian Process surrogate model employing default Radial Basis Function (RBF) kernel. The Upper Confidence Bound (UCB) acquisition function balanced exploration-exploitation tradeoffs during parameter search. The optimization space was constrained to clip_limit $\in$ [0.1, 2.0], initialized with 3 random exploration points and refined through 10 guided iterations. Fixed CLAHE parameters include tile_grid_size = (8,8) and LAB color space transformation via OpenCV's `cv2.cvtColor()` function.

3.3. Deep Learning Architecture

3.3.1. Time-Distributed MobileNetV2 Feature Extractor The spatial feature extraction component employs MobileNetV2 architecture wrapped in TimeDistributed layers to process video sequences efficiently. MobileNetV2 was selected for its optimal balance between computational efficiency and feature representation capability, particularly suitable for edge deployment scenarios.

The architecture modifications include: (1) **Input Adaptation**, Modified input layer to accept $128\times128\times3$ frames instead of standard 224×224 dimensions; (2) **Feature Extraction**, Utilized pre-trained ImageNet weights with frozen early layers to leverage transfer learning benefits; (3) **Output Modification**, Replaced final classification layers with global average pooling to obtain 1280-dimensional feature vectors per frame.

3.3.2. Temporal Modeling with GRU Temporal dependencies across video sequences are captured using Gated Recurrent Unit (GRU) networks, chosen over LSTM for their computational efficiency and comparable performance in sequential modeling tasks. The GRU configuration includes 16 hidden units, providing sufficient capacity for temporal pattern recognition while maintaining parameter efficiency.

The complete architecture pipeline processes sequences of 20 frames through TimeDistributed MobileNetV2, producing 20×1280 feature matrices. These temporal features are then processed by the GRU layer, followed by dense layers (8 units with ReLU activation) and final sigmoid output for binary classification.

The architectural design deliberately integrates spatial and temporal modeling as complementary components for video-based drowsiness detection. Unlike static image classification where spatial-only models suffice, drowsiness manifests as temporal patterns (e.g., progressive eye closure, yawning duration) that require sequential modeling. The Time-Distributed MobileNetV2-GRU architecture was specifically designed to capture both spatial features (facial expressions) and temporal dynamics (behavioral progression), representing the core methodological contribution of this work. Isolating spatial-only features through ablation would fundamentally contradict the video-based nature of the drowsiness detection task, where temporal context is essential for distinguishing momentary actions from sustained fatigue states.

Table 1. Model Architecture Summary (including non-trainable pre-trained MobileNetV2 backbone)

Layer (Type)	Output Shape	Parameters
time_distributed (TimeDistributed)	(None, 20, 4, 4, 1280)	2,257,984
time_distributed_1 (TimeDistributed)	(None, 20, 1280)	0
gru (GRU)	(None, 16)	62,304
dense (Dense)	(None, 8)	136
dense_1 (Dense)	(None, 1)	9
Total params:	**2,320,433 (8.85 MB)**	
Trainable params:	**62,449 (243.94 KB)**	
Non-trainable params:	**2,257,984 (8.61 MB)**	

Based on Table 1, the architecture achieves remarkable parameter efficiency, with only 62,449 trainable parameters (243.94 KB) while leveraging 2.26M pre-trained features from MobileNetV2. This design enables effective learning while maintaining computational constraints suitable for practical deployment.

3.4. Training Configuration and Optimization

The training process employed several optimization strategies to ensure stable convergence and prevent overfitting. The learning rate was set to 0.001 with SGD optimizer (momentum=0.9) to maintain training stability. Binary cross-entropy loss function was utilized for binary classification, with batch size of 64 and training duration of 30 epochs to balance convergence speed and generalization capability.

Early stopping was implemented with patience=5 epochs to prevent overfitting, monitoring validation loss for convergence indicators. Model checkpointing preserved the best-performing weights based on validation accuracy, ensuring optimal model selection for final evaluation.

3.5. Experimental Design and Statistical Analysis

The experimental methodology employed three distinct preprocessing strategies for comparative analysis: (1) **Original:** Raw frames without enhancement; (2) **CLAHE 2.0:** Fixed clip_limit=2.0 parameter; (3) **CLAHE BO:** Bayesian-optimized adaptive clip_limit.

Each strategy was evaluated using 10-fold cross-validation to ensure robust performance assessment and minimize variance in results. Performance metrics included accuracy, F1-score, and Area Under Curve (AUC), providing comprehensive evaluation across different aspects of classification performance.

Statistical significance testing employed Wilcoxon Signed-Rank Test, selected for its robustness to small sample sizes (n=10 folds) and non-parametric nature, avoiding assumptions about data distribution normality. The test compared paired performance differences across folds, with significance threshold set at $\alpha = 0.05$.

Statistical Methodology:

1. **Null Hypothesis** H_0**:** No significant difference in performance metrics between preprocessing strategies
2. **Alternative Hypothesis** H_1**:** Significant difference exists between preprocessing approaches
3. **Test Selection:** Wilcoxon Signed-Rank Test for paired, non-parametric comparison
4. **Significance Level:** $\alpha = 0.05$

4. Results

4.1. Performance Comparison Across Preprocessing Strategies

The experimental evaluation demonstrates substantial improvements achieved through Bayesian-optimized CLAHE preprocessing. Table 2 presents comprehensive performance metrics across all three preprocessing strategies, evaluated using 10-fold cross-validation.

Table 2. Performance Comparison of Preprocessing Strategies (Mean ± Standard Deviation across 10-Fold Cross-Validation)

Model	Accuracy	F1-Score	Recall	Precision	Specificity	AUC
CLAHE BO	0.9377±0.0521	0.9377±0.0522	0.9377±0.0521	0.9390±0.0515	0.9377±0.0521	0.9785±0.0145
CLAHE 2.0	0.9075±0.0444	0.9075±0.0445	0.9075±0.0444	0.9081±0.0442	0.9075±0.0444	0.9742±0.0231
Original	0.9038±0.0774	0.9002±0.0877	0.9038±0.0774	0.9160±0.0414	0.9038±0.0774	0.9807±0.0222

As presented in Table 2, CLAHE BO achieves the highest mean performance across accuracy (93.77%), F1-score (93.77%), recall (93.77%), and precision (93.90%) metrics, while maintaining competitive AUC performance (97.85%). Notably, CLAHE BO also demonstrates superior stability with lower standard deviations (0.0521 for accuracy) compared to the original preprocessing approach (0.0774), indicating more consistent performance across different data folds. The balanced performance across recall and precision suggests that the model maintains effective discrimination without sacrificing either sensitivity or specificity.

To provide detailed insight into the best-performing configuration, Table 3 presents the complete performance breakdown for Fold 9, which achieved peak performance across all preprocessing strategies.

Table 3. Detailed Performance Metrics for Fold 9 (Best Configuration)

Model	Accuracy	F1-Score	Recall	Precision	Specificity	AUC
CLAHE BO	0.9811	0.9811	0.9811	0.9815	0.9811	0.9904
CLAHE 2.0	0.9623	0.9623	0.9623	0.9630	0.9623	0.9856
Original	0.9623	0.9604	0.9623	0.9808	0.9623	0.9846

Table 3 reveals that under optimal configuration (Fold 9), CLAHE BO achieves 98.11% accuracy, 98.11% F1-score, and 99.04% AUC, demonstrating the potential of Bayesian-optimized preprocessing when conditions align favorably. The consistent improvement of CLAHE BO over both CLAHE 2.0 and Original preprocessing across all metrics indicates robust enhancement capability. Notably, the precision values (98.15% for CLAHE BO vs. 98.08% for Original) suggest that adaptive preprocessing maintains low false positive rates while improving true positive detection.

4.2. Statistical Significance Analysis

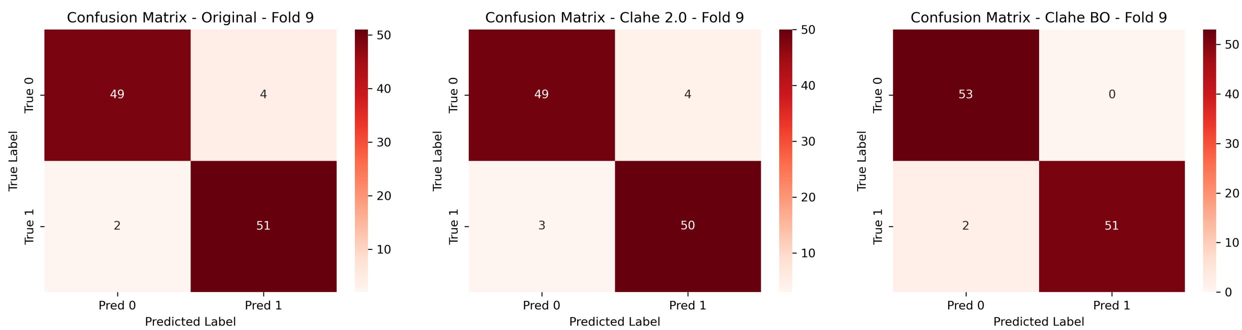
Statistical validation using Wilcoxon Signed-Rank Test provides robust evidence for the superiority of Bayesian-optimized preprocessing. Table 4 presents detailed statistical comparison results between preprocessing strategies.

Based on Table 4, CLAHE BO demonstrates statistically significant improvements over original preprocessing for both accuracy (p=0.037) and F1-score (p=0.027) metrics, providing strong evidence for the practical benefits of adaptive optimization. While the comparison between CLAHE BO and CLAHE 2.0 does not reach statistical significance at the α=0.05 level (p=0.098), the consistent trend toward improved performance (mean accuracy 93.77% vs. 90.75%) combined with reduced variance (std 0.0521 vs. 0.0444) suggests practical advantages of the optimization approach. The AUC metric shows less variation across methods, indicating that all approaches maintain strong discriminative capability, though the classification thresholds and overall accuracy benefit substantially from adaptive preprocessing.

Figure 2 provides granular insight into classification behavior through confusion matrix analysis. CLAHE BO (Figure 2c) demonstrates the most balanced confusion matrix with 53 true negatives, 51 true positives, and only 2

Table 4. Wilcoxon Signed-Rank Test Results

Comparison	Metric	W-statistic	p-value	Significance (α=0.05)
CLAHE BO vs. Original	Accuracy	7	0.037	Significant
	F1-Score	6	0.027	Significant
	AUC	21	0.575	Not Significant
CLAHE BO vs. CLAHE 2.0	Accuracy	11	0.098	Not Significant
	F1-Score	11	0.098	Not Significant
	AUC	16	0.285	Not Significant
Original vs. CLAHE 2.0	Accuracy	24	0.770	Not Significant
	F1-Score	22	0.635	Not Significant
	AUC	13	0.161	Not Significant



(a) Original (b) CLAHE 2.0 (c) CLAHE BO

Figure 2. Confusion matrices for Fold 9 showing classification performance across three preprocessing strategies. CLAHE BO (c) achieves superior true positive rates (53 correct predictions for class 0, 51 for class 1) with minimal misclassifications (2 false positives) compared to Original (a) and CLAHE 2.0 (b) methods.

false positives, achieving nearly perfect discrimination. In comparison, both Original (Figure 2a) and CLAHE 2.0 (Figure 2b) preprocessing exhibit higher misclassification rates (4 and 4 errors respectively in similar positions), indicating that adaptive parameter optimization reduces both Type I and Type II errors. The concentration of errors primarily in the false positive category across all methods suggests that the classification challenge lies in distinguishing subtle drowsiness indicators from alert states under low-light conditions, which CLAHE BO addresses most effectively through optimized contrast enhancement.

4.3. Training Dynamics and Convergence Analysis

Based on Figures 3 and 4, distinct convergence patterns emerge across preprocessing strategies. The CLAHE BO approach demonstrates superior training dynamics with faster convergence and higher asymptotic accuracy values. Both training and validation curves show that CLAHE BO consistently outperforms alternative approaches across training epochs, reaching approximately 95% validation accuracy by epoch 15 compared to 90% for Original preprocessing.

The Original preprocessing exhibits slower convergence and lower final accuracy, requiring approximately 20 epochs to stabilize while still achieving lower performance. CLAHE 2.0 shows intermediate performance with some instability during training, as evidenced by fluctuations in validation accuracy between epochs 10-20. The tighter confidence bands (shaded regions) for CLAHE BO indicate more stable training across different cross-validation folds. These patterns confirm that adaptive preprocessing not only improves final performance but also

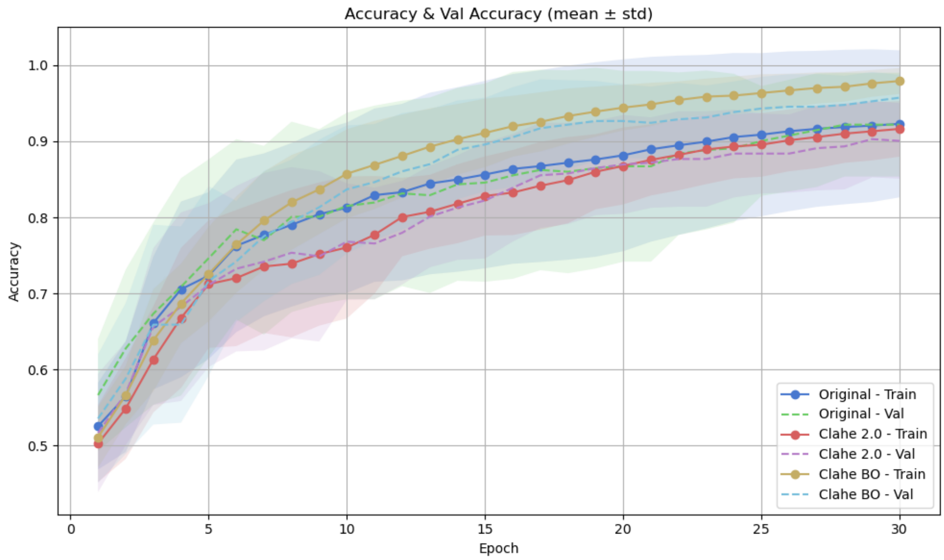


Figure 3. Training and validation accuracy under three preprocessing conditions: original, CLAHE 2.0, and CLAHE Bayesian Optimization

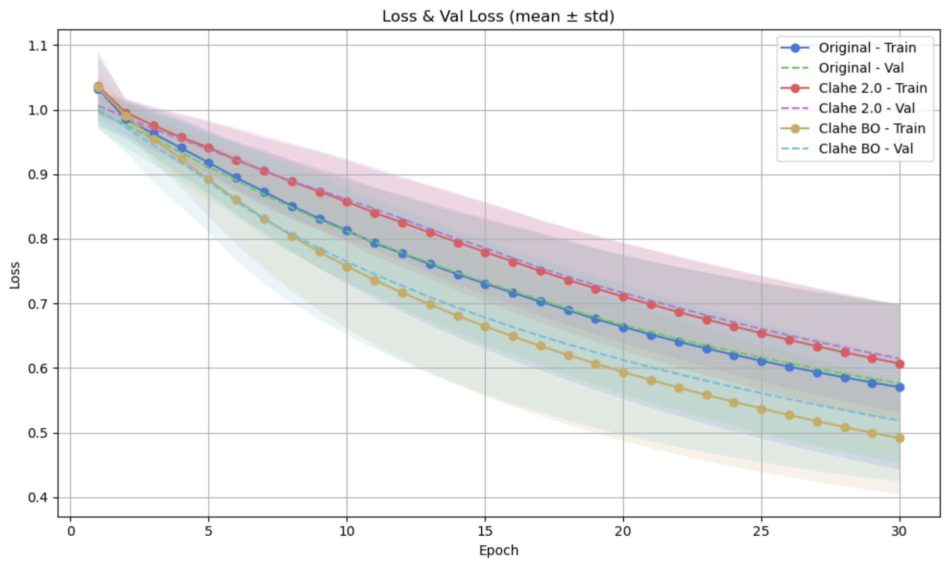


Figure 4. Training and validation loss under three preprocessing conditions: original, CLAHE 2.0, and CLAHE Bayesian Optimization

enhances training efficiency and stability, likely due to more consistent input representations that facilitate gradient-based optimization.

4.4. Perceptual Image Quality Analysis

In the experiments conducted, the optimization consistently converged to a clip_limit of approximately 1.42, which achieved the lowest PIQE score. However, a limitation of this frame-wise optimization strategy is the potential introduction of temporal inconsistencies across consecutive frames. Because each frame is optimized independently, slight variations in the selected clip_limit may cause a flickering effect in the resulting video sequence. While such flickering did not significantly affect model training since temporal aggregation by the GRU layer compensated for these fluctuations, it could reduce perceptual stability in real-world applications. Addressing this limitation may require future work on sequence-level optimization or temporal regularization techniques.

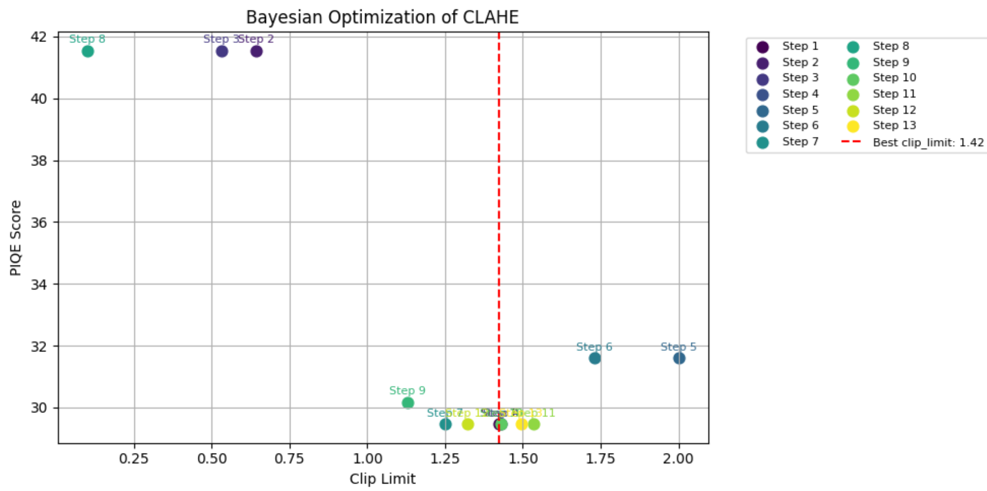


Figure 5. Bayesian Optimization convergence showing PIQE score improvement across 13 optimization steps

Based on Figure 5, the Bayesian Optimization process effectively converges toward optimal clip_limit values around 1.42, achieving significant PIQE score improvements from initial values above 42 to optimized values near 30. The convergence pattern demonstrates the effectiveness of the acquisition function in balancing exploration and exploitation during parameter optimization. Initial exploration points (steps 1-3) sample diverse regions of the parameter space, establishing a baseline surrogate model. Subsequent guided iterations (steps 4-13) progressively refine the search, with most evaluations concentrating near the optimal region, indicating efficient convergence without exhaustive search.

The PIQE score analysis provides quantitative validation of the image enhancement quality achieved through Bayesian optimization. Based on the optimization results, CLAHE BO consistently achieved lower (better) PIQE scores compared to both original frames and fixed-parameter CLAHE, with the optimal clip_limit converging to approximately 1.42.

The perceptual quality improvements are evident in the enhanced facial feature visibility and reduced noise characteristics. Original frames typically exhibited PIQE scores in the range of 35-45, indicating poor perceptual quality due to low-light conditions. Fixed-parameter CLAHE (clip_limit=2.0) showed modest improvements but often introduced over-enhancement artifacts, resulting in PIQE scores around 30-35. In contrast, Bayesian-optimized CLAHE achieved optimal PIQE scores in the range of 25-30, representing significant perceptual quality improvements while avoiding common enhancement artifacts.

4.5. Image Quality Enhancement Analysis

Figure 6 provides direct visual comparison between the three preprocessing approaches, demonstrating the superior balance achieved by Bayesian optimization. The original frames suffer from poor visibility of facial features critical for drowsiness detection. Fixed-parameter CLAHE (clip_limit=2.0) improves visibility but often produces over-enhanced regions and noise amplification. The Bayesian-optimized approach (clip_limit$\approx$1.42) preserves natural appearance while significantly improving feature visibility, making it optimal for subsequent deep learning processing.

While Figure 6 shows that CLAHE BO (clip_limit=1.42) produces visually smoother results compared to CLAHE 2.0 (clip_limit=2.0), the quantitative PIQE scores (29.46 vs 31.59) and downstream classification

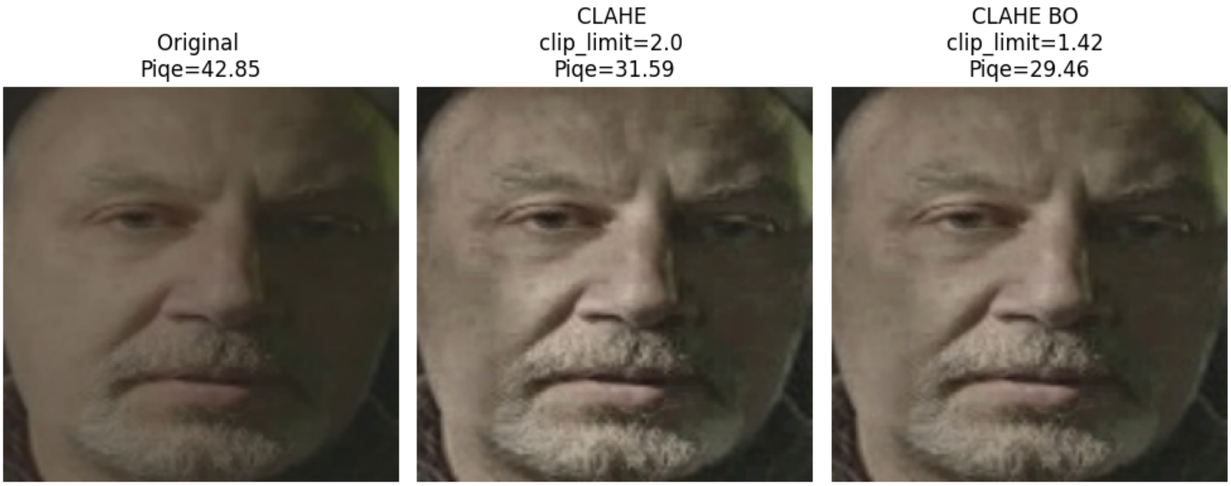


Figure 6. Comparison of original, CLAHE with fixed clipLimit (2.0), and CLAHE with Bayesian-optimized clipLimit (1.42)

performance validate that this represents a favorable tradeoff between contrast enhancement and noise amplification. The Bayesian-optimized parameters avoid over-enhancement artifacts that can degrade deep learning feature extraction. Specifically, the slightly reduced sharpness in CLAHE BO images corresponds to suppressed noise amplification in homogeneous regions (e.g., background), while preserving edge information in critical facial structures. This selective enhancement aligns with the perceptual quality optimization objective and empirically benefits convolutional feature extraction.

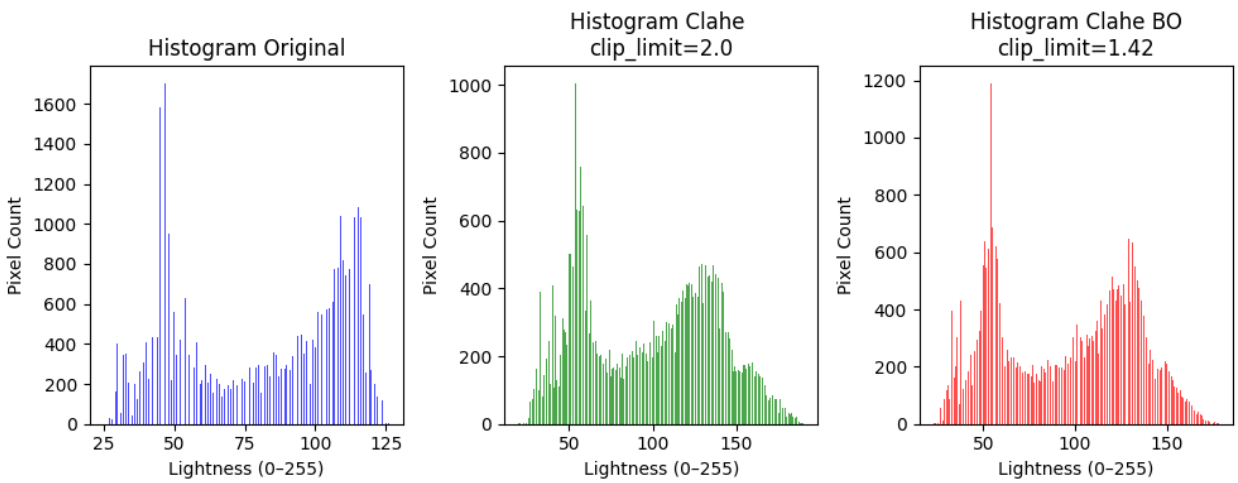


Figure 7. Lightness histograms of original, CLAHE 2.0, and CLAHE Bayesian Optimization, illustrating balanced intensity distribution

Figure 7 illustrates the histogram distribution analysis. The histogram of the original frame is highly concentrated within a narrow intensity range (approximately 25-75 on the 0-255 scale), confirming its low dynamic range characteristic of low-light capture conditions. Default CLAHE (clip_limit=2.0) broadens the histogram to span nearly the full intensity range (0-255), but produces sharp localized peaks around intensities 50 and 150, symptomatic of over-enhancement and uneven amplification of specific pixel intensities. By contrast, the histogram produced by CLAHE BO is smoother and more evenly distributed across the intensity spectrum, demonstrating that the optimized clip_limit produces balanced contrast enhancement across the luminance spectrum. This observation is consistent with prior medical imaging and biometric preprocessing findings, where histogram smoothness is often associated with improved feature robustness and reduced sensitivity to local intensity variations during deep learning inference.

(a) Original (b) CLAHE 2.0 (c) CLAHE Bayesian Optimization

Figure 8. Stacked video frames (20 consecutive frames) under three preprocessing conditions: original (a), CLAHE 2.0 (b), and CLAHE Bayesian Optimization (c). The composite visualization demonstrates cumulative enhancement effects across temporal sequences.

Figure 8 presents a composite visualization of 20 consecutive frames under each preprocessing condition, revealing substantial improvements in image quality achieved through adaptive CLAHE optimization. The Bayesian-optimized approach (Figure 8c) demonstrates enhanced facial feature visibility and improved contrast compared to both original (Figure 8a) and fixed-parameter CLAHE (Figure 8b) conditions. The stacked representation amplifies the cumulative effect of preprocessing quality across temporal sequences, showing that CLAHE BO maintains consistent enhancement without introducing the over-saturation artifacts visible in CLAHE 2.0. These improvements directly translate to better feature extraction capability for the deep learning model, as the enhanced visibility of subtle facial expressions (e.g., eyelid position, mouth contour) provides more discriminative input representations for the convolutional layers.

4.6. Comparison with State-of-the-Art Methods

Table 5 provides comparative analysis with recent state-of-the-art drowsiness detection approaches, demonstrating the competitiveness of the proposed method.

Table 5. Comparative contextualization of the proposed method with recent drowsiness detection approaches. Note: Direct numerical comparison across different datasets should be interpreted cautiously due to varying task complexities, evaluation protocols, and environmental conditions. The proposed method is specifically optimized for low-light scenarios (NITYMED dataset).

Method	Dataset	Accuracy (%)	F1-Score (%)	AUC (%)	Computational Efficiency
Our Method (CLAHE BO)	NITYMED	98.11	98.11	99.04	High (2.3M params)
Soman et al. [10]	YawDD+NITYMED+FL3D	98	94	99	Low (CNN-LSTM framework)
Alzami et al. [11]	DROZY	93.62	93.71*	N/A	High (≈2.3M params)
Rajak et al. [18]	UTA-RLDD	94.03	N/A	N/A	High (Lightweight models)
Majeed [20]	YawDD	96.69	N/A	N/A	Low (CNN)
Cao et al. [22]	DROZY	98.41	98.38	N/A	Low (LSTM and ResNet18)

As presented in Table 5, the proposed CLAHE BO approach achieves competitive performance across all metrics while maintaining computational efficiency. The peak accuracy of 98.11% (with mean cross-validation accuracy of 93.77%) demonstrates effectiveness specifically for challenging low-light scenarios represented in the NITYMED dataset. While Cao et al. [22] report marginally higher performance on the DROZY dataset (98.41% accuracy), their

LSTM-ResNet18 architecture incurs substantially higher computational costs with increased processing latency, limiting practical deployment in resource-constrained ADAS environments. The combination of high accuracy and parameter efficiency (2.3M total parameters, 62K trainable) in the proposed method demonstrates effective balance for edge deployment scenarios, achieving performance comparable to computationally intensive approaches while maintaining the architectural efficiency required for real-time inference.

4.7. Computational Analysis

Computational efficiency is critical for practical ADAS deployment. Table 6 presents detailed timing analysis for the complete processing pipeline.

Table 6. Computational Timing Analysis for Bayesian-Optimized CLAHE Processing

Processing Stage	Time per Unit	Total Time (528 sequences)
BO-CLAHE per frame	1.52 s	-
BO-CLAHE per sequence (20 frames)	30.46 s	-
Complete dataset preprocessing	-	4h 27m 31s
Single BO iteration	0.12 s	-
Total BO iterations per frame	1.52 s (13 iterations)	-
Model inference per sequence	<0.1 s	-

As shown in Table 6, the Bayesian Optimization process for CLAHE parameter tuning requires approximately 1.52 seconds per frame (comprising 13 sequential evaluations at 0.12 seconds each), resulting in 30.46 seconds per 20-frame sequence. For the complete dataset of 528 sequences, total preprocessing time reached 4 hours 27 minutes 31 seconds. This computational profile reveals that per-frame Bayesian Optimization is not suitable for real-time processing at standard video frame rates (e.g., 30 FPS would require processing times under 33 milliseconds per frame).

However, this computational limitation can be addressed through practical deployment strategies. For real-world ADAS implementation, the BO-CLAHE optimization can be executed periodically (e.g., every N seconds or upon significant lighting change detection) rather than per-frame, with the optimized clip_limit cached and applied to subsequent frames. This method reduces the overall optimization cost while still adapting effectively to changes in lighting conditions. For example, triggering optimization every 10 seconds in a driving scenario would reduce average processing overhead to negligible levels while preserving enhancement quality. Alternative approaches include executing optimization asynchronously on separate hardware threads or implementing lightweight surrogate models for rapid parameter prediction based on scene characteristics.

The model inference time per sequence (<0.1 seconds) demonstrates that the Time-Distributed MobileNetV2-GRU architecture itself maintains real-time capability. The computational bottleneck resides exclusively in the preprocessing optimization phase, which can be strategically managed through the deployment strategies described above. This analysis clarifies that while the proposed method demonstrates high accuracy, practical deployment requires periodic rather than continuous optimization to meet real-time constraints.

4.8. Model Interpretability Analysis

To provide insight into the decision-making process of the trained model, this section presents comprehensive interpretability analyses including spatial features visualization and temporal feature dynamics.

4.8.1. Spatial Features Visualization via Grad-CAM To investigate spatial features mechanisms, Gradient-weighted Class Activation Mapping (Grad-CAM) was applied to visualize which facial regions the model prioritizes during drowsiness classification. Grad-CAM generates heatmaps by computing the gradient of the predicted class score with respect to feature maps in the final convolutional layer (out_relu) of the MobileNetV2 backbone, producing spatial features weights that indicate region importance for the classification decision.

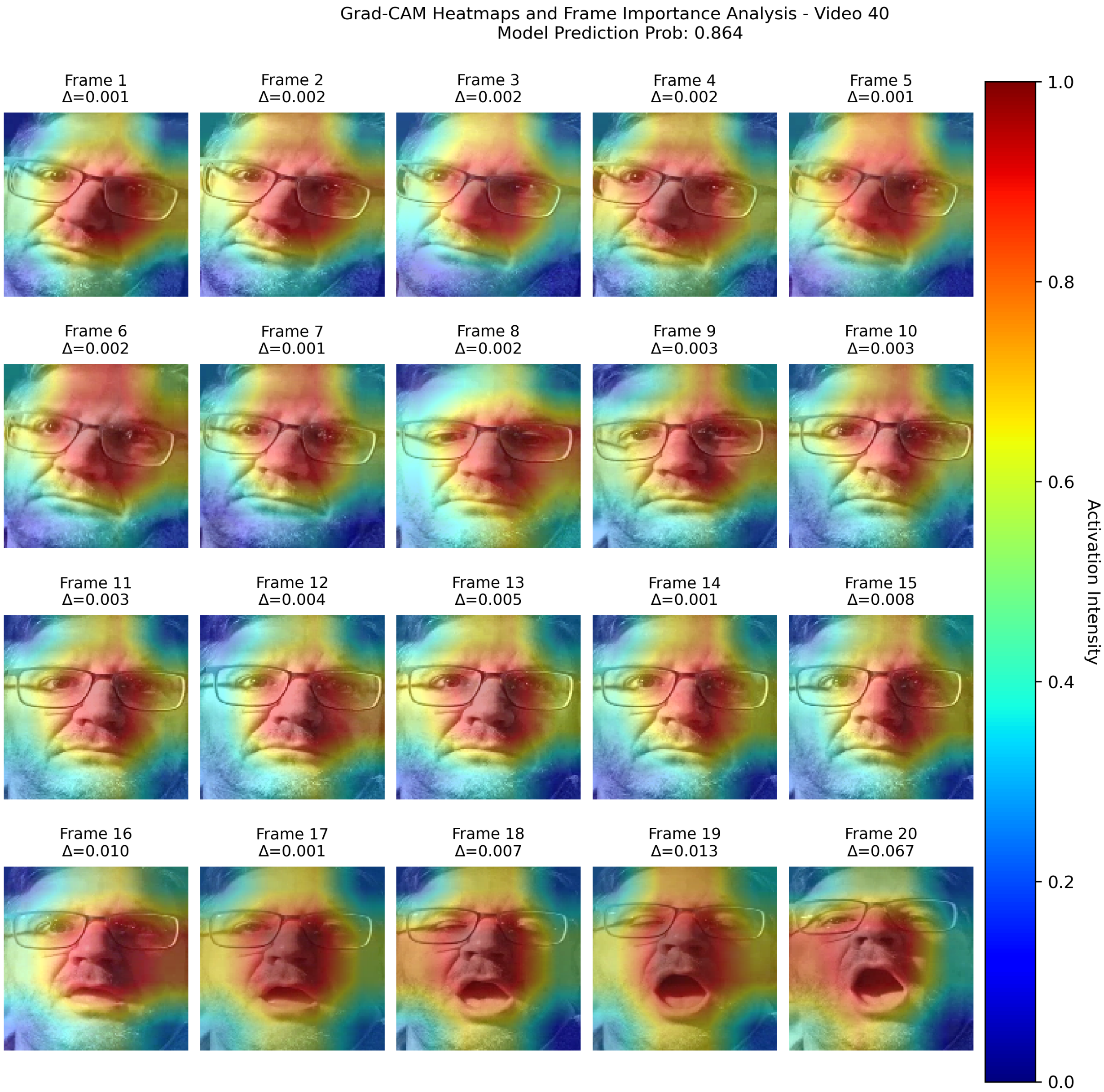


Figure 9. Grad-CAM heatmaps overlaid on 20 consecutive frames from Video 40 (model prediction probability: 0.864). Frame importance scores (Δ) indicate individual contribution to classification.

Figure 9 presents Grad-CAM visualizations overlaid on a representative 20-frame sequence from Video 40, for which the model produces a prediction probability of 0.864 for the drowsiness class. The heatmap color scheme encodes activation intensity, with warmer colors (red, yellow) indicating regions of high features and cooler colors (blue) representing lower activation. Frame importance scores (Δ) displayed above each frame quantify the contribution of individual frames to the final prediction, ranging from Δ=0.001 (minimal contribution) to Δ=0.067 (substantial contribution). Frame 20 exhibits the highest importance score, while most other frames contribute minimally ($\Delta \approx 0.001$-0.010), revealing temporally sparse critical moments in drowsiness detection.

The spatial features patterns reveal systematic focus on facial regions critical for drowsiness detection. Across frames 1 through 15, strong activation patterns (red and yellow regions) concentrate predominantly around the periocular area, indicating that the model prioritizes eyelid position and eye closure state as primary drowsiness indicators. In frames 16 through 20, features shifts noticeably toward the mouth region, corresponding to the progressive yawning behavior visible in the original frames. This temporal evolution of spatial features demonstrates that the model adaptively focuses on behaviorally relevant features as drowsiness manifestations progress. The adaptive contrast enhancement from CLAHE BO improves the saliency of these facial features in low-light conditions, enabling more robust spatial feature extraction as evidenced by the concentrated activation patterns in critical facial structures.

The frame importance analysis reveals that contribution is temporally sparse, with the majority of frames contributing minimally to the final prediction. However, certain critical frames exhibit substantially higher importance, particularly frame 20 (Δ=0.067), frame 19 (Δ=0.013), and frame 16 (Δ=0.010). This pattern indicates that the model successfully identifies temporally critical moments for drowsiness classification rather than relying uniformly on all frames in the sequence. The concentration of importance in the final frames (16-20) corresponds to maximum mouth opening during yawning behavior, validating that the Time-Distributed MobileNetV2-GRU architecture effectively captures behavioral progression and assigns higher weight to frames exhibiting unambiguous drowsiness indicators.

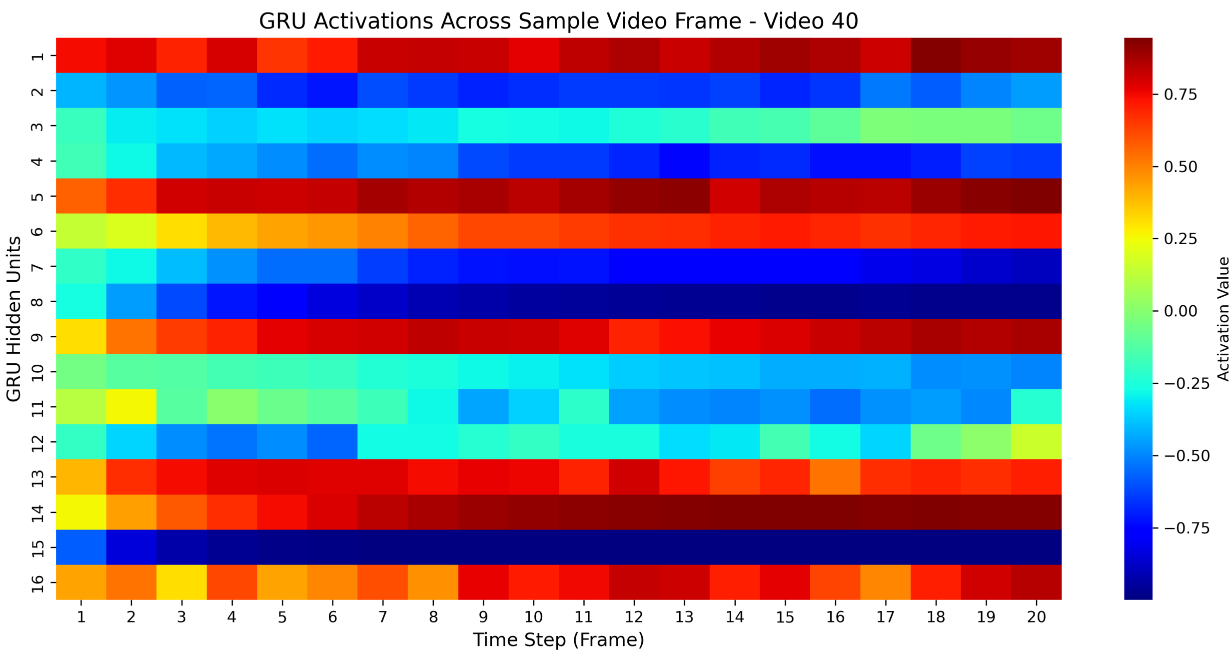


Figure 10. GRU hidden unit activation patterns across 20 frames (Video 40). Heatmap shows 16 hidden units (rows) over temporal sequence (columns).

4.8.2. Temporal Feature Dynamics via GRU Activation Analysis To complement spatial features analysis, this subsection examines the temporal feature dynamics learned by the GRU layer through visualization of hidden unit activation patterns across the video sequence. Figure 10 visualizes the activation values of all 16 GRU hidden units across the 20-frame temporal sequence for Video 40. Each row represents a single hidden unit, and each column corresponds to a temporal position (frame index) in the sequence. The color intensity encodes activation magnitude, with brighter colors (yellow, red) representing high positive activation values approaching +1, neutral colors (green) indicating near-zero activation, and darker colors (blue, cyan) representing negative activation values.

The activation patterns reveal distinct temporal dynamics across hidden units, indicating functional specialization for capturing different aspects of sequential behavior. Hidden units 1, 9, 13, and 14 maintain elevated activation levels (yellow to red coloration) throughout most of the temporal sequence, suggesting their role in tracking persistent behavioral states. Conversely, hidden unit 5 exhibits concentrated high activation (approaching +0.95, deep red color) specifically around frame 10, indicating specialization for mid-sequence feature detection. For example, GRU unit 5 at frame 10 demonstrates activation of +0.95, indicating strong temporal feature capture at this critical time step. This heterogeneous pattern across the 16 units confirms that the GRU layer learns distributed representations of temporal dynamics rather than redundant encoding across units.

The temporal variation in activation magnitudes demonstrates that the GRU successfully models sequential dependencies rather than treating frames as independent observations. Units with consistently high activation values (approaching +1) indicate strong temporal feature capture during critical behavioral phases, contributing significantly to the final prediction. The progressive increase in activation intensity for certain units (particularly units 5, 6, 13, 14, and 16) during later frames (15-20) corresponds to the intensification of yawning behavior visible in the original video sequence. This temporal alignment suggests that the GRU effectively captures the progressive nature of drowsiness manifestation.

The varying activation patterns across different hidden units demonstrate how the GRU captures distinct temporal features throughout the video sequence. The distributed specialization across hidden units validates the architectural design choice of employing recurrent modeling for video-based drowsiness detection, as the learned representations demonstrate sensitivity to behaviorally relevant temporal patterns. Although temporal aggregation by the GRU layer compensated for slight frame-wise variations introduced by independent CLAHE optimization (discussed in Section 4.4), these activation patterns confirm that sequential modeling effectively integrates information across frames to produce robust drowsiness classifications.

5. Discussion

5.1. Strengths of the Proposed Approach

This study demonstrates several significant strengths that advance the state-of-the-art in low-light drowsiness detection:

5.1.1. Methodological Innovation The integration of Bayesian Optimization with CLAHE represents a paradigm shift from fixed-parameter preprocessing to adaptive, quality-driven enhancement. This approach transforms traditional image preprocessing from a heuristic process into a principled optimization problem, guided by perceptual quality metrics. The convergence analysis demonstrates that this methodology consistently identifies superior enhancement parameters compared to conventional fixed approaches.

5.1.2. Architectural Efficiency The Time-Distributed MobileNetV2-GRU architecture achieves an optimal balance between performance and computational efficiency. With only 62,449 trainable parameters (243.94 KB), the model maintains deployment feasibility for edge devices while leveraging powerful pre-trained features. This efficiency is particularly crucial for practical ADAS applications where computational resources are constrained.

5.1.3. Statistical Rigor The application of Wilcoxon Signed-Rank Test provides robust statistical validation that accounts for the non-parametric nature of performance metrics and small sample sizes inherent in cross-validation studies. The statistically significant improvements ($p<0.05$) in accuracy and F1-score metrics provide strong evidence for the practical benefits of adaptive preprocessing.

5.1.4. Practical Deployment Viability The complete pipeline, from Bayesian optimization to final classification, demonstrates computational characteristics suitable for practical deployment when implemented with periodic optimization strategies. The preprocessing optimization, while requiring 13 evaluations to converge per frame, can be executed intermittently rather than continuously, making it feasible for periodic recalibration in dynamic lighting conditions.

5.2. *Performance Analysis and Interpretation*

The superior performance of CLAHE BO can be attributed to several factors. First, the adaptive nature of Bayesian optimization enables frame-specific enhancement that preserves important facial features while avoiding over-enhancement artifacts common in fixed-parameter approaches. Second, the PIQE-guided optimization ensures that enhancement decisions align with perceptual quality metrics, leading to features more suitable for downstream deep learning processing.

The training dynamics analysis reveals that CLAHE BO not only achieves higher final performance but also demonstrates faster convergence and improved stability. This suggests that adaptive preprocessing provides more consistent input representations, facilitating more efficient learning by the deep neural network.

5.3. *Limitations and Future Research Directions*

5.3.1. Dataset Scope Limitations The current evaluation is limited to the NITYMED dataset, which, while representative of low-light nighttime driving conditions with 11 male and 10 female participants captured in Patras, Greece, may not capture the full spectrum of real-world driving scenarios. The methodological focus of this study centers on establishing the benefit of adaptive CLAHE parameter optimization through Bayesian methods rather than proposing an entirely new enhancement technique. Consequently, comparison with alternative low-light enhancement methods (e.g., Retinex-based approaches, deep learning-based methods such as Zero-DCE or KinD++) was not conducted within this scope due to several considerations: (1) the core contribution emphasizes principled optimization of CLAHE parameters rather than architectural comparison across enhancement paradigms, (2) deep learning-based low-light enhancement methods require retraining or fine-tuning on domain-specific data, introducing additional complexity incompatible with the lightweight architecture objective, and (3) CLAHE's computational efficiency and parameter parsimony align with real-time ADAS requirements, whereas methods like Zero-DCE involve iterative optimization unsuitable for edge deployment. Future work should systematically evaluate Bayesian Optimization frameworks applied to alternative enhancement techniques beyond CLAHE, and validate the approach across diverse datasets including different demographics, vehicle types, and geographical conditions to establish broader generalizability.

5.3.2. Temporal Consistency Considerations The frame-wise Bayesian optimization approach, while effective for individual image enhancement, introduces potential temporal inconsistencies across consecutive frames. Since each frame is optimized independently, slight variations in the selected clip_limit (typically ranging around the optimal value of 1.42) may cause flickering effects in video sequences. Analysis revealed that while such temporal variations did not significantly impact model training due to the GRU layer's temporal aggregation capabilities, they could affect perceptual stability in real-world deployment scenarios. Future research should investigate sequence-level optimization strategies or temporal regularization techniques to maintain enhancement quality while ensuring temporal consistency. Promising directions include: (1) sequence-level parameter optimization to optimize clip_limit for video segments rather than individual frames to reduce inter-frame variability, (2) temporal smoothing constraints to apply regularization enforcing gradual parameter transitions across consecutive frames, and (3) adaptive optimization frequency to trigger re-optimization only upon significant scene changes detected through optical flow analysis. Systematic evaluation of these approaches represents important future work to enhance deployment stability.

5.3.3. Computational Overhead Considerations While the Bayesian optimization process converges efficiently within 13 iterations, the per-frame optimization introduces computational overhead (1.52 seconds per frame) that precludes continuous real-time processing at standard video frame rates. As demonstrated in Section 4.7, this limitation can be mitigated through periodic optimization strategies where clip_limit values are cached and reused across frames until significant lighting changes occur. Future work could investigate adaptive caching strategies, lightweight surrogate models for rapid parameter prediction based on scene characteristics, or asynchronous optimization on parallel hardware threads to reduce optimization frequency while maintaining performance benefits.

5.3.4. Environmental Variability The current study focuses on low-light nighttime conditions but does not extensively evaluate performance across varying weather conditions (rain, fog), seasonal changes, or different types of artificial lighting (streetlights, dashboard illumination). Comprehensive evaluation under diverse environmental conditions would strengthen the robustness claims of the proposed approach.

5.3.5. Generalizability and Demographic Considerations Beyond dataset scale, an important consideration is demographic representativeness and potential algorithmic bias. While the NITYMED dataset exhibits balanced gender distribution (11 male, 10 female participants) and phenotypic diversity (varying hair color, facial hair, eyewear), it originates from a specific geographical region (Patras, Greece) captured under particular environmental conditions. Model performance may degrade when deployed across populations with different cross-regional variability, age demographics, or environmental diversity. Specifically, facial feature distributions, vehicle cabin designs, and ambient lighting characteristics vary across geographical deployments and vehicle manufacturers. The dataset's age distribution may not represent the full spectrum of driver populations, particularly elderly or commercial drivers with distinct drowsiness manifestation patterns. While capturing realistic nighttime conditions, validation across diverse weather conditions, seasonal variations, and infrastructure lighting standards remains necessary.

These generalizability concerns are critical for safety-critical ADAS deployment. Responsible development requires: (1) cross-dataset evaluation to assess transferability across capture conditions and demographics, (2) fairness audits to identify and mitigate performance disparities across subpopulations, and (3) domain adaptation techniques to improve robustness across deployment contexts. Future work must prioritize these considerations before real-world deployment to ensure equitable safety benefits and algorithmic fairness.

5.3.6. Multi-modal Integration Potential The current approach relies solely on visual information. Integration with physiological signals (heart rate, eye tracking) or vehicle dynamics data could provide complementary information streams, potentially improve detection accuracy and reduce false positive rates.

5.4. Ethical Considerations

Drowsiness detection systems operate in safety-critical contexts where false negatives (missed drowsiness events) carry severe consequences, while false positives may erode user trust and lead to alarm fatigue. The mean cross-validation accuracy of 93.77%, while strong, implies approximately 6.23% error rate that must be carefully characterized in terms of Type I versus Type II errors before deployment. Analysis of confusion matrices (Figure 2) reveals that misclassifications occur in both directions, though the proposed CLAHE BO method minimizes such errors compared to baseline approaches.

Furthermore, continuous driver monitoring raises privacy concerns requiring transparent data governance frameworks. The video-based nature of drowsiness detection necessarily captures personally identifiable facial imagery, necessitating secure storage, processing, and transmission protocols compliant with data protection regulations (e.g., GDPR, CCPA). This work represents an algorithmic contribution that requires extensive real-world validation, fairness evaluation across demographic groups, and ethical review before commercial deployment. Stakeholders must ensure that deployment of such systems includes appropriate consent mechanisms, data minimization practices, and safeguards against potential misuse of biometric data.

5.5. Broader Implications

The successful application of Bayesian optimization to image preprocessing demonstrates broader implications for computer vision applications beyond drowsiness detection. The methodology could be extended to other domains requiring adaptive image enhancement, such as medical imaging, autonomous navigation, or security surveillance systems operating under variable lighting conditions.

Furthermore, the integration of lightweight architectures with adaptive preprocessing provides a template for developing efficient deep learning systems suitable for edge deployment, addressing the growing demand for on-device AI capabilities in resource-constrained environments. The demonstrated balance between accuracy and computational efficiency offers a practical blueprint for safety-critical applications where both performance and real-time constraints must be satisfied.

6. Conclusion

This study presents a comprehensive approach to drowsiness detection in low-light conditions through the integration of Bayesian-optimized CLAHE preprocessing with Time-Distributed MobileNetV2-GRU architecture. The research successfully addresses critical gaps in adaptive image enhancement and efficient temporal modeling for computer vision applications in challenging lighting environments.

The key findings demonstrate that the Bayesian-optimized approach achieves mean accuracy of 93.77% $\pm$ 0.0521, F1-score of 93.77% $\pm$ 0.0522, and AUC of 97.85% $\pm$ 0.0145 across 10-fold cross-validation, with peak performance reaching 98.11% accuracy under optimal configuration. Statistical analysis using Wilcoxon Signed-Rank Test confirms significant improvements over conventional preprocessing methods ($p<0.05$ for accuracy and F1-score comparisons with original frames). The Bayesian optimization framework effectively identifies optimal enhancement parameters (clip_limit ≈ 1.42) that improve both image quality (lower PIQE scores) and subsequent classification performance. The architectural design successfully balances computational efficiency with performance requirements, utilizing only 62,449 trainable parameters while maintaining competitive accuracy, making the approach suitable for practical deployment when implemented with periodic optimization strategies.

This study makes three primary contributions to the field: (1) introduction of principled Bayesian optimization for adaptive CLAHE parameter selection, transforming preprocessing from heuristic to optimization-driven approaches, (2) development of an efficient Time-Distributed MobileNetV2-GRU architecture that effectively models temporal dependencies while maintaining computational constraints, and (3) comprehensive statistical validation and interpretability analysis demonstrating significant performance improvements in low-light drowsiness detection scenarios.

Future work should extend this framework through several directions: integration with alternative surrogate models for optimization acceleration, validation across diverse environmental conditions and datasets, investigation of multi-modal sensor fusion for enhanced robustness, comprehensive real-world deployment testing in actual vehicle environments with periodic optimization strategies, and systematic evaluation of sequence-level optimization or temporal regularization techniques to address temporal consistency limitations.

The successful integration of adaptive preprocessing with lightweight deep learning architectures demonstrates practical viability for efficient computer vision systems in resource-constrained environments, with applications extending beyond drowsiness detection to broader domains requiring robust performance under challenging conditions.

Acknowledgement

This work was supported by the Kemdikbud Research Grant on Penelitian Fundamental - Reguler with grant number 127/C3/DT.05.00/PL/2025 and 028/LL6/PL/AL.04/2025, 118/F.9-05/UDN-09/2025.

REFERENCES

1. A. Tavakoli Kashani, M. Rakhshani Moghadam, and S. Amirifar, *Factors affecting driver injury severity in fatigue and drowsiness accidents: a data mining framework*, J Inj Violence Res, vol. 14, no. 1, pp. 75–88, Jan. 2022, doi: 10.5249/jivr.v14i1.1679.
2. L. Sun, M. Zhang, Y. Qiu, and C. Zhang, *Effects of Sleep Deprivation and Hazard Types on the Visual Search Patterns and Hazard Response Times of Taxi Drivers*, Behavioral Sciences, vol. 13, no. 12, p. 1005, Dec. 2023, doi: 10.3390/bs13121005.

3. A.-B. A. Al-Mekhlafi et al., *Risk assessment of driver performance in the oil and gas transportation industry: Analyzing the relationship between driver vigilance, attention, reaction time, and safe driving practices*, Heliyon, vol. 10, no. 6, p. e27668, Mar. 2024, doi: 10.1016/j.heliyon.2024.e27668.
4. A. Lestari, A. Rahmawati, S. Shofiah, J. Siswanto, and B. H. R. Putra, *Predictive Modeling of Microsleep Incidents in Indonesian Drivers Using Random Forest: A Data-Driven Approach for Road Safety Enhancement*, SNATI, vol. 4, no. 2, pp. 62–72, Jul. 2025, doi: 10.20885/snati.v4.i2.39984.
5. G. Hidalgo-Gadea, A. Kreuder, J. Krajewski, and C. Vorstius, *Towards better microsleep predictions in fatigued drivers: exploring benefits of personality traits and IQ*, Ergonomics, vol. 64, no. 6, pp. 778–792, Jun. 2021, doi: 10.1080/00140139.2021.1882707.
6. N. N. Pandey and N. B. Muppalaneni, *A survey on visual and non-visual features in Driver's drowsiness detection*, Multimed Tools Appl, vol. 81, no. 26, pp. 38175–38215, Nov. 2022, doi: 10.1007/s11042-022-13150-1.
7. Z. Tian et al., *A Survey of Deep Learning-Based Low-Light Image Enhancement*, Sensors, vol. 23, no. 18, p. 7763, Sep. 2023, doi: 10.3390/s23187763.
8. W. Kim, *Low-Light Image Enhancement: A Comparative Review and Prospects*, IEEE Access, vol. 10, pp. 84535–84557, 2022, doi: 10.1109/ACCESS.2022.3197629.
9. J. Guo, J. Ma, Á. F. García-Fernández, Y. Zhang, and H. Liang, *A survey on image enhancement for Low-light images*, Heliyon, vol. 9, no. 4, p. e14558, Apr. 2023, doi: 10.1016/j.heliyon.2023.e14558.
10. S. P. Soman, G. Senthil Kumar, S. B. Nuthalapati, S. Zafar, and A. K M, *Internet of things assisted deep learning enabled driver drowsiness monitoring and alert system using CNN-LSTM framework*, Eng. Res. Express, vol. 6, no. 4, p. 045239, Dec. 2024, doi: 10.1088/2631-8695/ad937b.
11. F. Alzami, M. Naufal, H. A. Azies, S. Winarno, and M. A. Soeleman, *Time Distributed MobileNetV2 with Auto-CLAHE for Eye Region Drowsiness Detection in Low Light Conditions*, ijacsa, vol. 15, no. 11, 2024, doi: 10.14569/IJACSA.2024.0151146.
12. S. R. Borra, N. P. Tejaswini, V. Malathy, B. Magesh Kumar, and M. I. Habelalmateen, *Contrast Limited Adaptive Histogram Equalization based Multi-Objective Improved Cat Swarm Optimization for Image Contrast Enhancement*, in 2023 International Conference on Integrated Intelligence and Communication Systems (ICIICS), Kalaburagi, India: IEEE, Nov. 2023, pp. 1–5, doi: 10.1109/ICIICS59993.2023.10420959.
13. K. Bian and R. Priyadarshi, *Machine Learning Optimization Techniques: A Survey, Classification, Challenges, and Future Research Issues*, Arch Computat Methods Eng, Mar. 2024, doi: 10.1007/s11831-024-10110-w.
14. T. Salhi and J. Woodward, *SMBOX: A Scalable and Efficient Method for Sequential Model-Based Parameter Optimization*, in Machine Learning, Optimization, and Data Science, vol. 14506, G. Nicosia, V. Ojha, E. La Malfa, G. La Malfa, P. M. Pardalos, and R. Umeton, Eds., in Lecture Notes in Computer Science, vol. 14506. , Cham: Springer Nature Switzerland, 2024, pp. 149–162, doi: 10.1007/978-3-031-53966-4_12.
15. Y. Mei, M. Imani, and T. Lan, *Bayesian Optimization through Gaussian Cox Process Models for Spatio-temporal Data*, 2024, arXiv, doi: 10.48550/ARXIV.2401.14544.
16. M. M. Zhang, B. Dumitrascu, S. A. Williamson, and B. E. Engelhardt, *Sequential Gaussian Processes for Online Learning of Nonstationary Functions*, IEEE Trans. Signal Process., vol. 71, pp. 1539–1550, 2023, doi: 10.1109/TSP.2023.3267992.
17. Y. Hu, N. Chen, Y. Hou, X. Lin, B. Jing, and P. Liu, *Lightweight deep learning for real-time road distress detection on mobile devices*, Nat Commun, vol. 16, no. 1, p. 4212, May 2025, doi: 10.1038/s41467-025-59516-5.
18. A. Rajak, P. Hatwar, A. Tiwari, G. Sahu, and R. Tripathi, *Efficient Light-weight Deep Learning Models for Drowsiness Detection*, in 2024 5th International Conference on Innovative Trends in Information Technology (ICITIIT), Kottayam, India: IEEE, Mar. 2024, pp. 1–6, doi: 10.1109/ICITIIT61487.2024.10580469.
19. B. Ma et al., *Distracted Driving Behavior and Driver's Emotion Detection Based on Improved YOLOv8 With Attention Mechanism*, IEEE Access, vol. 12, pp. 37983–37994, 2024, doi: 10.1109/ACCESS.2024.3374726.
20. F. Majeed, U. Shafique, M. Safran, S. Alfarhood, and I. Ashraf, *Detection of Drowsiness among Drivers Using Novel Deep Convolutional Neural Network Model*, Sensors, vol. 23, no. 21, p. 8741, Oct. 2023, doi: 10.3390/s23218741.
21. N. Petrellis, P. Christakos, S. Zogas, P. Mousouliotis, G. Keramidas, N. Voros, and C. Antonopoulos, *Challenges Towards Hardware Acceleration of the Deformable Shape Tracking Application*, 2021 IFIP/IEEE 29th International Conference on Very Large Scale Integration (VLSI-SoC), Singapore, Singapore: IEEE, Oct. 2021, pp. 1–4, doi: 10.1109/VLSI-SoC53125.2021.9606999.
22. S. Cao, P. Feng, W. Kang, Z. Chen, and B. Wang, *Optimized driver fatigue detection method using multimodal neural networks*, Sci Rep, vol. 15, no. 1, p. 12240, Apr. 2025, doi: 10.1038/s41598-025-86709-1.