Comparison of Histogram Equalization and Multi-Scale Retinex Methods for Near-Infrared Image Enhancement in Drowsiness Detection

Moh Hadi Subowo[1], Pulung Nurtantio Andono[2], Guruh Fajar Shidik[3], Heru Agus Santoso[4]
Doctoral Program in Computer Science, Universitas Dian Nuswantoro, Semarang, Indonesia 50131[1,2,3,4]
Information Technology-Faculty of Science and Technology, UIN Walisongo Semarang, Indonesia[1]

***Abstract*—Computer vision-based drowsiness detection faces significant challenges in low-light conditions, particularly when using near-infrared (NIR) sensors for driver monitoring systems. Appropriate image enhancement methods are crucial to improve detection accuracy. This study systematically evaluates five enhancement methods: Histogram Equalization (HE), Adaptive Histogram Equalization (AHE), Contrast-Limited Adaptive Histogram Equalization (CLAHE), Brightness Preserving Dynamic Histogram Equalization (BPDHE), and Multi-Scale Retinex with Color Restoration (MSRCR). The evaluation was conducted on 4,272 frames from the University of Liège (ULg) Multimodality Drowsiness Database (DROZY) using four no-reference metrics: Natural Image Quality Evaluator (NIQE), Perception-based Image Quality Evaluator (PIQE), Shannon Entropy, and Lightness Order Error (LOE). Additional validation was performed by measuring the face detection rate using MediaPipe. The results show that CLAHE achieves an optimal balance with an NIQE of 4.61 (best natural quality), a detection rate of 97.9%, and an LOE of 0.058 (superior structural preservation). MSRCR produces the highest entropy (6.58) but the lowest detection rate (75.6%), indicating structural distortion in the NIR context. Statistical validation using the Wilcoxon signed-rank test and the Friedman test confirmed the significance of the findings ($p < 0.05$). CLAHE is recommended for NIR surveillance-based drowsiness detection systems.**

Keywords*—Image enhancement; near-infrared; drowsiness detection; histogram equalization; multi-scale retinex; CLAHE; no-reference quality metrics

I. Introduction

Traffic accidents caused by drowsiness contribute to 20-30% of global road deaths. Automatic drowsiness detection systems are a critical component of modern automotive safety. Computer vision technology offers non-invasive solutions through analysis of facial expressions and driver behavior patterns. However, the performance of these systems is highly dependent on the quality of the input image.

Near-infrared (NIR) sensors are increasingly used for driver monitoring because they can operate in low-light conditions [1]. Unlike visible-light RGB cameras, NIR sensors are unaffected by variations in ambient light and can operate at night without additional lighting distracting the driver [2]. The DROZY dataset, a leading benchmark for drowsiness detection research, uses a Kinect v2 NIR sensor, which represents realistic low-light surveillance conditions [3].

NIR images have different characteristics than conventional RGB images. The spectral response of NIR is limited to wavelengths of 700-1000 nm, resulting in essentially monochrome images with a narrow dynamic range [4]. The lack of chrominance information eliminates the advantages of color restoration-based enhancement methods. Furthermore, NIR sensors are sensitive to thermal noise, especially at low light intensities [5]. These unique characteristics mean that enhancement methods designed for natural RGB images may not be optimal for NIR surveillance scenarios.

Image enhancement is a crucial preprocessing step in the drowsiness detection pipeline. Lin et al. demonstrated that an appropriate enhancement algorithm can improve detection accuracy by up to 5% on a realistic driving dataset [6]. Aprilia et al. demonstrated that integrating CLAHE with YOLOv5 resulted in a Mean Average Precision of 0.959 for drowsiness detection [7]. Face detection rate, a necessary condition for subsequent facial landmark analysis, is significantly affected by the quality of preprocessing. Methods that fail to preserve facial geometry hinder the extraction of features such as eye closure ratio and mouth opening, which are key indicators of drowsiness [8].

The literature identifies two main categories of enhancement methods: histogram-based and Retinex-based. Histogram-based methods, including HE and its variants (AHE, CLAHE, BPDHE), modify the pixel intensity distribution to enhance global or local contrast [9]. CLAHE is particularly popular for its ability to limit noise amplification through a clip limit parameter [7]. Retinex-based methods, with MSRCR as a representative example, use a multi-scale approach to separate the illumination and reflectance components in images [10]. MSRCR theoretically excels in natural image enhancement due to its sophisticated color restoration component [11].

However, systematic comparisons of these methods specifically for the context of NIR drowsiness detection are limited. Wang et al. conducted a comprehensive review of enhancement methods but focused on RGB images [9]. Zhan et al. surveyed modern low-light enhancement techniques, including deep learning approaches, but did not specifically analyze NIR surveillance scenarios [12]. Zhang et al. explored NIR-RGB fusion for denoising but focused on inconsistency handling, not enhancement comparison [4]. Liu et al. examined the NIR-VIS domain gap for face recognition, not drowsiness detection tasks [13].

This gap is significant because practitioners need empirical guidance to select enhancement methods that fit their deploy-

ment constraints. Real-time drowsiness systems face a trade-off between enhancement quality and computational efficiency. Methods that are too complex are not feasible for embedded automotive platforms. Furthermore, over-aggressive enhancement can distort facial structure, causing false negatives in the face detection module [14]. Comprehensive evaluation using multiple quality metrics and practical validation through detection rate can provide actionable insights for system designers.

This study fills this gap by systematically evaluating five enhancement methods on 4,272 NIR frames from the DROZY dataset. The evaluation focuses on frame-level image quality assessment to establish empirical foundations for preprocessing method selection, with face detection serving as a practical bridge between enhancement quality and downstream task requirements. Specific contributions include: 1) a comprehensive comparison of histogram-based methods (HE, AHE, CLAHE, BPDHE) versus the Retinex-based method (MSRCR) in the NIR context, 2) a multi-dimensional evaluation using four no-reference quality metrics that capture different aspects of image quality, 3) practical validation using face detection rate as a proxy for downstream task suitability, and 4) a rigorous statistical analysis using non-parametric tests to confirm the significance of the findings.

This paper is organized as follows. Section II reviews the literature related to enhancement methods, NIR imaging challenges, and quality assessment metrics. Section III describes the experimental methodology, including dataset characteristics, implementation details, and statistical analysis procedures. Section IV presents the quantitative and qualitative evaluation results. Section V discusses the interpretation of the findings, practical implications, and limitations. Section VI concludes with recommendations and future work directions.

II. Related Works

A. Drowsiness Detection Systems

Computer vision has become the dominant approach for detecting driver drowsiness. Camera-based systems analyze physiological facial changes such as eye closure rate, yawn frequency, and head pose deviation. Cao et al. [15] integrated visual data with polysomnographic signals, achieving 95% accuracy in a sleep-deprived scenario. This multimodal approach is robust but requires additional sensors. Single-camera systems are more practical for mass deployment, as demonstrated by Winarno et al. [8] using EfficientNet-KNN with Haar Cascade preprocessing on the DROZY dataset.

Preprocessing quality determines the success of downstream analysis. Lin et al. demonstrated that an appropriate enhancement algorithm increased detection accuracy to 95.3% compared to 90.1% without preprocessing [6]. Face detection is a necessary condition for landmark extraction. Methods that fail to preserve facial geometry will hinder the calculation of Eye Aspect Ratio (EAR) and Mouth Aspect Ratio (MAR), two key indicators of drowsiness [8]. Aprilia et al. demonstrated CLAHE integration with YOLOv5 resulting in a mAP of 0.959, explicitly outperforming the model without enhancement [7]. Enhancement not only improves visibility; it preserves structural information for feature extraction.

B. Low-Light Enhancement Methods

Histogram-based methods modify the pixel intensity distribution. HE applies a global transformation to flatten the histogram, increasing the overall contrast of the image [9]. The simplicity of HE makes it popular for real-time applications, although Wang et al. identified noise amplification as a major drawback in low-light images. Over-enhancement of bright regions results in a washed-out appearance, while dark regions remain underexposed.

AHE overcomes this limitation of HE through local histogram equalization. Each pixel is transformed based on the distribution of its neighborhood. The exposure equalization approach of AHE produces better detail in dark regions [16]. The computational cost of AHE is higher because it requires calculating a local histogram for each pixel. Aprilia et al. found AHE effective for data enrichment in drowsiness detection datasets [7].

CLAHE introduces a clip limit to limit contrast amplification. This parameter prevents the excessive noise amplification that occurs in standard AHE [17]. Tile size determines the granularity of local enhancement. Small tiles produce superior local detail, while large tiles provide a more natural appearance. Wang et al. used CLAHE for nighttime road image enhancement, increasing visibility by up to 23% [18].

Ibrahim and Kong developed BPDHE for medical imaging applications which separates images based on mean brightness, performing independent equalization on sub-images preserves the original mean brightness, avoiding over-enhancement artifacts [19]. This method is less popular than CLAHE, although it theoretically offers better brightness preservation.

Retinex theory is based on human perception models. Jobson et al. developed Multi-Scale Retinex with Color Restoration (MSRCR) to separate illumination and reflectance components [10]. Multiple Gaussian scales capture detail at various spatial frequencies. The color restoration function prevents the graying effect that occurs in basic Retinex [11]. Jobson et al. standardized MSRCR parameters for automatic enhancement without manual tuning [11].

MSRCR excels on natural outdoor images. Li et al. applied MSRCR for underwater image enhancement, addressing color cast and backscatter issues [20]. Wang et al. used MSRCR on maize leaves to improve maize leaf disease recognition visualization [21]. Zhang demonstrated that MSRCR is effective for dim nighttime images, although it requires post-processing denoising [22].

Known limitations exist. Petro et al. documented halo artifacts and graying issues in spatially extensive homogeneous regions [23]. MSRCR's gray-world assumptions violate scenes with single-color dominance. Rahman et al. identified monochrome images as a pathological case where the color restoration component is ineffective [24]. Liu et al. demonstrated that MSRCR can aggravate scattering effects in turbid water conditions [25]. Computational intensity is also a concern for real-time systems. Wang et al. noted that multiple Gaussian filtering operations require significant processing power [26].

C. NIR-Specific Challenges

Near-infrared imaging differs fundamentally from visible-light RGB. Its spectral response is limited to 700–1000 nm, resulting in essentially monochrome images [4]. The absence of color information eliminates the benefits of chromatic-based enhancement methods. NIR sensors also have different sensitivity characteristics, responding more strongly to reflected light from skin tissue than to ambient illumination [24]. Content inconsistency between NIR and RGB modalities poses major challenges for fusion approaches [4], as shadows visible in RGB are often invisible in NIR due to differences in wavelength penetration. Shadows visible in RGB are often invisible in NIR due to differences in wavelength penetration. Direct fusion produces ghosting artifacts.

Yang et al. showed that multiple RGB colors can map to a single NIR intensity, creating mapping ambiguity for colorization tasks [27]. Liu et al. argued that the domain gap is not only spectral but also distributional [13]. The feature distributions of NIR and visible light differ significantly, requiring domain adaptation approaches.

Driver monitoring systems increasingly adopt NIR sensors. Nowara et al. extracted heart rate signals using NIR imaging photoplethysmography, demonstrating its feasibility for non-invasive physiological monitoring [1]. Tu et al. benchmarked deep learning methods for automotive imaging PPG, validating NIR narrow-band filters robustness against extreme lighting variations [2]. Borghi et al. established baselines for driver verification using depth sensors, tackling in-the-wild cockpit lighting challenges [28].

DROZY dataset represents realistic NIR surveillance conditions. Massoz et al. documented Kinect v2 sensor characteristics, noting 512×424 resolution at 30 fps [3]. Pre-annotated 68 facial landmarks provided to bypass face detection difficulties. Bodaghi et al. compared DROZY with newer datasets, highlighting DROZY primary limitation as a fixed low-light environment without diverse in-the-wild conditions [14]. Ghoddoosian et al. used UTA-RLDD dataset, noting real-life drowsiness differs significantly from acted drowsiness [29]. Asdyo et al. found facial landmarks (EAR/MAR) remain robust across light conditions once face detected, but initial detection remains a bottleneck [30].

D. Image Quality Assessment

No-reference metrics are essential because pristine reference images are not available in surveillance contexts. NIQE measures statistical distance between image features and natural scene statistics model [31]. Lower NIQE indicates closer alignment with natural image characteristics. Sabry et al. validated NIQE for nighttime driving images, showing correlation with human perceptual quality [32].

PIQE quantifies block-wise distortion based on human perception models [33]. Lower scores indicate higher perceptual quality. Wang et al. established interpretation standard: PIQE effectively detects artifacts introduced by aggressive enhancement algorithms [34]. Saleem et al. demonstrated PIQE robustness for underwater image evaluation, confirming utility across diverse domains [35].

Shannon Entropy measures information content or detail richness in an image [36]. Higher entropy suggests more visible details, although extremely high values can indicate excessive noise. Daway et al. explored the trade-offs between Entropy and Lightness Order Error (LOE), arguing good enhancement should maximize entropy while minimizing lightness order reversals [37]. Li et al. established Entropy as a critical metric for enhanced evaluation in comprehensive surveys [38].

LOE quantifies preservation of relative brightness ordering [39]. Algorithm samples pixel pairs, comparing whether brightness relationships preserved after enhancement. Lower LOE indicates better structural preservation. Wang et al. emphasized the importance of LOE for ensuring enhancement does not invert natural lighting structure, critical for object recognition tasks [38].

E. Gap Analysis

Existing studies compare enhancement methods in RGB contexts or apply NIR for different tasks. Wang et al. provided comprehensive categorization of enhancement algorithms, but evaluation focused on visible-light natural images [9]. Jingchun et al. surveyed modern low-light techniques including deep learning, but did not specifically analyze NIR surveillance scenarios [12]. NIR domain received attention for facial recognition and expression analysis. Liu et al. examined domain gap for heterogeneous face recognition [13]. Luo et al. proposed stochastic differential equations for NIR-to-VIS translation, preserving facial details [40]. Yang et al. explored colorization challenges using cooperative learning [27].

Different tasks require different quality criteria; recognition prioritizes feature distinctiveness; drowsiness detection prioritizes structural preservation for landmark extraction. No systematic evaluation exists for histogram variants versus Retinex methods specifically for NIR drowsiness datasets with comprehensive metrics plus detection validation. This gap is significant because practitioners need empirical guidance to select enhancement methods suitable for deployment constraints. The trade-offs between enhancement quality, computational efficiency, and downstream task compatibility remain underexplored. Our work addresses this gap through focused evaluation in the NIR business context, providing actionable insights for automotive safety system designers.

III. Methods

A. Dataset and Research Methodology Flow

This study used the DROZY (ULg Multimodality Drowsiness Database). The dataset contains videos from 14 subjects across three recording sessions using Microsoft's Kinect v2 NIR sensor. The resolution is 512×424 pixels with a frame rate of 30 fps. Subjects 1 to 8 were recorded at 15 fps for sessions 2 and 3 due to a recording error in the dark [3].

All videos were sampled every five seconds. This interval balances dataset size with temporal independence, avoiding duplicate frames while capturing sufficient variation across drowsiness states. A total of 4,272 frames were extracted. Each frame underwent identical preprocessing before enhancement: no resizing, no filtering, and the original bit depth was maintained to preserve sensor characteristics. The NIR nature of

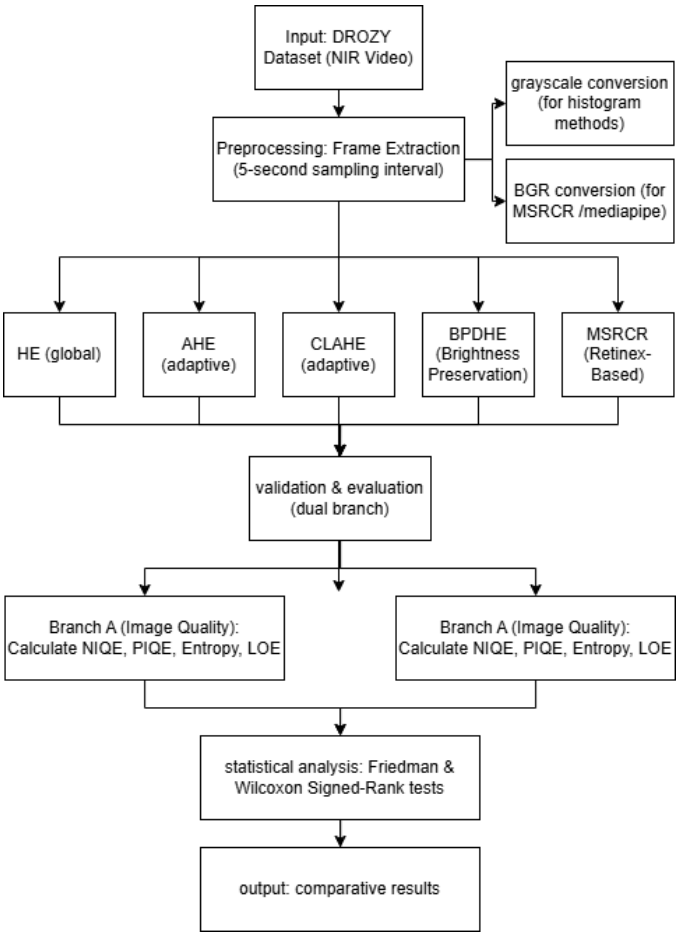


Fig. 1. Research methodology pipeline.

DROZY makes it suitable for evaluating enhancement methods in realistic surveillance conditions where visible-light cameras struggle.

The research methodology flow is visualized in Fig. 1. The pipeline begins with the DROZY dataset as input, consisting of NIR video. Frame extraction is performed with 5-second sampling intervals, resulting in 4,272 independent samples. Preprocessing includes two different format conversions as needed. Grayscale conversion is applied to the majority of enhancement methods (HE, AHE, CLAHE, BPDHE). BGR conversion is required specifically for MSRCR compatibility and MediaPipe face detection, as both tools are designed for multi-channel inputs.

The five enhancement methods are processed in parallel. HE operates globally without local adaptation. AHE and CLAHE apply an adaptive approach with different clip limiting. BPDHE separates images based on a brightness preservation strategy. MSRCR uses multi-scale Retinex decomposition. Each frame is processed through all five methods, resulting in five enhanced versions per original frame.

Validation and evaluation follow a dual-branch architecture. The first branch calculates four no-reference quality metrics: NIQE for natural quality assessment, PIQE for perceptual distortion quantification, Shannon Entropy measuring information content, and LOE evaluating structural preservation through lightness ordering. The second branch performs face detection validation using MediaPipe, recording success rates and confidence scores. Both branches run independently, but their outputs are combined for comprehensive analysis.

Statistical analysis uses non-parametric tests. The Friedman test compares all methods simultaneously across metrics. When global significance is detected, pairwise Wilcoxon signed-rank tests follow to identify specific differences between the original and enhanced conditions. The final output is comparative results, including quantitative metrics, detection rates, statistical validation, and qualitative visual comparisons.

B. Implementation of Enhancement Method

Five methods were evaluated. Implementations followed established conventions to ensure reproducibility.

1) Histogram Equalization (HE): HE applies a global transformation using OpenCV's `equalizeHist` function. The input format is converted to an 8-bit unsigned integer if necessary. This method redistributes pixel intensities across the dynamic range without local considerations.

$$s = T(r) = (L-1)\int_0^r p_r(w)dw \tag{1}$$

where, s is the output intensity, r is the input intensity, $p_r(w)$ is the normalized histogram of the input image, and L denotes the total number of intensity levels (256).

2) Adaptive Histogram Equalization (AHE): AHE utilizes scikit-image's `exposure.equalize_adapthist` with a clip limit of 0.01. The transformation of each pixel depends on the distribution of its local neighborhood, providing spatially adaptive contrast enhancement.

$$s(x,y) = T_{local}(r(x,y)) = (L-1)\int_0^{r(x,y)} p_R(w \mid \Omega(x,y))dw \tag{2}$$

where, $s(x,y)$ and $r(x,y)$ denote the output and input intensities at pixel location (x,y), respectively. The term $\Omega(x,y)$ represents the neighborhood window centered at (x,y), and $p_R(w \mid \Omega(x,y))$ is the local normalized histogram computed within this window. The parameter L denotes the total number of intensity levels, set to 256 for 8-bit grayscale images.

3) Contrast Limited Adaptive Histogram Equalization (CLAHE): CLAHE is implemented through OpenCV's `createCLAHE` with a clip limit of 2.0 and a tile grid size of 8×8 as default parameter value from OpenCV's library. Clip limiting prevents excessive noise amplification in homogeneous regions. The 8×8 tile size was chosen based on the trade-off between spatial resolution and local adaptation. Smaller tiles work more aggressively, while larger tiles produce a more natural appearance. The DROZY resolution of 512×424 pixels produces 64×53 tiles in this configuration.

The histogram clipping function is defined as:

$$h_{clip}(i) = \min(h_i, \text{clip\_limit}) \tag{3}$$

$$\text{clip\_limit} = \frac{N_{pixels\_per\_tile}}{N_{bins}} \left(1 + \frac{\text{clip\_parameter}}{100}\right) \quad (4)$$

Then, the CLAHE Transformation Function is:

$$s(x, y) = T_{CLAHE}(r(x, y)) \quad (5)$$

where the transformation function is derived from the clipped histogram h_{clip}.

4) Brightness Preserving Dynamic Histogram Equalization (BPDHE): BPDHE separates images based on their mean brightness. Equalization is performed independently on sub-images. Low-intensity pixels receive one transformation, while high-intensity pixels receive another, preserving the overall mean luminance, as described by Ibrahim and Kong [19].

5) Multi-Scale Retinex with Color Restoration (MSRCR): MSRCR uses three Gaussian scales: 15, 80, and 250 pixels. The parameters follow Jobson's specifications: gain $G = 192.0$, offset $B = 30.0$, alpha $\alpha = 125.0$, beta $\beta = 46.0$ [10], [11]. Each scale captures different spatial frequency information. Color restoration is applied even though the NIR is monochrome because we convert grayscale to BGR format for consistency with the evaluation metrics.

The MSRCR is defined as:

$$MSRCR(x, y) = \sum_i w_i(\log I_i(x, y) - \log(F(x, y) * G_i(x, y))) \quad (6)$$

where multiple Gaussian surround functions $G_i(x, y)$ with different scales σ_i are used to enhance image details at various spatial frequencies.

The color restoration function is formulated as:

$$CR(x, y) = \beta \log(\alpha \cdot I(x, y)) - \log\left(\sum_c I_c(x, y)\right) \quad (7)$$

The final enhanced output image is obtained by:

$$Output(x, y) = MSRCR(x, y) \times CR(x, y) + B \quad (8)$$

where, α, β, and B are tuning parameters controlling the color restoration process and dynamic range adjustment.

Finally, all enhanced images are saved in PNG format to avoid compression artifacts. Processing occurs on the Kaggle platform using P100 GPUs, although metric computations are executed on CPU due to CUDA compatibility constraints with the PyIQA library. Parameter settings are summarized in Table I.

TABLE I. PARAMETER SETTINGS FOR ALL ENHANCEMENT METHODS FOLLOW CANONICAL VALUES FROM THE LITERATURE

Method	Parameter	Value	Source
HE	-	Default	OpenCV
AHE	Clip limit	0.01	scikit-image
CLAHE	Clip limit	2.0	OpenCV
	Tile Size	8x8	OpenCV
BPDHE	Separation point	Mean Brightness	[19]
MSRCR	Gaussian scales (σ)	[15, 80, 250]	[10], [11]
	Gain (G)	192.0	[10], [11]
	Offset (B)	30.0	[10], [11]
	Alpha (α)	125.0	[10], [11]
	Beta (β)	46.0	[10], [11]

C. Quality Metrics Computation

Four no-reference metrics quantify enhancement quality. Pristine reference images are absent in surveillance contexts, making NR-IQA essential:

1) Natural Image Quality Evaluator (NIQE): It measures the statistical distance between image features and natural scene statistics. The PyIQA library provides an implementation. Lower scores indicate closer alignment to natural image characteristics, interpreted as better perceptual quality [31].

2) Perception-based Image Quality Evaluator (PIQE): It quantifies block-wise distortion based on a human perception model. Lower scores indicate fewer perceptual artifacts such as blocking, blur, or noise [33].

3) Shannon entropy: It is calculated from grayscale histograms. For color images, conversion to grayscale precedes histogram generation using 256 bins.

$$H = -\sum_i p(i) \log_2 p(i) \quad (9)$$

where, $p(i)$ represents the probability of intensity level i. Higher entropy indicates richer information content [36].

4) Lightness Order Error (LOE): samples 4,096 random pixel pairs per image. For each pair, the algorithm checks whether the relative brightness is preserved after enhancement. A lower LOE indicates better structural preservation [39].

D. Face Detection Validation

MediaPipe Face Detection serves as a practical validation tool [41]. Detection success indicates whether the enhanced image remains suitable for downstream drowsiness analysis tasks requiring facial landmarks. We chose MediaPipe over classical methods such as Haar cascades or HOG due to its superior robustness, single-shot detector architecture, and widespread adoption. The configuration uses a selection model of 1 (far range, optimized for surveillance distances) with a minimum detection confidence of 0.5. The detection rate is calculated as the percentage of frames where at least one face is detected.

E. Statistical Analysis

Non-parametric tests analyzed the results because image quality metrics typically violate the normality assumption. Following established practice for enhancement comparisons [42], we applied the Friedman test for overall significance across all methods, then pairwise Wilcoxon signed-rank tests when the global null hypothesis was rejected. The significance level was set at $\alpha = 0.05$. Effect sizes were quantified using rank-biserial correlation [43].

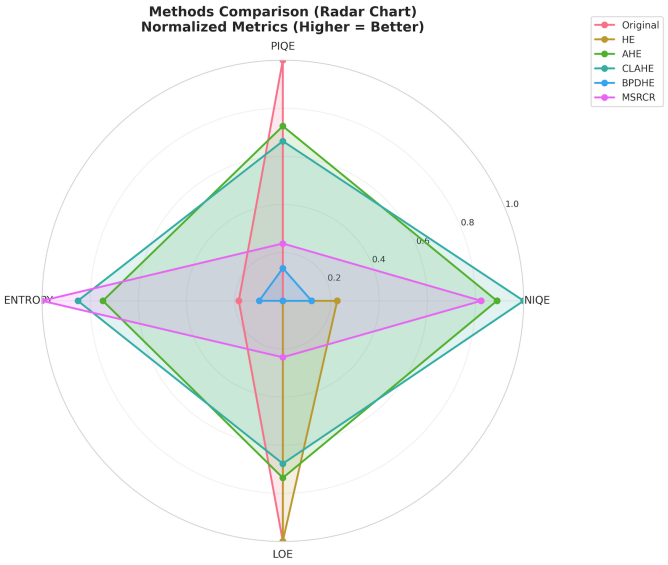


Fig. 2. Radar chart for normalized metrics of comparison methods.

IV. Results

A. Overall Performance

Table II summarizes the descriptive statistics for the six conditions. CLAHE achieved the lowest NIQE (4.6 ± 0.39), indicating the best natural quality. AHE followed with 4.76 ± 0.44. The original images had a NIQE of 5.96 ± 0.59. In contrast, BPDHE produced the worst NIQE at 5.80 ± 0.68. HE and MSRCR were in the middle with 5.66 and 4.84, respectively.

PIQE showed a different pattern. Original had the lowest score of 47.40 ± 5.53, indicating minimal perceptual distortion due to the lack of processing. All enhancement methods improved PIQE, meaning they introduced artifacts. HE performed the worst with 71.41 ± 3.09. BPDHE achieved 68.15 ± 2.16. MSRCR was 65.69 ± 3.21. AHE and CLAHE were more moderate, at 53.98 ± 4.07 and 55.48 ± 3.87, respectively.

The highest entropy was achieved by MSRCR (6.58 ± 0.20). CLAHE produced 6.37 ± 0.12. AHE was 6.22 ± 0.11. The original was only 5.43 ± 0.18. BPDHE was even lower than the original, at 5.31 ± 0.16. HE was the lowest at 5.17 ± 0.13, indicating a loss of detailed information.

The original LOE was 0.00 due to no transformation. AHE retained the best structure with an LOE of 0.047 ± 0.007. CLAHE was slightly higher at 0.058 ± 0.005. The MSRCR reached 0.137 ± 0.008. The BPDHE was highest with 0.179 ± 0.013, indicating significant structural distortion. The HE was almost perfect with 0.000 ± 0.000 because the global transformation preserved the relative ordering.

As Fig. 2 shows, the radar chart provides a comprehensive visualization of the trade-offs. CLAHE forms a balanced area, excelling in NIQE and competitive in other dimensions. MSRCR dominates entropy but is weak in PIQE. Original is superior in PIQE but inferior in entropy.

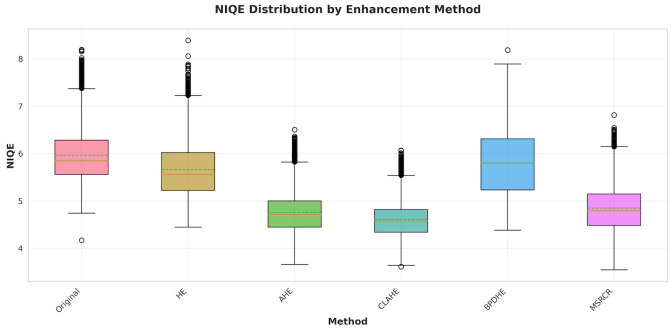


Fig. 3. NIQE distribution boxplot.

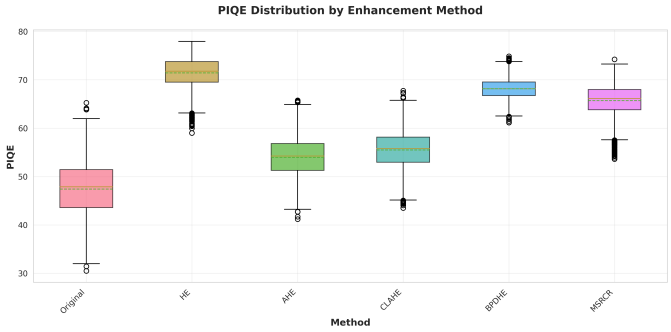


Fig. 4. PIQE distribution boxplot.

B. Per-Metric Analysis

1) NIQE distribution: The boxplot in Fig. 3 reveals substantial variability. CLAHE shows a narrow interquartile range, indicating consistency across frames. AHE has a slightly wider spread. Original and BPDHE display numerous outliers toward high scores (poor quality). MSRCR has a more compact distribution than Original, although the median is lower.

2) PIQE distribution: Fig. 4 illustrates the impact of enhancement on perceptual distortion. Original has the lowest median and the narrowest spread. Once enhancement is applied, PIQE jumps. HE shows high consistency but at a poor-quality level. CLAHE and AHE maintain moderate variability. BPDHE and MSRCR experience outliers toward very high scores in some frames.

3) Entropy distribution: MSRCR dominates with a median of 6.54, as shown in Fig. 5. The range of CLAHE is concentrated between 6.30 and 6.44. AHE partially overlaps with CLAHE but has a lower median. Original, HE, and BPDHE cluster in the low-entropy region. Some outliers appear in Original toward the high values, likely frames with high noise content.

4) LOE distribution: Fig. 6 confirms HE and Original have near-zero LOE. AHE and CLAHE display low values with minimal spread. MSRCR experiences a higher but consistent LOE. BPDHE shows the highest median and the greatest variability, with some frames reaching LOE above 0.20.

The heatmap in Fig. 7 integrates all metrics. Colors indicate normalized scores, with darker values indicating better performance. CLAHE achieves the best balance across rows.

TABLE II. DESCRIPTIVE STATISTICS OF COMPARISON METHOD

Method	NIQE	PIQE	Entropy	LOE
AHE	4.76 ± 0.4397	53.98 ± 4.0719	6.22 ± 0.1062	0.047 ± 0.0069
BPDHE	5.80 ± 0.6831	68.15 ± 2.1582	5.31 ± 0.1648	0.179 ± 0.0125
CLAHE	4.60 ± 0.3946	55.48 ± 3.8747	6.37 ± 0.1222	0.058 ± 0.0055
HE	5.66 ± 0.5875	71.41 ± 3.0934	5.17 ± 0.1331	0.000 ± 0.0002
MSRCR	4.84 ± 0.5133	65.69 ± 3.2075	6.57 ± 0.1956	0.137 ± 0.0085
Original	5.96 ± 0.5944	47.40 ± 5.5287	5.43 ± 0.1754	0.000 ± 0.0000

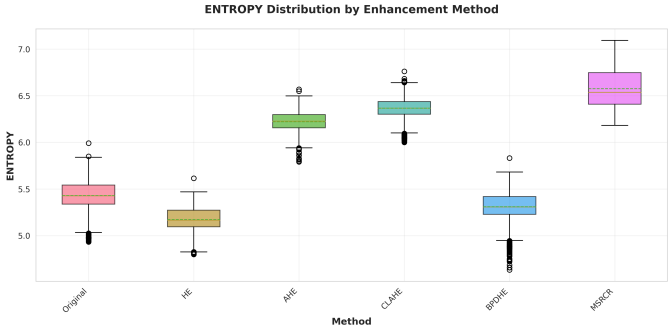


Fig. 5. Entropy distribution boxplot.

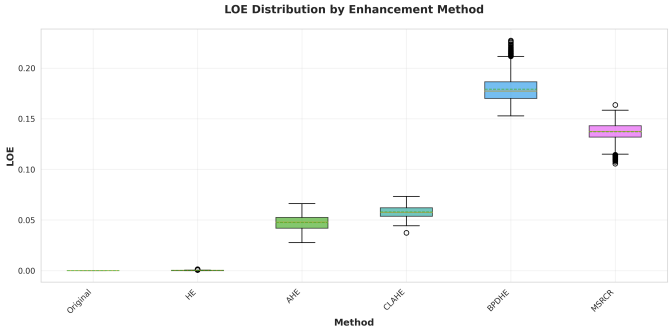


Fig. 6. LOE distribution boxplot.

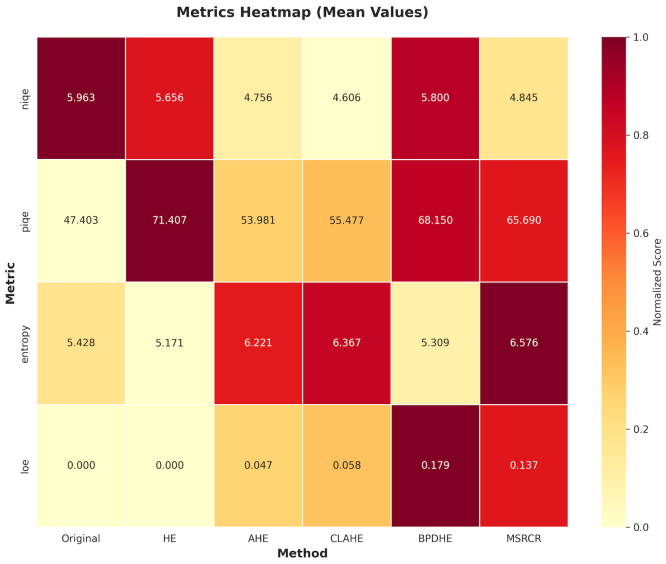


Fig. 7. Metric heatmap.

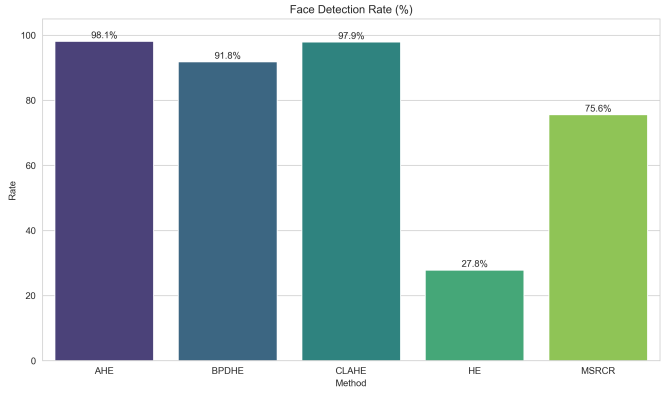


Fig. 8. Face detection rate for 5 methods using mediapipe.

MSRCR excels only in entropy. HE exhibits an inconsistent pattern, excelling in LOE but failing in other metrics. Original serves as the baseline with mixed performance.

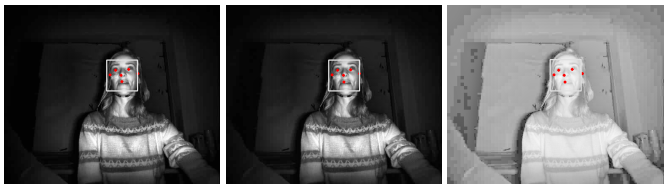
C. Face Detection Rates

Face detection rates provide practical validation of enhancement suitability for drowsiness detection systems, where successful face localization is prerequisite for landmark-based feature extraction. Detection success varied dramatically. AHE achieved the highest rate of 98.1%. CLAHE was very close at 97.9%. Original was not evaluated because it did not undergo enhancement, but BPDHE achieved 91.8%. MSRCR dropped to 75.6%. HE experienced catastrophic failure with only 27.8% detection success.

Fig. 8 illustrates this difference visually. The bar chart shows a substantial gap between histogram-based adaptive methods versus the global method (HE) and the Retinex-based method (MSRCR). The difference of more than 20 percentage points between CLAHE and MSRCR is practically significant.

Qualitative validation using MediaPipe detection boxes confirms the quantitative pattern. Frame 1-1_t0000 illustrates the success cases in Fig. 9(a-c). AHE (Fig. 9a) detects faces with high confidence, with accurate localization of facial landmarks. CLAHE (Fig. 9b) produces similar detections, with stable bounding boxes. MSRCR (Fig. 9c) is successful despite the different visual appearances; the MediaPipe algorithm still recognizes facial structures.

Failure patterns reveal method-specific issues. HE over-enhancement in frame 14-1_t0340 (Fig. 10a) removes facial boundaries. The washed-out appearance confuses the detector; no recognizable structures remain. BPDHE fails even on



(a) AHE (b) CLAHE (c) MSRCR

Fig. 9. Detection success - frame 1-1_t0000.

(a) HE (b) BPDHE (c) BPDHE

Fig. 10. Detection failures.

(a) MSRCR (b) CLAHE (c) AHE

Fig. 11. Additional failure cases.

relatively bright frames like 1-1_t0000 (Fig. 10b). Brightness discontinuity between sub-images creates artificial edges, disrupting facial geometry. Frame 14-1_t0340 exacerbates this problem (Fig. 10c), with brightness separation producing a patchy appearance.

MSRCR also experienced failures, although the overall rate was 75.6%. Frame 14-1_t0340 (Fig. 11a) shows subtle degradation; facial features are visually visible, but structural distortion is severe enough to trigger detection failure. Occasional CLAHE failures (2.1% of frames) occur in extreme cases. Frame 14-1_t0340 (Fig. 11b) represents a rare failure; subject position or motion blur interact with enhancement, although the method is generally robust. AHE failures (1.9%) occurred under similar conditions. Frame 14-1_t0340 (Fig. 11c) demonstrates an edge case where adaptive histogram equalization is inadequate. Detection failures are not random; they cluster on frames with challenging characteristics such as extreme head angles or motion artifacts. Methods with lower failure rates handle edge cases better.

Visual inspection confirms metrics are not misleading. Detection success correlates with perceptible facial structure preservation. The low scoring method on LOE (BPDHE 0.179) indeed distorts geometry visibly. High PIQE scores (HE 71.41) manifest as blocking artifacts and washed-out regions. Quantitative and qualitative findings align.

Fig. 12. Representative visual frame comparison.

Fig. 13. Second frame visual comparison.

For qualitative context, Fig. 12 shows a grid comparison of a single frame. Original exhibits limited visibility. HE overenhances significantly, washing out facial features to the point of near-detection. AHE improves visibility while preserving structure. CLAHE produces similar results with slightly reduced noise. BPDHE introduces brightness inconsistencies. MSRCR produces a more natural appearance visually, but facial boundaries become less distinct. Fig. 13 and 14 present additional comparison cases, demonstrating the consistency of patterns across diverse frames.

D. Statistical Significance

The Friedman test rejected the null hypothesis for all four metrics ($p < 0.001$), confirming that at least one method differs significantly. Table III summarizes the results for each metric with sample sizes, test statistics, and p-values.

Wilcoxon signed-rank tests compared each enhanced method with the original method. All comparisons yielded p-values below 0.05, except for HE on LOE, where perfect preservation yielded a near-zero difference. Effect sizes ranged from moderate to large. Table IV details the number of pairs, mean differences, test statistics, p-values, and significance indicators.

Fig. 14. Third frame visual comparison.

TABLE III. FRIEDMAN TEST RESULTS

Metric	Chi-Square (χ^2)	p-value	Significant
NIQE	16565.8768	< 0.001	Yes
PIQE	20710.6073	< 0.001	Yes
Entropy	21134.4112	< 0.001	Yes
LOE	21207.5235	< 0.001	Yes

TABLE IV. PAIRED WILCOXON TESTS (COMPARISON AGAINST ORIGINAL)

Metric	Method	Mean Diff.	W-statistic	p-value
NIQE	HE	-0.3065	1,669,917	< 0.001
	AHE	-1.2065	0	< 0.001
	CLAHE	-1.3569	0	< 0.001
	BPDHE	-0.1624	3,073,535	< 0.001
	MSRCR	-1.1179	2,504	< 0.001
PIQE	HE	24.0037	0	< 0.001
	AHE	6.5779	4,753	< 0.001
	CLAHE	8.0745	451	< 0.001
	BPDHE	20.747	0	< 0.001
	MSRCR	18.2877	0	< 0.001
Entropy	HE	-0.2571	0	< 0.001
	AHE	0.793	0	< 0.001
	CLAHE	0.9391	0	< 0.001
	BPDHE	-0.1189	0	< 0.001
	MSRCR	1.148	0	< 0.001
LOE	HE	0.0002	0	< 0.001
	AHE	0.0474	0	< 0.001
	CLAHE	0.0579	0	< 0.001
	BPDHE	0.1792	0	< 0.001
	MSRCR	0.1371	0	< 0.001

CLAHE showed significant improvements for NIQE (mean difference -1.36, $p < 0.001$) and entropy (mean difference +0.94, $p < 0.001$). Degradation occurred for PIQE (mean difference +8.07) and LOE (mean difference +0.058), both statistically significant. These trade-offs are expected because enhancement inherently modifies image characteristics. MSRCR achieved the largest entropy gain (+1.15) but also introduced substantial LOE (+0.137) and PIQE degradation (+18.29). Statistical significance does not equate to practical superiority; interpretation requires application context.

V. DISCUSSION

A. Advantages of Balanced CLAHE

CLAHE achieved an optimal balance across evaluation dimensions. A NIQE score of 4.61 indicates the best natural quality, outperforming all other methods, including the original (5.96). Local adaptation allows enhancement of dark regions without overprocessing bright areas. The clip limit parameter prevents excessive noise amplification, a problem with standard AHE. A detection rate of 97.9% validates practical suitability. Only 89 of 4,272 frames failed detection. This is important because face detection is a prerequisite for landmark extraction in drowsiness analysis. Methods that distort facial geometry, no matter how good their metrics, are not applicable to real-world systems. CLAHE preserves structure while improving visibility.

CLAHE is suitable for moderate lighting conditions. The Kinect v2 NIR sensor produces limited intensity but not complete darkness, typical of indoor surveillance conditions with minimal ambient light. Enhancement needs to improve local detail without over-amplification. The tile-based approach processes 8×8 regions independently, providing adaptive gain according to the local intensity distribution. A clip limit of 2.0 limits the maximum amplification, preventing noise explosion in near-dark regions. Extremely dark lighting is different. Aggressive enhancement is needed to reveal nearly invisible content. Retinex methods are designed for such scenarios, where the illumination component needs to be drastically adjusted to separate reflectance.

Multi-scale decomposition assumes large dynamic range variations. DROZY NIR falls in the middle zone, not extremely dark but also not well-lit. Aggressive enhancement is counterproductive here, introducing artifacts as seen in the MSRCR results. CLAHE lightweight adaptation is more suitable for moderate low-light surveillance contexts. A LOE score of 0.058 confirms the preservation of brightness ordering. While relative intensities are preserved, spatial relationships remain intact. Eye regions remain darker than the forehead, and mouth boundaries remain distinct from cheeks. This stability facilitates reliable calculation of EAR and MAR, two key drowsiness indicators.

The observed performance gap, CLAHE at 97.9% versus MSRCR at 75.6% face detection, empirically validates theoretical concerns regarding MSRCR's design assumptions. The color restoration component, foundational to MSRCR for RGB images [10], [11], operates on spectral diversity absent in single-band NIR data. This mismatch manifests not merely as reduced image quality metrics but as tangible impact on downstream task performance, as evidenced by the 22-percentage-point detection rate difference. The incompatibility of MSRCR is not an indication of poor methodology. The design assumptions do not match the operational conditions of DROZY. MSRCR is designed for scenarios where illumination variations are large and require multi-scale decomposition for normalization. The parameter $G = 192$ plus three scales [15, 80, 250] sets an aggressive enhancement suitable for extremely dim or high dynamic range scenes. Images with severe backlighting or deep shadows benefit from this approach. Zhang et al. demonstrated its effectiveness for underwater scenes where scattering creates extreme attenuation gradients

[23]. Outdoor nighttime scenes with headlight glare are also suitable [20]. DROZY is in the middle zone. Not extremely dark, not well-lit. Enhancement requirements are moderate. CLAHE's lightweight local adaptation is more appropriate than MSRCR's heavy multi-scale processing. The computational cost also differs significantly. Three Gaussian convolutions per scale require substantial processing power, while CLAHE's tile-based histogram operations are much more efficient. For real-time automotive systems processing 30 fps streams, efficiency matters. Here's a simple analogy. MSRCR is like a sledgehammer for breaking up large rocks. CLAHE is like a regular hammer for nails. DROZY NIR requires a hammer, not a sledgehammer. Excessive power can damage structures that need to be preserved for face detection.

B. Trade-Off MSRCR

MSRCR achieved the highest entropy (6.58), confirming its theoretical strength in detail revelation. Jobson et al. designed a multi-scale approach to capture information across spatial frequencies [10]. The color restoration component should preserve saturation while compressing dynamic range [11]. In theory, MSRCR is superior for natural images. A detection rate of only 75.6% reveals practical limitations. Over 1,000 frames failed detection. The inconsistency between high entropy and low detection rates requires explanation. We argue the problem is contextual, not methodological.

NIR imaging is essentially monochrome. DROZY uses a single-channel sensor at 850 nm wavelength [3]. The color restoration step in MSRCR, designed for RGB channels, becomes less relevant. Jobson et al. identified monochrome scenes as a pathological case where assumptions are violated [24]. The parameter $\beta = 46.0$ we used to govern the colour restoration strength. In grayscale NIR, this component may introduce artifacts rather than benefits. LOE score 0.137 versus CLAHE 0.058 indicates structural distortion. MSRCR multi-scale processing with aggressive parameters can cause brightness ordering reversals. Petro et al. documented halo artifacts around high-contrast edges [26]. In facial images, halos blur boundaries between features, confusing detection algorithms. MediaPipe relies on learned facial structure patterns; distortions degrade recognition.

Liu et al. demonstrated MSRCR aggravates scattering effects in turbid water [25]. A similar principle applies to limited spectral range NIR. Zhang noted MSRCR requires post-processing denoising for dim images, adding computational overhead [22]. Real-time drowsiness systems process 30 fps streams; extra denoising steps reduce throughput. We do not claim MSRCR is universally inappropriate. Wang et al. showed benefits for maize leaf disease recognition [9]. Zhang demonstrated that MSRCR is effective for dim nighttime images [22]. Context determines suitability. For NIR surveillance with structural preservation requirements, simpler methods like CLAHE offer a better balance.

C. HE and BPDHE Failure

HE experienced catastrophic failure with a detection rate of 27.8%. Global equalization overstretches the dynamic range without local considerations. Bright regions become washed out, and dark regions are excessively amplified. Visual inspection of Fig. 10 shows that facial features are almost obliterated. The PIQE score of 71.41 is the highest among all methods. Block-wise distortion is severe. Noise in the original images is amplified throughout the intensity range. Wang et al. identified noise amplification as the primary weakness of HE [9]. In NIR images with inherent thermal noise, HE exacerbates the problem.

Interestingly, the LOE is nearly zero (0.000). Global transformation preserves relative ordering perfectly. Every pixel undergoes an identical mapping function. Brightness relationships remain unchanged. However, structure preservation does not translate into detection success. Apparently, absolute intensity levels matter more than purely relative ordering for face detection algorithms. HE remains valuable for certain applications. Its computational simplicity makes it attractive for resource-constrained environments. Its processing speed is superior due to single-pass histogram computation. But for drowsiness detection requiring reliable face localization, HE is unsuitable.

BPDHE preserves mean brightness, theoretically avoiding over-enhancement. Detection rate of 91.8% is decent but still inferior to CLAHE and AHE. The highest LOE was 0.179, indicating substantial brightness ordering reversals. Separation in mean brightness creates discontinuity. Upper and lower sub-images receive different transformations, potentially causing inconsistent enhancement across facial regions. Ibrahim and Kong developed BPDHE for medical imaging [19]. Different context from surveillance. Medical images typically have controlled acquisition conditions; surveillance deals with variable, uncontrolled environments. Methods designed for one domain are not necessarily optimal for another.

D. Implications for Drowsiness Detection Systems

Face detection is a critical bottleneck. Landmark extraction algorithms assume face already localized. Calculation of EAR requires precise eye corner positions. MAR calculation needs accurate mouth boundary points. Methods failing detection render subsequent analysis impossible. Lin et al. demonstrated preprocessing impact, reporting 5% accuracy gain [6]. Our findings suggest the impact can be much larger when choosing between fundamentally different enhancement approaches. The gap between CLAHE (97.9%) and HE (27.8%) represents a 70-percentage point difference in usable frames.

Real-time systems process continuous video streams. Intermittent detection failures disrupt temporal analysis. Drowsiness often manifests through gradual changes over time, not discrete events. Missing frames due to enhancement-induced detection failures breaks temporal continuity, hampering longitudinal monitoring. Computational efficiency also matters. CLAHE operates through local histogram operations, computationally efficient. Wang et al. noted MSRCR multi-scale processing is more intensive [21]. For embedded automotive platforms with limited processing power, efficiency considerations are not trivial. CLAHE offers a favorable quality-efficiency trade-off.

E. Limitations

Several limitations provide context for interpreting results and identifying future research directions

Evaluation deliberately focused on the NIR surveillance context. DROZY represents realistic low-light conditions but is a single sensor type with a fixed wavelength (850 nm). Generalizability to visible-light scenarios or different NIR sensors requires validation. The frame-level analysis approach, while appropriate for comparative image quality assessment [42], treats samples as independent; temporal modeling of drowsiness patterns across video sequences represents complementary future investigation.

MediaPipe served as detection validator. Classical methods like Haar cascades or HOG detectors might show different patterns. Deep learning-based face detectors are generally more robust, but comparison across detector architectures could reveal method-detector interaction effects.

Parameter settings followed canonical configurations. CLAHE used clip limit 2.0 and tile size 8×8 based on OpenCV's documentation. MSRCR parameters from Jobson specifications [10], [11]. No optimization performed specifically for DROZY characteristics. Tuning might improve individual method performance, potentially altering relative rankings.

Statistical tests assumed independence between frames sampled every 5 seconds. Temporal autocorrelation may exist given video source. Stronger independence could be achieved through longer intervals or sampling from different video segments, though the trade-off is reduced sample size. The 4,272 frames provided sufficient statistical power for detecting observed differences, as confirmed by highly significant p-values ($p < 0.001$) across all comparisons.

F. Future Directions

Future work should explore adaptive enhancement frameworks. Parameters can be tuned dynamically based on local image characteristics. Machine learning approaches for parameter selection might optimize trade-offs automatically. Our findings suggest CLAHE as a strong starting point; optimization builds on solid foundation.

Comparison with deep learning-based enhancement methods (EnlightenGAN, Zero-DCE, RetinexNet) would benchmark traditional approaches against modern alternatives, though computational constraints for real-time automotive deployment remain practical considerations. Extension to end-to-end drowsiness detection pipelines would validate whether preprocessing improvements translate proportionally to classification accuracy gains.

Validation across multiple NIR sensors with different wavelength ranges (700-1000 nm) would establish generalizability beyond Kinect v2's specifications. Investigation of temporal modeling approaches would complement frame-level quality assessment with sequence-based drowsiness pattern analysis, addressing the temporal dynamics noted as a limitation of the current frame-independent approach.

VI. Conclusion

A. Summary of Findings

Five enhancement methods were evaluated on 4,272 DROZY frames. CLAHE was optimal. A NIQE score of 4.61 indicates the best natural quality, and a detection rate of 97.9% demonstrates practical suitability. A low LOE (0.058) preserves facial structure for landmark extraction. Trade-offs exist; PIQE increases by 8 points compared to the original, but improvements in other dimensions justify this cost.

MSRCR achieved the highest entropy (6.58) but the lowest detection rate (75.6%). The color restoration component, designed for RGB channels, is less relevant in a monochrome NIR context. Structural distortion (LOE 0.137) explains detection failures. For NIR surveillance, theoretical sophistication does not guarantee practical superiority. Context matters more than complexity.

HE failed catastrophically with a detection rate of 27.8%. Global transformation over-enhances without local adaptation. BPDHE is moderate at 91.8%, but an LOE of 0.179 indicates brightness ordering problems. AHE performs competitively (98.1%), but CLAHE's advantage in NIQE plus comparable detection makes it preferable for most scenarios.

B. Statistical Validation

Statistics validate the findings. Friedman tests reject the null hypothesis across all metrics ($p < 0.001$). Wilcoxon comparisons confirm significance for pairwise differences. Effect sizes are moderate to large. This is not an artifact of sample size; the differences are practically meaningful.

C. Practical Implications

This study focused on NIR drowsiness contexts. A single dataset limits generalizability, although DROZY is representative for low-light surveillance scenarios. MediaPipe is appropriate as a validator due to its representation of widely deployed deep learning methods, but detector choice influences results. Parameter settings followed canonical configurations without optimization specifically for DROZY, meaning tuning could potentially improve individual method performance.

The practical implications are clear. System designers selecting preprocessing for automotive monitoring should prioritize CLAHE. The balance between quality metrics, detection reliability, and computational efficiency makes it suitable for real-time embedded platforms. AHE is a viable alternative if a marginally higher detection rate is prioritized.

D. Contributions and Recommendations

The observed balance in CLAHE suggests potential for adaptive optimization frameworks. Clip limit and tile size can be tuned dynamically based on image characteristics. The parameters we use (2.0 and 8×8) work well across diverse DROZY frames, yet adaptive approaches might achieve even better trade-offs for specific conditions or subjects.

Beyond drowsiness detection, findings applicable to NIR-based face analysis broadly. Expression recognition, attention monitoring, gaze tracking all require reliable face localization as foundation. Enhancement method choice impacts entire pipeline performance, not isolated preprocessing steps.

This research provides empirical guidance for practitioners navigating enhanced method selection in NIR surveillance contexts. Systematic evaluation with multiple quality metrics

plus practical detection validation offers actionable insights. CLAHE is recommended as a robust baseline for drowsiness detection systems operating under low-light conditions.

ACKNOWLEDGMENT

Computations were performed using the Kaggle platform with P100 GPUs. The DROZY dataset is available through the Open Repository and Bibliography (ORBi) of the University of Liège.

REFERENCES

[1] E. M. Nowara, T. K. Marks, H. Mansour, and A. Veeraraghavan, "Near-Infrared Imaging Photoplethysmography During Driving," *IEEE Trans. Intell. Transport. Syst.*, vol. 23, no. 4, pp. 3589–3600, Apr. 2022, doi: 10.1109/TITS.2020.3038317.

[2] Y. Tu, S. Fernando, and M. van Gastel, "Internship Report: Benchmark of Deep Learning-based Imaging PPG in Automotive Domain," *arXiv preprint arXiv:2411.00919*, Nov. 2024.

[3] Q. Massoz, T. Langohr, C. Francois, and J. G. Verly, "The ULg multimodality drowsiness database (called DROZY) and examples of use," in *2016 IEEE Winter Conference on Applications of Computer Vision (WACV)*, Lake Placid, NY, USA, 2016, pp. 1–7.

[4] R. Xu, Z. Zhang, R. Wu, and W. Zuo, "NIR-Assisted Image Denoising: A Selective Fusion Approach and a Real-World Benchmark Dataset," *IEEE Trans. Multimedia*, vol. 27, pp. 2543–2555, 2025.

[5] S. Jee and M. G. Kang, "Sensitivity Improvement of Extremely Low Light Scenes with RGB-NIR Multispectral Filter Array Sensor," *Sensors*, vol. 19, no. 5, p. 1256, Mar. 2019.

[6] N. Lin and Y. Zuo, "Advancing driver fatigue detection in diverse lighting conditions for assisted driving vehicles with enhanced facial recognition technologies," *PLoS ONE*, vol. 19, no. 7, p. e0304669, Jul. 2024.

[7] S. J. N. Aprilia and D. Fitrianah, "Automatic drowsiness detection system to reduce road accident risks," *Bulletin of Electrical Engineering and Informatics*, vol. 14, no. 4, pp. 2674–2683, 2025.

[8] S. Winarno, F. Alzami, M. Naufal, H. Al Azies, M. A. Soeleman, and N. H. A. H. Malim, "EfficientNet-KNN for Real-Time Driver Drowsiness Detection via Sequential Image Processing," *ISI*, vol. 30, no. 7, pp. 1703–1713, Jul. 2025.

[9] W. Wang, X. Wu, X. Yuan, and Z. Gao, "An Experiment-Based Review of Low-Light Image Enhancement Methods," *IEEE Access*, vol. 8, pp. 87884–87917, 2020.

[10] D. J. Jobson, Z. Rahman, and G. A. Woodell, "A multiscale retinex for bridging the gap between color images and the human observation of scenes," *IEEE Trans. on Image Process.*, vol. 6, no. 7, pp. 965–976, Jul. 1997.

[11] D. J. Jobson, "Retinex processing for automatic image enhancement," *J. Electron. Imaging*, vol. 13, no. 1, p. 100, Jan. 2004.

[12] Z. Jingchun, G. Eg Su, and M. Shahrizal Sunar, "Low-light image enhancement: A comprehensive review on methods, datasets and evaluation metrics," *Journal of King Saud University - Computer and Information Sciences*, vol. 36, no. 10, p. 102234, Dec. 2024.

[13] M. Tarasiou, J. Deng, and S. Zafeiriou, "Rethinking the Domain Gap in Near-infrared Face Recognition," in *2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, Seattle, WA, USA, 2024, pp. 940–949.

[14] M. Bodaghi et al., "UL-DD: A Multimodal Drowsiness Dataset Using Video, Biometric Signals, and Behavioral Data," *arXiv preprint arXiv:2507.13403*, 2025.

[15] 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.

[16] S. Liu, Q. Lu, and S. Dai, "Adaptive histogram equalization framework based on new visual prior and optimization model," *Signal Processing: Image Communication*, vol. 132, p. 117246, Mar. 2025.

[17] F. Alzami, S. Winarno, M. Naufal, and H. Al Azies, "Enhancing Driver Drowsiness Detection through GMM-Optimized CLAHE," in *2024 International Seminar on Application for Technology of Information and Communication (iSemantic)*, Semarang, Indonesia, 2024, pp. 212–217.

[18] Y. Han et al., "Low-Illumination Road Image Enhancement by Fusing Retinex Theory and Histogram Equalization," *Electronics*, vol. 12, no. 4, p. 990, Feb. 2023.

[19] H. Ibrahim and N. Pik Kong, "Brightness Preserving Dynamic Histogram Equalization for Image Contrast Enhancement," *IEEE Trans. Consumer Electron.*, vol. 53, no. 4, pp. 1752–1758, Nov. 2007.

[20] T. Li and T. Zhou, "Multi-scale fusion framework via retinex and transmittance optimization for underwater image enhancement," *PLoS ONE*, vol. 17, no. 9, p. e0275107, Sep. 2022.

[21] P. Wang, Y. Xiong, and H. Zhang, "Maize leaf disease recognition based on improved MSRCR and OSCRNet," *Crop Protection*, vol. 183, p. 106757, Sep. 2024.

[22] S. Zhang, M. Zhu, and K. Meng, "An Automated Multi-scale Retinex for Dim Image Enhancement," in *2022 IEEE 2nd International Conference on Power, Electronics and Computer Applications (ICPECA)*, Shenyang, China, 2022, pp. 647–651.

[23] A. B. Petro, C. Sbert, and J.-M. Morel, "Multiscale Retinex," *Image Processing On Line*, vol. 4, pp. 71–88, Apr. 2014.

[24] Z. Rahman, D. J. Jobson, and G. A. Woodell, "Multi-scale retinex for color image enhancement," in *Proc. 3rd IEEE Int. Conf. Image Process.*, Lausanne, Switzerland, 1996, pp. 1003–1006.

[25] R. Liu, X. Fan, M. Zhu, M. Hou, and Z. Luo, "Real-World Underwater Enhancement: Challenges, Benchmarks, and Solutions Under Natural Light," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 30, no. 12, pp. 4861–4875, Dec. 2020.

[26] W. Sun, L. Han, B. Guo, W. Jia, and M. Sun, "A fast color image enhancement algorithm based on Max Intensity Channel," *Journal of Modern Optics*, vol. 61, no. 6, pp. 466–477, Mar. 2014.

[27] X. Yang, J. Chen, and Z. Yang, "Cooperative Colorization: Exploring Latent Cross-Domain Priors for NIR Image Spectrum Translation," in *Proc. 31st ACM Int. Conf. Multimedia*, Ottawa ON Canada, 2023, pp. 2409–2417.

[28] G. Borghi, S. Pini, R. Vezzani, and R. Cucchiara, "Driver Face Verification with Depth Maps," *Sensors*, vol. 19, no. 15, p. 3361, Jul. 2019.

[29] R. Ghoddoosian, M. Galib, and V. Athitsos, "A Realistic Dataset and Baseline Temporal Model for Early Drowsiness Detection," in *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, Long Beach, CA, USA, 2019, pp. 178–187.

[30] B. Asdyo, B. Kanigoro, and Rojali, "Drowsy Detection System by Facial Landmark and Light Gradient Boosting Machine Method," *Procedia Computer Science*, vol. 227, pp. 500–507, 2023.

[31] A. Mittal, R. Soundararajan, and A. C. Bovik, "Making a 'Completely Blind' Image Quality Analyzer," *IEEE Signal Process. Lett.*, vol. 20, no. 3, pp. 209–212, Mar. 2013.

[32] M. Sabry, G. Schroeder, J. Varughese, and C. Olaverri-Monreal, "Shadow Erosion and Nighttime Adaptability for Camera-Based Automated Driving Applications," *arXiv preprint arXiv:2504.08551*, 2025.

[33] S. Higashiyama et al., "Usefulness of a No-Reference Metric for Evaluation of Images in Nuclear Medicine - A Comparative Study with Visual Assessment," 2021.

[34] B. Fei et al., "A diffusion model for universal medical image enhancement," *Commun Med*, vol. 5, no. 1, p. 294, Jul. 2025.

[35] A. Saleem, S. Paheding, N. Rawashdeh, A. Awad, and N. Kaur, "A Non-Reference Evaluation of Underwater Image Enhancement Methods Using a New Underwater Image Dataset," *IEEE Access*, vol. 11, pp. 10412–10428, 2023.

[36] Y. Karaca and M. Moonis, "Shannon entropy-based complexity quantification of nonlinear stochastic process," in *Multi-Chaos, Fractal and Multi-Fractional Artificial Intelligence of Different Complex Systems*, Elsevier, 2022, pp. 231–245.

[37] H. Daway et al., "Colour Image Enhancement by Fuzzy Logic Based on Sigmoid Membership Function," *IJIES*, vol. 13, no. 5, pp. 238–246, Oct. 2020.

[38] C. Li, J. Zhu, L. Bi, W. Zhang, and Y. Liu, "A low-light image enhancement method with brightness balance and detail preservation," *PLoS ONE*, vol. 17, no. 5, p. e0262478, May 2022.

[39] M. Akai, Y. Ueda, T. Koga, and N. Suetake, "Low-Artifact and Fast Backlit Image Enhancement Method Based on Suppression of Lightness Order Error," *IEEE Access*, vol. 11, pp. 121231–121245, 2023.

[40] B. Luo et al., "Multi-Energy Guided Image Translation with Stochastic Differential Equations for Near-Infrared Facial Expression Recognition," *AAAI*, vol. 38, no. 1, pp. 565–573, Mar. 2024.

[41] C. Lugaresi et al., "MediaPipe: A Framework for Building Perception Pipelines," *arXiv preprint arXiv:1906.08172*, 2019.

[42] Y. Sun et al., "Low-Illumination Image Enhancement Algorithm Based on Improved Multi-Scale Retinex and ABC Algorithm Optimization," *Front. Bioeng. Biotechnol.*, vol. 10, p. 865820, Apr. 2022.

[43] E. E. Cureton, "Rank-Biserial Correlation," *Psychometrika*, vol. 21, no. 3, pp. 287–290, Sep. 1956.