

AI-Based Driver Drowsiness Detection Using Image Enhancement and Swin Transformer

Shravani Suresh Kotal¹, P. A. Sawant², Dr. H. B. Jadhav³

Student, Department of Computer Engineering¹

Guide, Department of Computer Engineering²

Co-Guide, Department of Computer Engineering³

Adsul's Technical Campus, Chas, Ahilyanagar, India

Abstract: This study introduces a smart system designed to detect driver drowsiness in real time, focusing mainly on checking whether a driver's eyes are closed for an unusually long period. While conventional AI image detection methods (like standard CNNs) often fail when faced with poor lighting, motion blur, noisy images, or deliberate visual distortions (adversarial attacks), this solution solves these issues by combining two advanced techniques:

Diffusion-Based Denoising (U-Net Architecture): Acts as a preprocessing cleanup layer. Before analyzing the image, it strips away visual noise, blur, and distortion, turning corrupted camera frames into clear, high-quality images.

Swin Transformer Model: Serves as the primary classification engine. It uses a multi-scale, window-based attention mechanism to accurately detect eye states (open vs. closed) across both fullface images and close-up eye patches.

To make the AI more resilient to unpredictable driving conditions, the researchers trained it using artificial image modifications (like rotation, zooming, and shifting) along with simulated visual attacks.

Keywords: Driver Drowsiness Detection, Swin Transformer, Diffusion Models, Image Denoising, Deep Learning, Eye-State Detection, Adversarial Training, U-Net, Image Processing, Driver Safety.

I. INTRODUCTION

Driving late at night or for long stretches can easily lead to fatigue, making driver drowsiness one of the leading causes of serious road accidents worldwide. When a driver starts to nod off, their reaction times slow down dramatically, which often results in severe collisions. To tackle this, researchers have been developing real-time driver monitoring systems designed to spot early signs of tiredness—most notably, tracking whether a driver's eyes stay closed for longer than normal.

While modern AI tools have made great strides in monitoring eye movements, they still struggle in real-world driving conditions. Low lighting inside the car, camera shake, bad weather, or sudden visual glare can easily corrupt video frames, causing these systems to misinterpret a closed eye as open—or vice versa. On top of that, digital visual distortions (known as adversarial noise) can trick traditional vision models into making dangerous errors.

Diffusion-Based Image Restoration: Acting as an automated digital cleanup layer, a U-Net diffusion model takes blurry, noisy, or distorted camera frames and cleans them up before any detection happens.

Swin Transformer Vision Model: Once the image is restored, a Swin Transformer model analyzes it. By looking at both the whole face and close-up patches around the eyes using advanced self-attention mechanisms, it accurately determines whether the eyes are open or closed—even under tough conditions.



II. OBJECTIVES AND CONTRIBUTIONS

- To detect driver drowsiness accurately.
- To improve image quality and to strengthen the model against adversarial attacks.
- To improve the model's ability to handle different images.
- Combination of Swin Transformer and diffusion models.
- Improved robustness against image distortions.
- Comprehensive performance evaluation and Improved detection performance.

III. RELATED WORK

Over the years, researchers have tried various strategies to detect driver drowsiness, which generally fall into three main categories: Physiological Systems:

Vehicle Behavior Monitoring: Other researchers looked at driving patterns, such as sudden lane drifting or erratic steering wheel movements. However, these methods are easily influenced by bad road conditions, weather, or individual driving habits, often delaying alerts until after a driver has already lost control.

Computer Vision and AI Approaches: To overcome these drawbacks, non-intrusive camera systems became the preferred choice. Later, deep learning models like Convolutional Neural Networks (CNNs) were introduced, significantly improving eye-state detection under controlled conditions.

Sensitivity to Bad Lighting and Blur: Most existing vision-based systems perform well only under bright, clear conditions. They struggle when camera feeds are degraded by low night lighting, heavy shadows, or vehicle vibrations.

Vulnerability to Visual Noise: Most current models are not built to withstand digital distortions or adversarial attacks, making them prone to misinterpreting eye states in unpredictable environments.

By addressing these limitations, recent research has shifted toward combining generative image restoration with modern vision transformers, creating a much more resilient defense against real-world driving conditions.

IV. RESEARCH GAP

Existing emphasis	Observed limitation	Proposed approach
Inability to Handle Real-World Image Noise	They fail to account for how real-world driving issues—such as motion blur, low nighttime lighting, shadows, or digital interference—severely degrade image quality and trick traditional AI models into making wrong decisions.	Most current studies test their models on clean, high-quality images
Separation of Cleaning and Detection	There is a lack of integrated frameworks that restore degraded images using advanced generative techniques before passing them to visual classifier.	try to detect eye closure directly from raw video feeds without cleaning them first.
Over-Reliance on Standard CNNs	They lack the powerful context-awareness that modern vision transformers provide.	Many current solutions still rely solely on traditional Convolutional Neural Networks (CNNs), which struggle with fine visual details and localized features when an image is partially obscured or distorted
Weak Defense Against Visual Distortions	current models lack the necessary resilience needed for safety-critical automotive applications	Very few drowsiness detection frameworks are designed to resist deliberate or unexpected digital distortions

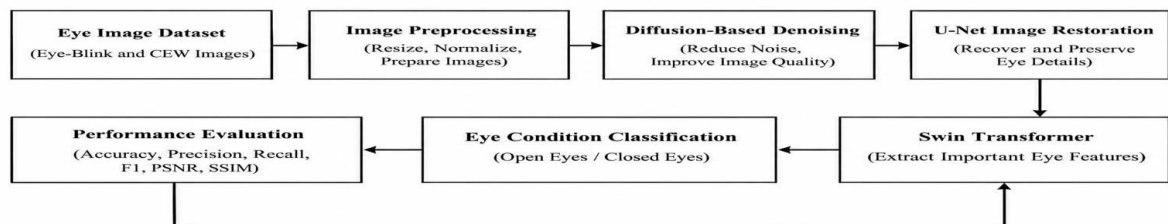


V. PROPOSED SYSTEM

The proposed system aims to detect driver drowsiness more accurately by improving the quality of images captured by a camera before identifying whether the driver's eyes are open or closed. In real-world situations, images of a driver's face may be affected by poor lighting, noise, blur, or low resolution. These issues can make it difficult for a normal drowsiness detection system to identify eye conditions correctly.

To overcome this problem, the system combines diffusion-based image denoising with a Swin Transformer model. First, the input image is processed to reduce noise and improve its visual quality. The diffusion-based denoising method helps restore image details, while the U-Net structure supports the reconstruction process. The improved image is then analyzed using the Swin Transformer to identify important visual features related to eye conditions and help determine whether the driver may be drowsy.

- Proposed layered architecture of the AI-Based Driver Drowsiness Detection Using Image Enhancement and Swin Transformer :



VI. HARDWARE COMPONENTS

Component	Function	Design consideration
Camera / Webcam	A camera captures images of the driver's face and eye region	Captures real-time images of the driver
Computer / Laptop	It processes the captured images and displays the prediction results	Runs the software and processes images
GPU (Graphics Processing Unit)	A GPU helps perform the complex calculations required by the Swin Transformer and diffusion-based denoising models	Accelerates deep-learning computations
RAM	Sufficient memory helps the computer handle image datasets and deep-learning tasks smoothly	Supports efficient data processing and multitasking
Storage Device (SSD or HDD)	The storage device keeps the Eye-Blink and CEW datasets, trained model files, program code, and experimental results for future use	Stores datasets, models, and project files
Power Supply	Provides electricity to operate the camera and computing devices	Supplies reliable power to the system's hardware components

VII. SOFTWARE COMPONENTS

- Matplotlib helps display images and create graphs of the experimental results. It can be used to compare original and denoised images and visualize the model's performance during testing
- NumPy helps handle image data in the form of numerical arrays. It supports calculations and data manipulation required during image processing and model evaluation.
- OpenCV helps the system capture images from a camera and prepare them for analysis. It can be used to resize images, adjust their format, and perform basic image-processing operations before they are passed to the mode.



- Python is the main programming language used to develop the project. It connects the different parts of the system, including image preprocessing, image denoising, and driver drowsiness detection.

VIII. SYSTEM METHODOLOGY

- **Data Collection**

The first step is to collect images of eyes for training and testing the system. The project uses the Eye-Blink and CEW datasets, which contain eye images useful for studying eye conditions. These images help the model learn to distinguish between open and closed eyes

- **Image Preprocessing**

After collecting the images, they are prepared for model processing. Preprocessing may include resizing images into a suitable format, normalizing pixel values, and organizing the data for training and testing. This helps maintain consistency when the images are passed to the model

- **Image Denoising Using the Diffusion Model**

In this stage, the system uses a diffusion-based denoising model to improve the quality of input images. Images captured under poor lighting or containing unwanted noise may make eye detection difficult. The denoising process helps reduce these disturbances and recover useful visual details

- **Eye Condition Detection**

The extracted features are used to identify eye conditions, such as open or closed eyes. These predictions can help support driver drowsiness detection. However, eye condition alone does not always confirm drowsiness; a practical system may need additional information, such as eye closure duration, to make a more reliable decision.

IX. PROPOSED ALGORITHM

1. Initialize the legal document collection, processing module, multilingual query interface, retrieval system, language model, and citation verification module.
2. Load Eye-Blink and CEW datasets.
3. Preprocess the images. Resize and normalize the input images.
4. Apply diffusion-based image denoising.
5. Use the U-Net structure to restore image details.
6. Pass the enhanced images to the Swin Transformer.
7. Train the model using the prepared dataset. Apply data augmentation and suitable robustness techniques, including FGSM and PGD.
8. Analyze the extracted eye features.
9. Predict the eye condition: Open eyes or closed eyes
10. Evaluate the model using: Accuracy, Precision, Recall, F1-score, PSNR and SSIM.
11. Display the prediction and evaluation results.

X. SECURITY AND PRIVACY DESIGN

The proposed security and privacy design aims to protect driver images, restrict unauthorized access, and improve the reliability of the drowsiness detection system. By combining responsible data handling with secure storage, access control, and robustness testing, the system can be developed with greater attention to user privacy and safety.

1. The system should handle driver images responsibly because facial images may contain personal information. Only the images needed for drowsiness detection should be collected and processed. Unnecessary personal information should not be stored.



2. The images used for training and testing, along with trained model files and experimental results, should be stored in a secure location. Access should be limited to authorized users to reduce the risk of data leakage or misuse
3. Access to the system and its stored data should be restricted to authorized users. Password protection and appropriate user permissions can help prevent unauthorized people from viewing or modifying project files.
4. The proposed system considers adversarial techniques such as Fast Gradient Sign Method (FGSM) and Projected Gradient Descent (PGD) to study the effect of small, deliberately introduced changes to input images. These techniques can help evaluate and improve model robustness, but they do not guarantee complete protection against attacks.

XI. EVALUATION METHODOLOGY

The evaluation methodology checks whether the proposed system can correctly identify eye conditions while improving the quality of input images. Accuracy, precision, recall, and F1-score measure the classification performance. PSNR and SSIM help determine whether the denoising process produces clearer images while preserving important details.

Metric	Measurement method	Target interpretation
Accuracy	Measure the percentage of eye images correctly classified by the model	Higher accuracy is better
Precision	Measure how many images predicted as a particular eye condition are correctly classified	Higher precision is better
PSNR (Peak Signal-to-Noise Ratio)	Compare the denoised image with the reference image to measure image reconstruction quality	Higher PSNR generally indicates better image quality
SSIM (Structural Similarity Index)	Compare the structural details and visual similarity of the original and denoised images	A value closer to 1 indicates greater similarity.
Denoising Effectiveness	Compare images before and after denoising to check whether noise is reduced while important eye details are preserved	Better noise reduction with fewer lost details is preferred.
Generalization	Evaluate the trained model on unseen test images from the selected datasets, such as Eye-Blink and CEW.	Better performance on unseen images is preferred

XII. TEST SCENARIOS

- Test the system using clear images of open and closed eyes to check whether it identifies eye conditions correctly.
- Provide images containing noise to check whether the diffusion model improves image quality.
- Use PSNR and SSIM to measure how effectively the denoising model improves image quality
- Test the model on new images that were not used during training to evaluate its ability to generalize.
- Check whether all components work together correctly, from image input and denoising to eye-condition prediction and performance evaluation

XIII. RESULTS AND DISCUSSION

The diffusion-based denoising model aims to reduce unwanted noise in input images while preserving important details around the eyes. The U-Net encoder-decoder structure helps reconstruct image details, making the images more suitable for further analysis. The quality of the restored images can be evaluated using PSNR and SSIM. Higher PSNR and SSIM values generally indicate better reconstruction quality and greater similarity to the reference images.



The combination of diffusion-based denoising and Swin Transformer feature extraction offers two complementary benefits. The denoising stage aims to improve the visual quality of the input, while the transformer learns useful patterns for eye-condition recognition.

The system's overall effectiveness depends on how well these components work together. Improving image quality does not automatically guarantee better classification, so both denoising quality and detection performance must be evaluated experimentally

XIV. ADVANTAGES AND LIMITATIONS

- **Advantages**

1. Improves eye-image quality by reducing noise through diffusion-based image denoising.
2. Uses the Swin Transformer to identify important eye features and distinguish between open and closed eyes
3. Improves the system's ability to handle noisy and distorted images
4. Adversarial training using FGSM and PGD can help improve the model's robustness against small input perturbations.
5. Combines image denoising and deep learning to support more reliable driver drowsiness monitoring.

- **Limitations**

1. Performance depends on the quality and diversity of the eye images used for training and testing.
2. Diffusion models and Swin Transformers may require high computational power and memory.
3. Poor lighting, severe blurring, or partially covered eyes may affect detection accuracy.
4. Identifying open and closed eyes alone may not be sufficient to confirm driver drowsiness.
5. Real-time performance may be limited by image-processing time and available hardware resources.

XV. FUTURE SCOPE

1. The system can be extended to detect drowsiness in real time using a camera installed inside the vehicle..
2. Additional facial features, such as yawning, head movement, and facial expressions, can be included to improve drowsiness detection.
3. The model can be optimized to reduce computational requirements and support deployment on embedded devices and low-cost hardware.
4. More diverse datasets containing different lighting conditions, face angles, and driver appearances can be used to improve the model's reliability.
5. Future improvements can focus on faster image denoising and prediction to support real-time driver monitoring.

XVI. CONCLUSION

This paper proposed AI-Based Driver Drowsiness Detection Using Image Enhancement and Swin Transformer aims to improve driver drowsiness detection by combining image denoising and deep-learning techniques. The diffusion model helps reduce noise and improve image quality, while the Swin Transformer extracts important features from eye images to identify open and closed eyes.

The system also considers data augmentation and adversarial techniques to improve the model's robustness under different image conditions. Its performance can be evaluated using accuracy, precision, recall, F1-score, PSNR, and SSIM. Although challenges such as poor lighting, computational requirements, and image quality may affect performance, the proposed approach has the potential to support more reliable driver monitoring.



REFERENCES

1. S. Abd El-Nabi, A. F. Ibrahim, E.-S. M. El-Rabaie, O. F. Hassan, N. F. Soliman, K. F. Ramadan, and W. El-Shafai, "Driver Drowsiness Detection Using Swin Transformer and Diffusion Models for Robust Image Denoising," IEEE Access, vol. 13, 2025.
2. D. P. Kingma, T. Salimans, B. Poole, and J. Ho, "Variational Diffusion Models," in Advances in Neural Information Processing Systems, 2021, pp. 21696–21707
3. X. Li, X. Li, Z. Shen, and G. Qian, "Driver Fatigue Detection Based on Improved YOLOv7," Journal of Real-Time Image Processing, vol. 21, no. 3, Art. no. 75, 2024
4. D. Salem and M. Waleed, "Drowsiness Detection in Real-Time via Convolutional Neural Networks and Transfer Learning," Journal of Engineering and Applied Science, vol. 71, no. 1, Art. no. 122, 2024.
5. R. Daza, A. Morales, J. Fierrez, R. Tolosana, and R. Vera-Rodriguez, "MEBAL2 Database and Benchmark: Image-Based Multispectral Eyeblink Detection," Pattern Recognition Letters, vol. 182, pp. 83–89, 2024.

