Hybrid CNN and Temporal Analysis Approach for Robust Driver Drowsiness Detection

Manith Ratnayake   Ms. Rasanjalee Rathnayake

Problem

Driver drowsiness is a major contributing factor to road accidents worldwide, resulting in severe social and substantial economic losses. Most existing computer vision-based solutions are primarily evaluated in **simulated environments**. This fails to capture the long-term temporal patterns inherent in **real world driving conditions**.

- Expensive to purchase detection products due to commercial reasons.
- Commercial products' model weights are not publicly available.
- Level 5 self-driving cars are still far from real world deployment.

Detection Methods

Physiological | **Behavioral** | Vehicular | Subjective

Objectives

- Test sim-trained models on real driving data, measure accuracy drop and reduce it through finetuning with a small real dataset
- Use images from dashboard, mirror and steering column viewpoints. Compare model stability across them
- Benchmark different model architectures for the same dataset to identify the best performer
- Evaluate long term fatigue trends by analyzing multi seconds or multi minutes driving sessions
- Experiment the optimal balance between model accuracy and inference speed for deployment
- Test the model on videos with different resolutions and frame rates, note accuracy changes, and train with mixed quality inputs for robustness
- Use attention visualization to check where the model focuses and ensure it tracks key facial cues correctly
- Use an ensemble of models for predictions to reduce false positives and false negatives
- Examine head pose sensitivity by measuring accuracy under different driver head orientations
- Analyze the effect of occlusions such as sunglasses, masks, hats, and hands covering the face

Timeline

- **1. Data Collection** September - October
- **2. Literature Review** October - November
- **3. Implementation** December - February
- **4.Evaluation** March

Data Strategy

Training & Testing — National Tsing Hua University Driver Drowsiness Detection Dataset

Deployment Testing — Dataset Owned by a Private Company

Innovation & Significance

- Custom Light Weight Architecture (Fu et al. 2025, *Flores-Monroy et al. 2021)*
- Architecture that performs better with a GPU (Jarndal et al. 2025)

Literature Review

Architecture	Accuracy
Integrated MobileNetV3 as a backbone for feature extraction with a Feature Pyramid Network (FPN) to capture multi-scale facial details.	99.4%
FU, G. et al., 2025. A Lightweight MobileNetV3-FPN Framework for Driver Drowsiness Detection. In: 2025 International Symposium on Signals, Circuits and Systems (ISSCS). 2025 International Symposium on Signals, Circuits and Systems (ISSCS), Iasi, Romania, 17 July 2025. Iasi, Romania: IEEE. pp. 1–4. Available from: https://ieeexplore.ieee.org/document/11105647/ [Accessed 7 Oct 2025].	
Feature extraction is done using Viola & Jones algorithm. A Shallow CNN (SS-CNN) architecture with around 600K parameters is introduced.	98.95%
FLORES-MONROY, J. et al., 2021. Visual-based Real Time Driver Drowsiness Detection System Using CNN. In: 2021 18th International Conference on Electrical Engineering, Computing Science and Automatic Control (CCE). 2021 18th International Conference on Electrical Engineering, Computing Science and Automatic Control (CCE), Mexico City, Mexico, 10 November 2021. Mexico City, Mexico: IEEE. pp. 1–5. Available from: https://ieeexplore.ieee.org/document/9633082/ [Accessed 7 Oct 2025].	
A Multi-task Convolutional Neural Network (ConNN*) is proposed to detect drowsiness by simultaneously analyzing eye and mouth features, using PERCLOS and FOM to determine fatigue levels across three classes	98.91%
SAVAS, B.K. and BECERIKLI, Y., 2020. Real Time Driver Fatigue Detection System Based on Multi-Task ConNN. IEEE Access, 8, pp. 12491–12498.	
A Practical Facial Landmark Detector (PFLD) detects 106 key points and uses a transformer encoder fused with EAR, MAR, and 3D head pose features, along with an LSTM to model temporal dynamics, for detection.	93.15%
TAIGUO LI and CHAO LI, 2024. A Deep Learning Model Based On Multi-granularity Facial Features And LSTM Network For Driver Drowsiness Detection. 淡江理工學刊, 27(7).	
The architecture consists of a Vision Transformer backbone that encodes the face image into patch embeddings, processes them through multiple transformer encoder layers with self-attention to extract global features, and finally uses an MLP classifier for prediction	98.89%
JARNDAL, A. et al., 2025. A Real-Time Vision Transformers-Based System for Enhanced Driver Drowsiness Detection and Vehicle Safety. IEEE Access, 13, pp. 1790–1803.	

Note: Accuracy based only on models trained and tested on the NTHUDDD dataset

Commercial Products

Research Gap

The primary limitation in the current literature is the lack of comprehensive, real world validation datasets and the insufficient focus on generalizability beyond simulated environments. This limits the identification of drowsiness patterns over long durations. Reasons causing the issue are,

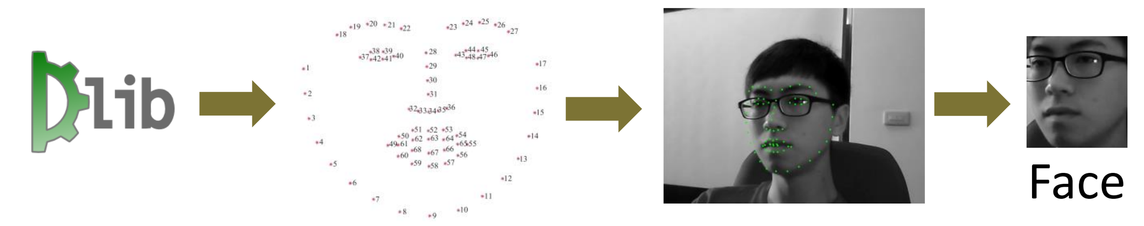
- Competitive pressure among companies in the drowsiness detection market
- Concerns over personal data privacy in driver monitoring datasets
- Regulatory and compliance challenges related to data collection and usage

Because of this, many proposed driver drowsiness detection systems remain untested in realistic driving conditions, leading to uncertainty about their practical reliability.

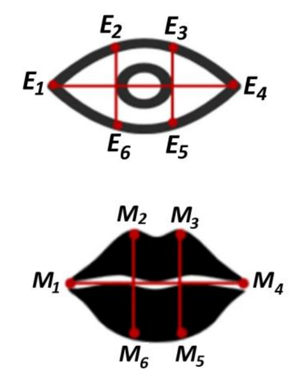
Proposed Methodology

1. Image Processing

2. Locating Eye & Mouth Key Points



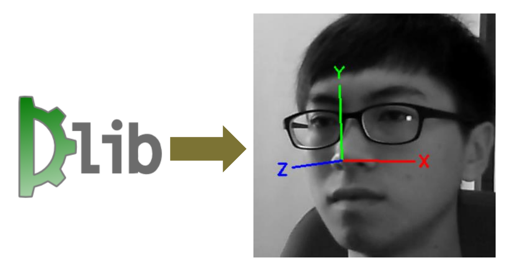
3. Eye & Mouth Openness Level



$$\text{EAR} = \frac{||E_2 - E_6|| + ||E_3 - E_5||}{2||E_1 - E_3||}$$

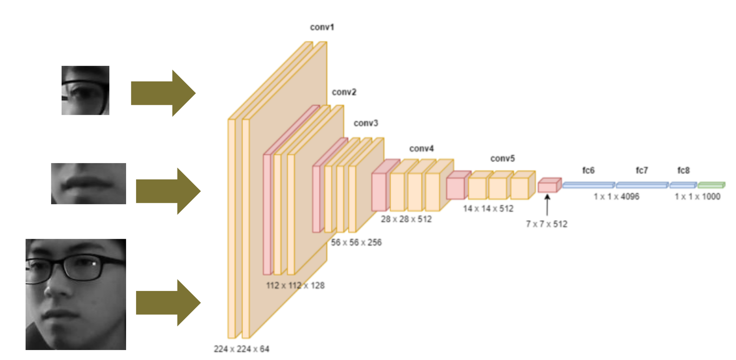
$$\text{MAR} = \frac{||M_2 - M_6|| + ||M_3 - M_5||}{2||M_1 - M_3||}$$

4. Head Pose Estimation



6 Degree of Movement

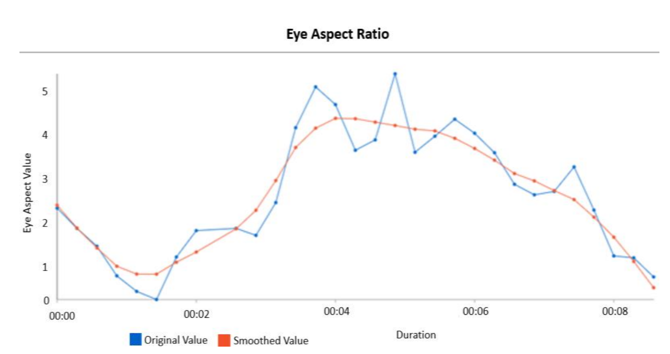
5. Spatial Model Prediction



- CNN
- ViTs
- Attention + CNN

PyTorch, Hugging Face

6. Temporal Smoothing



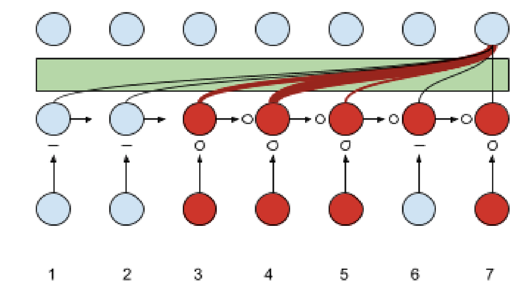
- Exponential Smoothing
- Kalman Filter
- Savitsky Golay

7. Eye Closure & Mouth Opening Duration

$$\text{PERCLOS} = \frac{T_{\text{closed}}}{T} \times 100\%$$

$$\text{FOM} = \frac{N_{\text{open}}}{N}$$

8. Temporal Model Prediction



- LSTM
- HMM
- Attention

9. Model Evaluation

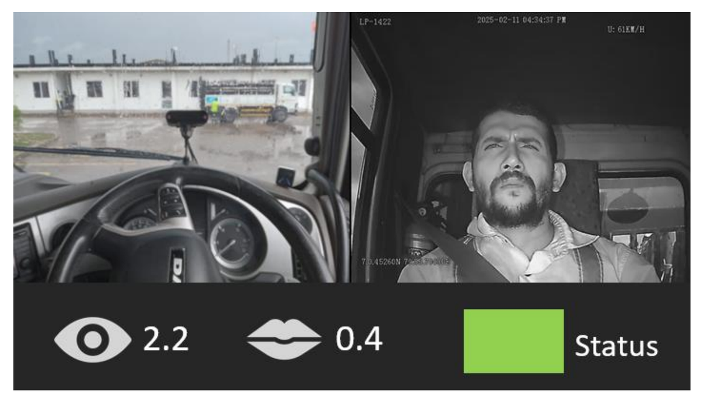
scikit learn

- Accuracy
- Precision
- Confusion Matrix
- Recall
- AUC-ROC
- F1 Score

10. Benchmarking

- Model Parameters Count
- Inference Speed
- Resource Usage (CPU/GPU)

Expected Contribution



- Deliver a **highly reliable, interpretable, and deployment ready** driver drowsiness detection system suitable for real time operation in intelligent transportation systems.
- Offer validated insights into how models trained on public datasets can be adapted for personalized driver drowsiness contexts to improve real world applicability.