International Scientific Journal of Engineering and Management (ISJEM) ISSN: 2583-6129
Conference Proceedings: Volume: 05 ACME – 2026 DOI: 10.55041/ISJEM XXXX
An International Scholarly || Multidisciplinary || Open Access || Indexing in all major Database & Metadata

# Hybrid GNN-Based Driver Monitoring System using Facial Landmarks

**M Poojitha[1], M Shobha Rani[2], K Saketh Ram[3] , A Bhagyaraja[4], Mrs L Lavanya [5]**

[1] *Department of Computer Science and Engineering(AI&ML), Student, Sri Venkateswara College of Engineering, Tirupati*

[2] *Department of Computer Science and Engineering(AI&ML), Student, Sri Venkateswara College of Engineering, Tirupati*

[3] *Department of Computer Science and Engineering(AI&ML), Student, Sri Venkateswara College of Engineering, Tirupati*

[4] *Department of Computer Science and Engineering(AI&ML), Student, Sri Venkateswara College of Engineering, Tirupati*

[5] *Department of Computer Science and Engineering(AI&ML), Assistant Professor, Sri Venkateswara College of Engineering, Tirupati*

---------------------------------------------------------------------***---------------------------------------------------------------------

***Abstract---***Driver fatigue and stress are major causes of road accidents and unsafe driving conditions. Traditional methods that rely on Eye Aspect Ratio (EAR) and Mouth Aspect Ratio (MAR) often fail under real-world conditions such as lighting variations, head movements, and subtle facial changes. This paper presents a hybrid driver monitoring framework based on insights from existing studies for detecting both driver drowsiness and stress using facial landmarks. The system utilizes MediaPipe Face Mesh to extract 468 facial landmarks and computes classical features such as EAR and MAR for fatigue detection. In addition, a Graph Neural Network (GNN) is employed to model the relationships between facial landmarks and capture complex behavioral patterns. The outputs of classical and graph-based methods are fused to determine the driver's state as SAFE, WARNING, or CRITICAL. The proposed hybrid approach aims to balance computational efficiency and intelligent feature learning, making it suitable for real-time applications.

This paper not only reviews existing driver monitoring techniques but also proposes a hybrid framework based on identified research gaps.

**KEYWORDS**
Driver Monitoring System, Drowsiness Detection, Stress Detection, Graph Neural Networks, Facial Landmarks, Hybrid Approach

## 1. INTRODUCTION

Driver fatigue and psychological stress are critical factors that negatively influence driving safety. Reduced alertness, delayed reactions, and poor decision-making caused by fatigue significantly increase accident risk. Similarly, stress affects concentration and cognitive performance, making driving more hazardous. Many existing driver monitoring systems rely on simple visual cues such as eye closure or yawning detection. Although these approaches are computationally efficient, they are not reliable under real-world conditions, where lighting variations, head movements, and occlusions are common. Additionally, most systems are limited to drowsiness detection and do not consider stress analysis. Recent developments in computer vision have enabled the use of facial landmarks for non-intrusive monitoring. However, conventional approaches often treat these features independently and fail to capture their interrelationships. To overcome this limitation, this work utilizes Graph Neural Networks to model structured dependencies among facial landmarks.

The proposed system combines classical feature extraction with graph-based learning to achieve a balance between efficiency and intelligent pattern recognition, making it suitable for real- time driver monitoring.

## 2. PROBLEM STATEMENT

Driver drowsiness and stress significantly contribute to road accidents worldwide. Most existing systems focus only on detecting fatigue using basic indicators such as eye blinking or yawning, which are highly sensitive to environmental conditions like lighting changes and head movements. Advanced deep learning models can improve detection accuracy but require large datasets and

International Scientific Journal of Engineering and Management (ISJEM) ISSN: 2583-6129
Conference Proceedings: Volume: 05 ACME – 2026 DOI:10.55041/ISJEM.ACME038
An International Scholarly || Multidisciplinary || Open Access || Indexing in all major Database & Metadata

high computational resources, making them less suitable for real-time applications. On the other hand, physiological signal-based approaches such as EEG provide reliable results but are intrusive and not practical for everyday usage. Therefore, there is a need for a system that is lightweight, non- intrusive, and capable of detecting both drowsiness and stress efficiently using visual information alone.

## 3. LITERATURE SURVEY

### 3.1 CLASSICAL METHODs

Classical approaches rely on geometric relationships between facial landmarks. Eye Aspect Ratio (EAR) is used to detect eye closure, while Mouth Aspect Ratio (MAR) identifies yawning behavior. These methods are computationally efficient and suitable for real-time applications. However, they are highly sensitive to environmental conditions and require manually defined thresholds.

### 3.2 DEEP LEARNING-BASED METHODS

Convolutional Neural Networks (CNNs) have been widely used for driver monitoring by analyzing facial images. These models achieve high accuracy but require large datasets and computational resources. Recurrent Neural Networks (RNNs) and Long Short-Term Memory (LSTM) models are used to capture temporal dependencies in video sequences, but they introduce additional latency.

### 3.3 MULTIMODAL SYSTEMS

Multimodal approaches combine visual features with physiological signals such as heart rate or EEG. While these systems improve accuracy, they depend on additional hardware, limiting their practicality for real-world deployment. Camera-based multimodal detection remains relatively underexplored.

### 3.4 GRAPH NEURAL NETWORK-BASED METHODS

Recent research has explored the use of Graph Neural Networks to model relationships between facial landmarks. By representing landmarks as nodes and their connections as edges, GNNs can capture spatial dependencies more effectively than traditional methods. However, their application in real-time driver monitoring systems is still limited.

### 3.5 COMPARATIVE INSIGHTS

Classical methods such as EAR-based detection [10] provide real-time performance but are sensitive to environmental variations. Deep learning approaches [2], [6] improve detection accuracy but introduce computational overhead. Recent studies using Graph Neural Networks [1], [3] demonstrate improved capability in capturing spatial relationships among facial landmarks. However, their application in real-time systems remains limited. Furthermore, stress detection techniques [7]–[9] highlight the feasibility of visual emotional analysis but are rarely integrated with drowsiness detection in a unified framework.

## 4 RESEARCH GAPS

Based on the reviewed literature, several research gaps are identified. Most existing systems focus primarily on drowsiness detection and neglect stress monitoring. Multimodal systems often rely on physiological sensors, limiting their practicality for real-world deployment. Although Graph Neural Networks have been explored for facial analysis [1], [3], their integration into real-time driver monitoring systems remains limited. Additionally, there is a lack of hybrid frameworks that combine lightweight classical methods with relational modeling techniques for efficient and robust performance.

## 5. LIMITATIONS OF EXISTING SYSTEMS

- High false positives in classical methods
- Sensitivity to lighting and environmental conditions
- Inability to detect subtle fatigue patterns such as micro-sleeps
- High computational requirements for deep learning models
- Lack of stress detection in most systems
- Intrusive nature of physiological monitoring systems

## 6. PROPOSED HYBRID SYSTEM

Based on the identified research gaps, a hybrid system architecture is proposed as a potential solution. The proposed system integrates classical feature extraction with Graph Neural Networks to create a robust and efficient driver monitoring system.

International Scientific Journal of Engineering and Management (ISJEM) ISSN: 2583-6129
Conference Proceedings: Volume: 05 ACME – 2026 DOI:10.55041/ISJEM.ACME038
An International Scholarly || Multidisciplinary || Open Access || Indexing in all major Database & Metadata

### 6.1 SYSTEM ARCHITECTURE

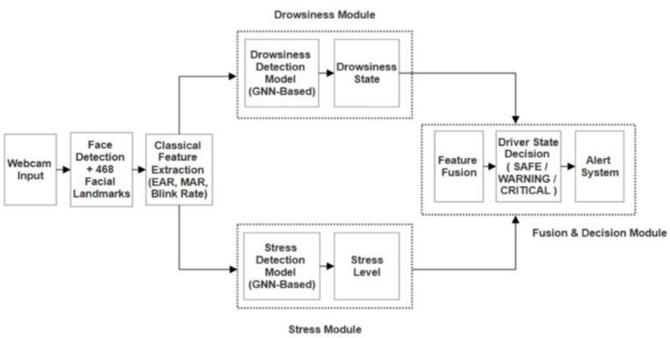


***Fig 1:** Architecture*

The proposed architecture illustrates the complete workflow of the driver monitoring system. The input video stream is processed to extract facial landmarks, which are used for both classical feature extraction and graph construction. The Graph Neural Network (GNN) modules analyze drowsiness and stress independently, and their outputs are fused to determine the final driver state. An alert mechanism is triggered when unsafe conditions are detected.

### 6.2 FACE AND LANDMARK DETECTION MODULE

A webcam is used to capture real-time video input. The system detects the driver's face and extracts 468 facial landmarks using the MediaPipe Face Mesh model. These landmarks provide detailed information about facial structure and movement.

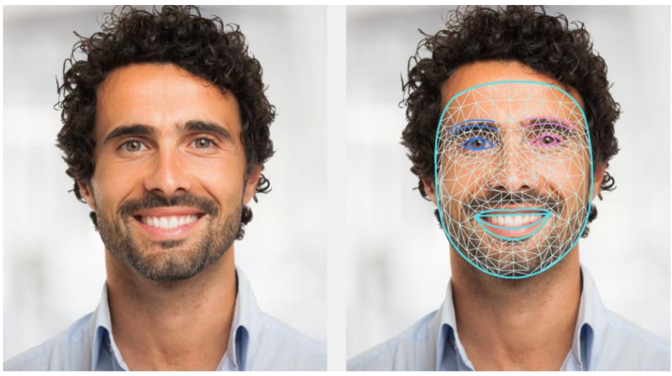

***Fig 2:** Facial landmarks mesh*

### 6.3 CLASSICAL FEATURE EXTRACTION MODULE

From the extracted landmarks, the system computes:

- Eye Aspect Ratio (EAR)
- Mouth Aspect Ratio (MAR)
- Blink rate and eye closure duration

These features provide quick and interpretable indicators of driver fatigue.

### 6.4 DROWSINESS DETECTION MODULE (GNN-BASED)

Facial landmarks are converted into a graph structure where nodes represent landmarks and edges represent spatial relationships. A Graph Neural Network processes this graph to detect patterns associated with drowsiness. The module classifies the driver as ALERT or DROWSY.

### 6.5 STRESS DETECTION MODULE (GNN-BASED)

A parallel GNN-based module analyzes facial expressions and relationships between landmarks to detect stress levels. The system classifies stress into LOW, MEDIUM, or HIGH categories. This module explores the use of relational modeling for emotional state detection.

Stress detection using facial landmarks has been explored in [7]–[9], indicating the feasibility of vision-based emotional state analysis without additional sensors.

Facial landmarks inherently form a structured representation, making them well-suited for graph-based modeling approaches.

### 6.6 FEATURE FUSION AND DECISION MODULE

The outputs from the drowsiness and stress modules are combined using a fusion mechanism. Based on predefined rules, the driver state is classified into:

- SAFE
- WARNING
- CRITICAL

### 6.7 ALERT SYSTEM

When the driver enters a WARNING or CRITICAL state, the system generates real-time alerts to prevent potential accidents.

## 7. METHODOLOGY

### 7.1 DATASETS USED

The proposed system utilizes two publicly available datasets for training and evaluation of the Graph Neural Network (GNN) modules.

- **NTHU Driver Drowsiness Detection (NTHU-DDD) Dataset**:

This dataset is used for training the drowsiness detection module. It contains video sequences of drivers under various conditions such as yawning, eye closure, head movement, and different lighting environments.

- **Cohn-Kanade (CK+) Dataset**:

The CK+ dataset is used for training the stress detection module. It includes facial expression sequences representing different emotional states, which are useful for identifying stress-related patterns.

### 7.2 DATA PREPROCESSING

Before feeding the data into the model, several preprocessing steps are performed to ensure consistency and improve model performance.

1. **FrameExtraction**

Video sequences from the NTHU-DDD dataset are converted into individual frames for analysis.

2. **Face Detection and Landmark Extraction**

Facial landmarks (468 points) are extracted using a face mesh model. These landmarks represent key facial regions such as eyes, mouth, and eyebrows.

3. **DataSplitting**

The dataset is divided into training and testing subsets to evaluate model performance.

The dataset was divided into training and testing sets as follows:

- **Total samples**: 2782
- **Training samples**: 736
- **Testing samples**: 184

A subset of the dataset was used for training and testing due to preprocessing constraints.

### 7.3 DATA PROCESSING PIPELINE

The proposed system processes a real-time video stream captured from a webcam. Each frame undergoes face detection followed by extraction of 468 facial landmarks using a facial mesh model. These landmarks are utilized in classical feature extraction and then graph-based modeling.

The extracted landmarks are first used to compute geometric features such as Eye Aspect Ratio (EAR), Mouth Aspect Ratio (MAR), and blink rate. Simultaneously, the landmarks are transformed into a graph structure and processed using Graph Neural Networks (GNNs). The outputs from both modules are fused to determine the final driver state.

### 7.4 CLASSICAL FEATURE EXTRACTION

#### EYE ASPECT RATIO (EAR)

The Eye Aspect Ratio is used to detect eye closure and is defined as:

$$EAR = \frac{|p2 - p6| + |p3 - p5|}{2 * |p1 - p4|}$$

where p1,p2,...,p6 represent eye landmark coordinates.

#### MOUTH ASPECT RATIO (MAR)

The Mouth Aspect Ratio is used to detect yawning behavior:

$$MAR = \frac{|p3 - p9| + |p4 - p8|}{2 * |p1 - p7|}$$

### 7.5 GRAPH CONSTRUCTION

Facial landmarks are modeled as a graph G=(V,E)G = (V, E)G=(V,E), where:

- V represents nodes corresponding to facial landmarks
- E represents edges connecting spatially related landmarks

Edges are defined based on facial regions such as eyes, mouth, and eyebrows to capture structural relationships.

### 7.6 GRAPH NEURAL NETWORK PROCESSING

The constructed graph is processed using a Graph Neural Network, where node features are iteratively updated using message passing:

$$h_i^{(k+1)} = \sigma\left(\sum_{j \in N(i)} \quad W \cdot h_j^{(k)}\right)$$

where:

- hi^{(k)} - hi^(k) is the feature vector of node i at layer k
- N(i) represents neighboring nodes
- W is the learnable weight matrix
- σ is an activation function

This enables the model to capture spatial dependencies between facial features.

Unlike CNN-based approaches that treat images as grid structures, GNNs explicitly model relationships between facial landmarks, making them more suitable for structured facial data.

### 7.7 DROWSINESS DETECTION MODULE

The GNN processes the facial graph to classify the driver's state as **Alert or Drowsy**, based on learned relational patterns such as eye closure and facial fatigue indicators.

### 7.8 STRESS DETECTION MODULE

A parallel GNN module analyzes facial expressions and structural changes to estimate stress levels. The output is categorized as **Low, Medium, or High stress**.

### 7.9 FEATURE FUSION AND DECISION MAKING

The outputs from classical features and GNN modules are combined using a rule-based fusion mechanism. The driver state is classified into:

- **SAFE** → Normal behavior
- **WARNING** → Early signs of fatigue/stress
- **CRITICAL** → High risk

An alert system is triggered when unsafe conditions are detected.

## 8. PERFORMANCE EVALUATION

The performance of the proposed system can be evaluated using standard classification metrics such as Accuracy, Precision, Recall, and F1-score.

### 8.1 ACCURACY

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

Where:

- **TP (True Positive)** → Correctly detected drowsy/stress state
- **TN (True Negative)** → Correctly detected alert state
- **FP (False Positive)** → Incorrectly predicted drowsiness
- **FN (False Negative)** → Missed detection

The performance of the proposed system is evaluated based on classification accuracy for both drowsiness and stress detection.

- **Drowsiness Detection Accuracy**: **94.43%**
- **Stress Detection Accuracy**: **95.11%**
- **Overall System Accuracy**: **94.60%**

The experimental results highlight the effectiveness of combining classical feature-based methods with graph-based learning. The hybrid model achieves an overall accuracy of **94.60%**, demonstrating its suitability for real-time driver monitoring applications. The system performs consistently under varying conditions, although further improvements can be achieved by incorporating larger datasets and optimizing the GNN architecture.

### 8.2 COMPARISON WITH EXISTING METHODS

To evaluate the effectiveness of the proposed hybrid approach, its performance is compared with existing methods reported in the literature.

| METHOD | ACCURACY(%) |
|---|---|
| EAR- based[10] | ~85-90 |
| CNN-based [2],[6] | ~90-95 |
| GNN based [1],[3] | ~91-93 |
| **Hybrid (proposed)** | **94.60** |

***Table 1: Comparative*** *analysis of accuracies*

## 9. COMPARATIVE ANALYSIS

| Method | Advantages | Limitations |
|---|---|---|
| EAR/MAR | Fast, simple | Sensitive to noise |
| CNN | High accuracy | Computationally expensive |
| RNN/LSTM | Captures temporal behavior | High latency |
| GNN | Captures relationships | Less explored, complex |
| **Hybrid (Proposed)** | **Balanced and efficient** | **Needs optimization** |

*Table 2: Comparision table*

Note: Performance values vary across datasets and experimental conditions; hence qualitative comparison is emphasized.

## 10. ADVANTAGES OF PROPOSED SYSTEM

- Works with standard webcam (low cost)
- Detects both drowsiness and stress
- Captures fine facial dynamics using landmarks
- Combines speed of classical methods with intelligence of GNN
- Suitable for real-time applications

## 11. FUTURE ENHANCEMENTS

- Integration with vehicle control systems
- Use of larger and diverse datasets
- Detection of driver distractions (mobile usage)
- Deployment on embedded systems
- Integration with cloud-based monitoring systems

## 12. DISCUSSION

The comparison of existing methods highlights a trade-off between computational efficiency and detection accuracy. Classical methods such as EAR and MAR are suitable for real-time applications but lack robustness under varying conditions. Deep learning models provide improved performance but require high computational resources and large datasets. Graph Neural Networks introduce a novel perspective by modeling relationships between facial landmarks. However, their practical implementation in real-time systems remains challenging. The proposed hybrid approach attempts to bridge this gap by combining the efficiency of classical methods with the representational power of GNNs. Additionally, the inclusion of a stress detection module enhances the system's capability beyond traditional drowsiness detection systems. Direct comparison based solely on accuracy is challenging due to variations in datasets, evaluation metrics, and experimental conditions across different studies. The proposed hybrid framework aligns with emerging research trends that emphasize lightweight, interpretable, and real-time intelligent driver monitoring systems. The experimental results highlight the effectiveness of combining classical feature-based methods with graph-based learning.

## 13. CONCLUSION

This paper reviewed various approaches for driver drowsiness and stress detection and identified key limitations in existing systems. Classical methods are efficient but lack robustness, while deep learning models are accurate but computationally expensive. Graph Neural Networks offer a promising solution by capturing relational information between facial landmarks. The proposed hybrid system combines classical and graph-based methods to achieve a balance between efficiency and accuracy. This approach has strong potential for real-time driver monitoring applications and can contribute to reducing road accidents.

The hybrid model achieves an overall accuracy of **94.60%**, demonstrating its suitability for real-time driver monitoring applications. The system performs consistently under varying conditions, although further improvements can be achieved by incorporating larger datasets and optimizing the GNN architecture.

The integration of classical feature-based methods with graph-based learning presents a promising direction for developing efficient, scalable, and real-time driver monitoring systems.

## ACKNOWLEDGEMENT

The authors would like to thank the project guide and institution for their support.

## REFERENCES

[1] V. Agarwal and Rajneesh, "An Accurate Detection of Drowsiness Using a Graph-Based Neural Network," in Proc. 2023 International Conference on Applied Intelligence and Sustainable Computing (ICAISC), 2023.

[2] N. Patil, N. Patole, C. Thorat, and A. Phapale, "Deep Learning Based Driver Yawning Detection," Journal of Emerging Technologies and Innovative Research (JETIR), 2023.

[3] Q. T. Ngoc, S. Lee, and B. C. Song, "Facial Landmark-Based Emotion Recognition via Directed Graph Neural Network," 2023.

[4] A. Ritesh, N. Jagatia, and P. Deshmukh, "Driver Drowsiness Detection Using Machine Learning," in Proc. 2023 6th International Conference on Advances in Science and Technology (ICAST), 2023.

[5] M. E. Shaik, "A Systematic Review on Detection and Prediction of Driver Drowsiness," Transportation Research Interdisciplinary Perspectives, vol. 21, 2023.

[6] H. S. Gujjar, "A Deep Learning Approach to Automated Drowsiness Detection in Drivers Using CNN Models for Enhanced Road Safety," International Journal on Science and Technology (IJSAT), 2025.

[7] T. Jeon, H. Bae, Y. Lee, S. Jang, and S. Lee, “Stress Recognition Using Face Images and Facial Landmarks,” in Proc. 2020 International Conference on Electronics, Information, and Communication (ICEIC), IEEE, 2020.

[8] M. Faizal, Nikhil, Prajwal, and A. R. B. P., “Recognition of Stress Using Face Image and Facial Landmarks,” International Advanced Research Journal in Science, Engineering and Technology, vol. 8, no. 5, May 2021.

[9] R. M. S. Parvathi, M. Sanjay, P. S. Santhosh Kumar, and P. Solai Murugan, “Stress Detection Using Machine Learning and Image Processing,” International Journal of Creative Research Thoughts (IJCRT), 2025.

[10] V. Varghese, A. Shenoy, S. K. Sreeram, and R. K. P., “EAR Based Driver Drowsiness Detection System,” IOSR Journal of Engineering (IOSRJEN), 2018.