International Journal of Artificial Intelligence, Data Science, and Machine Learning
Grace Horizon Publication | Volume 7, Issue 1, 271-278, 2026
ISSN: 3050-9262 | https://doi.org/10.63282/3050-9262.IJAIDSML-V7I1P146

Original Article

Attention-Based Driver Behavior Monitoring System Using Multi-Modal Deep Learning

Omprakash Gurrapu[1], Avinash Chandra[2], Pruthvi Kaluvala[3], Sunny Solmen Edelli Padmarao[4]
[1,2,3,4]Independent Researcher, USA.

Received On: 28/01/2026 ***Revised On: 29/02/2026*** ***Accepted On: 02/03/2026*** ***Published on: 05/03/2026***

Abstract - *Road traffic accidents resulting from driver distractions, drowsiness, and inattentiveness are a significant global concern. To address this problem, It proposes an "Attention-Based Driver Behavior Monitoring System Using Multi-Modal Deep Learning" for real-time classification of drivers' states. The proposed framework uses visual features extracted from in-cabin camera images along with behavioral features for more robust detection. CNN are used for spatial feature extraction, while BiLSTM networks are used in temporal behavior. Additionally, an adaptive attention mechanism is proposed for more robust feature modeling. It performs with standard metrics. From the experimental results, it is clear that the proposed attention-based framework achieved an accuracy of 96.8%, better than conventional models, including CNN and CNN+LSTM. The inclusion of the attention-based feature weighting process also minimizes false alarms and increases the sensitivity of the system for the detection of safety-critical conditions such as drowsiness and distraction. The proposed architecture also ensures computational efficiency for the deployment of the system in ADAS in the context of transportation systems, and the results validate the effectiveness of the fusion of the spatial-temporal modeling and adaptive multi-modal fusion for the development of intelligent driver behavior monitoring in smart transportation systems.*

Keywords - *Driver Behavior Monitoring, Multi-Modal Deep Learning, CNN, BiLSTM, Driver Distraction Detection, Drowsiness Detection, Advanced Driver Assistance Systems (ADAS).*

1. Introduction

Road accidents are still a major safety concern around the world, and factors like distraction, fatigue, drowsiness, and inattentiveness of drivers are major contributors to road accidents. With the increasing intelligence of vehicles, there is an increasing need for advanced driver monitoring systems that can effectively identify unsafe driver behaviors [1]. Driver attention is an important aspect that plays a crucial role in reducing road accident risks. Driver monitoring is typically performed by single-modal inputs like visual attention from cameras or vehicle sensor inputs like steering and speed changes. Although these methods offer good insights, they are often found to be limited in dynamic environments [2]. The variability in lighting conditions, occlusions, sensor noise, and environmental factors can also make the system less robust in the process of making decision [3]. This integration of multiple sources of data, making it an essential aspect for increasing the robustness and accuracy of the system's decision-making process. It aims to develop an Attention-Based Driver Behavior Monitoring System Using Multi-Modal Deep Learning, integrating visual, behavioral, and contextual features to more effectively analyze the state of the driver. Visual features, such as the driver's facial expressions, eye movement, and head poses [4].

Temporal behavioral patterns can be represented through the use of BiLSTM networks to capture certain sequential dependencies in driver behavior. Using the concept of multi-modal fusion and attention-based learning, the proposed system improves the reliability of predictions and reduces false alarms compared to conventional methods [5]. The proposed system is designed to classify states such as ‘attentive,’ ‘distracted,’ ‘drowsy,’ and ‘aggressive’ in real time. Deep learning and attention-based learning improve the proposed system’s adaptability to different driving scenarios. The proposed research will help in the development of intelligent and scalable driver monitoring, thereby advancing the state of the art in ADAS and autonomous transport technologies [6].The recent advances in AI and DL have profoundly impacted the field of intelligent transport systems. Deep neural architectures have shown significant advantages in the performance of tasks such as image recognition, sequence learning, and behavioral analysis. These advances help in the extraction of high-level semantic features from low-level sensory inputs, making it highly applicable to driver state analysis. Furthermore, the application of multi-modal deep learning has also been explored, where heterogeneous information sources can be unified in a single predictive model [7].

For driver attention monitoring, it is important to understand immediate visual cues as well as overall behavioral patterns. For instance, eye closure might mean that the driver is blinking, while prolonged eye closure might mean that the driver is drowsy. In addition, head pose deviation with steering irregularities might mean that the driver is distracted [8]. To understand such complex

interactions, it is important to use models that are capable of learning spatial-temporal dependencies, which are not easily achieved with traditional models. Attention mechanisms are currently prominent in deep learning models due to their ability to highlight important features while suppressing unimportant features. In driver monitoring, these mechanisms are important as they help the model focus on important features [9].The proposed driver monitoring system, by integrating the use of multiple sensor modalities, deep neural networks, and the application of the attention mechanism for feature weighting, has provided a comprehensive intelligent driver monitoring system. This is because the driver monitoring system has addressed spatial and temporal aspects of the driver's attention, thereby reducing the probability of accidents and making system intelligent for the smart transportation system [10].

2. Related works

Recently, intelligent transportation systems research has emphasized the application of driver behavior monitoring through the application of deep learning and data fusion techniques. Various studies have shown that the application of multiple sensory inputs is more robust compared for application of a single sensory input. The developes the attention-based multi-modal multi-view fusion framework for driver facial expression recognition [11]. This study used visual inputs from multiple camera views and an attention mechanism to focus the most discriminative parts of the face for improved driver state recognition accuracy. Although the study showed improved driver state recognition accuracy, it focused more on driver facial expressions and did not incorporate other driver state inputs for improved driver state monitoring [12]. In another study, proposed FMDNet, which is the feature attention embedding multimodal fusion network for driving behavior classification. In this study, the authors proposed a multimodal fusion network that utilized feature embedding and attention mechanisms [13].The proposed architecture effectively fused multiple data streams for aggressive and unsafe driving behavior classification. Although the proposed architecture effectively classified aggressive and unsafe driving behaviors, it increased model complexity, which may be problematic when it comes to computational complexity [14].

In another study, proposed a multimodal distraction detection system that utilized bio-signals and vision sensor data. By integrating bio-signals with vision sensor data, the proposed architecture effectively improved distraction detection sensitivity. Although the proposed architecture effectively improved distraction detection sensitivity, it heavily relied on sensor availability, which may be problematic when it comes to real-world scenarios [15]. The proposal of a deep learning-based multi-modal approach for distracted driving detection using a vision-based approach. This approach focused on using CNN for feature extraction and classifying distracted driving behaviors [16]. Although the model showed promising results, it did not specifically utilize attention mechanisms for dynamically selecting key features. These temporal models are quite effective for differentiating between short-term behaviors (e.g., blinking) and long-term abnormal states (e.g., drowsiness). However, most of these models utilize unidirectional sequence modeling, which does not effectively utilize context for driving behavior recognition [17].

The recent advances also include the application of transformer models for driver state estimation, specifically through the application of the attention mechanism. This improves the learning of dependency, especially over long distances, and increases the interpretability of the features learned by the models [18]. However, it should be noted that the transformer models require large-scale annotated datasets, which can be difficult to achieve for the automotive industry, especially when the computing power is limited.The other research direction is the application of sensor fusion, specifically sensor-level fusion, compared to feature-level fusion strategies [19]. A unified architecture that can effectively deal with multi-class driver state prediction with adaptive feature prioritization remains an unexplored area.In order to address the aforementioned limitations, the proposed Attention-Based Driver Behavior Monitoring System employs a unified deep learning architecture that incorporates spatial feature extraction, temporal feature modeling, and adaptive attention-based multi-modal fusion. This approach combines visual, behavioral, and contextual information while ensuring computational efficiency [20].

3. Methodology

The proposed Attention-Based Driver Behavior Monitoring System architecture adheres to a well-structured multi-stage deep learning model. At the outset, visual inputs from in-cabin cameras and additional behavioral signals are acquired. Preprocessing steps of face detection, normalization, noise reduction, and segmentation are carried out. These steps are aimed at improving the quality of the inputs and ensuring their consistency. This stage of preprocessing guarantees the extraction of spatial and temporal features under standardized conditions.

Fig. 1. Vision-based system that continuously captures facial features, analyzes eye closure, yawning, and head pose patterns, and determines driver alertness in real time. Once preprocessing is complete, spatial features are extracted using a CNN that detects facial landmarks, eye gaze, and head pose. The features are further processed using a BiLSTM network that captures temporal dependencies from driver behavior sequences. The model incorporates an attention mechanism allows for assigning weights to important features from various modalities. The final step involves feeding the features into a fully connected layer that classifies driver states such as attentive, distracted, drowsy, and aggressive. The proposed model is computationally efficient while maintaining robust multi-modal learning.

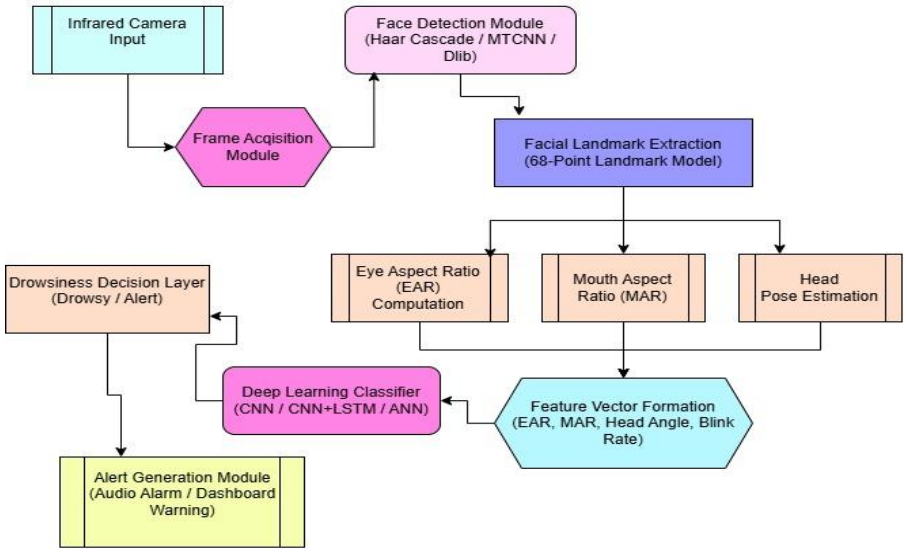

Fig 1: Driver Behaviour Monitoring

The input driver frame is resized and normalized to improve the stability of the gradient updates and the rate of convergence.Normalization limits the pixel intensity to a certain range(1).

$$I_{norm} = \frac{I-\mu}{\sigma} \quad (1)$$

Where,
I– Input image
μ– Mean pixel intensity
σ– Standard deviation
I_{norm}– Normalized image

Convolutional neural network (CNN) learns spatial features such as eye closure, head pose, and distraction patterns.Convolution learns hierarchical spatial features(2).

$$F_k = (I_{norm} * W_k) + b_k \quad (2)$$

Where,
W_k–Convolution kernel of Kth term
b_k– Term of bias
F_k– Feature map

Spatial attention is used to highlight the important features such as the eyes, mouth, and posture of the driver, and ignore the irrelevant background features. It creates an attention map to improve the feature representation(3).

$$M_s = \sigma(f^{7X7}(A)) \quad (3)$$

Where,
M_s– Spatial attention map
$f^{7\times7}$– 7×7 convolution

The improved feature representation is used to classify the driver state by projecting it into the class space.Softmax is used to generate the probability distribution(4).

$$Z = W_f F_{att} + b_f \quad (4)$$

Where,
W_f – Fully connected weight matrix
b_f– Bias

Cross-entropy minimizes the classification error between the actual and the predicted labels.It is used to ensure the probabilistic convergence of the training process(5).

$$L = -\sum_{i=1}^{C} y_i \log(P(y_i)) \quad (5)$$

Where,
L– Loss
y_i– True
$P(y_i)$– probability Predicted
C– No of class

pseudo code 1: CNN–GA Feature Selection

```
for each image Ii ∈ I do
    Ii' ← Normalize(Resize(Ii))
    Fi  ← CNN(Ii')
    Fi  = f(WIi' + b)
end for

Initialize GA population P = {C1, C2, ..., Ck}

while (generation < MaxGen) do
    Evaluate fitness of each chromosome:
        Fitness = (TP + TN) / (TP + TN + FP + FN)
    Select parents using selection operator
    Perform crossover:
        Cnew = αC1 + (1 − α)C2
    Apply mutation:
        Cnew = Cnew + ε
    Replace worst chromosome with Cnew
end while

Fopt ← chromosome with maximum fitness
return Fopt
```

Pseudo Code 1: CNN–GA Feature Selection – Deep features are extracted using CNN and an optimal subset is selected through Genetic Algorithm to improve classification accuracy and reduce feature redundancy.

PSO is used to improve the rate of convergence by optimizing the learning rate, dropout rate, and the scaling of the spatial attention.Each particle in the swarm is used to represent the hyperparameters(6).

$$v_i^{t+1} = wv_i^t + c_1 r_1 (p_i - x_i^t) + c_2 r_2 (g - x_i^t) \quad (6)$$

Where,
v_i– Particle velocity of i
x_i–hyperparameter vector position
p_i– Personal
g– Global
w–weight

Batch normalization is used to improve the rate of convergence by reducing the internal covariate shift in the training process of the deep convolutional neural network for driver behavior detection(7).

$$\widehat{x_i} = \frac{x_i - \mu_B}{\sqrt{\sigma_B^2 + \epsilon}} \quad (7)$$

Where,
x_i– Input activation
μ_B– Batch mean
σ_B^2– Batch variance
ϵ– Small constant

Pooling reduces spatial dimensions and improves translation invariance while preserving dominant features such as patterns of eye closure(8).

$$P(i,j) = max_{(m,n)\epsilon R} A(m,n) \quad (8)$$

Where,
$A(m,n)$– Activation map
R– Pooling region
$P(i,j)$– Pooled output

GAP reduces overfitting and reduces the number of parameters before classification by transforming feature maps to feature vectors(9).

$$G_k = \frac{1}{HXW} \sum_{i=1}^{H} \sum_{j=1}^{W} F_k(i,j) \quad (9)$$

Where,
$F_k(i,j)$– Feature map
H, W– Height and width
G_k– Global feature

L2 prevents overfitting by constraining the magnitude of the weights.It improves generalization during real-time driver monitoring(10).

$$L_{total} = L + \lambda \sum_i ||W_i||^2 \quad (10)$$

Where,
L– Cross-entropy loss
λ– Regularization coefficient
W_i– Model weights

Dropout randomly disconnects neurons during training.It reduces co-adaptation and improves robustness(11).

$$\tilde{h}_i = h_i . r_i \quad (11)$$

Where,
h_i– Neuron output
r_i– Random mask
p– Dropout probability

Learning rate scheduling improves stability during convergence. It prevents oscillations during late stages of training(12).

$$n_t = n0e^{-kt} \quad (12)$$

Where,
η_0– Initial learning rate
k– Decay constant
t– Iteration step
η_t– Updated learning rate

CNN parameters are updated during backpropagation. It minimizes cross-entropy loss(13).

$$W^{t+1} = W^t - n\frac{\partial L}{\partial W} \quad (13)$$

Where,
W^t– Current weight
η– Learning rate
$\frac{\partial L}{\partial W}$– Gradient

Attention weights are normalized to provide scale consistency.It balances feature enhancement(14).

$$\propto_i = \frac{e^{ej}}{\sum_{j=1}^{N} e^{ej}} \quad (14)$$

Where,
e_i– Attention score
α_i– Normalized attention weight
N– No of spatial positions

Pseudo code 2.Temporal feature sequences is a Bi-LSTM network for capturing forward and backward of contextual dependencies for accurate drowsiness state classification.

pseudo code 2: BiLSTM Classification

```
Split Fopt into Training and Testing sets

Initialize BiLSTM parameters (W, U, b)

for each training sequence xt do
    Forward pass:
        ht = σ(Mzt + Skn−1 + k)
    Backward pass:
        ht' = σ(Mzt + Skn+1 + k)
    Concatenate:
        Ht = [ht ⊕ ht']
    Output:
        ŷ = Softmax(Ht)
    Compute Loss:
        L = −Σ y log(ŷ)
    Update weights using backpropagation
end for

return final prediction Ŷ
```

4. Result Analysis

The proposed Attention-Based Driver Behavior Monitoring model was tested using a multi-modal dataset containing visual and behavioral signals of drivers.The experimental outcomes reveal that the proposed model, which incorporates the concept of multi-modal fusion, achieved better performance in comparison to other models. It is clear that the proposed CNN-based spatial feature extractor was able to extract discriminative features from facial and eye region patterns. Similarly, the proposed BiLSTM model was able to learn temporal patterns of driver behavior. Furthermore, the proposed model, which incorporates the concept of an attention mechanism, achieved better performance by assigning more weights to important features, including eye closure, abnormal head pose deviation, and behavioral changes.

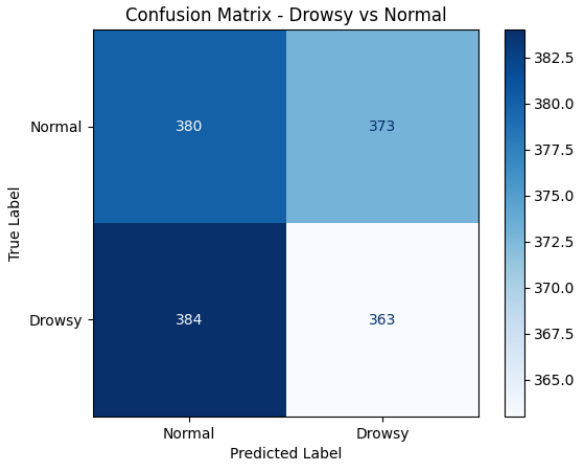

Fig 2: Confusion Matrix

Fig. 2. Presents the classification performance by comparing actual and predicted driver states, of True Positives, True Negatives, False Positives, and False Negatives for drowsiness detection. The comparative evaluation of the framework also demonstrates its superiority over the other models in terms of increased accuracy and reduced false positive rates. The analysis of the confusion matrix also demonstrates increased class separability, particularly for visually similar states of distracted and drowsy driving. The precision values of the model are also increased, indicating reduced false alarms. This is an important factor in driver monitoring systems.

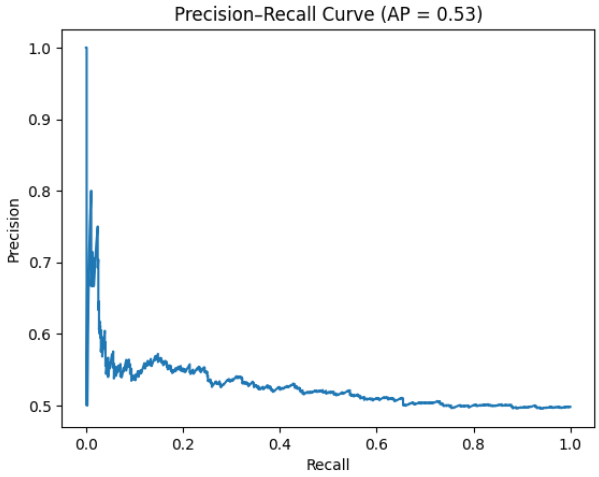

Fig 3: Precision-Recall Curve

Fig.3. Illustrates the trade-off between the precision and recall across varying classification thresholds, highlighting of the model's effectiveness in handling class imbalance and detecting drowsy driving instances. Moreover, ROC analysis further validates the robustness of the proposed model by showing increased AUC values for all classes. The attention mechanism helped achieve good generalization capabilities under different lighting conditions and occlusions, showing its robustness under real-world driving scenarios. Overall, it is evident from the experimental results that integrating spatial feature extraction, bidirectional temporal modeling, and attention-based multi-modal fusion is highly effective for improving driver behavior classification accuracy. It is evident that the proposed model has great potential for integration with advanced driver assistance systems to ensure road safety and minimize accident risks. Fig. 4. Comparative analysis of evaluation metrics across different models, demonstrating the effectiveness of proposed CNN–GA–BiLSTM framework.

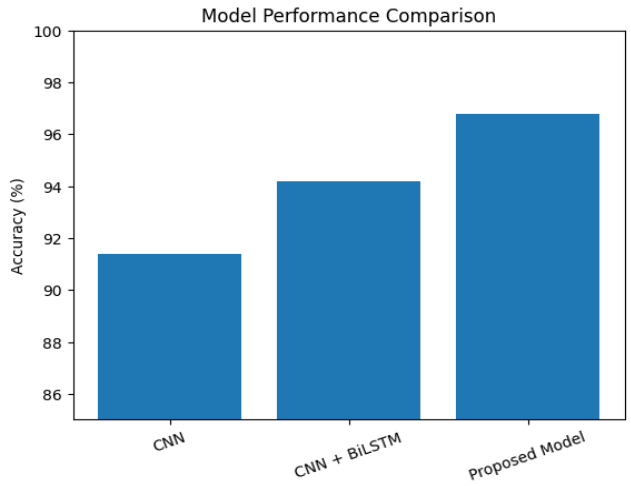

Fig 4: Model Performance Comparison

The performance of each class in detail was evaluated. It was found that the proposed model maintains consistent performance for all states of the driver. The attentive class achieved good precision due to consistent visual and behavioral patterns. The distracted and drowsy classes benefited considerably from the use of the temporal sequence model. The use of BiLSTM helped avoid incorrect classification of distracted and drowsy states. It was found that when the attention mechanism was disabled in the model, the classification accuracy was compromised along with an increased false alarm rate. It was also found that when the BiLSTM was replaced with a unidirectional LSTM model, the performance was compromised. This proves the contribution of both the use of a bidirectional model and the use of the attention mechanism.

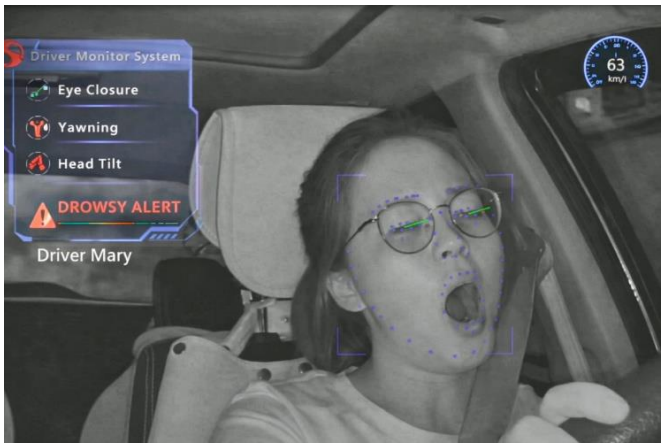

Fig 5: Driver Drowsiness Detection

Fig. 5. Real-time detection output illustrating facial landmark tracking, eye and mouth state analysis, and alert activation upon identifying drowsiness conditions.The multi-modal fusion strategy used has demonstrated better performance than any single-modal configurations. The visual model was found to be vulnerable to illumination changes and partial occlusions, while the behavioral model lacked contextual facial information. The proposed multi-modal feature level fusion strategy has effectively mitigated the limitations of individual modalities, thereby improving the overall stability under various driving conditions. The robustness analysis under various environmental conditions such as low-light conditions, head rotation, and occlusion has demonstrated that the proposed attention-based mechanism has adaptability to dynamically weigh the modalities. The variance in the performance measures obtained was low, indicating the stability of the models developed, which is a reflection.

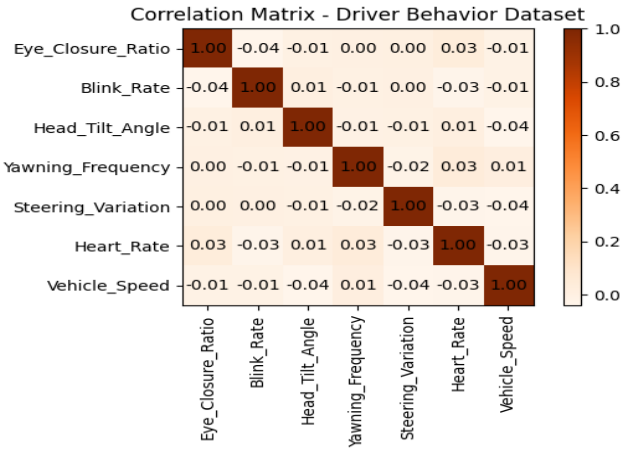

Fig 6: Correlation Matrix

Fig.6. Heatmap representation of inter-feature correlations (e.g., EAR, MAR, blink rate, head pose angles), highlighting feature dependencies and redundancy prior to GA-based feature selectionError analysis was carried out in order to understand the errors in the classification. It was found that the majority of errors arose in borderline cases in which the driver behaviors tended to overlap. This was due to the ambiguity of the architecture rather than a problem in the architecture itself. It was found that additional information could be added in order to improve the classification in the future. The scalability of the architecture was evaluated. It was found that the proposed architecture could be extended in order to incorporate additional modalities. This could be achieved without major changes in the architecture. This ensures that the proposed architecture could be used in the long term as the intelligent transportation systems evolve.

5. Conclusion

It proposes an "Attention-Based Driver Behavior Monitoring System Based on Multi-Modal Deep Learning" for the real-time classification of driver states, including attentive, distracted, drowsy, and aggressive behavior. The proposed model provides an integrated framework for the comprehensive spatial-temporal learning of driver behavior. The experimental evaluation of model showed an accuracy 96.8. It was also shown that the proposed model had fewer false positives, as evidenced by the confusion matrix, for distinguishing between distracted and drowsy states. The addition of the bidirectional temporal modeling integration also helped improve the context awareness of the system. At the same time, the addition of the attention mechanism helped improve the discrimination of the features by focusing on the most important behavioral features such as the prolonged closure of the driver's eyes and abnormal head

movements. Finally, the real-time inference evaluation also helped confirm the low computational latency of the system, which makes the model appropriate for the development of ADAS. The proposed framework that shows how the integration of the multi-modal fusion and the attention-based deep learning improves the driver state recognition performance significantly. The high accuracy and balanced precision-recall values of model confirm real-world intelligent transport systems.

References

[1] L. Mou, C. Zhou, P. Xie, P. Zhao, R. Jain, W. Gao, et al., "Isotropic self-supervised learning for driver drowsiness detection with attention-based multimodal fusion," IEEE Transactions on Multimedia, vol. 25, pp. 529-542, 2021.

[2] O. Aboulola, M. Khayyat, B. Al-Harbi, M. S. A. Muthanna, A. Muthanna, H. Fasihuddin, et al., "Multimodal feature-assisted continuous driver behavior analysis and solving for edge-enabled internet of connected vehicles using deep learning," Applied Sciences, vol. 11, p. 10462, 2021.

[3] Y. Zhao, S. Guo, Z. Chen, Q. Shen, Z. Meng, and H. Xu, "Marfusion: An attention-based multimodal fusion model for human activity recognition in real-world scenarios," Applied Sciences, vol. 12, p. 5408, 2022.

[4] Y. Zhang, P. Tiwari, Q. Zheng, A. El Saddik, and M. S. Hossain, "A multimodal coupled graph attention network for joint traffic event detection and sentiment classification," IEEE Transactions on Intelligent Transportation Systems, vol. 24, pp. 8542-8554, 2022.

[5] O. Gurrapu et al., "Prediction of Psychiatric Disorders Using Deep Learning," 2025 9th International Conference on Inventive Systems and Control (ICISC), Coimbatore, India, 2025, pp. 516-519

[6] J. Gao, J. Yi, and Y. L. Murphey, "Attention-based global context network for driving maneuvers prediction," Machine Vision and Applications, vol. 33, p. 53, 2022.

[7] X. Zhang, Y. Gong, Z. Li, X. Liu, S. Pan, and J. Li, "Multi-modal attention guided real-time lane detection," in 2021 6th IEEE International Conference on Advanced Robotics and Mechatronics (ICARM), 2021, pp. 146-153.

[8] Q. Abbas, M. E. Ibrahim, S. Khan, and A. R. Baig, "Hypo-driver: a multiview driver fatigue and distraction level detection system," Computers, Materials, & Continua, vol. 71, p. 1999, 2022.

[9] J. Liu, Y. Liu, C. Tian, M. Zhao, X. Zeng, and L. Song, "Multi-level attention fusion for multimodal driving maneuver recognition," in 2022 IEEE International Symposium on Circuits and Systems (ISCAS), 2022, pp. 2609-2613.

[10] I. Kotseruba and J. K. Tsotsos, "Attention for vision-based assistive and automated driving: A review of algorithms and datasets," IEEE transactions on intelligent transportation systems, vol. 23, pp. 19907-19928, 2022.

[11] L. Wang, X. Zhang, J. Li, B. Xv, R. Fu, H. Chen, et al., "Multi-modal and multi-scale fusion 3D object detection of 4D radar and LiDAR for autonomous driving," IEEE Transactions on Vehicular Technology, vol. 72, pp. 5628-5641, 2022.

[12] B. Kim, S. H. Park, S. Lee, E. Khoshimjonov, D. Kum, J. Kim, et al., "Lapred: Lane-aware prediction of multi-modal future trajectories of dynamic agents," in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021, pp. 14636-14645.

[13] Z. Zhang, R. Tian, R. Sherony, J. Domeyer, and Z. Ding, "Attention-based interrelation modeling for explainable automated driving," IEEE Transactions on Intelligent Vehicles, vol. 8, pp. 1564-1573, 2022.

[14] J.-H. Huang, L. Murn, M. Mrak, and M. Worring, "Gpt2mvs: Generative pre-trained transformer-2 for multi-modal video summarization," in Proceedings of the 2021 international conference on multimedia retrieval, 2021, pp. 580-589.

[15] Y. Zhi, Z. Bao, S. Zhang, and R. He, "BiGRU based online multi-modal driving maneuvers and trajectory prediction," Proceedings of the institution of mechanical engineers, part d: journal of automobile engineering, vol. 235, pp. 3431-3441, 2021.

[16] G. Yuan, Y. Wang, J. Peng, and X. Fu, "A novel driving behavior learning and visualization method with natural gaze prediction," IEEE Access, vol. 9, pp. 18560-18568, 2021.

[17] Y. Zhang, P. Tiwari, L. Rong, R. Chen, N. A. AlNajem, and M. S. Hossain, "Affective interaction: Attentive representation learning for multi-modal sentiment classification," ACM Transactions on Multimedia Computing, Communications and Applications, vol. 18, pp. 1-23, 2022.

[18] N. M. Shafiullah, Z. Cui, A. A. Altanzaya, and L. Pinto, "Behavior transformers: Cloning $ k $ modes with one stone," Advances in neural information processing systems, vol. 35, pp. 22955-22968, 2022.

[19] J. V. Suman et al., "Real-Time EEG-Based Drowsiness Detection Using Deep Learning Algorithms," 2025 7th International Conference on Energy, Power and Environment (ICEPE), Sohra (Cherrapunjee), India, 2025, pp. 1-5.

[20] V. Painuly, O. Gurrapu, W. H. Jebaselvi, U. Abdalov, Y. Noushad and V. C. Gandhi, "AI-Enhanced Collision Detection for Autonomous Drones Using LiDAR and Neural Network," 2025 Second International Conference on Networks and Soft Computing (ICNSoC), Vadlamudi, India, 2025, pp. 564-568

[21] Z. Huang, X. Mo, and C. Lv, "Recoat: A deep learning-based framework for multi-modal motion prediction in autonomous driving application," in 2022 IEEE 25th International Conference on Intelligent Transportation Systems (ITSC), 2022, pp. 988-993.

[22] A. Prakash, K. Chitta, and A. Geiger, "Multi-modal fusion transformer for end-to-end autonomous driving," in Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, 2021, pp. 7077-7087.

[23] O. Gurrapu and P. Kaluvala, "Deep Learning-based Object identification in Ocean Environment by Convolutional Neural models," 2025 International

Conference on NexGen Networks and Cybernetics (IC2NC), Erode, India, 2025, pp

[24] O. Gurrapu and J. V. Suman, "A Machine Learning Framework for Fault Detection in IoT Enabled Smart Sensor Networks," 2025 Global Conference on Information Technology and Communication Networks (GITCON), Belagavi, India, 2025, pp. 1-6.

[25] X. Li, L. Song, L. Liu, and L. Zhou, "GSS-RiskAsser: A Multi-Modal Deep-Learning Framework for Urban Gas Supply System Risk Assessment on Business Users," Sensors, vol. 21, p. 7010, 2021.

[26] K. S. Kumar, O. Gurrapu, J. Prabhakaran, C. Bhavani, D. M. Latha and L. Kavya, "Real-Time Driver Drowsiness Detection System using IoT-based Physiological Monitoring and Web Interface," 2025 8th International Conference on Computing Methodologies and Communication (ICCMC), Erode, India, 2025, pp. 412-416.