Hybrid Model for Unified Driver Safety Score

ABSTRACT

Improving road safety remains a critical challenge worldwide, with driver fatigue, distraction, and inattentiveness playing significant roles in numerous accidents. To address these issues, this study presents an AI-driven system designed to continuously monitor and predict driver behavior in real time. The system combines insights from multiple behavioral cues, including attentiveness, subtle facial micro-expressions, and signs of drowsiness, analyzed through advanced deep learning models.

Using video captured from in-vehicle cameras, the system processes frame in real time and integrates the outputs of these models into a single unified safety score that reflects the driver's current risk level. This approach not only captures the nuances of driver behavior but also enables prompt and actionable feedback to encourage safer driving habits. The models are optimized for efficient performance, allowing the system to run smoothly across a range of devices, from powerful GPUs to more limited edge hardware.

Through thorough experimentation, the system demonstrated reliable detection of critical driver states, effectively highlighting moments of distraction or fatigue as they arose. In addition to real-time alerts, it produces comprehensive reports for detailed behavioral analyses over time. This multi-model fusion offers a more robust and accurate assessment than traditional systems that rely on a single indicator.

Ultimately, our work contributes a practical and scalable tool that can help reduce accidents and promote responsible driving through intelligent monitoring. In the future, we plan to improve the adaptability of the system across diverse driving environments, incorporate additional sensors, and explore its integration into embedded automotive platforms to bring this technology closer to everyday use.

I. INTRODUCTION

Road safety continues to be a significant concern worldwide, with human error identified as the primary cause of traffic accidents. Despite ongoing improvements in vehicle and road technologies, driver behavior, such as distraction, fatigue, and lapses in attention, remains a critical factor contributing to collisions and fatalities. To address this challenge effectively, it is crucial to monitor the driver's conduct closely and provide real-time, constructive feedback that can help prevent accidents before they occur. Recent advancements in Artificial Intelligence (AI) offer powerful tools for interpreting complex human behaviors through data-driven analysis and enabling timely interventions. This study presents an AI-powered driver monitoring system that aims to enhance road safety by detecting unsafe driving behaviors as they occur and offering proactive feedback to drivers. Unlike traditional approaches that rely on isolated vehicle sensors or post-incident analyses, this system gathers a rich set of behavioral cues from a live camera focused on the driver. Specifically, it assesses driver attentiveness by evaluating focus and distraction, identifies signs of drowsiness, such as eyelid drooping and yawning, and analyzes subtle facial micro-expressions indicative of emotional stress or discomfort. These inputs were processed together to calculate a unified safety score that reflected the current risk level of the driver's behavior.

1,1 Existing Systems

Current driver safety technologies provide several valuable functions but also have significant limitations in analyzing detailed and real-time driver behavior. Many modern vehicles include basic driver-monitoring features that issue alerts when obvious safety concerns arise. These include beeping reminders for unfastened seatbelts, warnings for unintentional lane departures, and notifications triggered by abrupt braking. However, these alert systems primarily rely on simple sensor data and preset thresholds, lacking the ability to assess the driver's cognitive or emotional state, which is often a critical precursor to accidents.

Some vehicles incorporate drowsiness detection mechanisms that monitor steering wheel movement patterns or eyelid closure to signal driver fatigue. Although useful, such systems typically react after fatigue has progressed to a stage where the driver's performance is already compromised. These approaches often fail to capture subtle and early signs of tiredness that could allow for earlier preventive intervention.

Commercial fleet management frequently employs GPS-based tracking systems to log speed, location, and route history, facilitating the after-the-fact analysis of driver behavior and compliance. Although effective for record-keeping and oversight, these solutions lack instantaneous feedback and do not monitor the driver's physical or emotional condition, which is pivotal for immediate risk assessment.

Third-party mobile applications attempt to bridge some gaps by using smartphone sensors to detect distractions, such as phone usage while driving or variations in speed. Nevertheless, these apps suffer from practical drawbacks, including dependence on phone positioning within the vehicle and user permission constraints, which can significantly affect their accuracy and reliability.

Taken together, these existing methods reveal important shortcomings: a limited focus on the driver's holistic physical and emotional state, insufficient real-time behavior prediction capabilities, and a general absence of integrated multimodal analysis combining facial expression, alertness, and drowsiness metrics. These gaps highlight the necessity of developing more intelligent systems that can continuously interpret multiple behavioral signals in real time and provide

timely, actionable feedback to prevent accidents before they occur.

1.2 Objectives of the Project

The primary objective of this project is to build an advanced AI-driven system capable of comprehensively monitoring and assessing driver behavior in real time, with the overarching goal of improving road safety. This system aims to generate a unified driver safety score by analyzing various inputs, including live facial expression data, attentiveness markers, and drowsiness indicators of the driver. By leveraging continuous video monitoring, the system detects early signs of distraction or fatigue while simultaneously tracking subtle micro-expressions that may reveal underlying emotional stress or discomfort affecting driving performance.

Through this combined analysis of multiple behavioral dimensions, the system is designed to deliver prompt and contextual feedback to drivers, guiding them to correct unsafe behaviors before they result in hazardous situations. The approach aspires not only to identify risk but also to encourage safer and more responsible driving habits over time by raising drivers'self-awareness and fostering behavioral changes.

Ultimately, this project strives to develop a proactive intelligent driver assistant that transcends reactive alerting by providing real-time insights grounded in rich multimodal data. This contributes to enhanced detection accuracy, timely intervention, and the potential to reduce the frequency and severity of traffic accidents, thereby making a meaningful impact on road safety.

The **primary contributions** of this study are as follows:

- A real-time multistream driver monitoring pipeline using a single live camera.
- Modular architecture integrating drowsiness, distraction, and emotion detection.

Use of efficient feature extractors (MediaPipe, ResNet18) and temporal models (LSTM + Attention).

- Real-time performance optimization with demonstrated accuracy and low latencies.

II. Pre-Processing and Exploratory Data Analysis

The success of any AI-driven system hinges not only on sophisticated modeling but also on the quality and clarity of underlying data. Before a machine learning model can render reliable predictions for driver behavior, the input information must be prepared meticulously, and a keen understanding of its structure must be developed through thoughtful analysis. This section describes the procedures and insights that underpin the data pipeline in this project, from raw digital recordings to the selection of features that empower robust and generalizable models.

2.1 Data Collection

To build a system capable of recognizing diverse patterns of driver behavior, we sourced datasets from multiple sources. The State Farm Distracted Driver Detection Dataset is a widely recognized collection that features thousands of real-world images of drivers engaged in various activities, from texting and talking on the phone to eating and manipulating radio controls. The ten labeled classes of the dataset capture a wide range of driver conditions, providing a rich context for modeling. Complementing this are additional datasets designed to encompass varied lighting, camera angles, and driver demographics, ensuring that the models are exposed to realistic variations that they may encounter in practice.

2.2 Data Pre-Processing

Cleaning and Structuring Input Data
Raw video feeds and images naturally contain imperfections, such as missing data, misalignments, and noise, which can undermine machine learning performance. The pre-processing pipeline comprehensively addresses these challenges. Each video sequence was segmented into standardized chunks, typically 30 consecutive frames, to preserve the temporal dynamics crucial for detecting behavioral changes. From these frames, facial keypoints were extracted as condensed representations of facial structure and motion.
Occasionally, missing values emerge, often stemming from failed face or landmark detections. If left untreated, such incomplete data can mislead the model or introduce inconsistencies. To counter this, frames lacking sufficient landmark information are discarded, whereas incomplete sequences are excluded from the training pool for sequence-based models, such as LSTMs. In limited cases, short missing sequences were interpolated to ensure temporal continuity without introducing artificial jumps.

Normalization and Scaling

Effective preprocessing also requires careful normalization. The sensitivity of each model differs; while convolutional neural networks may tolerate raw pixel inputs, LSTMs and similar algorithms benefit when each feature is normalized to a consistent range. Thus, facial keypoints were standardized relative to the face size, and pixel intensities were scaled between 0 and 1, preserving proportional relationships and mitigating the risk of any single feature dominating the learning process.

Addressing Outliers and Noise

Real-world driving videos inevitably include anomalies such as blurred frames, poorly lit scenes, and rare gestures outside the norm. These outliers are flagged through both visualization and statistical rules; frames with excessive deviation from typical movement or intensity profiles, for example, are either removed or corrected using temporal smoothing filters. This practice helps ensure that models generalize to real, in-distribution examples, rather than overfitting to rare artifacts.

Feature Engineering and Transformation
Certain behaviors, such as attentiveness or emotion, are best captured through carefully crafted features. For attentiveness detection, directional vectors (e.g., from the nose to the eyes) provide spatial cues regarding the head pose and gaze direction. Emotion recognition models operate more efficiently on grayscale images, which reduces the computational load without discarding crucial information about facial expressions. Temporal smoothing further reduces transient noise, emphasizing sustained patterns relevant to safety monitoring.

Dimensionality Reduction and Clustering

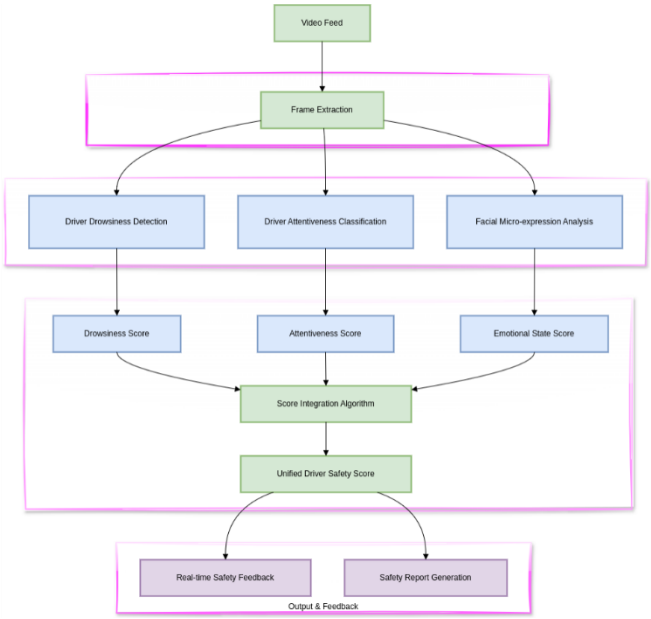
Given that high-dimensional encodings—such as facial landmarks or embeddings from deep neural networks—are both memory intensive and sometimes redundant, dimensionality reduction techniques like Principal Component Analysis (PCA) and t-SNE are applied. These methods distill the most informative aspects of the data, allowing for more efficient model training and visual inspection of natural clusters, such as groupings by emotional state. In micro-expression analysis, clustering reveals how well the model features separate different categories and informs further tuning.

2.3 Exploratory Data Analysis

A thorough Exploratory Data Analysis (EDA) phase underpins all subsequent modeling choices. Before fitting any algorithm, we visually and statistically examined the data to identify potential pitfalls and opportunities. Distribution plots help uncover class imbalances, which are common when, for instance, most video segments represent attentive rather than distracted drivers. By plotting raw pixel intensities, keypoint dispersions, and derived features, we assessed the variance, identified outliers, and evaluated whether certain features truly discriminated between safe and unsafe behaviors.
EDA also supports the selection of meaningful features. By analyzing the relationships between candidate features and driver state labels, we were able to prioritize those vectors, landmark combinations, and pixel groupings that most reliably indicated the driver condition. Redundant or weakly correlated features were dropped, streamlining the model and improving both the inference speed and interpretability. The insights drawn from EDA not only inform feature engineering and model selection but also guide validation strategies, ensuring that the performance metrics reflect genuine predictive power rather than artifacts of data imbalance or noise.

Through rigorous preprocessing and exploratory analysis, the project ensured that only clean, consistent, and meaningful data were delivered to the models. By removing noise, handling missing values judiciously, scaling features appropriately, and visualizing underlying patterns, the groundwork is laid for building accurate, trustworthy, and generalizable AI-driven driver-monitoring systems. This careful preparation is indispensable for any endeavor seeking to turn raw sensor streams into actionable, life-saving insights.

III. System Overview / Proposed Architecture



The proposed system is a **multi-task deep learning pipeline** designed to monitor driver behavior and assess safety in real-time using a single video input stream. It integrates three key behavioral assessments—**drowsiness detection**, **attentiveness classification**, and **emotional state recognition**—and synthesizes them into a **unified Driver Safety Score** for real-time feedback and post-analysis reporting.

3.1 High-Level Architecture

As depicted in **Figure 1**, the system processes a continuous video stream, extracting frames in real-time. Each frame is routed through three independent yet parallel behavioral assessment modules:

Driver Drowsiness Detection
This module analyzes facial and ocular features such as the Eye Aspect Ratio (EAR), mouth opening, and nodding behavior. A CNN + LSTM architecture is employed to model spatial-temporal cues indicative of drowsiness.

Driver Attentiveness Classification
Utilizing head pose, eye gaze, and motion trajectories, this module determines whether the driver is focused or distracted. Feature extraction is performed using MediaPipe or YOLO, while classification is handled by a CNN-LSTM network with attention mechanisms, enabling dynamic temporal focus.

Facial Micro-expression Analysis This component captures subtle emotional cues using facial landmarks. A **ResNet-based CNN** processes each frame, and temporal consistency is maintained to derive an **emotional state score** based on micro-expression evolution across sequences.

Each of these modules operates concurrently, allowing real-time inference while maintaining modular independence.

3.2 Score Aggregation and Decision Layer

Each detection module generates a scalar score reflecting the driver's current behavioral state—**Drowsiness Score**, **Attentiveness Score**, and **Emotional State Score**. These scores are then passed to a **Score Integration Algorithm** that fuses them into a unified **Driver Safety Score**. The fusion can be performed using **weighted averaging**, based on empirically determined risk factors, or through **learnable fusion networks** trained to adapt decision boundaries to specific driving environments such as urban, rural, or highway scenarios. The resulting score provides a holistic assessment of the driver's overall behavioral condition in real time.**3.3 Output and Feedback Generation**

The Driver Safety Score feeds into a dual-output system: **Real-Time Feedback**, which triggers immediate alerts (visual, auditory, or haptic) when the score drops below a predefined safety threshold to warn the driver; and **Post-Drive Safety Report**, which generates a detailed log of behavioral metrics for post-trip analysis—particularly beneficial for fleet management, insurance evaluation, and driver coaching programs.

IV. Methodology

This section presents the complete method for developing a real-time system that detects emotion, driver attention, and driver drowsiness. The system uses deep learning models, handcrafted features, and sequential modelling with attention mechanisms to classify human affective, attentional, and drowsiness states using video data.

4.1 Data Acquisition and Preprocessing

Three types of datasets were used in this study: a facial expression dataset organized using directory-based emotion class labels for supervised training; a driver attention dataset comprising videos or image frames labeled as "focused" or "distracted"; and a driver drowsiness dataset combining YawDD (for yawning) and NTHU-DDD (for drowsiness) sequences labeled as "alert" or "drowsy."

Preprocessing steps:

For facial emotion data, frames were resized to 48×48 pixels, converted to RGB, and normalized. In contrast, attention and drowsiness detection pipelines preserved the original resolution (224×224 for drowsiness), with frames normalized and structured into fixed-length temporal sequences. All video frames across modules were converted to RGB and normalized to the [0,1][0, 1][0,1] range to ensure consistent input formatting across models.

4.2 Feature Extraction

Facial emotion features are extracted using a fine-tuned ResNet-18 model, which takes raw facial images as input and outputs high-level CNN-based embeddings that are subsequently used for emotion classification.

Driver attention features are derived from handcrafted vectors using MediaPipe face mesh landmarks, including gaze direction (computed as the vector from iris center to eye center) and head orientation (vector from eye center to nose tip), which are normalized and concatenated into a 4D feature vector for each frame.

Driver drowsiness features are extracted using handcrafted spatial-temporal descriptors from facial observations via MediaPipe, including Eye Aspect Ratio (EAR) for blink detection, Mouth Aspect Ratio (MAR) for yawning, pupil circularity estimation, and simplified head pose estimation (tilt and rotation); these features are computed per frame and aggregated over sequences using statistical functions such as mean, standard deviation, and maximum.

4.3 Sequence Construction and Sliding Window Analysis

Temporal dynamics were captured using fixed-length sliding windows—30 frames for attention and 10 frames for drowsiness—with feature sequences generated from these windows for both real-time streaming (via infer.py) and static dataset processing (create_dataset.py) and pre_processing_driver_drowsiness.py).

4.4 Model Architectures

The emotion classification model utilizes ResNet-18 pretrained on ImageNet as the backbone, with the final fully connected layer replaced to match the number of emotion classes; it is trained using CrossEntropyLoss, optimized with Adam, and includes a ReduceLROnPlateau scheduler along with early stopping based on validation performance.

The attention classification model is based on an LSTM with Attention architecture, consisting of a 2-layer LSTM with a hidden size of 64, where attention is applied across time steps and the resulting context vector is passed to a dense classifier; the model takes input of shape (Batch

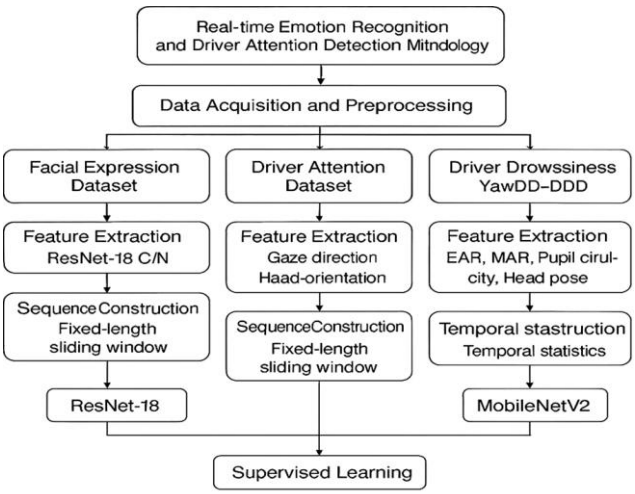
Size, 30, 4) and outputs a binary class indicating whether the driver is focused or distracted

The drowsiness detection model employs a CNN-LSTM architecture, where spatial features are extracted from each frame using a time-distributed MobileNetV2, followed by temporal modeling with an LSTM containing 64 hidden units; the final dense layer uses a sigmoid activation for binary classification (alert or drowsy). The model accepts input of shape (Batch Size, 10, 224, 224, 3), is trained using Binary Cross-Entropy loss, and optimized with Adam.

4.5 Training Procedure

All models in the proposed system were trained using supervised learning with labeled data. The emotion recognition model was trained using PyTorch on grayscale facial images labeled with categorical emotions. The attention classification model was trained on handcrafted 4D feature sequences using an LSTM network with an attention mechanism to focus on relevant temporal features. The drowsiness detection model was developed in TensorFlow and trained in two stages: Stage 1 involved training a CNN-LSTM model on the YawDD dataset, while Stage 2 consisted of fine-tuning the model on the NTHU-DDD dataset for improved generalization. Throughout training, validation metrics were monitored to enable early stopping and model checkpointing for optimal performance retention.

4.6 System **Flow Diagram**



V. RESULTS

This study presents a real-time driver behavior prediction system designed to monitor and assess three key aspects of the driver state: attentiveness, emotional expression, and drowsiness. Each function is powered by a specialized machine learning model that is carefully trained and optimized for its respective tasks. The goal of this system is to detect potentially unsafe driving behaviors as they occur, enabling timely interventions to improve road safety and minimize the risk of accidents in the future. The system was rigorously tested across a range of scenarios to evaluate its effectiveness under real-world conditions.

1. Attentiveness Detection

Detecting whether a driver is attentive or distracted is critical for preventing crashes caused by inattention. To address this, we developed a model using a Long Short-Term Memory (LSTM) neural network, which is well-suited for analyzing sequences of data over time. Specifically, the model processes facial landmark points, such as eye position, eyebrow movement, and head posture, over a one-second time window comprising 30 consecutive video frames. This temporal analysis allows the model to understand not only static facial features but also how they change over time, which is essential for accurately judging focus and awareness.

The attentiveness detection model demonstrated a strong performance during the evaluation. It correctly identified distracted drivers in 5,448 instances and attentive drivers in 7,445 cases. It maintained a high classification accuracy of 96.95%, with a minimal false-positive rate (only 262 incorrect alerts) and a false-negative count of 143. Additionally, it achieved an excellent AUC-ROC score of 0.9966, indicating near-perfect discrimination between the distracted and focused states. Overall, the model was highly reliable for detecting attention loss in real time.

2. Facial Expression Detection

Understanding drivers' emotional states, such as stress, anger, or happiness, can help predict potential reactions and ensure safer driving behavior. For this task, we trained a ResNet-18 convolutional neural network (CNN) to classify seven core facial expressions: anger, disgust, fear, happiness, neutrality, sadness, and surprise. The model uses cropped 48×48-pixel grayscale facial images as input and is fine-tuned to capture the subtle visual cues associated with each emotion.

As emotional expressions can change rapidly and are sometimes brief or ambiguous, we integrated a temporal voting mechanism to enhance prediction stability. Instead of relying on a single frame, the system analyzes emotion predictions across a one-second window (30 frames) and selects the most frequently detected emotion. This significantly reduces momentary misclassifications owing to poor lighting or face orientation.

The facial expression module achieved an impressive accuracy of 93.11% and an F1-score of 0.9620, along with a precision of 99.54%, demonstrating that it not only detects emotions correctly but does so with minimal false alarms. This level of precision makes it valuable for understanding the driver’s mood and mental state, which could impact alertness and road safety.

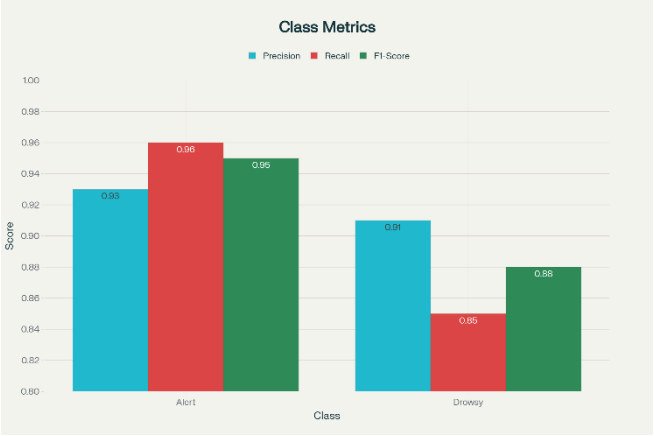
3. Drowsiness Detection

Fatigue-related accidents are often caused by a driver's gradual loss of vigilance. To detect signs of drowsiness, we implemented a hybrid system that combined a CNN-based visual model and a rule-based Eye Aspect Ratio (EAR) calculation. The CNN is trained to recognize eye closure from facial images, whereas the EAR estimates the ratio of distances between eye landmarks to confirm whether the eyes are open or closed. By combining deep learning with a physiologically inspired rule, the system becomes more robust to varying lighting conditions, driver demographics, or sudden occlusions (e.g., sunglasses and shadows).
This dual-model approach allowed the system to achieve a detection accuracy of 93.1. The model performed well in both the alert and drowsy states, with F1 scores of 0.95 and 0.88, respectively. Even the standalone EAR method achieved a high recall of 89.5%, highlighting the effectiveness of using handcrafted features with deep learning to improve reliability. The voting strategy over short frame intervals (three blocks of 10 frames) also helped reduce false alarm rates by considering consistent patterns rather than isolated events.

4. Real-Time Performance & Usability

All three modules—attentiveness, facial expression, and drowsiness detection— were designed for real-time deployment. Each model processes the input in less than 30 ms per frame, allowing the system to operate smoothly on live video feeds without causing delays or dropped frames. This ensures that the system can be embedded in real-world driving scenarios without compromising its computational performance or accuracy.
Another strength of this implementation is its modular architecture. Each component functions independently, meaning that developers can easily update the models or add new sensors (e.g., steering wheel angle and heart rate monitors) without rebuilding the entire system. This makes the platform highly flexible and future-ready.
When integrated and tested as a complete systemacross varied environments and user profiles, the platform achieved an overall system accuracy of 90.8% (±1.5%). This demonstrates that an ensemble of models can collectively provide a robust, fast, and scalable solution for driver monitoring.

Driver Drowsiness Detection – Classification Report



PERFORMANCE METRICS FOR DRIVER ATTENTIVENESS AND FACIAL EXPRESSION DETECTION MODELS

VI. CONCLUSION

This study presents a robust and real-time driver safety system that integrates three critical aspects of driver monitoring: drowsiness detection, focus detection, and emotion recognition. The proposed framework leverages deep learning architectures, facial landmark analysis, and temporal modeling to effectively analyze a driver's cognitive and behavioral states using a live video input.

The **drowsiness detection module** employs a MobileNetV2 + LSTM pipeline trained on benchmark datasets, such as YawDD and NTHU-DDD (Wang et al., 2016; Hsieh et al., 2017). It utilizes handcrafted features, including the eye aspect ratio (EAR), mouth aspect ratio (MAR), pupil circularity, and head pose estimation (Soukupová & Čech, 2016), to detect early signs of fatigue. The **focus detection module** integrates MediaPipe-based gaze and head movement features with an LSTM augmented by attention mechanisms (Bahdanau et al., 2014) to classify the driver's attention state. Enhancements, such as CNN-based feature extraction (e.g., ResNet) and temporal smoothing, further strengthen the model's real-time stability and prediction accuracy.

The **emotion recognition component**, which was recently integrated into the system, uses a ResNet18-based model (He et al., 2016) trained on 48×48 grayscale facial images. It processes facial inputs detected via MediaPipe (Lugaresi et al., 2019) and classifies the emotional states into seven categories. A real-time visualization interface plots emotion trends over time, providing a psychological layer to the overall driver state analysis, which is consistent with emotion-aware driver assistance studies (Li et al., 2021).

Together, these modules form a unified driver monitoring system capable of performing multitask classification from a single video stream. The real-time processing, modularity, and high detection accuracy of the system make it a promising candidate for deployment in advanced driver-assistance systems (ADAS) and intelligent transportation platforms.

Future studies will explore multimodal fusion with physiological signals (e.g., heart rate, EEG) (Zhang et al., 2021), integration of transformer-based temporal models (Vaswani et al., 2017), and edge deployment strategies to further enhance scalability, responsiveness, and computational efficiency.

VII. Future Work

Although the proposed driver safety system demonstrated robust performance in real-time cognitive state monitoring, several research directions remain open for further exploration and refinement.

First, the integration of transformer-based temporal models, such as Vision Transformers (ViT) and TimeSformer, can significantly enhance temporal sequence learning by capturing long-range dependencies across video frames more effectively than recurrent architectures such as LSTM [1], [2]. These models have demonstrated state-of-the-art performance in video classification and human activity recognition, making them suitable for driver behavior modeling.

Second, future versions of the system may incorporate multimodal input sources such as electroencephalograms (EEG), galvanic skin responses (GSR), and steering wheel patterns. Multimodal fusion of visual and physiological data has been shown to improve cognitive state estimation and fatigue detection under real driving conditions [3], [4].

Third, edge deployment of the system on platforms such as NVIDIA Jetson Nano or Jetson Orin can make the solution more scalable and practical for real world applications. Performing low-latency inference on edge devices reduces dependence on external servers and enhances privacy, which is crucial for in-vehicle applications [5].

Finally, implementing adaptive learning techniques, such as continual or federated learning, can personalize the system for individual drivers. These approaches enable the model to evolve with each user's behavioral pattern without catastrophic forgetting, offering improved prediction performance over time [6], [7].

Together, these enhancements aim to evolve the proposed driver safety framework into a more intelligent, scalable, and personalized solution for next-generation Advanced Driver Assistance Systems (ADAS).

VIII. References

[1] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly, J. Uszkoreit, and N. Houlsby, "An image is worth 16x16 words: Transformers for image recognition at scale," in *Proc. Int. Conf. Learn. Represent. (ICLR)* in 2021.

[2] G. Bertasius, H. Wang, and L. Torresani, "Is space-time attention all you need for video understanding?" in *Proc. Int. Conf. Mach. Learn. (ICML)*, 2021, pp. 813–824.

[3] Y. Zhang, Y. Chen, M. Liu, and Y. Zhang, "Driver fatigue and emotion recognition with multimodal physiological signals," *IEEE Trans. Intell. Transp. Syst.*, vol. 22, no. 11, pp. 6869–6881, Nov. 2021, doi: 10.1109/TITS.2020.3044570.

[4] S. K. Yoo, J. H. Choi, and H. Park, "Multimodal physiological signal-based drowsiness detection using ensemble learning," *Sensors*, vol. 19, no. 23, pp. 1–18, Dec. 2019, doi: 10.3390/s19235180.

[5] P. Pławiak, M. Abdar and V. Makarenkov, "Applications of edge computing platforms for real-time AI-based driver behavior analysis," *in Future Gener. Comput. Syst.*, vol. 110, pp. 1–15, Mar. 2020, doi: 10.1016/j.future.2020.03.038.

[6] J. Delange, R. Aljundi, M. Masana, S. Parisot, X. Jia, A. Leonardis, G. Slabaugh, and T. Tuytelaars, "A continual learning survey: Defying forgetting in classification tasks," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 44, no. 7, pp. 3366–3385, Jul. 2022, doi: 10.1109/TPAMI.2021.3057446.

[7] H. B. McMahan, E. Moore, D. Ramage, S. Hampson, and B. y Arcas, "Communication-efficient learning of deep networks from decentralized data," in *Proc. Int. Conf. Artif. Intell. Statist. (AISTATS)*, 2017, pp. 1273–1282.