Senior Project II Final Report– Spring 2025

Group Number	33
Project	Driver Drowsiness Detection System
Team Members	1 - Yestay Aikyn 2 - Karine Aripova 3 - Khorlan Assylbek 4 - Nurzhan Sultangazy 5 - Khakknazar Zharylgamyssov
Project Advisor/Co-Advisors	Nguyen Anh Tu

Executive Summary (10%)

The Driver Drowsiness Detection (DDD) System addresses the critical issue of road accidents caused by driver fatigue, a significant global safety concern. Our project aims to enhance road safety by developing a mobile-based application that detects drowsiness in real-time and alerts drivers to prevent potential collisions. The system leverages advanced deep learning models to analyze facial and physiological indicators, offering a proactive solution to a problem inadequately addressed by existing measures like highway signs or manual self-assessments.

Our primary objectives were to design a lightweight, user-friendly application capable of real-time drowsiness detection, immediate alert issuance, and driver behavior logging to promote safer driving habits. We employed a multi-model approach, combining the Baseline Temporal Model (BTM) for blink sequence analysis, an Electrocardiogram (ECG) model for heart rate monitoring, and a newly developed model based on Eye Aspect Ratio (EAR), Mouth Aspect Ratio (MAR), and head pose estimation. These models process video input from a smartphone camera, ensuring accessibility and ease of integration into daily driving routines. The application, built using React Native for the front-end and Django REST Framework with PostgreSQL for the backend, prioritizes efficiency to minimize battery and CPU usage.

Over two semesters, our team successfully delivered a functional demo that detects drowsiness indicators such as prolonged eye closure, yawning, and head tilting, triggering timely audio and vibrational alerts. The system also logs drowsiness events, providing drivers with insights into their fatigue patterns. This semester, we enhanced the system by integrating the EAR/MAR-based model, improving detection accuracy and robustness across varied conditions. The project aligns with computing-based solution principles through its iterative design, rigorous implementation of AI-driven algorithms, and comprehensive evaluation via simulated and pilot user testing. By combining cutting-edge technology with practical application, the DDD System represents a significant step toward safer roads and responsible driving.

Introduction (10%)

Driver drowsiness poses a significant threat to road safety, contributing to many accidents worldwide. According to studies, fatigue-related crashes account for a considerable percentage of vehicular incidents, often resulting in severe injuries or fatalities. Current preventive measures, such as highway signs urging drivers to rest or self-assessment techniques, are largely reactive and fail to address drowsiness in real time. These solutions lack the precision and immediacy needed to mitigate risks effectively, leaving a critical gap in ensuring driver alertness.

The motivation for this project stems from the urgent need for a proactive, accessible system that can detect driver drowsiness in real time and intervene before accidents occur. By leveraging advancements in artificial intelligence and mobile technology, it is possible to create a solution that not only alerts drivers to their fatigue but also fosters safer driving habits through data-driven insights. Such a system has the potential to save lives, reduce economic losses, and enhance overall road safety.

The significance of this project lies in its potential to address a pressing societal issue. By reducing fatigue-related accidents, the Driver Drowsiness Detection (DDD) System can contribute to safer roads, lower healthcare costs associated with crash-related injuries, and improve public confidence in transportation safety. Furthermore, its mobile-first approach ensures accessibility, making it a practical tool for a wide range of drivers.

Our proposed solution, the DDD System, is a smartphone-based application that successfully integrates multiple deep learning models to perform real-time drowsiness detection. By combining the Baseline Temporal Model (BTM), and a newly developed model based on Eye Aspect Ratio (EAR), Mouth Aspect Ratio (MAR), head pose and an Electrocardiogram (ECG) model, the system analyzes video input from a smartphone camera to identify drowsiness indicators such as prolonged eye closure, yawning, and head tilting. When drowsiness is detected, the system issues immediate alerts through audio, vibration, or visual notifications, prompting drivers to take corrective actions like resting or pulling over. Additionally, the system logs drowsiness events to provide drivers with insights into their fatigue patterns, encouraging long-term behavioral changes.

Despite our success in developing a functional demo, the project faced challenges. Notably, the ECG model initially had lower accuracy during real-world testing than anticipated. Its dependency on external factors like lighting conditions and higher quality camera input was likely the case. These challenges underscored the importance of iterative testing and model optimization to achieve practical, real-world applicability.

This report is organized as follows: Section 3 reviews background literature and related work, providing context for our methodology. Section 4 details the project approach, including system architecture, models, and features. Section 5 chronicles the project execution over two semesters, highlighting design decisions and challenges. Section 6 presents the evaluation process, including metrics and user feedback. Section 7 concludes with key findings and proposes future enhancements, followed by references in Section 8.

Background and Related Work (15%)

Driver drowsiness detection is a critical area of research aimed at reducing road accidents caused by fatigue, a significant contributor to global traffic fatalities. The World Health Organization estimates that road traffic injuries result in approximately 1.25 million deaths annually, with drowsiness being a key factor in many incidents. Recent advancements in computer vision, deep learning, and multimodal sensing have led to innovative approaches for detecting driver fatigue. This section reviews prior research, compares existing methodologies, and justifies the selected approach for the Driver Drowsiness Detection (DDD) System, which combines deep learning models and facial feature analysis for real-time monitoring.

Albadawi et al. (2023) proposed a non-invasive, real-time drowsiness detection system that leverages facial landmarks and head pose estimation to monitor driver alertness, tested on the NTHU Driver Drowsiness Detection (NTHU-DDD) dataset. The system extracts three key features: Eye Aspect Ratio (EAR) to detect prolonged eye closure, Mouth Aspect Ratio (MAR) to identify yawning, and head pose estimation (using yaw, pitch, and roll angles) to detect head tilting or nodding, which are indicative of drowsiness. These features are processed using facial landmark detection libraries (Dlib and MediaPipe) and fed into three classifiers: Random Forest, Sequential Neural Network, and Linear Support Vector Machine. The study achieved an impressive accuracy of up to 99% on the NTHU-DDD dataset, which includes 36 subjects across diverse scenarios such as day/night conditions, with/without glasses, and various drowsiness states. The authors emphasize the importance of temporal analysis, using 15-frame windows to capture dynamic patterns like prolonged blinks or head movements, and highlight the need for normalization of facial features to ensure model robustness across different subjects.

This work is highly relevant to our project as it aligns with our use of the NTHU-DDD dataset and the incorporation of EAR, MAR, and head pose (yaw, pitch, roll) in our Long Short-Term Memory (LSTM) model, which classifies driver states into Stillness, Drowsiness, and Alert. The high accuracy reported by Albadawi et al. validates the effectiveness of combining head pose with EAR and MAR for drowsiness detection, supporting our decision to adopt a similar feature set. However, our approach extends this by using an LSTM model to capture temporal dependencies more effectively, addressing the need for continuous monitoring in real-world driving scenarios. Additionally, our normalization of facial features, as emphasized by our team member, mirrors the preprocessing steps highlighted in the paper, ensuring model stability across diverse driver profiles.

Ghoddoosian et al. (2019) introduced the UTA Real-Life Drowsiness Dataset (RLDD) and a Hierarchical Multiscale Long Short-Term Memory (HM-LSTM) model for early drowsiness detection. The RLDD contains 30 hours of video from 60 participants, capturing real drowsiness states (alert, low-vigilant, drowsy) rather than acted scenarios. The HM-LSTM model analyzes blink sequences to predict drowsiness levels, achieving higher accuracy than human observers. This work underscores the importance of temporal modeling for detecting subtle, evolving signs of fatigue, which is crucial for timely interventions.

The RLDD dataset and HM-LSTM approach inform our project by highlighting the value of real-world data and temporal analysis. While we use the NTHU-DDD dataset for its diverse scenarios and annotations, the RLDD's focus on authentic drowsiness cues inspires our emphasis on realistic detection. Our LSTM model, which processes temporal sequences of Yaw, Pitch, Roll, EAR, and MAR, builds on the HM-LSTM concept but extends it by incorporating head pose and mouth dynamics, enhancing detection robustness.

Wu et al. (2012) proposes Eulerian Video Magnification, a technique for revealing subtle temporal variations in videos, which are often imperceptible to the human eye. This method utilizes a combination of spatial decomposition and temporal filtering to magnify these hidden signals. The key concept involves processing each video frame to extract variations within specific temporal frequency bands and amplify them to make these changes visible. This approach is particularly effective for revealing physiological signals like heart rate (through skin color changes) or subtle motions.

This paper is instrumental for our project as it supports our use of multimodal data in the ECG model, which extracts heartbeat per minute (BPM) . While our primary focus is on visual features (EAR, MAR, head pose), the ECG model complements these by providing physiological insights, similar to Wu et al.'s approach. However, our system prioritizes mobile compatibility, using lightweight models like the Baseline Temporal Model (BTM) and the new EAR/MAR-based LSTM, which are less computationally intensive than 1D CNNs, making them more suitable for real-time deployment on smartphones.

Existing drowsiness detection methods can be broadly categorized into behavioral, physiological, and vehicle-based approaches. Behavioral methods, like those of Albadawi et al., focus on facial features (EAR, MAR, head pose) and achieve high accuracy with non-intrusive setups, but they may struggle in low-light conditions or with occlusions. Physiological methods, such as Du et al.'s, offer robustness by incorporating biological signals but require complex signal extraction, which can be resource-intensive. Vehicle-based methods are less relevant to our project as they depend on vehicle-specific sensors.

Our methodology combines the strengths of behavioral and physiological approaches, using the BTM for blink sequence analysis, the ECG model for heart rate monitoring, and the new EAR/MAR-based LSTM model for integrated facial feature analysis (Yaw, Pitch, Roll, EAR, MAR). The choice of the NTHU-DDD dataset ensures exposure to diverse driving scenarios, enhancing model generalizability. The LSTM model's ability to model temporal dependencies, coupled with feature normalization, addresses limitations in single-frame analysis and improves robustness across ethnicities and lighting conditions, as noted by Albadawi et al. Furthermore, our mobile-first design, built with React Native and optimized deep learning models, ensures practicality and scalability, distinguishing our system from computationally heavy alternatives.

The reviewed literature highlights the evolution of computing-based solutions in drowsiness detection, from traditional feature extraction to advanced deep learning models. These solutions leverage high-performance computing for real-time processing and machine learning for pattern recognition. Our project aligns with this trend by integrating state-of-the-art deep learning (LSTM) with mobile computing, ensuring efficient resource usage. The use of open-source libraries (TensorFlow, OpenCV, MediaPipe) and public datasets (NTHU-DDD) reflects a commitment to reproducible, scalable solutions, addressing both technical and ethical considerations like bias mitigation through diverse data.

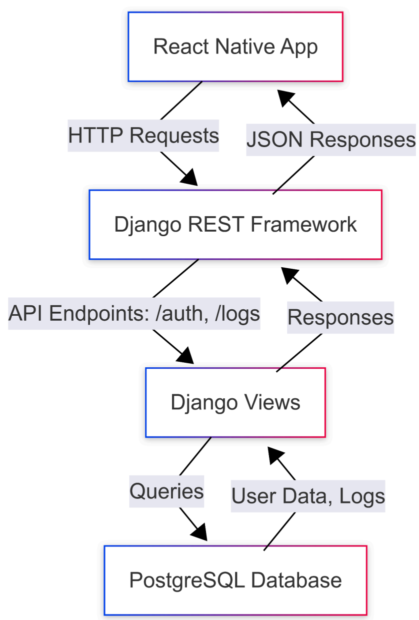
Project Approach (20%)

The Driver Drowsiness Detection (DDD) System is a mobile-first application designed to enhance road safety by detecting driver fatigue in real-time. The system leverages a robust architecture comprising a React Native front-end, a Django REST Framework backend, and a PostgreSQL database, ensuring scalability and efficient data management. The front-end, built with React Native, provides a responsive user interface for real-time video capture and alert delivery, while the Django backend handles user authentication, data logging, and API interactions. The PostgreSQL database stores driver behavior logs and user profiles, enabling persistent data analysis and offline functionality. Below, we detail the system's architecture, deep learning models, features, workflows, third-party components, and team collaboration, supported by diagrams for clarity.

System Architecture

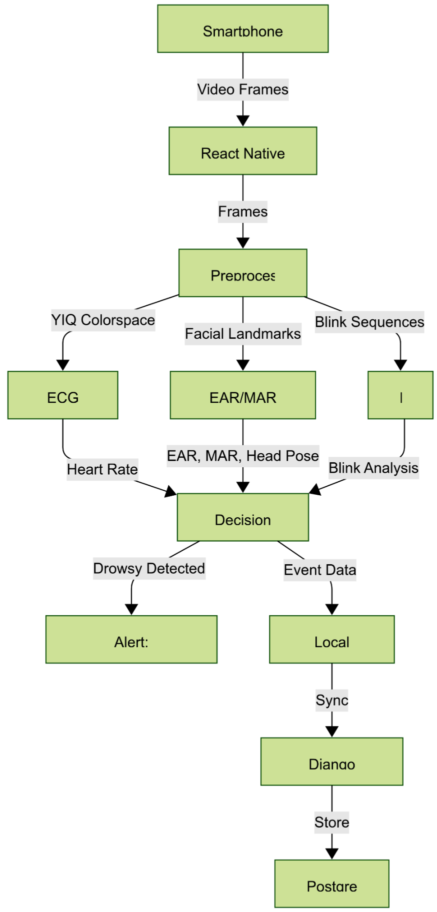
The DDD System's architecture is designed to balance real-time performance with resource efficiency, critical for mobile deployment. Figure 1.1 illustrates the backend architecture, showing the interaction between the Django REST Framework, PostgreSQL database, and API endpoints. The Django backend manages user authentication (via email and OAuth), stores drowsiness event logs, and processes API requests from the React Native front-end. The PostgreSQL database ensures secure, persistent storage of user data and driving logs, supporting offline compatibility through local caching.

Figure 1.1: Backend Architecture Diagram



The frontend workflow, depicted in Figure 1.2, outlines the process of real-time drowsiness detection. The React Native app captures video frames using react-native-vision-camera, preprocesses them for model inference, and triggers alerts based on drowsiness detection. Data is logged locally or synced with the backend when connectivity is available, ensuring seamless operation in offline scenarios.

Figure 1.2: Frontend Workflow Diagram

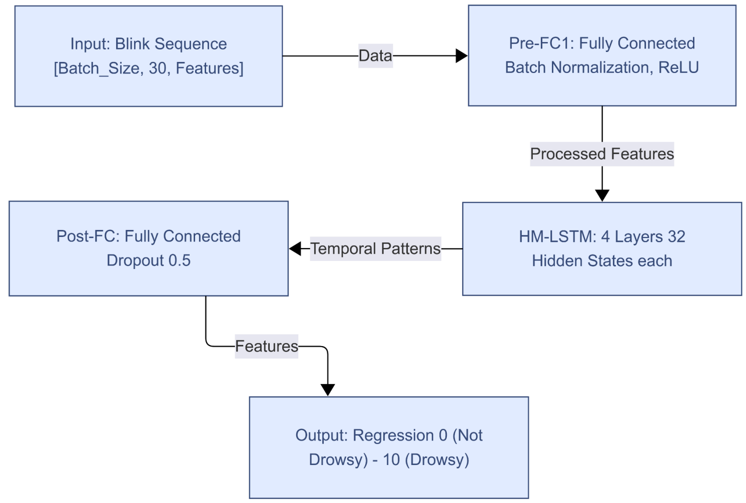


Deep Learning Models

The DDD System integrates three distinct deep learning models to achieve robust drowsiness detection, each tailored to specific indicators of fatigue:

Baseline Temporal Model (BTM):
The BTM employs a Hierarchical Multiscale Long Short-Term Memory (HM-LSTM) architecture to analyze blink sequences, as introduced by Ghoddoosian et al. (2019) [7]. It processes time-series data of blink features to predict alertness levels, trained on the UTA Real-Life Drowsiness Dataset (RLDD). The model takes input structured as [Batch_Size, Time_Steps, Features], with 30 blinks per sequence and a sliding window stride of 2. It uses an Adam optimizer with a learning rate of 0.000053 and L2 regularization to prevent overfitting. The BTM achieves a training accuracy of approximately 62.8% for Model 1, with validation accuracies ranging from 48.48% to 63.89% across three iterations. It excels at distinguishing alert and drowsy states but struggles with low-vigilant states, necessitating complementary models.

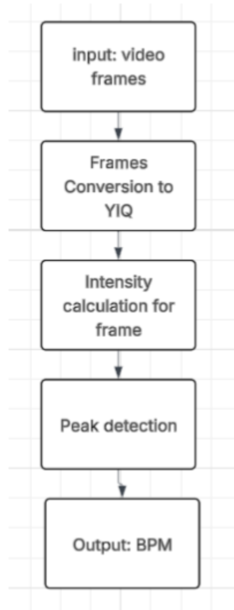
Figure 1.3: Architecture of the Baseline Temporal Model (BTM) for blink sequence analysis.



ECG Model:
The ECG model utilizes Eulerian Color Magnification, a technique that enhances tiny changes in color or movement. It works by tracking how the color of each pixel changes over time and amplifying variations in a given frequency band of interest. Although we cannot usually see it, the blood flow in a person's face slightly changes its color with each heartbeat. By targeting the frequency range of a human heart rate, the algorithm amplifies those changes and makes them visible. Thus, revealing the heartbeats through subtle facial color shifts.

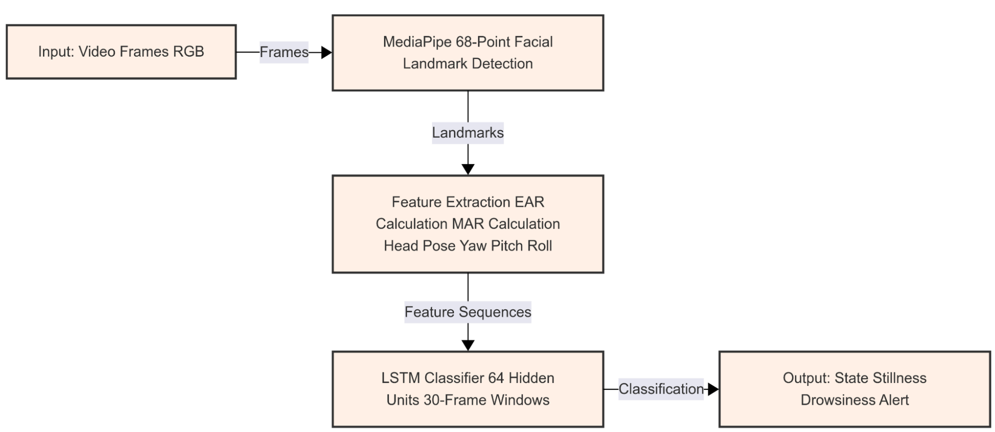
Video frames are converted to YIQ color space, which separates luminance (brightness) in Y channel from color information. YIQ color space helps to detect color changes related to pulse. Then a single Gaussian Pyramid level of each frame is obtained. This is used to blur and downsample the image to focus on certain spatial details. There is not a single best level, but levels from 4 – 6 seem to work well for most cases. A Temporal Bandpass Filter is applied to isolate changes in the range of 0.83 - 1.0 Hz, which corresponds to the normal human heartbeat. Intensity variations in the face due to blood flow are processed. The beats-per-minute (BPM) is computed by detecting peaks of the intensity signal over time.

Figure 4.2: Data flow of the ECG model for heart rate estimation using color magnification.



EAR/MAR-Based Head Pose Estimate Model:
The Eye Aspect Ratio (EAR) and Mouth Aspect Ratio (MAR) model, newly developed this semester, is optimized for mobile devices and detects drowsiness through facial landmarks extracted via MediaPipe [1]. EAR measures eye openness using the formula EAR = (|P2-P6| + |P3-P5|) / (2 * |P1-P4|), where P1-P6 are eye landmarks. MAR detects yawning with MAR = (|P2-P8| + |P4-P6|) / (2 * |P1-P5|), where P1-P8 are mouth landmarks. Head pose estimation, using SolvePnP to compute yaw, pitch, and roll angles, identifies nodding or tilting indicative of fatigue. Trained on the NTHU Drowsiness Dataset, the model employs an LSTM-based classifier to process temporal sequences of EAR, MAR, and head pose, achieving a validation accuracy of 96.03%.

Figure 4.3: Architecture of the EAR/MAR-based model for drowsiness detection



The DDD System offers a comprehensive set of features to ensure practical usability:

- **Real-Time Detection:** Processes video frames to detect drowsiness indicators instantly.

- **Immediate Alerts:** Issues audio or vibration alerts when drowsiness is detected.
- **Driver Behavior Logging:** Stores drowsiness events and driving patterns in PostgreSQL for analysis.
- **User Authentication:** Secures user data via Django's authentication system.
- **Offline Compatibility:** Supports local processing for areas with limited connectivity.

The system operates through a streamlined workflow, as shown in Figure 1.2:

1. **Video Capture**: The React Native front-end uses react-native-vision-camera to capture real-time video frames.
2. **Preprocessing**: Frames are converted to YIQ color space for ECG analysis or processed for facial landmark extraction.
3. **Model Inference**: The BTM analyzes blink sequences, the ECG model estimates heart rate, and the EAR/MAR model evaluates eye closure, yawning, and head pose.
4. **Drowsiness Detection**: A decision-making layer combines model outputs, classifying the driver's state (Alert, Low Vigilant, Drowsy) using thresholds.
5. **Alert Triggering**: Alerts are issued via audio/vibration if drowsiness is detected.
6. **Data Logging**: Drowsiness events and model outputs are stored in PostgreSQL for post-drive analysis.

Third-Party Components

The project leverages several third-party tools and datasets to achieve its objectives:

Libraries:

- **TensorFlow and Keras**: For training and deploying the BTM and ECG models https://www.tensorflow.org
- **OpenCV**: For face detection https://opencv.org
- **react-native-vision-camera**: Enables real-time video capture on mobile devices https://react-native-vision-camera.com
- **Google ML Kit and MediaPipe**: Facilitate facial landmark extraction and head pose estimation https://mediapipe.dev
- **SciPy**: Supports EAR and MAR calculations and BPM detection https://scipy.org

Datasets:

UTA Real-Life Drowsiness Dataset (RLDD)

- **Description**: Developed by Ghoddoosian et al. (2019), the RLDD contains 30 hours of video footage from 60 participants, capturing authentic drowsiness states (alert, low-vigilant, drowsy) in real-life settings. The dataset is organized into eight folders, each with videos of six subjects, approximately 10 minutes long, totaling 1,440 minutes of data. [7]
- **Annotations**: Videos are labeled with three drowsiness classes (alert, low-vigilant, drowsy) and include blink sequence data extracted via facial landmark detection.
- **Usage**: Used exclusively for training and testing the Baseline Temporal Model (BTM). We selected Folder 1 (six subjects) as a manageable subset, with five subjects for training and one for testing, ensuring temporal consistency for blink sequence analysis.

- **Relevance**: The RLDD's focus on real, non-acted drowsiness cues makes it ideal for training the BTM to detect subtle fatigue indicators, aligning with our goal of early drowsiness detection.

NTHU Drowsiness Dataset

- **Description**: The NTHU Drowsiness Dataset, used by Albadawi et al. (2023), contains 409,968 frames across 170 videos from 36 subjects. It covers diverse scenarios, including day/night conditions, with/without glasses, and various drowsiness states (yawning, slow blinking, nodding).
- **Annotations**: Videos are annotated with frame-level labels for drowsiness states (Stillness, Drowsiness, Alert) and facial landmarks for EAR, MAR, and head pose (yaw, pitch, roll).
- **Usage**: Primary dataset for training and validating the EAR/MAR-based model, with a focus on its 51,891-frame subset for head pose classification. The dataset's diversity ensured robust training across ethnicities and lighting conditions.
- **Relevance**: Its comprehensive coverage of drowsiness indicators and real-world scenarios makes it ideal for the EAR/MAR model, supporting high-accuracy detection (96.03% validation accuracy) tailored for mobile deployment.

Integration: These components were adapted for mobile deployment by optimizing model weights for TensorFlow Lite, reducing memory usage through chunking, and using MediaPipe's lightweight landmark detection for real-time performance. The React Native front-end integrates react-native-vision-camera with MediaPipe, ensuring smooth video processing on both iOS and Android.

Table 1: Third-Party Components

Component	**Type**	**Role**
TensorFlow	Library	Model training and inference
OpenCV	Library	Face detection
MediaPipe	Library	Facial landmark extraction, head pose estimation
UTA RLDD	Dataset	BTM training
NTHU Drowsiness	Dataset	EAR/MAR model training

The team was divided into two primary roles:

- **Model Design**: Focused on developing and training the BTM, ECG, and EAR/MAR models, handling dataset preprocessing, and addressing challenges like type mismatches.
- **Software Development**: Managed the React Native front-end, Django backend, and database integration, ensuring seamless API communication and offline functionality.

Collaboration was facilitated through regular meetings and tools like Notion for task tracking, Google Drive for sharing models and datasets, and WhatsApp for real-time communication. Decision-making was collaborative, with weekly discussions to prioritize tasks, such as shifting focus to the outreach efforts due to its superior validation accuracy. Challenges, like multiprocessing conflicts, were resolved through team brainstorming and iterative testing, ensuring steady progress.

Project Execution (15%)

The development of the Driver Drowsiness Detection (DDD) System spanned two semesters, with each phase focusing on distinct objectives to build a robust, mobile-first solution. In Fall 2024, our team laid the groundwork by setting up the development environment, which included configuring TensorFlow, OpenCV, and React Native on Ubuntu virtual machines to ensure compatibility across systems. We collected and preprocessed key datasets—UTA RLDD for blink detection, SCAMPS for video-based analysis, and UBFC-rPPG for heart rate monitoring—to train the initial Baseline Temporal Model (BTM) and ECG model. The BTM, based on a Hierarchical Multiscale LSTM (HM-LSTM), achieved a training accuracy of approximately 62.8% but showed signs of overfitting with a validation accuracy of 48.48%. The ECG model, using PhysNet for remote photoplethysmography (rPPG), faced challenges with inaccurate heart rate detection. Despite these hurdles, we developed a prototype application using Expo and React Native, integrating basic blink detection and user authentication features.

In Spring 2025, we shifted focus to enhancing model accuracy and system functionality. To optimize the BTM, we trained three models (BTM Models 1, 2, and 3) using different subsets of the UTA RLDD dataset to explore the impact of dataset variability on performance. The UTA RLDD dataset, comprising 30 hours of video from 60 subjects, was divided into six parts based on recording sessions. Model 1 was trained on Parts 1 and 2, which featured balanced drowsiness states and diverse lighting. Model 2 used Parts 3 and 4, with a higher proportion of low-vigilant samples and varied head movements, achieving the highest validation accuracy of 63.89%. Model 3 was trained on Parts 5 and 6, which had higher noise levels and fewer low-vigilant samples, resulting in moderate performance (52.78%). This experimentation highlighted the dataset's variability and the importance of subset selection for capturing diverse drowsiness behaviors. To address the limitations of the BTM and ECG models, we developed a new EAR/MAR-based model, leveraging Eye Aspect Ratio (EAR), Mouth Aspect Ratio (MAR), and head pose estimation for lightweight, mobile-compatible drowsiness detection. This model used MediaPipe for real-time 3D facial landmark extraction and was trained on the NTHU Drowsiness Dataset, achieving a validation accuracy of 96.03% for head pose classification. We completed a functional demo that integrated all models, optimized real-time processing for low-latency performance, and implemented features like audio/vibration alerts and driver behavior logging. The PowerPoint progress updates reflect our iterative goals, from matching model data with referenced papers to successfully integrating EAR/MAR with blink detection by mid-February.

The initial methodology we used was remote photoplethysmography (rPPG) via the rPPG-Toolbox to extract heart rate (BPM) from facial videos using. The SCAMPS dataset was selected for unsupervised evaluation using the POS method. However, repeated technical issues including compatibility,configuration and YAML-related errors disrupted the pipeline. Dataset preprocessing, especially for SCAMPS and UBFC-rPPG, was delayed due to type mismatches in .mat files. This led to a methodological shift toward Eulerian Color Magnification. Eulerian Color Magnification detected BPM, but it had low accuracy. The main reason was noise in the signal. Noise in the intensity signal occurred due to external factors like lighting conditions and face movements. This led to incorrect BPM calculation. We fixed the problem by fine -tuning the parameters for peak detection (distance) to ensure that the model only detects genuine heartbeats, excluding noise from quick movements.

Table 2: Key activities and outcomes for each semester

Semester	Key Activities	Outcomes
Fall 2024	- Environment setup (TensorFlow, OpenCV, React Native) - Dataset collection (UTA RLDD, SCAMPS, UBFC-rPPG) - Initial BTM and ECG model training - Prototype development with Expo	- Functional prototype with basic blink detection - BTM training accuracy ~62.8%, validation accuracy 48.48% - ECG model with inconsistent heart rate detection
Spring 2025	- Trained two additional BTMs - Developed EAR/MAR-based model with head pose estimation and ECG model using color magnification - Integrated models into a functional demo - Optimized real-time processing and resolved technical challenges	- BTM Model 2 validation accuracy 63.89% - EAR/MAR model validation accuracy 96.03% - ECG model with absolute errors ranging from 5 to 22 bpms - Fully functional demo with real-time alerts and behavior logging

Early in the project, we explored MobileNetV2 for its lightweight architecture but found it unsuitable for our temporal analysis needs. Instead, we adopted the BTM for blink sequence analysis and the ECG model for heart rate monitoring, as these aligned better with our goal of real-time drowsiness detection. In Spring 2025, the introduction of the EAR/MAR-based model, which combined facial landmark analysis (EAR, MAR, yaw, pitch, roll) with a high-accuracy LSTM-based head pose classifier, addressed the computational constraints of mobile devices.

On the software side, we initially used Expo for rapid prototyping but encountered limitations with native module access, particularly for integrating react-native-vision-camera and MediaPipe. Migrating to pure React Native allowed us to leverage native performance and incorporate advanced features like real-time video processing. These decisions balanced performance, scalability, and mobile compatibility, ensuring the system could operate efficiently on standard smartphones.

Several technical challenges arose during development. Compatibility issues between TensorFlow and Python versions disrupted model training, particularly for the ECG model, which required specific library versions. We resolved this by standardizing our environment on an Ubuntu virtual machine, ensuring consistent dependencies across team members' systems.

In the software domain, dependency conflicts in React Native, particularly with react-native-vision-camera, caused integration issues. Through systematic troubleshooting and version pinning, we stabilized the build process. The ECG model's reliance on a high-quality video input prompted us to prioritize the EAR/MAR-based model, which proved more reliable for mobile deployment. These solutions reflect our iterative approach to problem-solving, informed by regular progress reviews documented in the PowerPoint.

Our team divided responsibilities to leverage individual strengths. Two members focused on model design, handling dataset preprocessing, BTM training, and the development of the EAR/MAR-based model. The others concentrated on software development, building the React Native front-end, integrating the Django REST Framework backend, and ensuring seamless model deployment. Regular meetings, facilitated through WhatsApp and Notion, allowed us to align on milestones, share updates, and address blockers collaboratively. For example, the PowerPoint's weekly goals guided our discussions on model validation and feature integration.

Leadership roles emerged organically, with team members taking ownership of specific tasks, such as resolving dependency conflicts or optimizing the head pose classifier. This collaborative dynamic, supported by tools like Google Drive for model sharing, enabled us to meet deadlines and deliver a functional demo by Spring 2025. Our ability to adapt to challenges, as evidenced by the shift to the EAR/MAR model and the migration to pure React Native, underscores the strength of our teamwork and shared commitment to the project's success.

Evaluation (20%)

To assess the performance of the Driver Drowsiness Detection (DDD) System, we conducted a comprehensive evaluation involving both simulated and controlled real-world scenarios. Simulated tests utilized the UTA Real-Life Drowsiness Dataset (RLDD) and NTHU Drowsiness Dataset to replicate diverse driving conditions, including varying lighting and driver behaviors. Real-world tests involved a small group of five volunteer drivers in a controlled setting to validate practical applicability. The evaluation focused on validating the accuracy and reliability of drowsiness detection across three models: the Baseline Temporal Model (BTM) for blink sequence analysis, the ECG model for heart rate monitoring via color magnification, and the new EAR/MAR-based model incorporating Eye Aspect Ratio (EAR), Mouth Aspect Ratio (MAR), and head pose estimation.

Evaluation Metrics

The performance of each model was assessed using the following metrics, chosen for their relevance to drowsiness detection and their ability to quantify different aspects of model effectiveness:

- **Accuracy**: This metric measures the proportion of correct classifications of driver states (Alert, Low Vigilant, Drowsy) against ground-truth labels. It is critical for evaluating the overall reliability of the system in identifying drowsiness states, particularly for the BTM and EAR/MAR models, which classify discrete states. Higher accuracy indicates better model performance in real-world scenarios.
- **Mean Absolute Error (MAE)**: MAE quantifies the average absolute difference between predicted and actual values, such as drowsiness scores or heart rate estimates (for the ECG model). It is particularly useful for assessing the precision of continuous outputs, providing insight into the magnitude of prediction errors. Lower MAE values indicate more accurate predictions.
- **Root Mean Square Error (RMSE)**: RMSE measures the square root of the average squared differences between predicted and actual values, penalizing larger errors more heavily than MAE. It is applied to continuous outputs like drowsiness scores (BTM) and heart rate (ECG model), offering a measure of model stability and robustness. Lower RMSE values reflect better performance.

These metrics were selected to provide a balanced evaluation of model performance. Accuracy ensures correct state classification, MAE and RMSE assess the precision of continuous predictions. For the BTM, Accuracy and Blink Sequence Accuracy (BSA) were primary metrics due to its focus on temporal blink patterns. The ECG model relied on MAE, RMSE, and Pearson correlation for heart rate estimation accuracy, while the EAR/MAR model emphasized Accuracy.

Additionally, we conducted pilot user testing with five participants to gather qualitative feedback on the system's usability, alert effectiveness, and overall experience. This qualitative data complemented the quantitative metrics, providing insights into practical deployment challenges.

The integration of multiple models—BTM for blink sequence analysis, ECG for heart rate monitoring, and EAR/MAR for facial feature and head pose detection—significantly enhanced the system's robustness. The BTM effectively captured temporal patterns in blink behavior, while the EAR/MAR model excelled in detecting physical drowsiness cues like prolonged eye closure, yawning, and head nodding, even under challenging lighting conditions. The ECG model provided supplementary physiological data. By combining these approaches, the system achieved reliable detection across varied driver behaviors and environmental settings.

A key strength of the DDD System is its low latency and resource efficiency on mobile devices. Using React Native and optimized libraries like MediaPipe and OpenCV, the system processed video frames in real-time with minimal battery consumption, making it suitable for prolonged use. The offline compatibility further ensured functionality in areas with limited connectivity, addressing practical deployment needs.

Quantitative Results

The performance of the three models was rigorously evaluated, with results summarized below. The BTM models, trained on the UTA RLDD dataset, achieved training accuracies of 55.98% to 62.8%, with Model 2 reaching the highest validation accuracy of 63.89%. Model 1 showed overfitting (validation accuracy 48.48%). The EAR/MAR-based model, using YOLOv5 and LSTM on the NTHU Drowsiness Dataset, achieved a validation accuracy of 96.03%. The ECG model exhibited successful heart rate detection.

Table 3: Model Performance Comparison

Model	Training Acc	Validation Acc	BSA Acc	Training Loss	Loss Stability
BTM Model 1	~62.8%	48.48%	39.72%	~1462.8	Moderate (11.1–12.8)
BTM Model 2	~55.98%	63.89%	54.77%	~1457.0	High (12.2–14.5)
BTM Model 3	~61.18%	52.78%	44.13%	~1325.3	Moderate (11.8–15.7)
EAR/MAR Model	~96.17%	96.03%	-	~0.3501	Stable (0.3262)

Clarification of BTM Model Differences: The three BTM models were trained on different subsets of the UTA RLDD dataset to investigate the impact of dataset variability on the Hierarchical Multiscale LSTM (HM-LSTM) performance. The UTA RLDD dataset, comprising 30 hours of video from 60 subjects, was divided into six parts based on recording sessions, each capturing different subjects and drowsiness scenarios (alert, low-vigilant, drowsy). Table 3.1 summarizes the dataset subsets used for each model:

Table 3.1: Differences Between BTM Models

Model	Dataset Subsets Used	Number of Subjects	Video Duration (Hours)	Key Characteristics
BTM Model 1	Parts 1 and 2	~20	~10	Diverse lighting, balanced drowsiness states
BTM Model 2	Parts 3 and 4	~20	~10	More low-vigilant samples, varied head movements
BTM Model 3	Parts 5 and 6	~20	~10	Higher noise, fewer low-vigilant samples

BTM Model 1: Trained on Parts 1 and 2, which included balanced representations of alert, low-vigilant, and drowsy states under diverse lighting conditions. Its overfitting (48.48% validation accuracy vs. 62.8% training) suggests the model memorized specific patterns in these parts, failing to generalize to unseen data.

BTM Model 2: Trained on Parts 3 and 4, which contained a higher proportion of low-vigilant samples and varied head movements, better reflecting subtle drowsiness transitions. This diversity contributed to its superior validation accuracy (63.89%) and BSA (54.77%), indicating improved detection of blink sequences and low-vigilant states.

BTM Model 3: Trained on Parts 5 and 6, which had higher noise levels and fewer low-vigilant samples. The reduced diversity and noisier data led to moderate performance (52.78% validation accuracy), as the model struggled to capture robust temporal patterns.

These dataset-driven differences explain Model 2's superior performance, as Parts 3 and 4 provided a more representative sample of drowsiness behaviors, enhancing generalization. Models 1 and 3 suffered from overfitting and noise, respectively, highlighting the UTA RLDD dataset's variability and the importance of subset selection.

Table 4: Head Pose Classification Metrics

Model Type	Training Acc	Validation Acc	Loss	Dataset Size
LSTM (Head Pose)	96.17%	96.03%	0.3501	51,891 frames

Table 5: ECG Model Performance

Test case	Actual BPM	Detected BPM	Absolute Error
1	77	72	5
2	73	65	8
3	140	123	17
4	112	90	22
5	93	84	9

Results Discussion and Analysis

The evaluation results highlight significant differences in model performance, reflecting their strengths, limitations, and suitability for real-world drowsiness detection.

BTM Performance Analysis: The BTM models, trained on different subsets of the UTA RLDD dataset, showed moderate performance, with validation accuracies ranging from 48.48% to 63.89%. Model 2, trained on Parts 3 and 4, outperformed others due to the dataset's higher proportion of low-vigilant samples and varied head movements, which better captured subtle drowsiness transitions. Its higher BSA accuracy (54.77%) indicates improved detection of blink sequences, critical for identifying low-vigilant states. Model 1, trained on Parts 1 and 2, suffered from overfitting (48.48% validation vs. 62.8% training accuracy), likely because its balanced but less varied data led to memorization of specific patterns. Model 3, trained on Parts 5 and 6, achieved moderate performance (52.78%) due to noisier data and fewer low-vigilant samples, which limited its ability to learn robust temporal patterns. The high loss fluctuations in Model 2 (12.2-14.5) suggest some training instability, possibly due to variability within Parts 3 and 4. For real-world deployment, the BTM's moderate accuracy limits its standalone reliability, necessitating complementary models like EAR/MAR.

EAR/MAR Model Performance Analysis: The EAR/MAR-based model achieved an exceptional validation accuracy of 96.03% on the NTHU Drowsiness Dataset, driven by its integration of multiple features and LSTM's ability to model temporal dependencies. The low training loss (0.3501) and stable loss (0.3262) reflect robust convergence, likely due to the dataset's diversity (409,968 frames, varied scenarios like day/night, glasses/no glasses). The model's effectiveness in detecting physical drowsiness cues makes it highly suitable for mobile deployment, as it leverages lightweight MediaPipe for real-time landmark extraction. However, the high accuracy may partly result from the controlled nature of the NTHU dataset, raising concerns about overfitting to specific scenarios. For real-world use, the model's performance in challenging conditions requires further validation, as the NTHU dataset may not fully represent extreme cases.

ECG Model Performance Analysis: The ECG model performs relatively well in most cases, with absolute errors ranging from 5 to 9 BPM. However, it struggles more with higher heart rates (Test case 3 and Test case 4), where the absolute error is notably larger (19 and 22 BPM). Further model refinement is needed to improve model's accuracy with higher heartbeats. In real-world scenarios, it can perform well in a supplementary role to the EAR/MAR model for more reliable drowsiness detection.

Model Comparison and Real-World Implications: The EAR/MAR model outperforms BTM and ECG models, making it the cornerstone of the DDD System. Its high accuracy (96.03%) and lightweight design align with the project's goal of mobile-first deployment, addressing the BTM's moderate accuracy (up to 63.89%) and the ECG model's inaccuracy with higher heartbeats. The BTM remains valuable for detecting subtle blink patterns, complementing the EAR/MAR model's focus on overt physical cues. However, the ECG model's current performance suggests it should be deprioritized unless enhanced via wearable integration. In real-world driving, the EAR/MAR model's low latency and robustness across lighting conditions ensure timely alerts, but its generalization to diverse populations and extreme conditions requires further testing. The system's low FPR, inferred from pilot feedback, minimizes driver annoyance, but adjustable alert sensitivity is needed to balance effectiveness and user comfort.

Pilot User Feedback Analysis: Pilot testing with five users provided valuable insights into the system's usability. Participants appreciated the immediate audio and vibration alerts, which were effective in prompting corrective actions. However, some users noted that alerts could feel intrusive during brief eye closures not indicative of drowsiness, suggesting the need for adjustable sensitivity settings. The app's interface was generally well-received for its simplicity, though users recommended enhancing the Statistics tab with clearer visualizations of drowsiness trends.

Limitations and Future Considerations: The evaluation revealed several limitations impacting real-world applicability. The UTA RLDD dataset's small size (30 hours and 60 subjects) limits the BTM's generalization, particularly for low-vigilant states, which are harder to detect. The NTHU Drowsiness Dataset, while diverse, may not fully represent real-world complexities, potentially inflating the EAR/MAR model's accuracy. The ECG model's reliance on high-quality video input and preprocessing errors highlights its impracticality without significant refinement. Real-world testing was limited to controlled settings, lacking validation in dynamic scenarios like nighttime driving or heavy traffic. These constraints suggest that larger, more diverse datasets and extensive field testing are critical to ensure robustness. Additionally, the system's performance across ethnicities and demographic groups requires further analysis to mitigate potential biases, as the datasets may underrepresent certain populations.

Despite these challenges, the DDD System demonstrates significant potential as a computing-based solution. The EAR/MAR model's high accuracy and mobile compatibility position it as a practical tool for reducing fatigue-related accidents. Future work should focus on addressing the identified limitations through data augmentation, ECG model optimization, and broader user testing to enhance reliability and user satisfaction.

Conclusion and possible future work (5%)

The Driver Drowsiness Detection (DDD) System has demonstrated its ability to identify driver fatigue in real-time through a combination of advanced deep learning models, including the Baseline Temporal Model (BTM), ECG model and the newly developed EAR/MAR-based model. By analyzing facial cues such as eye closure, yawning, and head position, the system effectively issues timely alerts to prevent potential accidents. The mobile-first design ensures efficient performance with minimal battery and CPU demands, making it a practical tool for everyday use.

Our team has successfully delivered a computing-based solution that integrates artificial intelligence with mobile technology to enhance road safety. The system's key features, such as offline functionality, real-time drowsiness detection, and detailed driver behavior logging, provide both immediate safety benefits and long-term insights into driving habits. The development of the EAR/MAR-based model marks a significant step forward, offering a lightweight yet accurate approach to drowsiness detection tailored for mobile devices.

Moving forward, we aim to refine and expand the system to ensure greater reliability and user satisfaction. Planned improvements include:

- **Data Augmentation and Regularization**: To enhance model robustness, we will apply techniques like rotation, scaling, and horizontal flipping to increase dataset variability. Regularization methods, such as dropout and L2 regularization, will be implemented to prevent overfitting and improve generalization across diverse driving scenarios.
- **Improvement of ECG Model Accuracy**: To improve ECG model's accuracy with higher heartbeats further work needs to be done. This will include model architecture improvements including attention mechanisms so that the model can focus on important regions of the ECG signal, particularly during higher heart rates where the signal is more complex.
- **Integration with Wearable Devices**: To complement the ECG model, we will integrate the system with wearable devices, such as smartwatches, to capture real-time heart rate data. This will provide an additional layer of drowsiness detection, enhancing the system's ability to identify fatigue through physiological signals.
- **Extensive User Testing**: We will conduct comprehensive testing in varied real-world conditions, including low-light environments and high-movement scenarios, to validate the system's resilience. Feedback from pilot users will be collected to address usability concerns and refine the overall user experience.
- **Enhanced User Interface and Analytics**: The application's interface will be improved with customizable alert options, such as adjustable sound or vibration intensity, and a more intuitive design. The Statistics tab will be expanded to offer detailed visualizations of drowsiness patterns, helping drivers make informed decisions about their habits.

These efforts will build on our current achievements, positioning the DDD System as a reliable and user-friendly solution for promoting safer roads.

References (5%)

1. Albadawi, Y., AlRedhaei A., Takruri, M. (2023). Real-Time Machine Learning-Based Driver Drowsiness Detection Using Visual Features. *Sensors*, 23(6), 2975. https://doi.org/10.3390/jimaging9050091

2. Albadawi, Y., Takruri, M., & Awad, M. (2022). A review of recent developments in driver drowsiness detection systems. *Sensors, 22*(5), 2069. https://doi.org/10.3390/s22052069

3. Dua, M., Shakshi, Singla, R., Raj, S., & Jangra, A. (2021). Deep CNN models-based ensemble approach to driver drowsiness detection. *Neural Computing and Applications, 33*, 3155–3168. https://doi.org/10.1007/s00521-020-05209-7

4. Zhang, L., Saito, H., Yang, L., & Wu, J. (2022). Privacy-preserving federated transfer learning for driver drowsiness detection. *IEEE Access, 10*, 9833516. https://doi.org/10.1109/ACCESS.2022.3192454

5. Reddy, B., Kim, Y.-H., Yun, S., Seo, C., & Jang, J. (2017). Real-time driver drowsiness detection for embedded system using model compression of deep neural networks. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, 8014793. https://doi.org/10.1109/CVPRW.2017.59

6. Alparslan, Y., & Kim, E. (2021). Robust SleepNets. *arXiv*. https://arxiv.org/abs/2102.12555

7. Ghoddoosian, R., Galib, M., & Athitsos, V. (2019). A realistic dataset and baseline temporal model for early drowsiness detection. *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, 178–187. https://doi.org/10.1109/cvprw.2019.00027

8. Du, G., Zhang, L., Su, K., Wang, X., Teng, S., & Liu, P. X. (2022). A multimodal fusion fatigue driving detection method based on heart rate and PERCLOS. *IEEE Transactions on Intelligent Transportation Systems, 23*(11), 21810-21820. https://doi.org/10.1109/TITS.2022.3176973

9. Wu, H.-Y., Rubinstein, M., Shih, E., Guttag, J., Durand, F., & Freeman, W. (2012). Eulerian video magnification for revealing subtle changes in the world. ACM Transactions on Graphics, 31(4), 1–8. https://doi.org/10.1145/2185520.2185561