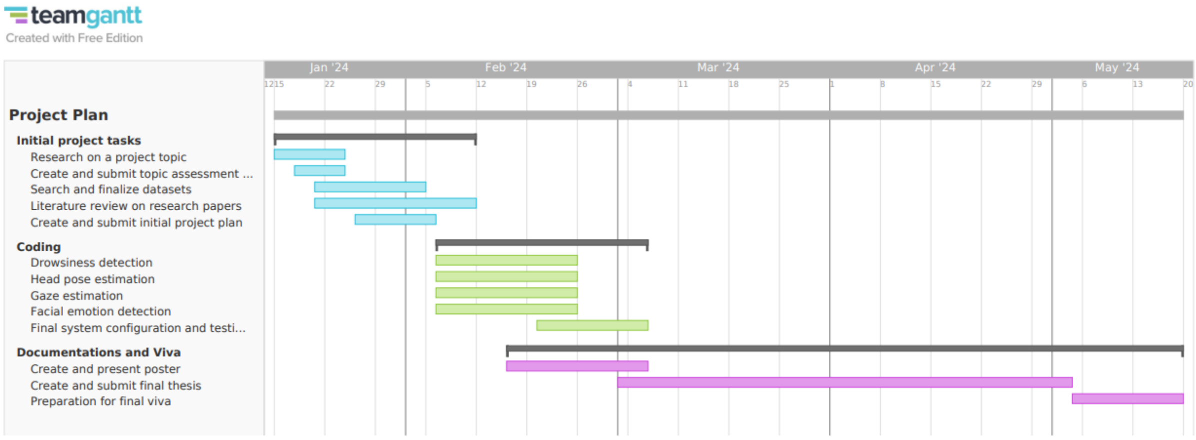
Multimodal Fusion for Comprehensive Driver Analysis

Introduction: With the continuous growth in transportation options, the rising number of road accidents poses a growing concern for passengers. The World Health Organisation estimates that there are between 20 and 50 million injuries and 1.3 million fatalities from traffic-related incidents annually [1]. One prominent element contributing to road accidents is driver distraction brought on by fatigue or drowsiness. According to the National Sleep Foundation, 60% of adult drivers in the United States have acknowledged driving when drowsy, with one-third falling asleep behind the wheel [2]. In response to the critical need for improved road safety, various research studies have focused on computer vision and machine learning for driver monitoring. Al-madani, A. M., et al. [3] designed and implemented a system to detect drowsiness based on eye movement and yawning using dlip and facial landmarks for the eyes and mouth. Huynh, X., et al. [4] introduced a drowsiness detection method using a 3D convolutional neural network (CNN) for feature extraction, followed by classification using gradient boosting, and enhanced system performance by incorporating semi-supervised learning. With the use of OpenCV and Dlib for feature extraction and state detection, Lashkov, I. et al. [5] developed a real-time driver monitoring system that evaluates the driver's dangerous state based on facial features such as eye state, head movements, and mouth state. According to [3], there are two approaches to detecting drowsiness: intrusive and non-intrusive methods. Intrusive methods include measuring heart rate and monitoring brain waves, which necessitate a direct connection to the driver's body and can be inconvenient and distracting [3]. Non-intrusive approaches, which include metrics such as eye closure length, frequent eye blinking, yawning detection, and head rotation, provide a more realistic solution because they do not disturb the driver and do not require direct connections to the driver's body [3]. The driver's mental state has a considerable impact on driving, with angry or furious states increasing the possibility of distractions, emphasising the critical significance of keeping focus and attention when driving [6]. According to Roidl, E. et al. [7], anger, stress, or anxiety may contribute to increased vehicle speed or acceleration. The primary objective of this research is to implement a real-time driver monitoring model that uses a non-intrusive approach that utilises sophisticated computer vision and machine learning techniques to identify possible instances of driver drowsiness or distraction based on the real-time analysis of head pose, gaze, drowsiness, and facial expressions.

Datasets and Ethical Consideration Statement: MRL Eye – The MRL Eye dataset is an extensive collection of infrared images of human eyes, including both high- and low-resolution images taken with various devices and under different lighting circumstances. The infrared ensures privacy by not including identifiable information or personal details. The use of MRL Eye dataset does not pose substantial privacy concerns because of the anonymised nature of the dataset. Eye Gaze Detection – The Eye gaze detection dataset consists of labelled eye images indicating the direction of gaze gathered for inattentive driver detection using UnityEyes eye-tracking simulator, to ensure precise and high-quality gaze data across a wide range of driving circumstances. This dataset is synthetically created and it does not involve real individuals. Ethical cinsiderations related to real-world surveillance are mitigated by the controlled nature of data gathering and labelling of this dataset. FER+ - The FER+ (Face Expression Recognition Plus) dataset was developed as an expansion of the FER2013 dataset in order to solve many of its shortcomings. The FER+ dataset includes facial images categorized into seven different facial expressions, similar to FER2013: neutral, happiness, surprise, sadness, anger, disgust, fear, and contempt. It was presented and published as part of the facial expression recognition research conducted by Microsoft Research researchers, indicating adherence to ethical research practices. To minimise privacy issues, the dataset consists of face image categories based on expressions and excludes personally identifiable information. The dataset is a component of a legitimate research project that advances standards of ethics and transparency in the field of facial expression recognition.

Methodology: Head pose detection: The project aims to accurately detect the driver's head pose by using OpenCV and the MediaPipe framework for facial landmark analysis. Drowsiness detection: Developing a neural network model from scratch to determine if the driver's eyes are open or closed. Similarly, a transfer learning model will be trained for the same task. Evaluating the performance of both neural network and transfer learning models and choosing the best performer for deployment. The chosen model will be used to generate real-time predictions about the driver's drowsiness level. Gaze detection: Train a model to identify the driver's gaze direction as forward, left, or right. Emotion detection: Training a model to identify emotion state of the driver using facial emotions. This contributes to a more complete understanding of the driver's mental state. After training each of the distinct models, the project intends to merge them and establish particular thresholds. The system will estimate whether the driver is distracted based on the outputs of the head posture, sleepiness, gaze, and emotion models.

Project Plan:



References:

[1] Mašanović, L., Vranješ, M., Džakula, R., & Lukač, Ž. (2019). Driver monitoring using the in-vehicle camera. In 2019 Zooming Innovation in Consumer Technologies Conference (ZINC) (pp. 33-38). Novi Sad, Serbia. doi:10.1109/ZINC.2019.8769377.
[2] Sleep Foundation. (2021). *Driving While Drowsy Can Be as Dangerous As Driving While Drunk*. [online] Available at: https://www.sleepfoundation.org/sleep-topics/drowsy-driving-vs-drunk-driving-how-similar-are-they. [Accessed 5 Feb. 2024].
[3] Al-madani, A. M., Gaikwad, A. T., Mahale, V., Ahmed, Z. A. T., & Shareef, A. A. A. (2021). Real-time Driver Drowsiness Detection based on Eye Movement and Yawning using Facial Landmark. In 2021 International Conference on Computer Communication and Informatics (ICCCI) (pp. 1-4). Coimbatore, India. doi:10.1109/ICCCI50826.2021.9457005.
[4] Huynh, XP., Park, SM., & Kim, YG. (2017). Detection of Driver Drowsiness Using 3D Deep Neural Network and Semi-Supervised Gradient Boosting Machine. In Computer Vision - ACCV 2016 Workshops. ACCV 2016. Lecture Notes in Computer Science (Vol. 10118).
[5] Lashkov, I., Kashevnik, A., Shilov, N., Parfenov, V., & Shabaev, A. (2019, August). Driver Dangerous State Detection Based on OpenCV & Dlib Libraries Using Mobile Video Processing. In 2019 IEEE International Conference on Computational Science and Engineering (CSE) and IEEE International Conference on Embedded and Ubiquitous Computing (EUC) (pp. 74-79).
[6] Tauqeer, M., Rubab, S., Khan, M. A., Naqvi, R. A., Javed, K., Alqahtani, A., Alsubai, S., Binbusayyis, A. (2022). Driver's emotion and behavior classification system based on Internet of Things and deep learning for Advanced Driver Assistance System (ADAS). Computer Communications, 194, 258-267. ISSN 0140-3664. https://doi.org/10.1016/j.comcom.2022.07.031.