FINDINGS REPORT

Computer Vision Monitoring for Work Fatigue

Client: Babcock Australasia – Karen Trezise, Aaron Boehm & Adrian Spinazzola

Project Mentor: Sira Yongchareon

Project Manager: Alana Mao
Team Members: Shim Chen, Nikisha Kumar

Prepared by: Shim Chen

Comparison Report: Fatigue Detection (using MAR and EAR) VS (Using yolov5)

Table of Contents

1. Objective 4
2. Methodology 4
   a. Tools compared 4
   b. Dataset 6
   c. Metrics Used 6
3. Results 7
4. Analysis 8
5. Conclusion 10

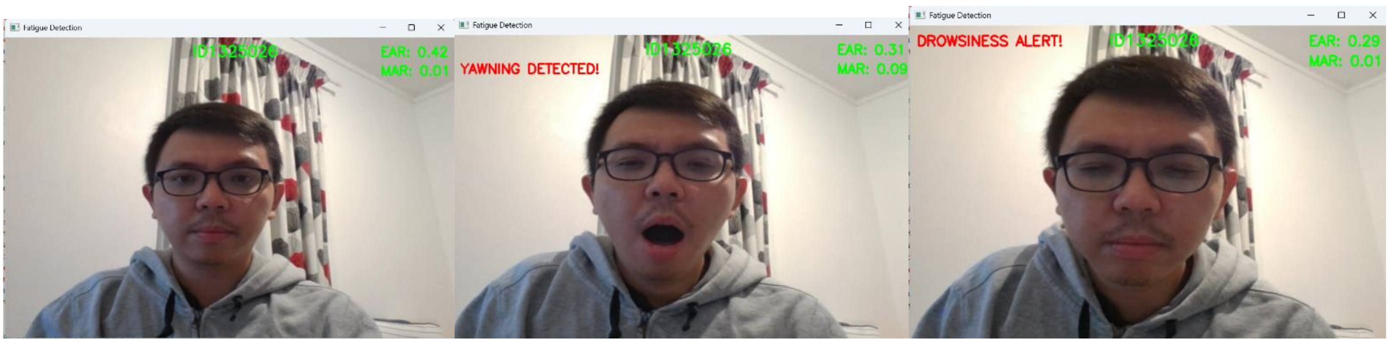
1. Objective

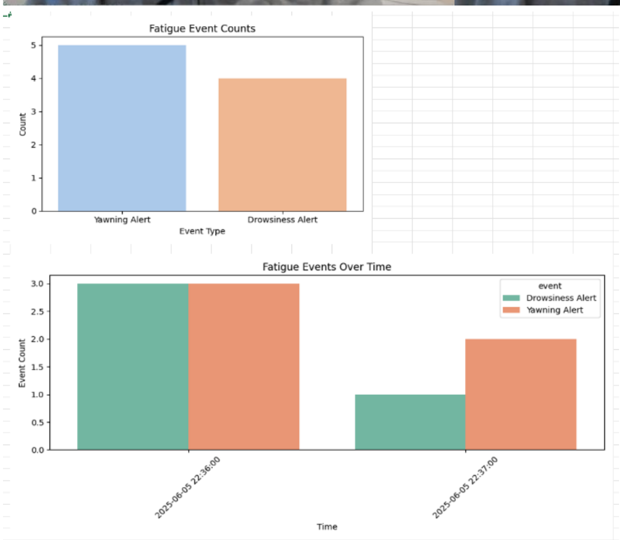
This report compares the performance of Experimenting with two different **Fatigue Detection using MAR, MAR** and **Fatigue Detection using yolov5** in identifying fatigue related behaviours using live feeds. The goal is to evaluate accuracy, and overall reliability to determine the more suitable tool for deployment in real-world scenarios.

2. Methodology

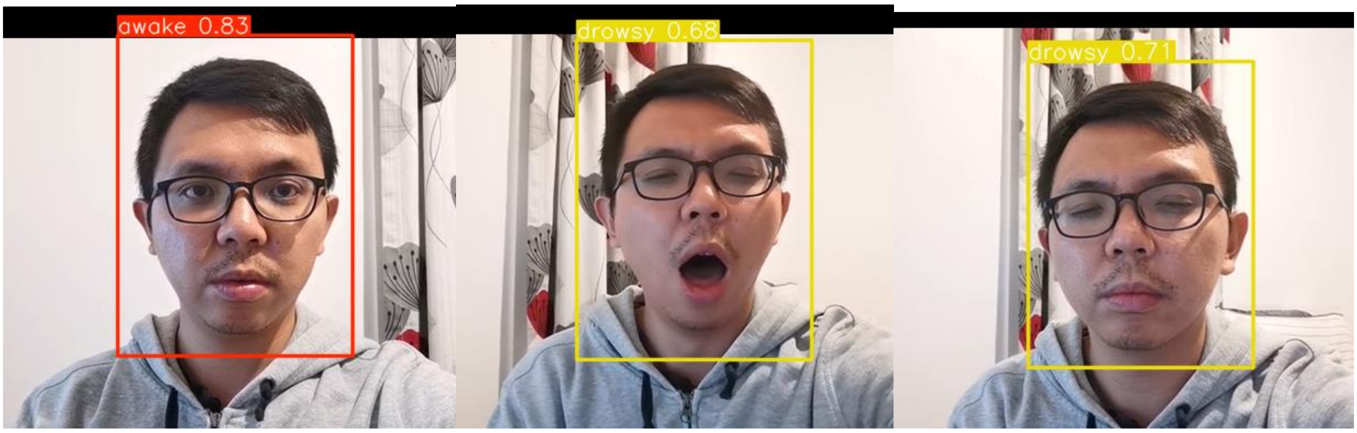
a. Tools compared

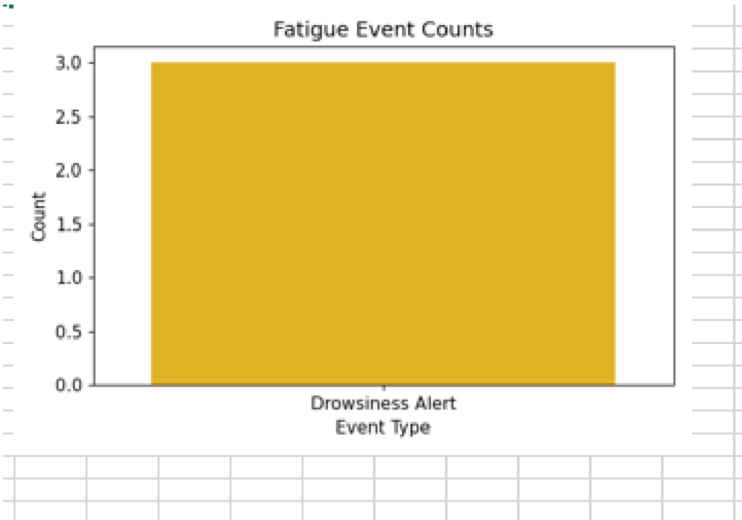
Fatigue Detection using MAR and EAR is a real-time fatigue detection system that uses a webcam to monitor a person's face and detect signs of drowsiness and yawning using facial landmarks. It employs MediaPipe Face Mesh to track eye and mouth movements and calculates EAR (Eye Aspect Ratio) and MAR (Mouth Aspect Ratio) to determine if the person is closing their eyes for too long or yawning frequently. DeepFace is used for face recognition, allowing the system to log fatigue events by individual. When thresholds are met, alerts are displayed on the screen, and the events are recorded with timestamps and names. The tool generates an Excel report (`fatigue_log.xlsx`) and a visual dashboard (`fatigue_dashboard.xlsx`) with a summary chart of fatigue events.

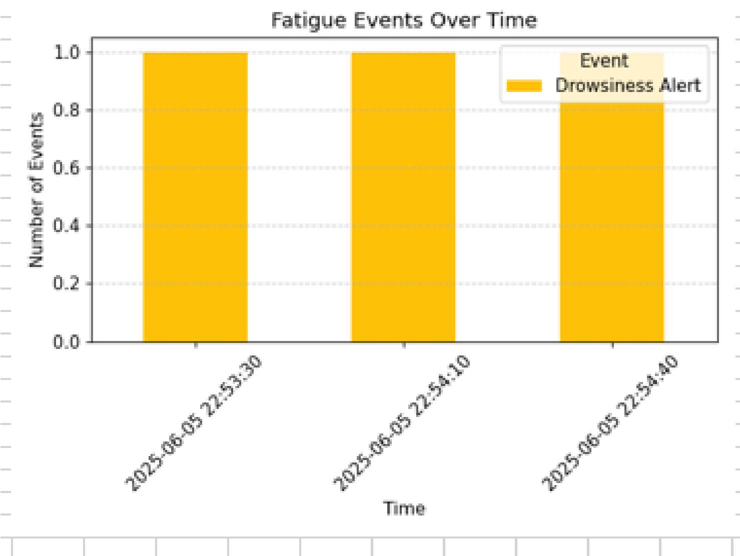




Fatigue Detection using YOLOv5 is a real-time fatigue detection system that uses object detection to identify and log prolonged drowsy behaviour in individuals. By analysing webcam frames, the system detects when a person is labelled as "drowsy" and only logs an event if the condition persists for a defined duration (e.g., 30 consecutive frames), effectively filtering out brief or accidental detections. It then visualizes the data through two types of plots: one showing the total number of fatigue events and another showing how those events are distributed over time in 10 second intervals. Both plots and the raw data are saved to a timestamped Excel report (`fatigue_dashboard_with_time_*.xlsx`) along with embedded charts.







b. Dataset

Dataset for **Fatigue Detection using MAR and EAR** is a self-recorded 3-5mins long video with different drowsy state. A ground truth log is created and manually labelled.

Dataset for **Fatigue Detection using YOLOv5** is a public dataset from Roboflow https://universe.roboflow.com/project/drosiness-detection/dataset/3 the dataset contains over 3000+ images of different race and people in drowsy and awake state.

c. Metrics Used

Accuracy:

- **Definition**: The proportion of total correct predictions out of all predictions.
- **Formula**: Accuracy = TP + TN / TP +TN + FP + FN

Precision:

- **Definition**: Of all the predicted positives, how many were actually correct?
- **Formula**: Precision = TP / TP + FP

Recall:

- **Definition**: Of all actual positives, how many were correctly predicted?
- **Formula**: Recall = TP / TP +FN

F1-score :

- **Definition**: The harmonic mean of precision and recall.
- **Formula**: F1 = 2 x (Precision x Recall) / (Precision + Recall)

Confusion matrix:

- **Definition**: A table showing the breakdown of predictions

3. Results

Fatigue Detection using MAR and EAR:

Training Set			
TARGET / OUTPUT	Class0	Class1	SUM
Class0	12 52.17%	3 13.04%	15 80.00% 20.00%
Class1	8 34.78%	0 0.00%	8 0.00% 100.00%
SUM	20 60.00% 40.00%	3 0.00% 100.00%	12 / 23 52.17% 47.83%

Class Name	Precision	1-Precision	Recall	False Negative Rate	F1 score
Class0	0.6000	0.4000	0.8000	0.2000	0.6857
Class1	0.0000	1.0000	0.0000	1.0000	0.0000

Fatigue Detection using YOLOv5:

Training Set			
TARGET / OUTPUT	Class0	Class1	SUM
Class0	98 49.00%	2 1.00%	100 98.00% 2.00%
Class1	2 1.00%	98 49.00%	100 98.00% 2.00%
SUM	100 98.00% 2.00%	100 98.00% 2.00%	196 / 200 98.00% 2.00%

Class Name	Precision	1-Precision	Recall	False Negative Rate	F1 score
Class0	0.9800	0.0200	0.9800	0.0200	0.9800
Class1	0.9800	0.0200	0.9800	0.0200	0.9800

4. Analysis

Fatigue Detection using MAR and EAR

- Moderate success in detecting **fatigue** cases (80% recall).
- Completely fails to detect **no_fatigue** cases (0% recall).
- Overall **accuracy is low** (52%) and indicates a strong **class imbalance**.
- Precision and recall disparity suggests **bias toward the fatigue class**.

Fatigue Detection using MAR and EAR demonstrates strong foundations in both functionality and extensibility. At its core, the tool integrates multiple modalities to monitor fatigue in real-time or from recorded videos. Specifically, it uses the Eye Aspect Ratio (EAR) to detect prolonged eye closure and the Mouth Aspect Ratio (MAR) to identify yawning, which are both common indicators of fatigue. These biometric signals are processed through MediaPipe's facial landmark detection, and identity recognition is handled using DeepFace, allowing for individual-level monitoring.

A key strength of the system lies in its logging feature is a major advantage it records detected events with associated timestamps and recognized identities, then saves them into an Excel report complete with charts. These visualizations, generated with seaborn and matplotlib, help users identify not just the quantity of fatigue events but also when they tend to occur.

However, there are some limitations that affect its robustness. One such challenge is the tool's sensitivity to tuning. The performance of the detection system is highly dependent on the specific threshold values you set for EAR, MAR, and the number of consecutive frames. These thresholds work differently under varying conditions, such as the position of the face and the camera. Small changes in lighting, camera angle, or user behaviours can lead to false positives or missed detections.

Additionally, while DeepFace is effective for face identification, it struggles under suboptimal conditions like side angles, poor lighting, or motion blur. This can lead to unreliable or missed identity matches. The system also lacks contextual awareness instances like talking or blinking can be misinterpreted as yawning or drowsiness, especially without further filtering or behavioural context.

An improvement to the fatigue detection tool is the alert system smarter by ignoring short, random blinks or yawns that don't mean someone is tired. Making face recognition more reliable in different lighting or when the person moves will help with consistent identity tracking. Allowing users to adjust sensitivity settings to better match their own facial features can also help make the system more accurate and user-friendly.

Fatigue Detection using YOLOv5

- Near-perfect performance across all metrics.
- Balanced and reliable detection of both **fatigue** and **no_fatigue**.
- High **precision, recall, and F1-score** confirm consistent and effective classification.

- **Robust model** suitable for real-time monitoring applications.

The YOLOv5-based fatigue detection tool demonstrates significant strengths, especially in terms of real-time performance and flexibility. Because YOLOv5 is a state-of-the-art object detection model known for its speed and accuracy, it enables quick detection of fatigue related states such as drowsiness by processing video frames almost instantly. This makes the tool highly suitable for live monitoring applications where latency matters, such as driver fatigue detection or workplace safety monitoring. Moreover, YOLOv5's capability to be fine-tuned on custom datasets allows the tool to be adapted specifically to the fatigue detection domain, making it more precise in recognizing relevant cues compared to generic fatigue detection systems.

However, the tool also carries some limitations. One major weakness is the reliance on the quality and quantity of the labelled training data. If the training dataset lacks diversity in terms of lighting conditions, angles, or subjects, the model may struggle to generalize well, leading to false positives or missed detections in real-world scenarios. Additionally, YOLOv5 detects objects based on visual features but does not inherently interpret physiological signals like eye movement dynamics or subtle micro expressions that might indicate fatigue, which more specialized models or sensors might capture. The system may also be sensitive to environmental factors such as low light, occlusions, or camera positioning, which can degrade detection reliability.

To improve the YOLOv5-based detection tool, one helpful step would be to train the model on a larger and more diverse set of videos, including different people, lighting conditions, camera angles and varied scenarios. This would help the tool perform better in real-life situations. Another improvement would be to add simple logic that avoids logging very short or accidental drowsy detections, so the system only records real, longer-lasting fatigue events. Finally, making the results easier to read and understand like showing clear alerts or summaries would make the tool more user-friendly and practical for everyday use.

5. Conclusion

Comparing the fatigue detection tools based using **MAR and EAR** with the fatigue detection using **YOLOv5** approach reveals distinct trade-offs in their design, performance, and applicability. The MAR and EAR method relies heavily on biometric signals derived from facial landmarks to detect eye closure and yawning, which are classical indicators of fatigue. While it demonstrates a reasonable ability to identify fatigue events, its performance is limited by sensitivity to environmental conditions, threshold tuning, and difficulties in accurately recognizing no-fatigue states. This leads to an imbalanced detection outcome with a tendency toward false positives and missed non-fatigue cases, resulting in lower overall accuracy and a less reliable system. Its strength lies in its modular, interpretable framework that can work in both live and recorded video modes and provides detailed logging with identity recognition, making it useful for individualized monitoring despite its sensitivity challenges.

On the other hand, the YOLOv5-based detection tool achieves near-perfect accuracy and balanced detection across both fatigue and no-fatigue classes, benefiting from the power of deep learning object detection to process visual cues rapidly and effectively. This approach offers robustness suitable for real-time applications where speed and responsiveness are critical, such as driver monitoring. However, its reliance on the quality and diversity of training data means its generalizability could suffer if exposed to unrepresented real-world conditions. Moreover, unlike the MAR and EAR system, YOLOv5 operates primarily on visual features without directly interpreting physiological signals, which may limit its ability to detect subtle fatigue indicators.

In conclusion, while the MAR and EAR-based tool provides explainable fatigue indicators and user specific tracking with flexibility across input types, it struggles with sensitivity and class imbalance that impacts reliability. The YOLOv5 tool delivers superior accuracy and real-time performance, making it more practical for immediate deployment, but it depends heavily on comprehensive data and lacks direct physiological insight. A combined or hybrid approach leveraging the interpretability of MAR and EAR with the speed and robustness of YOLOv5, along with improvements in data diversity and event filtering, could offer a more balanced and effective solution for fatigue detection in varied real-world settings.