Estimating Blood Alcohol Level Through Facial Features for Driver Impairment Assessment

Ensiyeh Keshtkaran[1], Brodie von Berg[2], Grant Regan[2], David Suter[1], Syed Zulqarnain Gilani[1,3]
[1]Centre for AI&ML, Edith Cowan University, Western Australia [2]MiX Telematics, Western Australia
[3]Nutrition and Health Innovation Research Institute, ECU, Western Australia
e.keshtkaran@ecu.edu.au

Abstract

Drunk driving-related road accidents contribute significantly to the global burden of road injuries. Addressing alcohol-related harm, particularly during safety-critical activities like driving, requires real-time monitoring of an individual's blood alcohol concentration (BAC). We devise an in-vehicle machine learning system that harnesses standard commercial RGB cameras to predict critical levels of BAC. Our system can detect instances of alcohol intoxication impairment as subtle as 0.05 g/dL (WHO recommended legal limit for driving), with an accuracy of 75%, by leveraging the physiological manifestations of alcohol intoxication on a driver's face. This system holds great promise for improving road safety. In tandem, we have compiled a data set of 60 subjects engaged in simulated driving scenarios, spanning three levels of alcohol intoxication. These scenarios were captured and divided into video segments labeled "sober","low', and "severe" Alcohol Intoxication Impairment (AII), constituting the basis for evaluating our system's performance. To the best of our knowledge, this study is the first to create a large-scale real-life dataset of alcohol intoxication and assess intoxication levels using an off-the-shelf RGB camera to detect drunk driving.

1. Introduction

Automatic identification of drunk driving presents a critical challenge in improving road safety. Extensive records indicate a strong link between increased alcohol consumption and compromised driving performance [5,21], resulting in an increased risk of fatal traffic accidents as Blood Alcohol Concentration (BAC) levels increase. The gravity of this issue is underscored by statistical evidence, which highlights the substantial contribution of alcohol-related crashes to road fatalities [24, 36, 38], unequivocally establishing intoxicated driving as the main cause of road-related fatalities. However, existing approaches to detect alcohol-impaired driving, primarily relying on random breath tests, do not adequately address this pressing problem.

Although efforts are underway to integrate driver alcohol detection systems into future vehicle generations [30], and the advent of autonomous cars is on the horizon [10], the persistent issue of drunk driving remains an urgent concern. Most of the research in the realm of detecting intoxicated driving predominantly centers around analyzing driving behavior such as driving and steering patterns, pedal usage, and vehicle speed [8, 11, 18, 19, 25, 27, 28, 39]. Some other approaches incorporate external sensors like alcohol detection or touch-based sensors [8, 44], however, there has been very limited exploration into the potential of leveraging computer vision techniques to identify signs of intoxication based on biobehavioral changes of drivers [23, 25].

One key limitation of using driving behavior to detect drunk driving is the requirement for the driver to be actively operating the vehicle for a considerable duration before their behavior can be assessed and identified as indicative of intoxication. This implies that a potentially impaired driver is already on the road, posing risks to themselves and other road users. Swift detection is crucial in identifying drunk drivers and preventing them from endangering public safety. In contrast, besides the need for regular calibration and maintenance in sensor-based approaches, it is important to note that sensor-based and behavior-based methods are confined to vehicles equipped with specific technologies, rendering them less practical for widespread adoption. A computer vision-based approach could potentially be integrated into road cameras in the future, akin to how these cameras currently detect seatbelt usage or mobile phone activity, making it applicable to various vehicle types without requiring specialized in-cabin installations.

In contrast, the few studies that have explored the use of computer vision techniques to detect intoxication based on drivers' biobehavioral shifts tend to heavily rely on specialized camera-based sensors such as eye-tracking systems and in-vehicle monitoring setups [23] or a combination of camera-based and vehicle-based sensors [25]. This reliance limits the applicability of these methods to only those vehicles that are already equipped with this technology.

In addition, the insufficiency of suitable datasets hampers progress in this field. A recently collected dataset

by [23] features data from 30 participants at the highest intoxication level of 0.05 g/dL. However, to gain greater relevance to real-world scenarios and account for individual variations in alcohol response, a larger sample size and a wider range of intoxication levels encountered in social contexts are essential. Furthermore, the findings from such a dataset cannot be readily applied to countries with higher legal alcohol limits for driving under the influence, such as the USA with a limit of 0.08 g/dL. On the other hand, the dataset curated by [25] comprises data from more than 100 participants with intoxication levels that exceed 0.08 g/dL. Nevertheless, this data collection involves sensor-based cameras and advanced infrared eye tracking systems, restricting its applicability to vehicles equipped with these technologies.

Our contributions manifest themselves in two ways. Firstly, we compile a dataset containing simulated driving videos involving 60 participants spanning three distinct levels of alcohol intoxication: sober (0.00 g/dL), low intoxicated (0.05 to 0.07 g/dL), and severe intoxicated (above 0.08 g/dL). Secondly, we present a machine-learning system that utilizes discernible cues from standard RGB videos of drivers' faces to gauge the degree of alcohol-related impairment. To the best of our knowledge, our system is the first to employ a standard RGB camera for detecting alcohol intoxication levels based on signs of impairment in drivers' faces. Additionally, our dataset is the first to include RGB video recordings of individuals at various alcohol intoxication levels while driving (simulated). Although deliberately not exploited in our baseline system, the dataset also contains 3D and infrared videos of the driver's face, rear-view RGB videos showing driver posture and steering interactions, driving simulation event logs, and screen recordings of driving behavior. The availability of this dataset not only enriches our research endeavors, but also provides the broader scientific community an invaluable resource for further exploration and study.

2. Data collection

We developed and assessed a machine learning (ML) framework designed to detect BAC thresholds in drivers. To accomplish this, data was collected by MiX Telematics, from alcohol-impaired drivers within a controlled, yet realistic, environment. The analysis of this data received approval from the Edith Cowan University Ethics Committee (ID: 2021-02805-KESHTKARAN). The data collection occurred between September 2022 and July 2023. Details about the data collection methodology, including participant demographics, simulator configuration, data collection procedures, and alcohol administration protocols, are outlined in the following sections.

	Category	n
Gender	Male	36
	Female	24
Age range	19-35 years old	20
	36-50 years old	25
	51 and above	15
Ethnicity	Caucasian	54
	Asian	5
	Other	1
Drinking habit	Moderate	37
	Low	21
	Heavy	2
Driving skills	Experienced	39
	Confident	21

Table 1. Summary of the participant's demographics, driving experience, and drinking habits.

2.1. Sample size and participants

The study involved 60 healthy men and women aged 19 to 76 (mean age: 43 $\pm$ 14.6), who participated in simulated driving scenarios on an urban roadway while experiencing three distinct BAC levels: 0.00 g/100ml (sober condition), 0.05 to 0.07 g/100ml (low alcohol intoxication) and above 0.08 g/100ml (severe alcohol intoxication). These thresholds were selected to represent varying levels of alcohol impairment commonly observed in real-life situations [21]. The choice was also informed by the World Health Organization's advised legal limit of 0.05 g/100ml in Australia (and more than 90 other countries), along with the prevalence of fatal accidents occurring at levels of BAC that exceed this threshold [5, 24]. The recruitment included a diverse participant population, with respect to age, gender, and drinking habits. Inclusion criteria for participation were as follows: (a) Possession of a valid driver's license and having driving experience for at least two years. (b) Absence of medications and/or health conditions incompatible with alcohol consumption. (c) Not being pregnant or lactating. (d) No experience of motion sickness when driving on the simulator. Detailed participant characteristics are reported in Tab. 1.[1]

2.2. Simulator setup

We developed our research simulator by integrating a research-grade simulator software, provided by Carnetsoft (BV, Groningen, The Netherlands), with a commercially available driving monitoring system by MiX Telematics (Perth, Western Australia). Carnetsoft's simulator software is well-recognized in medical research projects [26, 34], and the selected driver monitoring system is widely used glob-

[1] Data derived from participants' self-reported information.

(a) (b)

Figure 1. (a) Integration of the driving simulator with breathalyser and driver ID tag reader. (b) Top: simulation scenario and driving performance screen recording. middle right: 3D camera footage. middle left: driver's posture recording from the rear view. bottom right: infrared camera footage. bottom left: RGB camera footage.

ally for driver safety monitoring. The configured research simulator (depicted in Fig. 1) includes essential components: a panoramic 270-degree field of view through three monitors, an adjustable seat for optimal driving posture, a realistic force feedback steering wheel, gear shifting mechanism, indicator stalk, accelerator, and brake pedals. Moreover, the simulator is equipped with a two-stage ignition and engine start process that replicates genuine vehicle activation. We have also integrated an Autowatch 720 Tethered Alcohol Breathalyser module and driver ID tag reader. The simulator's integrated breathalyser uses a blood ratio constant of 2100 to convert alcohol values from mg/l to BAC, complying with Australian national law [21] as follows:

$$BAC = \text{mg/l} \times \left(\frac{\text{Blood Ratio}}{10000} \right) \quad (1)$$

This means that an alcohol concentration of 0.238 mg/l corresponds to a BAC of 0.05 g/100ml. Our simulator also features an RFID card reader for driver authentication, linking trip data to driver IDs and timestamps. This integration optimizes data registration, ensuring anonymous and error-free data recording.

2.3. Recruitment and data collection process

In response to the data collection advertisement, volunteers were contacted through phone and email communication. Comprehensive information outlining the data collection procedure and participant requirements was shared with potential participants. Their inquiries and concerns were addressed, and dates were scheduled for their visit to the data collection site. Participants were advised to avoid alcohol, caffeine, and nicotine prior to sessions. They were also advised to eat lightly within 3 to 2 hours before their appointment. Upon arrival, participants were required to complete a questionnaire to confirm their adherence to the stipulated instructions. Prior to data collection, each participant provided their informed consent. Unique RFID cards were issued to participants to verify their identity during the data collection session. These cards served for both completing the questionnaire and consenting through a mobile app. Following this, participants were instructed to tag their card for identification right after ignition activation, just before giving a breath sample and beginning the trip.

Data collection commenced with a practice drive to assess motion sickness and the comfort of the participants with the simulator. For those who did not feel motion sickness and were happy to continue, this was followed by sober driving tasks and two sessions involving controlled alcohol intake. The research simulator collected a diverse set of information, as depicted in Fig. 1. This dataset includes driver's ID, timestamps of trips, records of Blood Alcohol Concentration (BAC), facial video footage in full-color (RGB), infrared (IR), and 3D formats, as well as the driver's posture, head position, and hand placement from the rear-view (see supplementary material - section A for more details). Additionally, the driving behavior of the driver was captured through recordings of the simulator screen. Instances of hazardous driver maneuvers along with their associated alerts were also documented.

To ensure the accuracy of the Breathalyzer measurement and minimize the impact of ambient alcohol, the alcohol administration occurred in a separate room from the data collection setting. At the start of each driving task, breath samples were collected using the AlcoQuant 6020 breathalyzer (Envitec GmbH, Wismar, Germany) outside the alcohol administration room. Subsequently, within the simulator chamber, the Autowatch 720 Tethered Alcohol Breathalyzer (PFK Electronics, South Africa) was employed for confirmation and validation of readings. These breathalyzers were calibrated according to the standardized Australian Standard AS3547 and are akin to those utilized by law enforcement agencies in Australia, ensuring the reliability of collected data. At each level of alcohol intoxication, participants engaged in a 10-minute drive using the driving simulator. The simulated scenario involved navigating through a city environment with moderate to high traffic, including tasks such as yielding to pedestrians, obeying traffic lights, and responding to unexpected events like animals entering the road. Participants followed onboard navigation and adhered to Australian traffic regulations.

2.4. Alcohol administration

After completing a sober run, to attain the initial intoxication target level, participants were asked to start with consuming two standard drinks (Australian measure). To mitigate the possible residual alcohol in the oral cavity that could influence breath analysis, participants rinsed their mouths with water after consuming the drinks. Breath sampling began 7 to 10 minutes after drink consumption. Additional drinks were provided based on measured alcohol

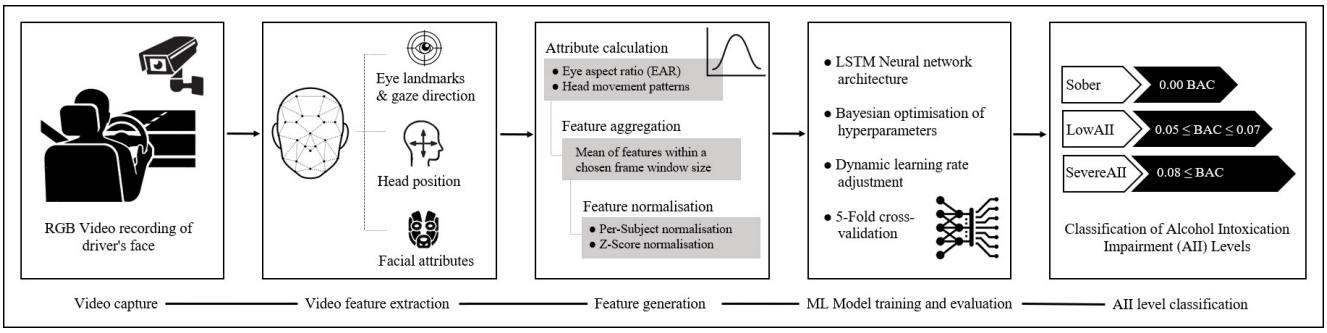


Figure 2. A diagram of our proposed system workflow.

concentration. Following each subsequent drink, a 7 to 10-minute waiting period and mouth rinse were observed before the next breath test, following the Australian Government standard drinks guidelines [22]. This process was repeated until target intoxication level was achieved. After reaching the second target BAC level ($0.05 \leq BAC \leq 0.07$ g/100ml), participants engaged in a 10-minute driving session. Following this, they returned to the alcohol administration room, consumed more beverages, and underwent the same breath sampling procedure, until the final BAC target ($BAC \geq 0.08$ g / 100 ml) was reached. Following this, the last 10-minute driving session was conducted (see supplementary material - section B for more details). Throughout the alcohol administration, participants were offered light refreshments and non-caffeinated beverages. After the conclusion of the final driving session, participants received a light meal and were encouraged to consume food and water to reduce their levels of intoxication. BAC levels were measured intermittently to ensure participant safety before leaving the data collection site. The selection of types of alcohol in our study was deliberately diverse. However, during the early stages of intoxication, we discouraged the consumption of beer due to its lower alcohol content. This approach was designed to mirror real-world scenarios, allowing participant-driven choices and avoiding coercion into unfamiliar substances.

The next section explains how we used the collected data to train and evaluate our machine learning system to estimate alcohol intoxication using facial features. While our dataset includes data from multiple sources, this study aims to detect impaired drivers using standard RGB videos, without utilizing additional data sources.

3. Machine learning framework

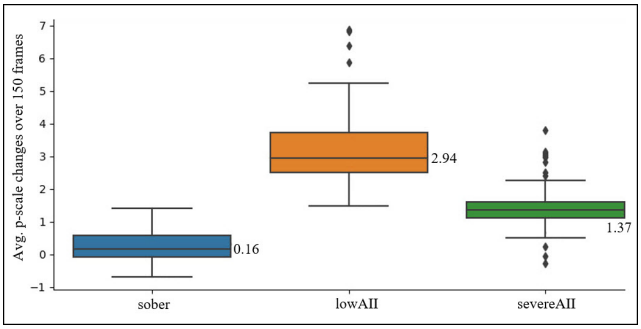
We propose a machine learning framework that unfolds as follows: Initially, the RGB camera captures video footage from the driver's face and head position. Following this, a feature generation step is performed to construct the essential attributes for training our machine learning model. Finally, our trained model classifies the data into three labels based on a range of BAC levels. The system workflow is shown in Fig. 2 and a detailed breakdown of each step is provided in the following sessions.

3.1. Feature generation

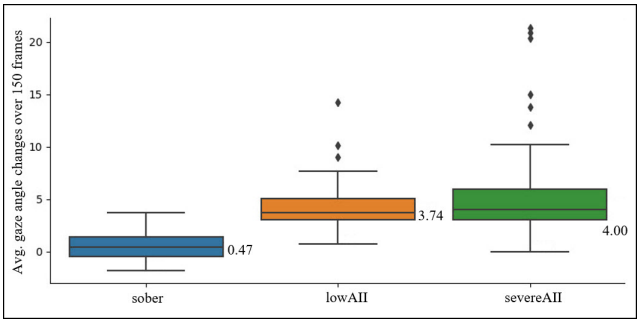
To generate features to train our model, we start by identifying attributes that have the capacity to reflect pathophysiological alterations resulting from alcohol intoxication, as documented in the existing literature [2, 6, 31, 35, 42]. Subsequently, we extracted the features from our RGB videos and visualized them across varying degrees of intoxication. The final features selected based on their strong alignment with the documented pathophysiological changes associated with alcohol intoxication, as well as their ability to exhibit significant variations across different levels of intoxication. The four selected features are (1) head movement, (2) gaze movement, (3) eye aspect ratio, and (4) scale factor of the face PDM (Parameters of a Point Distribution Model) [20]. This procedure is detailed below.

(a) Feature extraction Previous studies show that alcohol predominantly influences physiological alterations associated with ocular characteristics, gaze patterns, and facial expressions [2, 6, 31, 35, 42]. In light of this understanding, we extracted facial activity biometrics and landmarks from RGB videos of individuals across varying levels of intoxication, using the OpenFace software library [3]. The extracted features encompassed eye landmarks, eye gaze directions, PDM (Parameters of a Point Distribution Model) parameters, which quantifies the extent of face landmark shape variations [20], as well as facial Action Units (AUs) [13] and head pose. These feature values were extracted for each frame from RGB videos. The entire video length was utilized for this process.

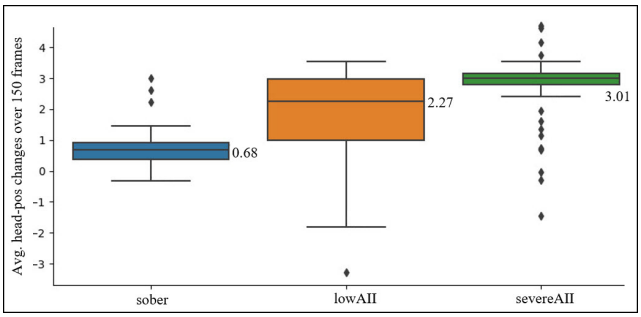
(b) Feature pre-processing and visualization We first calculate the aspect ratio of the eyes (EAR) using the extracted landmarks of the eyes for every frame, as defined in [7]. Subsequently, we project the three-dimensional



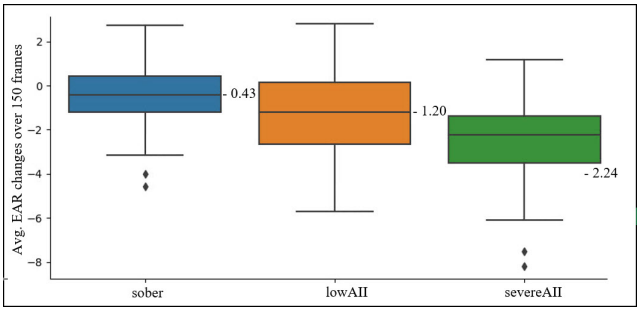
(a) Average p-scale variations.



(b) Average gaze-angle variations.



(c) Average head-pos variations.



(d) Average EAR variations.

Figure 3. Visualization of selected features across three levels of alcohol intoxication impairment.

spatial representations of head movement into a two-dimensional polar space. Then to capture meaningful variations in the selected features, we compute the average feature values over a window of frames. This approach was chosen because alcohol impairment detection requires the observation of behavioral patterns over time, such as gradual shifts in patterns of gaze and head movement. We examine the average values within *different* window sizes of consecutive video frames. The selected window sizes were 15, 30, 100, 150 and 200 frames. This testing of different window sizes allows us to encapsulate the temporal characteristics of these features within specific time frames, ranging from a few seconds (30 frames) to several seconds (about 7 seconds for 200 frames). We then employ t-test analyses, to compare the changes in the calculated averages across different levels of intoxication. Finally, the features that exhibit the most significant variations across various levels of intoxication are visualized for each participant and collectively for the entire group.

Illustrated in Fig. 3 and aligned with prior research [6, 42], our derived features reveal physiological alterations caused by alcohol intoxication. In particular, distinct attributes such as gaze movement, p-scale, head movement, and EAR exhibit noteworthy deviations in behavior across varying levels of intoxication. These differences are most pronounced when considering a window size of 150 frames, indicating its efficacy in capturing significant changes over a meaningful timeframe.

Fig. 3d shows that as the degree of alcohol intoxication increases, a corresponding reduction in the EAR can be observed, accompanied by heightened variability in head positioning (Fig. 3c) and gaze angle (Fig. 3b). Furthermore, a rise in p-scale variation is evident at a lower level of intoxication (Fig. 3a). This finding is consistent with established literature, suggesting that the primary indication of alcohol-induced intoxication in an impaired individual includes heightened emotional expressions (reflected in increased head movement and p-scale) and a tendency for their eyes to appear less open, conveying a perception of weariness due to intoxication (linked to reduced EAR) [6, 42].

(c) Feature normalization Assessing alcohol intoxication levels using visual biometric facial features presents a significant challenge due to the diverse ways alcohol impacts different individuals. Various factors such as drinking habit, age, gender, and ethnicity have been identified as contributors to this variability in research studies [12, 14, 29, 33, 41]. We first normalize the features across subjects to jointly train our model on this diverse data. To do this, we consider an initial one-third of features obtained from the sober state as a reference. For each individual, we calculat the mean ($\mu_{n,m}$) and standard deviation ($\sigma_{n,m}$) of each feature (n) within this sober state subset. We then use equation 2 to normalize the remaining features of the sober state, as well as those of the low AII and severe AII stages, all within the same individual:

$$\hat{F}_{n,m} = \frac{(F_{n,m} - \mu_{n,m})}{\sigma_{n,m}} \quad (2)$$

We apply this normalization to both the training and test data across all subjects and features. Note that the subset of features acquired from the sober state, which were used for calculating mean and standard deviation, are reserved for the normalization process. These features are excluded from the dataset employed for training and evaluation of the model. Moreover, we use Z-score normalization to center each feature with respect to the training data. Z-score normalization transforms the values of each feature so that they have a mean of 0 and a standard deviation of 1. This normalization method ensures that no single feature unduly influences the learning process due to its magnitude.

3.2. Model design and model training

We build and tune a Long Short-Term Memory (LSTM) neural network [37] with Keras tuner from TensorFlow [1]. LSTM was chosen for its capability to process sequential data, considering the temporal interdependencies in our features. For the tuner, we employ Bayesian optimization [15] to find the combination of hyperparameters that maximize the validation accuracy of our model. Then we train our LSTM model using the best hyperparameters found by the tuner along with the Sparse Categorical Cross-Entropy loss function. We randomly split our dataset of 60 subjects into 5 equal folds for training. During training, one fold was set aside for testing, and the remaining 4 folds were utilized for training and validation. This process was repeated until each fold served as a test set, and the final result was the average of test outcomes across all 5 folds. Training was carried out in 100 epochs, and the ReduceLROnPlateau learning rate scheduler was used to adjust the learning rate during training. The optimal learning rate was determined to be 6.9053×10^{-4}. The best hyperparameters obtained from the Bayesian optimizer resulted in a model with two LSTM layers: the first with 256 units and the second with 128 units, with a dropout rate of 0.1 in the second layer. Additionally, the model has two batch normalization layers, a dense layer with ReLU activation function [17], and L2 regularization [9] with a coefficient of 0.001. Python 3.9.16 [40], and TensorFlow 2.13.0 were used for experiments.

To evaluate the model performance, we employ a cross-validation approach and utilizes classification metrics including accuracy, precision, recall, and F1 score. Additionally, we present a confusion matrix to better understand how the model handles different classes. Evaluation details are provided in the following section.

4. Results and evaluation

We now present the results of our machine-learning framework for detecting BAC levels in drivers. Initially, we discuss its performance across three distinct classification assignments: sober, low, and severe AII. Next, we analyze the robustness of our machine learning model by studying how well it can generalize when subjected to different permutations of training and test sets. Finally, we assess the significance of each feature towards the model's predictive performance by scrutinizing the classification accuracy in the absence of each feature.

	Precision	Recall	$F1$-Score
Sober	0.79	0.85	0.82
LowAII	0.71	0.70	0.71
SevereAII	0.73	0.71	0.72
Accuracy			0.75
Macro Avg	0.75	0.75	0.75
Weighted Avg	0.74	0.74	0.74

Table 2. Evaluating our model's performance via averaged classification report

4.1. Performance evaluation

To evaluate the performance of our classifier, we utilize precision, recall, and F1-score evaluation metrics. Initially, individual evaluation metrics are computed for each class, and subsequently, the average of these class-specific metrics is determined. This technique, known as Macro-averaging [16], treats each class uniformly, regardless of its frequency or imbalance in the dataset. The macro-average for each evaluation metric (precision, recall, and F1-score) is calculated as below:

$$\text{MAvg}_m = \frac{m_{class_1} + m_{class_2} + \ldots + m_{class_n}}{N} \quad (3)$$

where *m* is the type of evaluation metric (precision, recall, or F1-score) and *N* is a number of classes.

Furthermore, we assess the model performance with regards to class distributions by computing a weighted average [16], given the under-representation of sober instances within our dataset. As outlined in section 3.1, one-third of each sober state is allocated to feature normalization per subject. Subsequently, this segment is excluded from both the training and testing phases. The Weighted Average approach takes into account the distribution of classes. It determines the average precision, recall, and F1-score, with the contribution of each class being weighted by the proportion of samples belonging to that specific class. The calculation for the weighted average is as follows:

$$\text{WAvg}_m = \frac{\text{m}_{class_1} \times \omega_{class_1} + \ldots + \text{m}_{class_n} \times \omega_{class_n}}{T}$$

where *m* is the type of evaluation metric (precision, recall, or F1-score). ω_{class_n} is the proportion of samples belonging to class n, and T is the total number of samples. These metrics are computed using the "classification report function", from the scikit-learn library [32].

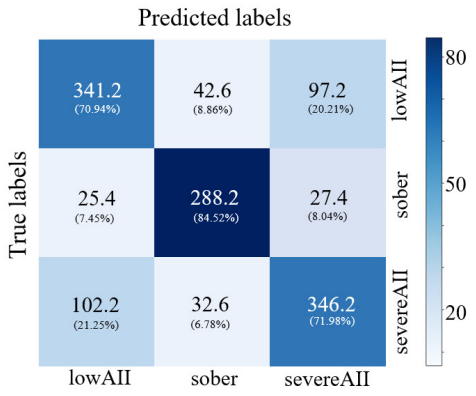


Figure 4. Confusion matrix showing predicted class distribution across true classes.

The classification performance for each individual class is shown in Tab. 2. Our model demonstrates a reasonable ability to classify alcohol intoxication levels, with notable strengths in the "sober" classes. Its performance in identifying "low AII" and "severe AII" is slightly lower, as evidenced by lower recall values. The balanced F1-scores and consistent macro and weighted averages suggest a generally stable and balanced performance across the three classes. The accuracy of 75% $\pm$0.01 exhibits a reasonable performance in categorizing alcohol intoxication levels solely based on observable manifestations of intoxication that can be captured through conventional RGB cameras.

We evaluate the efficacy of the model using a confusion matrix (Fig 4). Our model has a significant true positive (TP) rate of 84.52% for correctly identifying instances of the "sober" driving state. Similarly, the model achieves TP rates of 70.94% and 71.98% for the "low AII" and "severe AII" states, respectively. However, it is worth highlighting that there exists a moderate $21\% \pm 0.5$ degree of confusion between the "low AII" and "severe AII" states.

4.2. Robustness check

We execute a cross-validation strategy through a Stratified KFold cross-validation approach from the scikit-learn library to assess the robustness of the model performance [32]. Given that our dataset classes signify varying degrees of alcohol intoxication impairments (AII), Stratified K-Fold cross-validation is selected to ensure that the distribution of classes in the folds closely resembles the distribution in the overall dataset. The average accuracy of 75% suggests that, on average, the model predictions are accurate for around 75% of instances across the cross-validation folds. The small standard deviation of 0.01 indicates that the accuracy did not fluctuate significantly between folds, highlighting consistent performance. Moreover, we examine how consistent and dependable our model results are when exposed to different subsets of the data during the training and evaluation phases. This process involves executing the model multiple times, each instance initiated with a different random seed. The variability and uniformity of the results produced by the model across different instances of dataset partitioning is presented in Tab. 3. This analysis confirms that the model generalises successfully.

Instance	Seed	Accuracy (mean$\pm$SD)
1	42	0.74$\pm$0.01
2	132	0.74$\pm$0.00
3	random I	0.75$\pm$0.01
4	random II	0.74$\pm$0.00

Table 3. Model robustness check when exposed to different data subsets.

4.3. Analysis of feature significance

To understand the significance of each feature on the predictive performance of our model, we conduct a feature importance analysis. Through this analysis, we assess the impact of excluding each distinct feature from the dataset on the accuracy of the model classification. The outcome is summarized in Tab. 4. In particular, features are ranked in order of significance, with "p-scale", "gaze angle", "EAR" having pronounced effects on the model accuracy, while "head pose" exhibits minimal impact, suggesting its lesser influence.

5. Discussion

5.1. Contribution

Alcohol-related traffic safety concerns remain a pressing global issue, with combating drunk driving and alcohol intoxication road accidents posing significant challenges worldwide [24, 36, 38]. There exists an urgent need for scalable and cost-effective technologies that can efficiently and effectively detect alcohol intoxication impairment in drivers. This necessitates a technique that is not only low-cost but also adaptable to existing vehicles without extensive modifications to current technology. We have developed a machine learning system that harnesses visually observable cues of alcohol intoxication impairment from commercial RGB videos of drivers' faces, without relying on additional advanced camera-based sensor technologies. Our system detects varying levels of alcohol intoxication impairment, with an overall accuracy of 0.75% for the three-level classification (sober, low, and severe alcohol intoxication impairment). This not only benefits vehicles equipped with driver monitoring systems and eye-tracking technologies but also has the potential to extend to smartphones, making alcohol intoxication detection more practical, accessible, and widely applicable compared to other costly technologies such as breath-based sensors or sensor-based camera monitoring systems, which demand regular maintenance [8, 23, 25, 44]. It is also important to highlight

Omitted feature	Total accuracy	Accuracy per class	
Gaze angle	0.69 ± 0.00	0.79	Sober
		0.64	LowAII
		0.68	SevereAII
P-scale	0.61 ± 0.00	0.75	Sober
		0.55	LowAII
		0.57	SevereAII
Head pose	0.72 ± 0.01	0.84	Sober
		0.66	LowAII
		0.70	SevereAII
EAR	0.70 ± 0.00	0.84	Sober
		0.66	LowAII
		0.66	SevereAII
None	0.75 ± 0.01	0.85	Sober
		0.71	LowAII
		0.72	SevereAII

Table 4. Significance of selected features on model's predictive performance.

that our system has the capability to identify intoxication levels at the beginning of a drive, allowing for the potential prevention of impaired drivers from being on the road. This sets it apart from methods reliant on observable driving behaviors [8, 11, 18, 19, 25, 27, 28, 39], which require extended active vehicle operation to identify impairment (Additional information is available in supplementary section C). Additionally, we have compiled a dataset encompassing over 30 hours of driving videos across three intoxication levels: sober (0.00 BAC), low (above WHO's$\geq$0.05 and$\leq$0.07 BAC), and severe alcohol intoxication ($\geq$0.08 BAC). The dataset comprises RGB videos of the driver's face. Although deliberately not exploited in our baseline system, the dataset also contains IR, and 3D facial videos, rear-view RGB video capturing posture and steering interactions, driving event logs, and screen recordings. This dataset not only advances our research, but also serves as a valuable resource for the wider scientific community.

5.2. Limitations and future work

The limitations of this study include the impracticality of collecting data on individuals driving under the influence of alcohol due to legal restrictions. However, existing literature suggests that driver responses in simulators closely resemble real driving experiences and simulators can effectively replicate changes in driver behavior caused by alcohol consumption [4, 25, 43]. However, a constraint of simulator-based research is that participants can gradually improve their performance as they use the simulator. To mitigate this, we introduced an initial practice driving session aimed at acquainting participants with the simulator's interface.

Another limitation pertains to the normalization process, where we normalize the data for each individual using their corresponding *sober-state* data. This approach fits well in certain scenarios, such as when the system is installed in family vehicles or vehicles owned by organizations or companies with known drivers. In such cases, a calibration phase can be executed to establish a baseline representing the sober condition of the designated driver. Nevertheless, the applicability of our system encounters challenges when extended to more extensive use cases, such as rental car services, where the feasibility and assurance of a successful calibration process might be uncertain. One potential avenue for further investigation involves normalizing data using the average sober state across the entire population, rather than individualized states. However, determining the amount of population data required for this normalization and assessing its effectiveness when normalizing against a larger dataset are aspects that need to be investigated in future research. Furthermore, another constraint is our reliance on manually designed features. Transitioning from manually crafted features to an end-to-end learning approach holds significant potential and will be investigated in detail in our upcoming research endeavors (see supplementary section D for more information).

6. Conclusion

To the best of our knowledge, our system is the first to employ a standard RGB camera for detecting alcohol intoxication levels based on signs of impairment in drivers' faces. This system offers a simple yet effective approach for identifying intoxicated drivers which aligns with global need for cost-effective and seamlessly integratable system for widespread adoption to enhanced road safety and reduced incidents of alcohol-related accidents. While endeavors are being made to incorporate driver alcohol detection systems into upcoming generations of vehicles [44], and the advent of autonomous cars is on the horizon [10], the persistent issue of drunk driving remains an urgent concern. The tragic toll of fatal accidents and the resulting trauma for affected families necessitates immediate attention and action. This is precisely where our system plays a crucial role.

7. Data Availability

The data used in this paper will be made available on request, provided the use is for non-profit and research. A transfer agreement will be required and data may not be available before 2025.

8. Acknowledgment

The data collection process for this research was conducted and funded by MiX Telematics Australasia. SZG also wishes to acknowledge Raine Medical Research Foundation for partial salary support.

References

[1] Martín Abadi, Ashish Agarwal, Paul Barham, Eugene Brevdo, Zhifeng Chen, and et al. TensorFlow: Large-scale machine learning on heterogeneous systems, 2015. Software available from tensorflow.org. 6

[2] Anthony J Adams, Brian Brown, Gunilla Haegerstrom-Portnoy, Merton C Flom, and Reese T Jones. Marijuana, alcohol, and combined drug effects on the time course of glare recovery. *Psychopharmacology*, 56:81–86, 1978. 4

[3] Tadas Baltrusaitis, Amir Zadeh, Yao Chong Lim, and Louis-Philippe Morency. Openface 2.0: Facial behavior analysis toolkit. In *2018 13th IEEE international conference on automatic face & gesture recognition (FG 2018)*, pages 59–66. IEEE, 2018. 4

[4] Francesco Bella. Validation of a driving simulator for work zone design. *Transportation Research Record*, 1937(1):136–144, 2005. 8

[5] Robert F Borkenstein, Richard F Crowther, and RP Shumate. The role of the drinking driver in traffic accidents (the grand rapids study). *Blutalkohol*, 11(Suppl):1–131, 1974. 1, 2

[6] Eva Susanne Capito, Stefan Lautenbacher, and Claudia Horn-Hofmann. Acute alcohol effects on facial expressions of emotions in social drinkers: a systematic review. *Psychology research and behavior management*, pages 369–385, 2017. 4, 5

[7] Jan Cech and Tereza Soukupova. Real-time eye blink detection using facial landmarks. *Cent. Mach. Perception, Dep. Cybern. Fac. Electr. Eng. Czech Tech. Univ. Prague*, pages 1–8, 2016. 4

[8] Huiqin Chen and Lei Chen. Support vector machine classification of drunk driving behaviour. *International journal of environmental research and public health*, 14(1):108, 2017. 1, 7, 8, 13

[9] Corinna Cortes, Mehryar Mohri, and Afshin Rostamizadeh. L2 regularization for learning kernels, 2012. 6

[10] Federico Cugurullo, Ransford A Acheampong, Maxime Gueriau, and Ivana Dusparic. The transition to autonomous cars, the redesign of cities and the future of urban sustainability. *Urban Geography*, 42(6):833–859, 2021. 1, 8

[11] Jiangpeng Dai, Jin Teng, Xiaole Bai, Zhaohui Shen, and Dong Xuan. Mobile phone based drunk driving detection. In *2010 4th International Conference on Pervasive Computing Technologies for Healthcare*, pages 1–8, 2010. 1, 8, 13

[12] H De Wit, EH Uhlenhuth, J Pierri, and Chris E Johanson. Individual differences in behavioral and subjective responses to alcohol. *Alcoholism: Clinical and Experimental Research*, 11(1):52–59, 1987. 5

[13] P Ekman. Wv friesen, manual for the facial action coding system, 1977. 4

[14] Mark T Fillmore and M Vogel-Sprott. Expectancies about alcohol-induced motor impairment predict individual differences in responses to alcohol and placebo. *Journal of studies on alcohol*, 56(1):90–98, 1995. 5

[15] Peter I Frazier. A tutorial on bayesian optimization. *arXiv preprint arXiv:1807.02811*, 2018. 6

[16] Margherita Grandini, Enrico Bagli, and Giorgio Visani. Metrics for multi-class classification: an overview. *arXiv preprint arXiv:2008.05756*, 2020. 6

[17] Kazuyuki Hara, Daisuke Saito, and Hayaru Shouno. Analysis of function of rectified linear unit used in deep learning. In *2015 International Joint Conference on Neural Networks (IJCNN)*, pages 1–8, 2015. 6

[18] Hasanin Harkous and Hassan Artail. A two-stage machine learning method for highly-accurate drunk driving detection. In *2019 International Conference on Wireless and Mobile Computing, Networking and Communications (WiMob)*, pages 1–6. IEEE, 2019. 1, 8, 13

[19] Hasanin Harkous, Carine Bardawil, Hassan Artail, and Naseem Daher. Application of hidden markov model on car sensors for detecting drunk drivers. In *2018 IEEE International Multidisciplinary Conference on Engineering Technology (IMCET)*, pages 1–6. IEEE, 2018. 1, 8

[20] Tim J Hutton, BR Buxton, and Peter Hammond. Dense surface point distribution models of the human face. In *Proceedings IEEE Workshop on Mathematical Methods in Biomedical Image Analysis (MMBIA 2001)*, pages 153–160. IEEE, 2001. 4

[21] Alan Wayne Jones. The relationship between blood alcohol concentration (bac) and breath alcohol concentration (brac): a review of the evidence. *Road safety web publication*, 15:1–43, 2010. 1, 2, 3

[22] Agnieszka Kalinowski and Keith Humphreys. Governmental standard drink definitions and low-risk alcohol consumption guidelines in 37 countries. *Addiction*, 111(7):1293–1298, 2016. 4

[23] Kevin Koch, Martin Maritsch, Eva Van Weenen, Stefan Feuerriegel, Matthias Pfäffli, Elgar Fleisch, Wolfgang Weinmann, and Felix Wortmann. Leveraging driver vehicle and environment interaction: Machine learning using driver monitoring cameras to detect drunk driving. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, pages 1–32, 2023. 1, 2, 7, 13

[24] P Le Lièvre, D Adminaite, G Jost, and F Podda. Progress in reducing drink-driving and other alcohol-related road deaths in europe. *European Transport Safety Council: Brussels, Belgium*, 2019. 1, 2, 7

[25] John D. Lee, Dary D Fiorentino, Michelle L. Reyes, Timothy L. Brown, Omar Ahmad, James C Fell, Nicholas J. Ward, and R. Dufour. Assessing the feasibility of vehicle-based sensors to detect alcohol impairment. *Washington, DC: National Highway Traffic Safety Administration*, 2010. 1, 2, 7, 8, 13

[26] Vera Lehmann, Afroditi Tripyla, David Herzig, Jasmin Meier, Nicolas Banholzer, Martin Maritsch, Jörg Zehetner, Daniel Giachino, Philipp Nett, Stefan Feuerriegel, et al. The impact of postbariatric hypoglycaemia on driving performance: a randomized, single-blind, two-period, crossover study in a driving simulator. *Diabetes, obesity and metabolism*, 23(9):2189–2193, 2021. 2

[27] Zhenlong Li, Xue Jin, and Xiaohua Zhao. Drunk driving detection based on classification of multivariate time series. *Journal of safety research*, 54:61–e29, 2015. 1, 8, 13

[28] ZhenLong Li, HaoXin Wang, YaoWei Zhang, and XiaoHua Zhao. Random forest–based feature selection and detection method for drunk driving recognition. *International Journal of Distributed Sensor Networks*, 16(2):1550147720905234, 2020. 1, 8, 13

[29] M Linnoila, CW Erwin, D Ramm, and WP Cleveland. Effects of age and alcohol on psychomotor performance of men. *Journal of studies on alcohol*, 41(5):488–495, 1980. 5

[30] Scott E Lukas, Abdullatif Zaouk, Elizabeth Ryan, Jane McNeil, Justin Shepherd, Michael Willis, Neeraj Dalal, and Kelly Schwartz. Driver alcohol detection system for safety (dadss)-preliminary human testing results. In *Proceedings of the 25th International Technical Conference on the Enhanced Safety of Vehicles*, number 17-0304 in Enhanced Safety of Vehicles, pages 1–11, 2017. 1

[31] Pierre Maurage, Zoé Bollen, Nicolas Masson, and Fabien D'Hondt. Eye tracking studies exploring cognitive and affective processes among alcohol drinkers: a systematic review and perspectives. *Neuropsychology review*, 31(1):167–201, 2021. 4

[32] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830, 2011. 6, 7

[33] VA Ramchandani, WF Bosron, and TK Li. Research advances in ethanol metabolism. *Pathologie Biologie*, 49(9):676–682, 2001. 5

[34] Diogo Reis, Ricardo Tomás, Margarida C. Coelho, and Eloisa Macedo. Correlating driving behavior with safety performance: from the heart to the pedals with a driving simulator. *Transportation Research Procedia*, 69:217–224, 2023. AIIT 3rd International Conference on Transport Infrastructure and Systems (TIS ROMA 2022), 15th-16th September 2022, Rome, Italy. 2

[35] Steve Rubenzer. Judging intoxication. *Behavioral sciences & the law*, 29(1):116–137, 2011. 4

[36] Jennifer Schumann, Monica Perkins, Paul Dietze, Dhanya Nambiar, Biswadev Mitra, Dimitri Gerostamoulos, Olaf H Drummer, Peter Cameron, Karen Smith, and Ben Beck. The prevalence of alcohol and other drugs in fatal road crashes in victoria, australia. *Accident Analysis & Prevention*, 153:105905, 2021. 1, 7

[37] Ralf C Staudemeyer and Eric Rothstein Morris. Understanding lstm–a tutorial into long short-term memory recurrent neural networks. *arXiv preprint arXiv:1909.09586*, 2019. 6

[38] Timothy Stewart. Overview of motor vehicle crashes in 2020. Technical Report DOT HS 813, National Highway Traffic Safety Administration, 2022. 1, 7

[39] Yifan Sun, Jinglei Zhang, Xiaoyuan Wang, Zhangu Wang, and Jie Yu. Recognition method of drinking-driving behaviors based on pca and rbf neural network. *Promet-Traffic&Transportation*, 30(4):407–417, 2018. 1, 8, 13

[40] Guido Van Rossum and Fred L. Drake. *Python 3 Reference Manual*. CreateSpace, Scotts Valley, CA, 2009. 6

[41] M Vogel-Sprott and P Barrett. Age, drinking habits and the effects of alcohol. *Journal of Studies on Alcohol*, 45(6):517–521, 1984. 5

[42] Brian Wotring, Jonathan F Antin, and Ryan C Smith. Alcohol intoxication checklist: A naturalistic approach. Technical report, National Surface Transportation Safety Center for Excellence, 2021. 4, 5

[43] Ying Yao, Xiaohua Zhao, Hongji Du, Yunlong Zhang, Guohui Zhang, and Jian Rong. Classification of fatigued and drunk driving based on decision tree methods: a simulator study. *International journal of environmental research and public health*, 16(11):1935, 2019. 8

[44] Abdullatif K Zaouk, Michael Wills, Eric Traube, and Robert Strassburger. Driver alcohol detection system for safety (dadss)–a status update. In *24th Enhanced Safety of Vehicles Conference*. ESV Gothenburg, Sweden, 2015. 1, 7, 8