Investigating Physiological and Behavioural Sensing Modalities Towards Drowsiness Detection

Emma Perkins, *Student Member, IEEE*, Chiranjibi Sitaula, *Member, IEEE*, Michael Burke, *Member, IEEE*, Jessica Manousakis, Clare Anderson, and Faezeh Marzbanrad, *Senior Member, IEEE*

***Abstract*— Monitoring driver drowsiness is a crucial aspect of ensuring road safety. Many studies have explored a variety of physiological signals and behavioural monitoring of drivers using video, or a combination of these approaches. In this paper, we investigate and optimise the effectiveness of various modalities to monitor drowsiness. We developed a physiological model using Electrocenography, Electromyography, Electrooculography and Electrocardiography recordings. A video-based behavioural model was then developed, utilising a pre-trained ResNet-101, face landmarks and handcrafted features for feature extraction, followed by classification with two Long-Short-Term memory blocks. We also investigated a combination of these models using decision fusion to form a hybrid model. The proposed method was trained and evaluated on the publicly available database (DROZY) and compared to other methods on the same database. Our proposed physiological and behavioural models were separately compared with previous approaches where we demonstrated their superior performance when appropriately validated. We further improved the results by combining the physiological and behavioural models which detected drowsiness in 93.10% of trials through leave-one-subject-out cross-validation.**

***Index Terms*— drowsiness detection, fatigue, driver physiological monitoring, sensor applications, behavioural monitoring, hybrid model, road safety**

I. Introduction

DRIVER drowsiness poses a significant safety risk on urban and rural roads, endangering the lives of drivers and other road users, with similar numbers of crashes and fatalities to drink driving and speeding [1]. This results in both physical and economic losses for countries worldwide. Globally, it is predicted that 1.3 million people die per year because of car crashes, which cost the country approximately 3% of their GDP [2]. In Australia, driving with fatigue has contributed to 20-30% of road-related severe injuries and fatalities, which is similar to that of drink driving and speeding [1], [3]. It has been demonstrated that the effects of drowsiness are similar to that of being over the blood alcohol concentration limit of 0.05% [4]. Hence, accurate monitoring of driver drowsiness is important to road safety.

Current methods used to monitor drowsiness are categorised into physiological, behavioural, vehicle-based and hybrid monitoring, where hybrid is a combination of the first three methods. Physiological monitoring is the most accurate individual method; however, can incur high levels of intrusiveness [5]. Behavioural monitoring is the second most accurate individual method but can result in privacy concerns [6]. Vehicle-based monitoring is the least reliable but most widely used commercially, due to its simple implementation [7], [8]. Furthermore, authors in [8] showed that only around 2% of lane departure events captured by mobile eye were drowsiness related. Hybrid monitoring has shown to be the most accurate despite the least researched as it also carries an increased system complexity [5].

Considering the advantages and limitations of physiological and behavioural sensing modalities for driver drowsiness detection, we have developed separate models for each modality that significantly outperform previous works in the literature. Our work offers the flexibility of choosing the most suitable modality based on practical preferences or available resources. Furthermore, we contribute to the knowledge of hybrid systems by proposing the first hybrid model on the publicly available DROZY database [9]. DROZY is the only database that contains both behavioural and physiological data, making it possible to evaluate hybrid models. In this paper, we use the Karolinska Sleepiness Scale (KSS) ratings for drowsiness labels and real drowsiness. This work is cross-validated using a leave-one-subject-out approach in order to evaluate our models on unseen test data. We also explore the possibility of using this method without normalising the physiological data, in order to predict drowsiness without the need for a baseline for each driver.

A. Statement of Contributions:

The contributions of this work can be summarised as:

E. Perkins, C. Sitaula, M. Burke and F. Marzbanrad are with the Department of Electrical and Computer Systems Engineering, Monash University, Clayton, 3800, VIC, Australia. C. Anderson and J. Manousakis are with Turner Institute for Brain and Mental Health and School of Psychological Sciences, Monash University, Clayton, 3800, VIC, Australia. E. Perkins' research is supported by an Australian Government Research Training Program (RTP) Scholarship. Corresponding author: Faezeh Marzbanrad (e-mail: faezeh.marzbanrad@monash.edu).

- We proposed an improved physiological drowsiness detection model using multi-modal sources.
- We provided an extensive evaluation of model sensitivity to KSS driver drowsiness labelling, physiological model hyperparameters, and trial duration.
- We developed an improved behavioural drowsiness detection model which outperforms previous behavioural approaches.
- We explored a hybrid model that improves upon individual modalities
- We propose a clear testing protocol that considers the importance of data partitioning, and conduct a thorough analysis of the proposed models and existing driver drowsiness detection techniques

II. Background

Many studies involving driver drowsiness monitoring collect their own data for model development as there is not a large number of public databases available. Some databases used in the literature and their commonly used acronyms include:

- Physiological: A sustained-attention driving task database [10]
- Physiological and Behavioural videos: ULg multimodality drowsiness database (called DROZY) [9]
- Behavioural Videos: The National Tsing Hua University (NTHU) database [11]
- Yawning: Yawning detection dataset (YawDD) [12]
- Behavioural videos: UTA Real-Life Drowsiness Dataset (UTA-RLDD) [13]
- Behavioural Image datasets: Closed Eyes In The Wild (CEW) [14] and Eye-Chimera [15]
- Eyeblink videos: Zhejiang University Eyeblink Database (ZJU) [16]
- Vehicle and Behavioural: The Second Strategic Highway Research Program (SHRP2) [17] (restricted access)

NTHU is one of the more common databases for behavioural model development, but it is not based on real drowsiness. Furthermore, many databases do not contain video data but rather images or simulated drowsiness. The only databases containing multi-modal data include SHRP2 and the DROZY database. The SHRP2 database contains naturalistic on-road data; however, does not contain specific fatigue inducement or physiological data. Alternatively, DROZY contains behavioural and physiological data, captured through the inducement of real drowsiness via sleep deprivation. Hence, DROZY can be used to develop separate physiological and behavioural models as well as a hybrid drowsiness monitoring system that is validated on real drowsiness.

A. DROZY Database

The DROZY database includes 14 participants who underwent 3 trials each, lasting 10 minutes in duration, where each trial had near-infrared intensity videos. Participants underwent increased sleep deprivation. The first trial was collected between 10 and 11am, the second between 3:30 and 4am and the last trial between 12 and 12:30pm the following day. Data was missing for 6 of the trials, giving 36 trials of data in total. The participants had no reported sleep disorders, were imaged under single-scenario light settings and did not wear any form of glasses. There were 3 males and 11 females aged 22.7 $\pm$ 2.3 (mean $\pm$ standard deviation). Each participant allowed to self-report the KSS score before each trial in the laboratory environment, which was quiet and isolated. During the data collection, the participants were told to monitor a red rectangular box over a black background on a computer screen. They were told to press the response button soon after they noticed a yellow stimulus counter within that box in each trial.

The physiological data collected includes Electrocenography (EEG), Electromyography (EMG), Electrooculography (EOG) and Electrocardiography (ECG). The EMG recording was at the chin and EOG data included both vertical and horizontal recordings. As we were aiming to study the EOG data in terms of blinking, we only select features from the vertical EOG data. The EEG electrodes included in the DROZY database include Pz, C3, C4, Cz, and Fz.

The behavioural videos were collected using a near-infrared camera with around 10 minutes recorded per trial. The videos were recorded at 15 frames per second (fps) and 30 fps.

No simulated vehicle data was collected, therefore we developed our models based on the physiological and behavioural information only.

B. Related Work

Drowsiness is typically identified and labelled using subjective measures including observer ratings of drowsiness or KSS labelling. The data (e.g., behavioural or physiological) is pre-processed, features are extracted, in many cases feature selection is undertaken, and lastly, a classification decision is taken. In some cases, end-to-end processing is undertaken using deep learning, which can increase computation and introduces a 'black-box' solution where solutions are not always interpretable. More traditional methods for classification have included thresholding [18]–[20], Support Vector Machine (SVM) [21]–[23], k-Nearest Neighbour (kNN) [24] and more recently, deep learning approaches [25]–[27].

There has been a large amount of work undertaken in developing drowsiness models, with multi-modal physiological models and hybrid models performing the best [21], [28]. Despite this, fewer studies have focused on comparing these modalities, and optimising the performance of each modlaity and their combinations, on publicly available databases that contain real drowsiness.

Studies utilising the DROZY database focus on either physiological or behavioural data rather than a combination of both. Physiological studies use EEG, ECG, or a combination of the two and follow the process seen in Fig. 1. The work by Guarda et al. [29] used a Convolutional neural network (CNN) on EEG spectrograms to detect drowsiness. They reported using cross-validation for model development but did not specify what type. The data used for drowsiness labelling included using KSS less than or equal to 4 in the first trial as the alert data and the third trial where the KSS was 7 or higher for drowsy data. Guarda et al. also tested window sizes, where they found 13

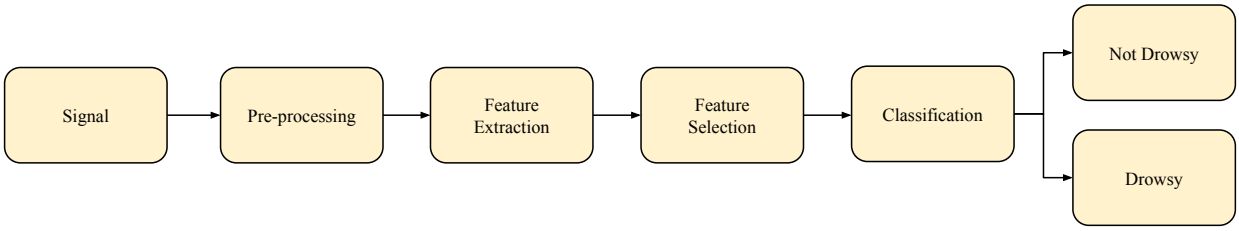

Fig. 1. Process for Physiological Drowsiness Detection.

seconds to be superior to 5 seconds, obtaining a classification accuracy of 86.74%. Maftukhaturrizqoh et al. [30] used ECG data and k-fold cross-validation to validate their work. Deep learning was also used in this work where a Radial Basis Function Artificial Neural Network was implemented. They obtained an accuracy of 81.96%, where KSS is said to be used as the definition of drowsiness, however the criteria for determining alert and drowsy states were not stated.

For behavioural studies using the DROZY database, García-García et al. [31] used MobileNets, a fast CNN architecture to classify drowsiness data. The alert label was given for the first trial (10 and 11am) and a drowsy label was given for the second trial (3:30 and 4am). They reported results for different methods of data partitioning for models including per participant models, singular split with the data combined and lastly with specific participants removed, in order to determine if the models will work for individuals not specifically trained by the database. They removed three participants' data from the training set in this case. García-García et al. demonstrated how the partitioning of data degrades the model – when using per person model an accuracy >99% is achieved, a 90.48% accuracy is achieved using combined data and lastly, an accuracy of 82.9% is seen on the three removed participants' data. Maior et al. [32] built a drowsiness model on their own data and used DROZY to validate their model. For alert labels, they used the first trial when the KSS was equal to 3 or less and for the drowsy labels they used the third trial when the KSS was 7 or more. Support vector machine (SVM) was used for classification where the method was declared to be correct when no drowsiness warnings were given for the alert trials and a minimum of one drowsiness warning was given in the drowsy trial per person, per trial used. This worked for 94.44% of the trials (17 out of 18). Ngxande [33] also looked at behavioural monitoring where the last few layers of the CNN-based, VGG-Face network were used for classification. Held-out data was used for testing purposes, where an accuracy of 83.86% was achieved.

Many current methods using the DROZY database are not generalisable due to the partitioning of data. The work of García-García et al. [31] and Maior et al. [32] are exceptions to this, where the application of their work to unseen driver's data was explored. This is important for driver drowsiness, as it is infeasible to require user-specific training data to detect drowsiness. Hence, when developing our models, we have focused on using leave-one-subject-out cross-validation in order to validate our models on unseen data. All reported accuracies in our work are based on unseen data.

Additionally, we have outlined in previous works the issues associated with drowsiness labelling [5], where we see that many of the studies have used varying forms of drowsiness labelling – either trial-based (time-based) or a combination of trial-based and KSS based. For our work, we present a KSS-based method, since KSS has been proven to be a viable drowsiness score [34].

To the best of our knowledge, a combination of the physiological and behavioural models has not yet been considered for the DROZY dataset. Hence, after proposing a physiological and a behavioural model, we apply decision fusion to combine the approaches on a per trial basis. Results show that this improves our accuracy of drowsiness detection at the cost of higher complexity.

III. Methodology

First a physiological model was developed where various parameters were explored to fine-tune the model. For this purpose we used both k-Nearest Neighbours (kNN) and Multi-layer Perceptrons (MLP). In order to develop our behavioural model, we explored pre-trained backbone architectures (VGG-16, VGG-19 [41], ResNet-50 and Res-Net101 [42]) for feature extraction, or landmark and handcrafted prediction, with a Long-Short-Term Model (LSTM) [43] classification head. The two models are also combined with a decision fusion method to evaluate a hybrid model combining the results of the behavioural and physiological models.

Driver drowsiness literature is frequently inconsistent when it comes to defining drowsiness, with subjective definitions and competing labelling definitions used across the literature. As a result, it is important to consider multiple potential definitions to avoid cherry-picking results. When exploring the physiological and behavioural data we tested different drowsiness labels using KSS labels considered in other studies including:

1) KSS 8 and 9 for drowsy labels
2) KSS 8 and 9 for drowsy labels with the 7 trials removed [23]
3) KSS 8 and 9 for drowsy labels with the 6 and 7 trials removed [44]
4) KSS 7, 8 and 9 for drowsy labels (KSS789) [45]
5) KSS 7, 8 and 9 for drowsy labels with the 6 trials removed (KSS789-6)

We also considered drowsiness definitions based on the duration of the trial (first half or whole trial). When KSS scores are collected at the start of a trial, this may indicate that the

TABLE I
FEATURES USED IN PHYSIOLOGICAL MODELS

Source	Features
EEG with 5 Electrodes, 24 features per electrode resulting in 120 features	Delta Power (0.5-4 Hz), Theta Power (4-7 Hz), Alpha Power (8-12 Hz), Beta Power (12.5-16 Hz), Approximate Entropy [35], Sample Entropy [35], Shannon Entropy [35], Fuzzy Entropy [35], Multiscale Entropy [35], Spectral Entropy [36], Wavelet Entropy [37], Mean/Median, Standard Deviation, Kurtosis, Variance, Skewness, Absolute Delta, Absolute Theta, Absolute Alpha, Absolute Beta, Relative Delta, Relative Theta, Relative Alpha, Relative Beta
EMG, 13 Features	Mean, Median, Variance, Low Frequency (1-15 Hz), Medium Frequency (15-30 Hz), High Frequency (30-50 Hz), Absolute Low Frequency, Absolute Medium Frequency, Absolute High Frequency, Relative Low Frequency, Relative Medium Frequency, Relative High Frequency, Shape Factor [24]
ECG, 20 features	Heart Rate (HR), Very Low Frequency (VLF) (0.003-0.04 Hz), Low Frequency (LF) (0.04-0.15 Hz), High Frequency (HF) (0.15-0.40), LF/HF Ratio, Power, Respiration Rate [35], Median Respiration Rate Variability (RRV) [35], Mean RRV, Approximate Entropy RRV [35], HR Standard Deviation, Shannon Entropy [35], Approximate Entropy [35], Fuzzy Entropy [35], Wavelet Entropy [37], Mean Heart Rate Variability (HRV), HRV Standard Deviation, HRV Kurtosis, HRV Variance, HRV Skewness
EOG, 17 Features	Blink Frequency [38], Low Frequency Energy* [39], Eye Closure Duration [39], [40], Time to close [39], Time to open [39], Delay of reopening [39], Velocity of closure [39], Velocity of opening [39], Amplitude [39], Delay ratio [39], Absolute Power (0-1 Hz), Relative Power (0-1 Hz), Mean, Standard Deviation, Kurtosis, Variance, Skewness

∗ Power ratio of 0 to 1 Hz and 1 to 20 Hz

first half of the trial is more relevant to the KSS recording than the whole trial.

For physiological data, two different window sizes for extracting features (15 and 30 seconds) were also tested, alongside different image sequence lengths for behavioural data including sequences of 5, 10, and 15.

A. Physiological Model Development

Our physiological model follows a standard pipeline, as shown in Fig. 1. Each stage is explained below:

1) Pre-processing: The EEG, EMG, ECG and EOG data underwent pre-processing to remove unwanted artifacts and noise. All the signals were re-sampled from 512 Hz to 200 Hz. Filtering was then undertaken including a 0.5 to 40 Hz bandpass filter for EEG, a 0.1 to 30 Hz bandpass filter for ECG, a 0.1 to 10 Hz bandpass filter for EOG and a 45 Hz lowpass filter for EMG. EEG data was also put through (modified) wavelet independent component analysis (w-ICA) in order to remove noise and extract components, using the method proposed in [46].

2) Feature Extraction: Features were then extracted across the different window frames, 15 and 30 seconds. This included 170 features across the four signals. A summary of features can be found in Table I. The data was then split, with one participant's data completely removed to perform leave-one-subject-out cross-validation, where the left-out participant's data was not involved in the feature selection.

3) Feature Selection: The features were put through a Shapiro–Wilk normality test in order to determine a statistical testing method appropriate to the data. The data was then analysed using independent T-tests for Gaussian and rank types. When the features had statistical significance in the training dataset, the features were kept. This was repeated for each of the window sizes, duration of trial and for each drowsiness label separately. The data was normalised per participant across all the trials, as physiological data often vary between individuals. In this work we also explore models without normalisation, in order to see if drowsiness can be monitored without the need for an individualised baseline.

The remaining data was split into a 70% training, 30% fine-tuning data set. The participants data were only ever completely in training or completely in testing, where no individual was split across the training and testing sets. When making the model, the final feature selection was completed using "SelectKbest" which utilises an ANOVA F-value scoring function. Different models were then tested with up to a maximum of 10 features to avoid overfitting.

4) Classification: The classification of data was completed using both kNN and MLP. The kNN model was fine-tuned based on the number of nearest neighbours considered. The nearest neighbour values tested from 2 to 50 with a step of 4 which were tested against the best features from "SelectKbest" where 1 to 10 features were tested. The mode accurate model on the 30% fine-tuning data was the model used. The model was then evaluated on the left-out participant's data. The results for both kNN and MLP methods are presented.

B. Behavioural model development

The DROZY video-based data had the same arrangement for drowsiness labels and trial duration as described above. Here we followed the process of video acquisition, image extraction, face detection and feature extraction, sequencing, and classification (Fig. 2). Throughout the classification process, we also undertook hyperparameter tuning.

1) Feature Extraction: Three sets of features were extracted including the 468 landmarks from MediaPipe [47] with x, y, and z locations, handcrafted features from predicting the head positioning (nodding or looking to the side), yawn occurrences and finally using an output from a pre-trained CNN-based model. There were many pre-trained models available, which we explored using VGG-16, VGG-19 [41], ResNet-50, and ResNet-101 [42] models. These models were selected based on past works where they had been used or suggested as a good model for drowsiness detection [33], [48], [49]. We used pool

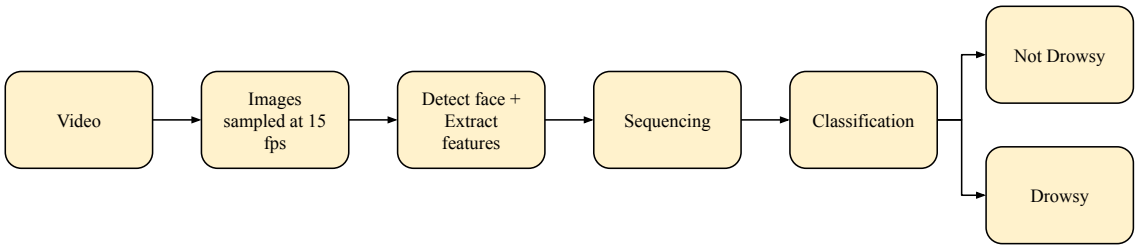

Fig. 2. Process for our Behavioural Drowsiness Detection.

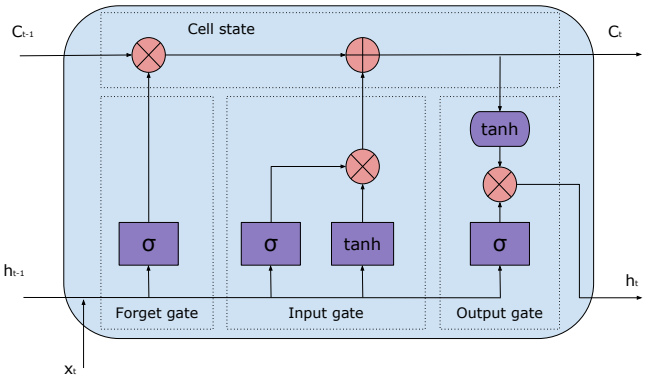

Fig. 3. LSTM cell structure where C_t is the cell state, x_t is the input and h_t is the output

5 layer for the feature extraction from VGG-16 and VGG-19, whereas the average pool layer was used for the feature extraction from ResNet-50 and ResNet-101. The images used in feature extraction were those with a face detected therein, as determined by MediaPipe.

2) Classification: Using videos for monitoring drowsiness is beneficial as drowsiness features become more evident over time. Hence, a temporal model can be applied to capture some of these features. Here we use an LSTM cell due to the temporal modelling benefits, following the the structure shown in Fig. 3 [43]. Multiple LSTM blocks were explored, with a dropout of 0.5 after each LSTM block, a recurrent regularizer and a kernel regularizer within each LSTM block, a mean squared error (MSE) loss function, and an Adam optimizer used for model training.

3) Hyper Parameter Tuning: The hyperparameter tuning was undertaken on the 70-30 split within the training data, where the left-out participant's data was not used in developing the models. A systematic approach was taken, where multiple parameters were explored including:

1) Sequence length: Number of frames used for temporal features (5, 10 and 15)
2) Batch size: 32, 6 and 128
3) Learning Rate: 0.01, 0.001, 0.0001, 0.00001
4) Pre-trained model: VGG-16, VGG-19, ResNet-50 and ResNet-101
5) Model parameters: The number of LSTM blocks (1, 2, 3 and 4)
6) Relevant features: Grouped by pre-trained model, MediaPipe annotations and handcrafted features

4) Evaluation: The data was partitioned and evaluated using leave-one-subject-out cross-validation - the same method used with the physiological data. In order to determine the best parameters, the models were developed for every participant before using the average accuracy across all participants to determine which parameter was best.

C. Hybrid Model

We combined the physiological and behavioral approaches by using the percentage of the trial labeled as drowsy for both classifications. We then calculated the average of the two percentages, and used this as the outcome for the hybrid model. This is repeated for when the physiological data is not normalised. The results for both kNN and MLP approaches are presented, alongside the behavioural method that utilises a ResNet101 pre-trained model (trained on Imagenet data) to extract features, as this was proven to be the most consistent behavioural method.

IV. Results

A. Physiological approach results

Results are presented for each of the tested parameters, including various drowsiness labelling conventions, window frame lengths and trial durations. We also present the most prominent features in detecting drowsiness, and a summary for each parameter tested is presented. The best model produced an accuracy of 91.43% with a sensitivity of 90.48% and specificity of 92.86%.

1) Drowsiness Labels: The most differentiating drowsiness labels were found using KSS789 and KSS789-6, achieving up to 91.43%. Similar to what discussed in Manousakis et al. [34], the detection of drowsiness was shown to be below 50% accuracy when the user was drowsy for cases of a KSS of 8 and 9 used for drowsiness, including when the participant's data for a KSS of 6 and 7 (or just 7) was removed. This occurred for all cases of window frames used and the duration of the trials. SMOTE was used to oversample the drowsiness data as we suspect a lack of data was an issue, however, this provided minimal improvement. Hence, only KSS789 and KSS789-6 were used for drowsiness. The global features, nearest neighbours and selected drowsiness labels were then tested, with the results presented in Table II.

2) Window Size: From Table II it is shown that the window size is not conclusively better in one case or the other. Using MLP the 30 second window performs better in 3 out of 4 cases, however using kNN the 15 second window performs

TABLE II
PHYSIOLOGICAL MODEL ACCURACY

			KSS789			KSS789-6		
Classification	Duration	Window	Accuracy	Sensitivity	Specificity	Accuracy	Sensitivity	Specificity
kNN	Whole Trial	30 s	69.69%	78.95%	57.14%	76.67%	62.50%	92.86%
	Half Trial	30 s	78.78%	89.47%	64.29%	75.86%	66.67%	85.71%
	Whole Trial	15 s	82.86%	90.48%	71.43%	76.67%	63.50%	92.86%
	Half Trial	15 s	**91.43%**	90.48%	92.86%	75.86%	62.50%	92.31%
MLP	Whole Trial	30 s	78.13%	88.89%	64.29%	80.65%	68.75%	93.33%
	Half Trial	30 s	81.82%	80.95%	83.33%	82.75%	73.33%	92.86%
	Whole Trial	15 s	74.29%	85.71%	57.14%	70.97%	68.75%	73.33%
	Half Trial	15 s	85.29%	90.48%	76.92%	79.31%	73.33%	85.71%

TABLE III
TOP 15 MOST COMMONLY DROWSINESS RELATED FEATURES, THEIR TREND WITH INCREASING DROWSINESS AND THE PERCENTAGE OF MODELS THAT INCLUDED THEM AS A FEATURE

Source	Feature	Trend	Percentage
EEG: Pz	Relative Alpha Power	↑	76.34%
EOG	Eye opening velocity	↑	57.14%
EMG	Relative Medium Frequency	↑	45.09%
EOG	Skewness	↓	43.75%
EEG: Pz	Sample Entropy	↓	37.50%
EEG: Pz	Fuzzy Entropy	↓	25.89%
EOG	Eye closure velocity	↓	24.11%
EEG: Pz	Approximate Entropy	↓	22.32%
EOG	Kurtosis	↓	22.32%
EEG: C3	Relative Alpha Power	↑	20.09%
EEG: C4	Relative Alpha Power	↑	16.07%
EOG	Delay Ratio	↑	16.07%
EOG	Delay of reopening	↑	13.39%
EEG: Cz	Relative Alpha Power	↑	11.16%
EMG	Relative High Frequency	↓	10.71%

better in 2 cases and equal in the other two. Moreover, the best results overall are seen using a 15 second window.

3) Duration of trial: Due to KSS being taken at the start of the trial we tested to see whether the whole or first half of the trials perform better, where we expected the half trial to be more related to the rating given by the user, due to the rating being taken before the trial. This was true in the majority of cases where the half trial preformed better in all instances (MLP and kNN), except when a KSS of 6 was removed and kNN classification was involved. This suggests that the models may be more linked to the KSS rating when the KSS is collected closer to the duration of trial.

4) KSS label performance: The removal of the bordering KSS value of the two labels was hypothesised to perform better in drowsiness labelling, allowing a greater separation of classes. This helped minimise the confusion over bordering KSS values for two labels. However, this was not the case, where 3 out of the 8 were better when excluding the data with KSS score of 6.

5) Prominent Features: The features used in the final drowsiness models can be found in Table III, where the feature trend is in relationship to increasing levels of drowsiness. These features were across all the different scenarios, and were used in the final models. The percentage represents how commonly the feature was deemed best across all the models. The main modalities found to contribute to drowsiness include EEG, EOG and EMG.

6) No normalisation: When participant's data did not undergo normalisation, the results were inferior, as presented in Table IV. However, up to 76.67% accuracy is able to be achieved, where possible improvements could allow non-normalised data to be used. The sensitivity without normalisation was 75.00% and the specificity was 78.57%.

B. Behavioural approach results

The best parameters found for behavioural classification are presented in Table V. From the hyper-parameter tuning, similar results were obtained for VGG-16 and ResNet-101. Instead of presenting the results for window sizes (as seen in the physiological approach), these two pre-trained models for feature extraction were compared. The results for classification are found in Table VI. The best results achieved and accuracy of 77.42%, sensitivity of 81.25% and specificity of 73.33%.

1) Pre-trained model: Overall, despite VGG-16 and ResNet-101 having the equal best model, it was noted that in the majority of cases, ResNet-101 performed better. Hence this was used in the hybrid model.

2) Duration of trial: The duration of trial results were different to that of the physiological data. It was found that the half trials performed much worse using the KSS labelling compared to the whole trial. This could be due to the reduction of data in training a deep learning model.

C. Hybrid approach results

With decision fusion, it was found that there was an improvement for the best overall drowsiness classification of close to 2% compared to physiological results and up to 16% improvement compared to behavioural alone. This brought the best model accuracy up to 93.10% for kNN classification across 31 trials using KSS789 with KSS trials of 6 removed, the whole trial and a window size of 15 seconds. The corresponding sensitivity was 85.71% and specificity of 100%. The accuracy using MLP improved to 87.10% with KSS789 with 6 removed and various combinations of windows and half/whole trials. The results are presented in Table VII.

There was also a significant improvement for instances where the physiological data was not normalised, where the best accuracy improved from 77.42% to 88.89%, using the

TABLE IV
NO NORMALISATION CLASSIFICATION ACCURACY

			KSS789			KSS789-6		
Classification	Duration	Window	Accuracy	Sensitivity	Specificity	Accuracy	Sensitivity	Specificity
kNN	Whole Trial	30 s	71.43%	71.43%	71.43%	74.19%	68.75%	80.00%
	Half Trial	30 s	68.75%	73.68%	61.54%	70.00%	68.75%	71.43%
	Whole Trial	15 s	73.53%	90.00%	50.00%	64.52%	68.75%	60.00%
	Half Trial	15 s	71.43%	75.00%	66.67%	65.52%	60.00%	71.43%
MLP	Whole Trial	30 s	62.86%	90.00%	26.67%	66.67%	81.25%	50.00%
	Half Trial	30 s	71.43%	80.95%	57.14%	70.97%	93.75%	46.67%
	Whole Trial	15 s	68.57%	71.43%	64.29%	70.97%	68.75%	73.33%
	Half Trial	15 s	70.59%	80.95%	53.85%	**76.67%**	75.00%	78.57%

TABLE V
HYPER-PARAMETERS FOR BEHAVIOURAL FEATURE EXTRACTION

Fine tuning elements	**Preset**	**Parameters Tested**	**Best Parameter**
Sequence length	N/A	5, 10, 15	10
Batch Size	32	32, 64, 128	64
Learning Rate	0.001	0.01, 0.001, 0.0001, 0.00001	0.001
Pre-trained model	VGG-16	ResNet-50, VGG-16, VGG-19, ResNet-101	VGG-16 and ResNet-101
Base LSTM Model	2 LSTM blocks	1, 2, 3, 4 LSTM blocks	2 LSTM blocks
Features	All	All, Pre-trained + MediaPipe, Pre-trained + Handcrafted, MediaPipe + Handcrafted, Pre-trained, MediaPipe, Handcrafted	All and ResNet + MediaPipe

TABLE VI
BEHAVIOURAL CLASSIFICATION ACCURACY

		KSS789			KSS789-6		
Duration	Pre-Trained Model	Accuracy	Sensitivity	Specificity	Accuracy	Sensitivity	Specificity
Whole Trial	VGG-16	65.71%	71.43%	57.14%	**77.42%**	81.25%	73.33%
Half Trial	VGG-16	54.28%	75.00%	26.67%	64.52%	68.75%	60.00%
Whole Trial	ResNet-101	75.00%	80.95%	66.67%	**77.42%**	81.25%	73.33%
Half Trial	ResNet-101	61.76%	80.00%	35.71%	67.74%	56.25%	80.00%

TABLE VII
HYBRID CLASSIFICATION ACCURACY

			KSS789			KSS789-6		
Classification	Duration	Window	Accuracy	Sensitivity	Specificity	Accuracy	Sensitivity	Specificity
kNN +	Whole Trial	15 s	80.56%	95.24%	60.00%	**93.10%**	85.71%	100.00%
LSTM	Half Trial	15 s	83.33%	100.00%	60.00%	86.67%	80.00%	93.33%
	Whole Trial	30 s	80.56%	90.48%	66.67%	83.87%	81.25%	86.67%
	Half Trial	30 s	75.00%	95.24%	46.67%	87.10%	81.25%	93.33%
MLP +	Whole Trial	15 s	80.56%	95.24%	60.00%	83.33%	80.00%	86.67%
LSTM	Half Trial	15 s	80.00%	100.00%	50.00%	87.10%	81.25%	93.33%
	Whole Trial	30 s	77.78%	90.48%	60.00%	87.10%	81.25%	93.33%
	Half Trial	30 s	82.35%	100%	57.14%	87.10%	81.25%	93.33%

KSS789 label for drowsy, the whole trial with a 30 seconds window and the kNN approach (Table VIII). The sensitivity was 95.24% and the specificity was 80%. This demonstrates that drowsiness may be accurately monitored without the need for individualised baseline, assisting with the application of a system in a vehicle.

D. Probability of drowsiness

To explore the characteristics of drowsiness, the probability of drowsiness was explored on a per-trial basis for the best combinations of Physiological (kNN)/(MLP) and Behavioural. This included the drowsiness label of KSS789-6 and the duration included the whole trial. The kNN approach used features extracted from a 15-second window and MLP a 30-second window. The trend of drowsiness was then calculated with a linear regression line across each trial. The start of the trial versus the end of the trial of the components of the regression lines was then used in a boxplot for both the alert (KSS < 6) and drowsy (KSS > 6) instances. According to Anderson et al. [50], there is a positive correlation between self-reported

TABLE VIII
HYBRID WITHOUT NORMALISATION CLASSIFICATION ACCURACY

			KSS789			KSS789-6		
Classification	Duration	Window	Accuracy	Sensitivity	Specificity	Accuracy	Sensitivity	Specificity
kNN + LSTM	Whole Trial	15 s	86.11%	100.00%	66.67%	86.21%	80.00%	92.86%
	Half Trial	15 s	72.22%	85.71%	53.33%	80.65%	75.00%	86.67%
	Whole Trial	30 s	**88.89%**	95.24%	80.00%	80.65%	75.00%	86.67%
	Half Trial	30 s	67.65%	85.00%	42.86%	77.42%	68.75%	86.67%
MLP + LSTM	Whole Trial	15 s	80.56%	85.71%	73.33%	76.67%	75.00%	78.57%
	Half Trial	15 s	72.22%	90.48%	46.67%	87.10%	87.50%	86.67%
	Whole Trial	30 s	66.67%	95.24%	26.67%	70.00%	87.50%	50.00%
	Half Trial	30 s	74.29%	95.00%	46.67%	78.57%	93.33%	61.54%

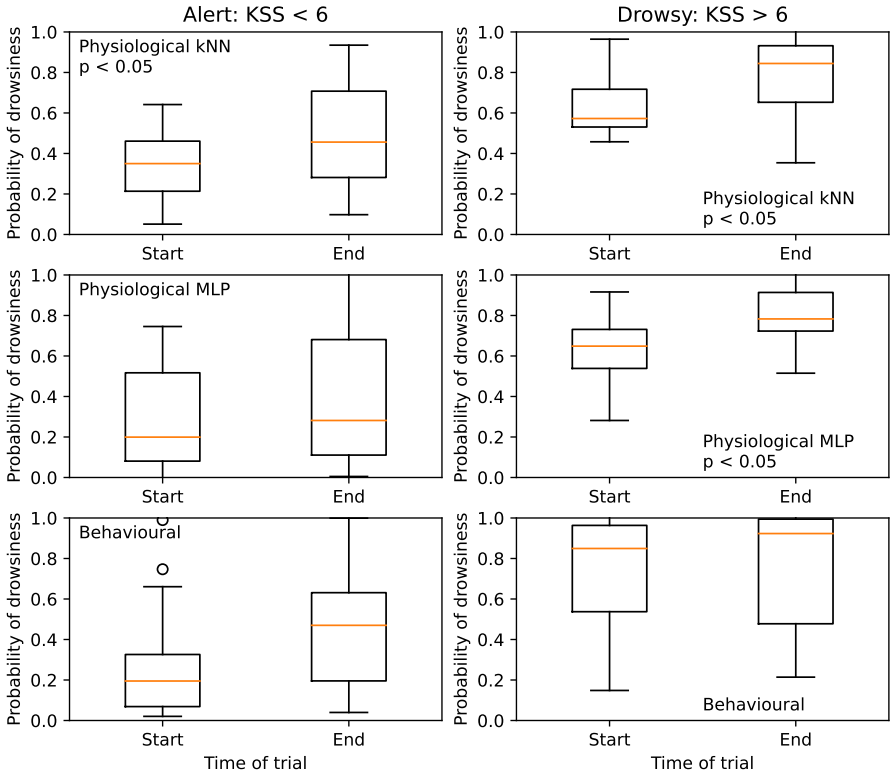

Fig. 4. Probability of drowsiness across the best hybrid model (KSS789-6)

sleepiness (KSS) and objective indices of drowsiness. The results are presented in 4, where both the alert and drowsy kNN instances proved to have statistically significant results, where the drowsiness probability increased over the trials. This also occurred when using the MLP classification method for the drowsy instances. These results demonstrated that in particular for physiological measures, drowsiness is predicted to increase over time.

E. Comparison with previous studies

Compared to previous physiological works on the DROZY database (Table IX), we demonstrate that with physiological data alone, our approach improved upon current methods. Our method is also validated on unseen data. This method is perhaps better than previous methods due to the inclusion of multiple physiological signals, rather than just EEG and ECG. This is further supported by the common features seen in making these models (refer Table III), where after EEG, the most common modalities include both EOG and EMG, which are not utilised for classification in other studies that involve the DROZY database.

In order to compare our behavioural model with other approaches, we had to apply their data partitioning to our method. The results are presented in Table X. We further validated on left-out participants' data and across all trials.

García-García et al. [31] implemented 3 approaches. We do not compare with the first approach which trains a model per person, as this is not related to our application. However, the second approach trains a model based on 1000 images per participant, 15% of this is used as validation during training. The remaining images are then used in testing. The alert data is from the first DROZY trial and the drowsy data is

TABLE IX
PHYSIOLOGICAL MODEL COMPARISON

Author	Year	Splitting Method	Classification	Accuracy
Guarda et al. [29]	2018	Cross-validation	ConvNets	86.74%
Maftukhaturrizqoh et al. [30]	2019	k-fold	RBFN	81.96%
Our method	2022	Leave-one-subject-out cross-validation	kNN	**91.43%**
Our method	2022	Leave-one-subject-out cross-validation	MLP	85.29%

TABLE X
PROPOSED BEHAVIOURAL MODEL WITH STATE-OF-THE-ART METHODS DATA PARITIONING

Author	Year	Splitting Method	Classification	Accuracy
García-García et al. [31]	2018	Single Split (combined data)	MobileNets	90.48%
Our method	2022	Single Split (combined data)	ResNet-101 + LSTM	$> 99\%$
García-García et al. [31]	2018	3rd Trial validation	MobileNets	77.82%
Our method	2022	3rd Trial validation	ResNet-101 + LSTM	83.60%
Maior et al. [32]	2020	Tested only*	SVM	94.44%
Our method	2022	Leave-one-subject-out cross-validation*	ResNet-101 + LSTM	94.44%
Ngxande [33]	2020	Held-out data	VGG-Face	83.86%
Our method	2022	Held-out data	ResNet-101 + LSTM	$> 99\%$
Our method	2022	Leave-one-subject-out cross-validation	ResNet-101 + LSTM	77.42%

$*$ Across 18 trials the using the first trial as alert and the third trial as drowsy, evaluated on the trials that suited the KSS and trial criteria

from the second trial. The third trial is held-out, in order to test the model. Our approach uses sequences in the LSTM model, where 10 images are used per input, hence we use 100 sequences in order to compare and train our model, allowing the equivalent of 1000 images per participant in the training phase. We then validate on the remaining sequences. Using this method, we obtain $> 99\%$ accuracy across 5 repetitions. We then validate on the third trial, where García-García et al. achieved 77.82%. Here, we obtained $83.60\% \pm 6.56\%$ (average and standard deviation) across five repetitions. García-García et al. also validate their model on participants 3, 8 and 14 specifically. Here they achieve 82.9% accuracy on the images. We compare our KSS leave-one-subject-out approach, where our model correctly labels all trials for participants 3 (trials 1, 2, and 3), 8 (trials 1 and 3), and 14 (trials 1, 2 and 3) based on KSS labels using our KSS789 with 6 removed, our best performing model.

Maior et al. [32] built a model external to DROZY and then tested their results on the first and last trials of DROZY, where KSS labels were below certain levels in trial 1 and above certain levels in trial 3. They determined their approach to work correctly on a trial if for the first trial there were no warnings given from their system and at least one warning was given in the last DROZY trial. They reported 1 trial to fail out of 18. Using our KSS method and a significantly larger number of trials, we found that for the specific 18 trials, 2 did not work. We then looked at using the first and last trials to train the model, as these were the trials used by Maior et al. Using leave-one-subject-out, we found that just 1 of the specific 18 trials failed and the rest succeeded, exactly comparable as Maior's. However, when evaluating our models, we consider a broader range of trials.

Using a percentage of random held-out data as seen in Nxgande [33], we obtain an average accuracy of $> 99\%$. The drowsiness labels in Nxgande's work were not specified, hence we applied the labels of García-García et al. Furthermore, we also test with our KSS labelling of drowsiness and random splitting, also achieving $> 99\%$ accuracy. Overall, our behavioural model also achived superior performance compared to previous works as presented in table X.

Unfortunately we were unable to compare our hybrid approach as no other hybrid studies were found using the DROZY database. However, using a simple hybrid approach we are able to achieve even greater accuracy than individual methods. Without normalisation, we can also see that drowsiness can still be detected up to 88.89% where an individual's baseline would not be required and up to 93.10% with a baseline.

V. DISCUSSION

As presented previously, many studies on the DROZY database do not encapsulate all available data, but rather select the first trial where the KSS is below a certain threshold and in the third trial where the KSS is above a certain threshold (Guarda et al. [29], Maior et al. [32]), or just use specific trials (García-Garcí et al. [31]). However, our work has used all available KSS labels, regardless of the trial they were in.

Based of these results, in particular, the behavioural results, the importance of data partitioning for model training is highlighted. In order to implement driver drowsiness monitoring on road, models cannot be validated on the participants' data on which they are trained, as models cannot be trained on all road users. Hence, we urge other researchers to consider leave-one-subject-out cross-validation when testing their models, in particular in the instance of small data sets. At first glance, the approach may appear inferior; however, this is a more reliable and realistic evaluation method. This paper demonstrates better accuracy than other approaches, when the same

data partitioning occurs, and that the model accuracy can be improved using a hybrid approach.

Furthermore, a hybrid approach can handle data loss, should a source be lost during driving. For example if video feed were lost, the system can still operate using the physiological signals and vice-versa. We hypothesise that other studies on the DROZY database have not been implemented as a hybrid approach due to the increased complexity of the model.

In the future, feature fusion should be considered for combining physiological and behavioural approaches, in order to evaluate on a real-time basis. We hypothesise that this will obtain more accurate results and be more reliable. The application of less intrusive acquisition systems is also important in order to progress this technology as drivers are more likely to use this life-saving technology. A more diverse range of participants and validation with on-road data is also required to progress this work, as the DROZY database is conducted within a lab, with a relatively small number of participants and a lack of diversity within them. Furthermore, participants do not wear glasses, meaning the behavioural component may not work for drivers wearing glasses. In addition, it is important to analyse how far in advance we can predict drowsiness.

VI. Conclusion

Using a real drowsiness database, DROZY, we proposed novel physiological and behavioral models that demonstrated superior performance compared to previously proposed methods in the literature. We further combined the two models as a hybrid approach which surpasses the accuracy achieved by each of the modalities. We utilised all 36 trials available across 14 participants where we labelled drowsiness with the KSS values of 7, 8, and 9 for the drowsy data. We also experimented using 31 of the trials, where the border KSS rating of 6 is removed. We validate our approach using leave-one-subject-out cross-validation, in order to evaluate the performance on unseen data. We are able to obtain accuracies up to 93.10% of detecting the drowsiness state using the combination of behavioural and physiological data. Due to the nature of physiological data, we normalised the recordings per person before model training and validation; however, we also explore the effect of not normalising physiological data in order to see if individualised baselines can be omitted. Here we obtained accuracies up to 88.89%. This encouraging result suggests that our work is useful to detect drowsiness in real driving scenarios.

References

[1] CARRS-Q, "Sleepiness and fatigue," 2018. [Online]. Available: https://research.qut.edu.au/carrsq/wp-content/uploads/sites/296/2020/12/Sleepiness-and-fatigue.pdf

[2] "Road traffic injuries," World Health Organisation, 2021. [Online]. Available: https://www.who.int/news-room/fact-sheets/detail/road-traffic-injuries

[3] M. D. Mulhall, T. L. Sletten, M. Magee, J. E. Stone, S. Ganesan, A. Collins, C. Anderson, S. W. Lockley, M. E. Howard, and S. M. W. Rajaratnam, "Sleepiness and driving events in shift workers: the impact of circadian and homeostatic factors," *Sleep*, vol. 42, no. 6, 2019. [Online]. Available: https://doi.org/10.1093/sleep/zsz074

[4] A. M. Williamson and A. M. Feyer, "Moderate sleep deprivation produces impairments in cognitive and motor performance equivalent to legally prescribed levels of alcohol intoxication," *Occupational and environmental medicine*, vol. 57, no. 10, pp. 649–655, 2000. [Online]. Available: https://pubmed.ncbi.nlm.nih.gov/10984335 https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1739867/

[5] E. Perkins, C. Sitaula, M. Burke, and F. Marzbanrad, "Challenges of driver drowsiness prediction: The remaining steps to implementation," *IEEE Transactions on Intelligent Vehicles*, vol. 8, no. 2, pp. 1319–1338, 2023.

[6] S. Martin, A. Tawari, and M. M. Trivedi, "Toward privacy-protecting safety systems for naturalistic driving videos," *IEEE Transactions on Intelligent Transportation Systems*, vol. 15, no. 4, pp. 1811–1822, 2014.

[7] M. Ngxande, J. Tapamo, and M. Burke, "Driver drowsiness detection using behavioral measures and machine learning techniques: A review of state-of-art techniques," in *2017 Pattern Recognition Association of South Africa and Robotics and Mechatronics (PRASA-RobMech)*, Conference Proceedings, pp. 156–161.

[8] M. D. Mulhall, J. Cori, T. L. Sletten, J. Kuo, M. G. Lenné, M. Magee, M.-A. Spina, A. Collins, C. Anderson, S. M. Rajaratnam *et al.*, "A pre-drive ocular assessment predicts alertness and driving impairment: a naturalistic driving study in shift workers," *Accident Analysis & Prevention*, vol. 135, p. 105386, 2020.

[9] Q. Massoz, T. Langohr, C. François, and J. G. Verly, "The ulg multimodality drowsiness database (called drozy) and examples of use," in *2016 IEEE Winter Conference on Applications of Computer Vision (WACV)*. IEEE, Conference Proceedings, pp. 1–7.

[10] Z. Cao, M. Chuang, J. King, and C.-T. Lin, "Multi-channel EEG recordings during a sustained-attention driving task," 6 2018.

[11] C.-H. Weng, Y.-H. Lai, and S.-H. Lai, "Driver drowsiness detection via a hierarchical temporal deep belief network," in *Asian Conference on Computer Vision*. Springer, Conference Proceedings, pp. 117–133.

[12] S. Abtahi, M. Omidyeganeh, S. Shirmohammadi, and B. Hariri, "Yawdd: Yawning detection dataset," 2020. [Online]. Available: https://dx.doi.org/10.21227/e1qm-hb90

[13] R. Ghoddoosian, M. Galib, and V. Athitsos, "A realistic dataset and baseline temporal model for early drowsiness detection," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, 2019.

[14] F. Song, X. Tan, X. Liu, and S. Chen, "Eyes closeness detection from still images with multi-scale histograms of principal oriented gradients," *Pattern Recognition*, vol. 47, no. 9, pp. 2825–2838, 2014.

[15] L. Florea, C. Florea, R. Vrânceanu, and C. Vertan, "Can your eyes tell me how you think? a gaze directed estimation of the mental activity." in *BMVC*, 2013.

[16] G. Pan, L. Sun, Z. Wu, and S. Lao, "Eyeblink-based anti-spoofing in face recognition from a generic webcamera," in *2007 IEEE 11th international conference on computer vision*. IEEE, 2007, pp. 1–8.

[17] J. M. Hankey, M. A. Perez, and J. A. McClafferty, "Description of the shrp 2 naturalistic database and the crash, near-crash, and baseline data sets," Virginia Tech Transportation Institute, Tech. Rep., 2016.

[18] D. SHIN, H. SAKAI, and Y. UCHIYAMA, "Slow eye movement detection can prevent sleep-related accidents effectively in a simulated driving task," *Journal of Sleep Research*, vol. 20, no. 3, pp. 416–424, 2011. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1365-2869.2010.00891.x

[19] O. Khunpisuth, T. Chotchinasri, V. Koschakosai, and N. Hnoohom, "Driver drowsiness detection using eye-closeness detection," in *2016 12th International Conference on Signal-Image Technology & Internet-Based Systems (SITIS)*. IEEE, Conference Proceedings, pp. 661–668.

[20] Y. Xu, W. Yang, X. Yu, H. Li, T. Cheng, X. Lu, and Z. L. Wang, "Real-time monitoring system of automobile driver status and intelligent fatigue warning based on triboelectric nanogenerator," *ACS nano*, vol. 15, no. 4, pp. 7271–7278, 2021.

[21] M. Awais, N. Badruddin, and M. Drieberg, "A hybrid approach to detect driver drowsiness utilizing physiological signals to improve system performance and wearability," *Sensors*, vol. 17, no. 9, p. 1991, 2017. [Online]. Available: https://www.mdpi.com/1424-8220/17/9/1991

[22] J. Jo, S. J. Lee, K. R. Park, I.-J. Kim, and J. Kim, "Detecting driver drowsiness using feature-level fusion and user-specific classification," *Expert Systems with Applications*, vol. 41, no. 4, Part 1, pp. 1139–1152, 2014. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0957417413006106

[23] S. Arefnezhad, S. Samiee, A. Eichberger, and A. Nahvi, "Driver drowsiness detection based on steering wheel data applying adaptive neuro-fuzzy feature selection," *Sensors*, vol. 19, no. 4, p. 943, 2019. [Online]. Available: https://www.mdpi.com/1424-8220/19/4/943

[24] M. Mahmoodi and A. Nahvi, "Driver drowsiness detection based on classification of surface electromyography features in a driving simulator," *Proceedings of the Institution of Mechanical Engineers, Part H: Journal of Engineering in Medicine*, vol. 233, no. 4, pp. 395–406, 2019.

[25] C. T. Lin, C. H. Chuang, Y. C. Hung, C. N. Fang, D. Wu, and Y. K. Wang, "A driving performance forecasting system based on brain dynamic state analysis using 4-d convolutional neural networks," *IEEE Transactions on Cybernetics*, pp. 1–9, 2020.

[26] R. Jabbar, M. Shinoy, M. Kharbeche, K. Al-Khalifa, M. Krichen, and K. Barkaoui, "Driver drowsiness detection model using convolutional neural networks techniques for android application," in *2020 IEEE International Conference on Informatics, IoT, and Enabling Technologies (ICIoT)*, Conference Proceedings, pp. 237–242.

[27] Z. Li, Q. Yang, S. Chen, W. Zhou, L. Chen, and L. Song, "A fuzzy recurrent neural network for driver fatigue detection based on steering-wheel angle sensor data," *International Journal of Distributed Sensor Networks*, vol. 15, no. 9, p. 1550147719872452, 2019.

[28] C. N. Watling, M. Mahmudul Hasan, and G. S. Larue, "Sensitivity and specificity of the driver sleepiness detection methods using physiological signals: A systematic review," *Accident Analysis & Prevention*, vol. 150, p. 105900, 2021. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0001457520317206

[29] L. Guarda, N. Astorga, E. Droguett, M. Moura, and M. Ramos, "Drowsiness detection using electroencephalography signals: A deep learning based model," *Proceedings of the Probabilistic Safety Assessment and Management PSAM, Los Angeles, CA, USA*, vol. 16, 2018.

[30] O. Maftukhaturrizqoh, N. Nuryani, and D. Darmanto, "Drowsiness detection using radial basis function network with electrocardiographic rr interval statistical feature," in *Journal of Physics: Conference Series*, vol. 1153, no. 1. IOP Publishing, 2019, p. 012049.

[31] M. García-García, A. Caplier, and M. Rombaut, "Sleep deprivation detection for real-time driver monitoring using deep learning," in *International conference image analysis and recognition*. Springer, 2018, pp. 435–442.

[32] C. B. S. Maior, M. J. das Chagas Moura, J. M. M. Santana, and I. D. Lins, "Real-time classification for autonomous drowsiness detection using eye aspect ratio," *Expert Systems with Applications*, vol. 158, p. 113505, 2020.

[33] M. Ngxande, "Correcting inter-sectional accuracy differences in drowsiness detection systems using generative adversarial networks (gans)," Ph.D. dissertation, 2020.

[34] J. E. Manousakis, N. Mann, K. J. Jeppe, and C. Anderson, "Awareness of sleepiness: temporal dynamics of subjective and objective sleepiness," *Psychophysiology*, vol. 58, no. 8, p. e13839, 2021.

[35] D. Makowski, T. Pham, Z. J. Lau, J. C. Brammer, F. Lespinasse, H. Pham, C. Schölzel, and S. H. A. Chen, "NeuroKit2: A python toolbox for neurophysiological signal processing," *Behavior Research Methods*, vol. 53, no. 4, pp. 1689–1696, feb 2021.

[36] Z. Mu, J. Hu, and J. Min, "Driver fatigue detection system using electroencephalography signals based on combined entropy features," *Applied Sciences*, vol. 7, no. 2, p. 150, 2017.

[37] U. R. Acharya, H. Fujita, V. K. Sudarshan, S. Bhat, and J. E. Koh, "Application of entropies for automated diagnosis of epilepsy using eeg signals: A review," *Knowledge-Based Systems*, vol. 88, pp. 85–96, 2015. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0950705115003081

[38] J. M. Cori, C. Anderson, S. S. Soleimanloo, M. L. Jackson, and M. E. Howard, "Narrative review: Do spontaneous eye blink parameters provide a useful assessment of state drowsiness?" *Sleep medicine reviews*, vol. 45, pp. 95–104, 2019.

[39] J.-X. Ma, L.-C. Shi, and B.-L. Lu, "An eog-based vigilance estimation method applied for driver fatigue detection," *Neuroscience and Biomedical Engineering (Discontinued)*, vol. 2, no. 1, pp. 41–51, 2014.

[40] C. Anderson, A.-M. Chang, J. P. Sullivan, J. M. Ronda, and C. A. Czeisler, "Assessment of drowsiness based on ocular parameters detected by infrared reflectance oculography," *Journal of Clinical Sleep Medicine*, vol. 9, no. 9, pp. 907–920, 2013.

[41] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv preprint arXiv:1409.1556*, 2014.

[42] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," 2015.

[43] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural computation*, vol. 9, no. 8, pp. 1735–1780, 1997.

[44] P. Ebrahim, "Driver drowsiness monitoring using eye movement features derived from electrooculography," 2016.

[45] L. Oliveira, J. S. Cardoso, A. Lourenço, and C. Ahlström, "Driver drowsiness detection: a comparison between intrusive and non-intrusive signal acquisition methods," in *2018 7th European Workshop on Visual Information Processing (EUVIP)*, Conference Proceedings, pp. 1–6.

[46] R. Mahajan and B. I. Morshed, "Unsupervised eye blink artifact denoising of eeg data with modified multiscale sample entropy, kurtosis, and wavelet-ica," *IEEE Journal of Biomedical and Health Informatics*, vol. 19, no. 1, pp. 158–165, 2015.

[47] Y. Kartynnik, A. Ablavatski, I. Grishchenko, and M. Grundmann, "Real-time facial surface geometry from monocular video on mobile gpus," *arXiv preprint arXiv:1907.06724*, 2019.

[48] M. E. Vaca-Recalde, J. Pérez, and J. Echanobe, "Driver monitoring system based on cnn models: An approach for attention level detection," in *International Conference on Intelligent Data Engineering and Automated Learning*. Springer, 2020, pp. 575–583.

[49] M. Dua, R. Singla, S. Raj, A. Jangra *et al.*, "Deep cnn models-based ensemble approach to driver drowsiness detection," *Neural Computing and Applications*, vol. 33, no. 8, pp. 3155–3168, 2021.

[50] C. Anderson, S. Ftouni, J. M. Ronda, S. M. Rajaratnam, C. A. Czeisler, and S. W. Lockley, "Self-reported drowsiness and safety outcomes while driving after an extended duration work shift in trainee physicians," *Sleep*, vol. 41, no. 2, p. zsx195, 2018.