Physiological Signal-based Drowsiness Detection using Machine Learning: Singular & Hybrid Signal Approaches

Md Mahmudul Hasan[a,b,*], Christopher N. Watling[a,b], Grégoire S. Laruea[a,b]

a Queensland University of Technology (QUT), Centre for Accident Research and Road Safety - Queensland (CARRS-Q), Australia

b Queensland University of Technology (QUT), Institute of Health and Biomedical Innovation (IHBI), Australia

*Corresponding author: Md Mahmudul Hasan
K Block, 130 Victoria Park Road
Kelvin Grove, QLD, 4059, Australia
Email: mahmudul.hasan.eee.kuet@gmail.com

Keywords: drowsiness, features, machine learning, physiological signals, ground truth, sensitivity, specificity, accuracy

Highlights

- Current drowsiness detection models are not accurate enough to be implemented in vehicles
- Combining multiple biosignals was investigated in a laboratory study inducing drowsiness
- Important physiological features correlated to the Karolinska Sleepiness Scale were identified
- Approaches based on a singular signal result in trade-offs between sensitivity and specificity
- Hybrid approaches reduce this trade-off and result in an improved overall accuracy

Abstract

Introduction: Drowsiness is one of the main contributors to road-related crashes and fatalities worldwide. To address this pressing global issue, researchers are continuing to develop driver drowsiness detection systems which utilise a variety of measures. However, most research on drowsiness detection uses approaches based on a singular metric and, as a result, fail to attain satisfactory reliability and validity to be implemented in vehicles. Method: This study examines the utility of drowsiness detection based on singular and a hybrid approach. This approach considered a range of metrics from three physiological signals – electroencephalography (EEG), electrooculography (EOG), and electrocardiography (ECG) – and used subjective sleepiness indices (assessed via the Karolinska Sleepiness Scale) as ground truth. The methodology consisted of signal recording with a psychomotor vigilance test (PVT), pre-processing, extracting, and determining the important features from the physiological signals for drowsiness detection. Finally, four supervised machine learning models were developed based on the subjective sleepiness responses using the extracted physiological features to detect drowsiness levels. Results: The results illustrate that the singular physiological measures show a specific performance metric pattern, with higher sensitivity and lower specificity or vice versa. In contrast, the hybrid biosignal-based models provide a better performance profile, reducing the disparity between the two metrics. Conclusions: The outcome of the study indicates that the selected features provided higher performance in the hybrid approaches than the singular approaches, which could be useful for future research implications. Practical Applications: Use of a hybrid approaches seems warranted to improve in-vehicle driver drowsiness detection system. Practical applications will need to consider factors such as intrusiveness, ergonomics, cost-effectiveness, and user-friendliness of any driver drowsiness detection system.

1. Introduction

Drowsiness is an important factor for road crash deaths and injuries, with significant efforts required to mitigate its effects. A case study conducted in Sweden found approximately 8-15% of crashes were caused by drivers experiencing acute sleepiness (Kecklund et al., 2011), while other research has highlighted the influence of sleep disorders on road crashes (Komada et al., 2013). According to the Australian National Road Safety Strategy during the 2011–2020 session, drowsiness while driving is responsible for 20-30% of road crash deaths and acute injuries (Australian Transport Council, 2011). Although drowsiness during monotonous driving environments is not unusual (Larue et al., 2011), the consequences of drowsy driving can be as dangerous as the impairment from alcohol intoxication (Dawson & Reid, 1997).

Recent research on driver drowsiness detection has been categorised into three key aspects of data usage, namely vehicle-based measures, behavioural measures, and physiological measures (Sahayadhas et al., 2012). Physiological or biosignal-based measures are more reliable and accurate (Doudou et al., 2019; Ramzan et al., 2019), and have been found to outperform vehicle and behavioural measures as they provide fewer false positives and are less affected by environmental and road conditions (Ji et al., 2004; Zilberg et al., 2007). In this aspect, there are three main signals used with the detection of drowsiness, namely electroencephalography (EEG), electrooculography (EOG), and electrocardiography (ECG) (Balandong et al., 2018; Barua, Ahmed, Ahlström, et al., 2019).

The EEG is used to measure the electric potentials in the brain which are associated with distinct levels of arousal, such as wakefulness, drowsiness, and being asleep (Sahayadhas et al., 2012). Several features can be extracted from the EEG signal in time and frequency domains for drowsiness detection. Among all the metrics, alpha, theta and beta band power (Awais et al., 2014; Dkhil et al., 2015; Goovaerts et al., 2014) and the ratio of EEG band powers such as $(\alpha+\theta)/\beta$ and $(\alpha+\theta)/(\alpha+\beta)$ and β/α (Eoh et al., 2005; Lee et al., 2014) have been associated with changes in arousal. EOG is the electrophysical recording of eye movements, with distinct eye movements occurring during periods of drowsiness (Khushaba et al., 2011). EOG-based driver drowsiness detection techniques consist of extracting certain eye movements or features in the EOG signal which have been indicative of changes in drowsiness levels, specifically blink duration (BD) (Barua, Ahmed, Ahlström, et al., 2019; Naurois et al., 2017), blinking rate (BR) (Barbato et al., 2000; Nguyen et al., 2017), blink amplitude (BA) (Barbato et al., 2000; Galley et al., 2004; Svensson, 2004), peak closing velocity (PCV) (Johns, 2003), and the amplitude/peak closing velocity ratio (AVR) (Johns, 2003). The ECG measures the electrical activity of the heart, which is generated by the bioelectric currents flowing through the heart at different stages of blood flow (Gupta et al., 2017). The ECG time-domain features show significant variations due to drowsiness, such as Heart Rate Variability (HRV) indices (Tran et al., 2009). Among all the HRV metrics, heart rate (HR) (Zhang et al., 2013) and R-R interval (RRI) (Tran et al., 2009) are time domain features associated with changes in drowsiness. On the other hand, frequency-domain features of the ECG-signal such as low frequency (LF) and high frequency (HF) power (Apparies et al., 1998; Tran et al., 2009), as well as the LF/HF ratio (Abe et al., 2014; Mahachandra et al., 2012), are prominent features related to drowsiness.

It is worth noting the importance of using varied measures as the ground truth for the drowsiness detection system. Several different measures of sleepiness such as response times via drivers' steering wheel corrections (Ko et al., 2020), observers' ratings (Vicente et al., 2016), arbitrary classification (Chen et al., 2019; Min et al., 2017), microsleeps, lane crossing events (Y. Liang et al., 2019), subjective sleepiness scales (e.g. Karolinska Sleepiness Scale (KSS)) (Persson et al., 2020) and Visual Analogue Scales (VAS) (Ingre et al., 2006) have been used with varying success. For instance, performance on reaction time tasks can be influenced by a participant's motivation levels (Watling, 2016), observer ratings have been shown to be unreliable (Anund et al., 2013), and microsleeps and lane crossing events can be sporadic despite drowsiness causing visible impairments in a driver (Watling et al., 2016).

Subjective sleepiness levels are assessed via self-reports directly from an individual. A widely used scale of subjective sleepiness is the Karolinska Sleepiness Scale (KSS) (Åkerstedt et al., 2014), with scores on this scale associated with physiological measures (e.g., EEG) and vigilance tests (e.g., psychomotor vigilance test- PVT). Specifically, higher KSS values have been found to be associated with increased EEG-α and EEG-θ band powers (Kaida et al., 2006; Putilov & Donskaya, 2013), slower reaction times and lapses with the PVT (Gorgoni et al., 2014a), and prolonged eye blink durations (Anderson et al., 2013). These findings suggest that the KSS is a reliable measure of subjective sleepiness and aligns with other well-validated measures of sleepiness and performance outcomes (Åkerstedt et al., 2014). It has also been used in previous studies as a ground truth or reference variable (e.g., Barua, Ahmed, Ahlström, et al., 2019; Persson et al., 2020).

It should be noted, however, that there are a number of shortcomings associated with previous studies using physiological measures to detect drowsiness. For instance, most of these studies have utilised a set of predefined features without applying any feature selection techniques (Liang et al., 2019; Min et al., 2017; Saleab et al., 2016), and have also used a particular set of features with limited justification for their inclusion. The use of a large number of features is not practical as it would lead to high-dimensionality issues in the training data (Zoubek et al., 2007), which could further cause creating an overfitting problem (Liu et al., 2005). Moreover, using a higher number of features sometimes required extra sensors to be utilised, which increases the system cost (Nakisa, 2019). Thus, it is important to consider feature selection procedures to ensure on the most relevant features are employed.

Another issue relates to the selection of the classification model. A classification model can predict a class label for a given input data, after the model is trained. Machine learning (ML) based classification has been used widely because it can work with large dimensionality of data, reducing it to lower dimensions, which is very appropriate for real-life applications (Bolón-Canedo et al., 2016). Most of all, machine learning techniques are being widely used for the interpretation and analysis of biosignals like EEG, EOG and ECG (Barua, Ahmed, Ahlström, et al., 2019; Chen et al., 2018; Guo et al., 2016; Mårtensson et al., 2018; Min et al., 2017; Persson et al., 2020) for detection of driver drowsiness. However, as there is limited guidance with the choice of the classification model (Chen et al., 2018), various models have been used with no systematic consideration of the specific application. Previous studies considered factors such as complexity, computation cost and flexibility of the classification model (Chen et al., 2018; Shih et al., 2009). To this point, the classifiers K-Nearest Neighbours (Chen et al., 2018), Support Vector Machine (Guo et al., 2016), Random Forest Classifier (Mårtensson, Keelan, & Ahlström, 2018) and Artificial Neural Networks (Min et al., 2017) have been used extensively. Therefore, comparing the most widely used classification models and their subsequent performance outcomes can provide a measure of understanding with the utility of these classifiers.

According to a recent systematic review of physiological-based driver drowsiness detection systems (Watling et al., 2021), 81% of studies included in this review had used a singular physiological signal, with the majority obtaining high sensitivity with lower specificity, or vice versa. On the other hand, the few studies based on hybrid approaches consisting of multi-signal combinations have achieved more consistent and increased sensitivity and specificity outcomes. Given that drowsiness can be multifaceted, it is possible that hybrid approaches are more successful in detecting drowsiness levels. Notably, direct comparisons between singular and hybrid approaches are rarely performed, thus limiting our understanding of the relative benefits of these two approaches in detecting drowsiness.

Hence, in light of the findings above, the aim of this study was to examine the utility of drowsiness detection using the most important features from EEG, EOG and ECG biosignals based on subjective sleepiness indices (assessed via the KSS as ground truth) during a laboratory task. The feature selection was performed using two filter methods, and the drowsiness detection was performed using four supervised classification models.

2. Method

2.1. Participants

The sample included 35 young individuals (19 male and 16 female) who were, on average, 21 years of age (*SD* = 2.33; range: 17-25). Exclusion criteria were used to prevent individuals who were poor sleepers, shift workers, had a habitual bedtime later than 12 am (i.e., midnight), had a known sleep disorder/s, or suffered from excessive daytime sleepiness, or pharmacological derived sleepiness or arousal (i.e., illicit drugs, medications, excessive coffee, or alcohol consumption), from participating in the study and unduly influencing the results.

2.2. Measures

2.2.1. Demographic and sleep-related questionnaire

Demographic and sleep-related information was collected, including age, sex, duration of holding a licence, number of hours driven per week, actual sleeping time (in hours), and usual sleep-wake time for weekdays.

2.2.2. Karolinska Sleepiness Scale (KSS)

The KSS (Åkerstedt & Gillberg, 1990) is a measure of subjective measure of sleepiness. Participants were asked to rate their sleepiness level on a 9-point Likert scale ranging from 1 ("*extremely alert*") to 9 ("*very sleepy, great effort to keep awake, fighting sleep*"). Typically, participants provide a KSS rating every 5 minutes (e.g., Åkerstedt & Gillberg, 1990), although other durations have been used (e.g., Åkerstedt et al., 2014). The KSS is a reliable and valid measure of subjective sleepiness (Åkerstedt et al., 2014) and is correlated with PVT outcomes (Kaida et al., 2006).

2.2.3. Physiological measurement

The physiological measurements of drowsiness included EEG, EOG, and ECG. The physiological data were recorded using the BioRadio 150 device, a wireless data acquisition device that includes integrated wireless and scaling technologies, and a signal processing software called BioCapture (Cleveland Medical Devices Inc., 2006). Data was recorded at 600 Hz and a 50 Hz Notch filter was applied. In line with standard Polysomnography (PSG) recording procedures to ensure signal data transmission, the minimum impedance of the electrode was less than five kilo-ohms (5 kΩ).

EEG data signify the electrical activity of the brain, which can indicate variations in drowsiness levels. The EEG data were recorded from the electrode sites C3-A2 and O1-A2 using Ag-Al electrodes, as these two sites have been shown to be more sensitive to variations in drowsiness (Eoh et al., 2005; Lee et al., 2014; Watling et al., 2014) than other electrode sites (Sanei & Chambers, 2008). Several frequency-based features were extracted from the EEG data. Following spectral analysis three EEG bands, being EEG-α, EEG-θ and EEG-β power spectra were calculated for the data analysis. Specifically, increases in EEG-α, EEG-θ are indicative of increases of drowsiness, whereas EEG-β is indicative of alertness and concentration, such that beta levels decrease during periods of drowsiness (Eoh et al., 2005). The combination of these frequency bands into band power ratios, i.e. $(\alpha+\theta)/\beta$ and $(\alpha+\theta)/(\alpha+\beta)$ and β/α have also been found sensitive to changes in drowsiness (Eoh et al., 2005; Lee et al., 2014).

Electrooculography (EOG) is the recording of the electrical potential between the cornea (positive charge) and retina (negative charge). The difference in potential signifies the eye movement or orientations, which is used to identify driver drowsiness (Hu & Zheng, 2009; Khushaba et al., 2011; Kurt et al., 2009). The electrode placement for the EOG recording used a vertical placement in line with the pupil. Several features from the EOG data were extracted. The time duration between the two blinks is referred to as blink duration (BD) (Barua, Ahmed, Ahlström, et al., 2019; Naurois et al., 2017) and the number of eye blinks per minute blinking rate (BR) (Barbato et al., 2000; Nguyen et al., 2017). Blink amplitude (BA) is the measure of the potential difference in µV unit (Barbato et al., 2000; Galley et al., 2004; Svensson, 2004), whereas the peak closing velocity (PCV) is defined as the highest speed achieved during the closing period of the eye (Johns, 2003). The above features with the amplitude/peak closing velocity ratio (AVR) (Johns, 2003).

The electrocardiogram (ECG) records the electrical potential resulted from the bioelectric current flowing through the heart at different stages of the cardiac cycle (Bonjyotsna & Roy, 2014; Gupta et al., 2017). For the ECG recording, the modified two-lead recording system was utilised. One electrode was placed approximately 3-5 cm below the collarbone, and the second electrode was placed on the left lower ribcage V6 location. The cardiac cycles obtained from the ECG signals are assigned as P, Q, R, S, and T, to represent the altered phases of a heartbeat. The time interval between two adjacent QRS complexes is referred to as the R-R interval (RRI) (Gupta et al., 2017) and the number of heartbeats per unit of time (minute) is measured by heart rate (HR) (Zhang et al., 2013). The ECG signal power at low frequency (LF) (0.04 - 0.15 Hz) and high frequency (HF) (0.15 - 0.4 Hz) (Apparies et al., 1998; Tran et al., 2009), as well as the LF/HF ratio (Abe et al., 2014; Mahachandra et al., 2012) are prominent features related to drowsiness and were extracted from the ECG data.

2.2.4. Psychomotor Vigilance Test (PVT)

It is important to use vigilance tests when conducting studies on drowsiness (Gorgoni et al., 2014a). The PVT (Dinges & Powell, 1985) is a reaction-time task requiring participants to respond quickly to the presentation of a stimulus (Dinges & Powell, 1985). The PVT was chosen as the stimuli for the laboratory task. A number of studies have demonstrated drowsiness increase while performing 5-, 10- and 30-minute versions of the PVT (Anderson, Wales, & Horne, 2010; Gorgoni et al., 2014b; Loh et al., 2004b). In line with Anderson and Horne (2006), Anderson, Wales and Home (2010), and Loh et al. (2004a), an extended 30-minutes PVT was utilised in this study. Increases in drowsiness can be measured through increases in reaction times (i.e., greater than 500 ms) and greater frequency of lapses (Gorgoni et al., 2014b).

2.3. Experimental Procedure

Ethical clearance was obtained from the Queensland University of Technology (QUT) Human Research Ethics Committee. Eligible participants attended two sessions in the study laboratory. The first session was an intake-screening session. Eligible participants were given an actigraph (Actiwatch®-2: Mini Mitter Co., Inc.) to wear and a simple sleep diary to fill out, for at least five days before the experiment to monitor their habitual sleep and wake patterns and ensure compliance with the study protocols. During the second session, participants arrived at the laboratory at 14:00 and were seated comfortably in front of a computer screen to complete some demographic and sleep-related questionnaires while the physiological electrodes were applied. Participants were then familiarised with the KSS and PVT task and, once comfortable, completed a 30-minute PVT task during which they rated their KSS levels every 5 minutes. A webcam was used in the laboratory to observe the participants actions in real-time (to ensure they were performing the tasks properly) throughout the study while the observer was located in an adjacent room.

2.4. Data Preprocessing and Analysis

The raw signals collected from the data acquisition device were stored for processing, analyses, and feature extraction (see Figure 1). The signals acquired usually contain noise or artefacts from different origins. Eyeblink artefacts were removed from the EEG data using independent component analysis contained in Acqknowledge®-4 software (Bipoac Systems, Goleta, CA, USA). In addition, all signals were visually inspected for artefacts, including movement artefact, from which the specific epoch was excluded from the data processing. Finite Impulse Response (FIR) bandpass filter (LF: 0.3 Hz, HF: 35 Hz (EEG/EOG), 70 Hz (ECG)) with a Hanning window was used for the filtering purpose, as this filtering window gives the best Signal to Noise ratio (SNR) (Hassan et al., 2015).

2.4.1. Feature Extraction and Dataset Design

Observations of the participant via the webcam noted some participants were not actively engaged in the task. These included some participants who were only intermittently responding to the on-screen stimuli and were displaying some extreme signs of sleepiness, while some participants appeared highly distracted and were not focusing or attending to the on-screen stimuli as required to complete the PVT. As such, these participants acquired a greater number of lapses (< 25) in any 5-minute period of the PVT task and were subsequently excluded from the main study. The remaining 26 participants' data were used for the subsequent analysis. In total, 22 features were extracted from EEG, EOG and ECG signals based on the existing literature, using a 5-second epoch length (except blinking rate from EOG and heart rate from ECG).

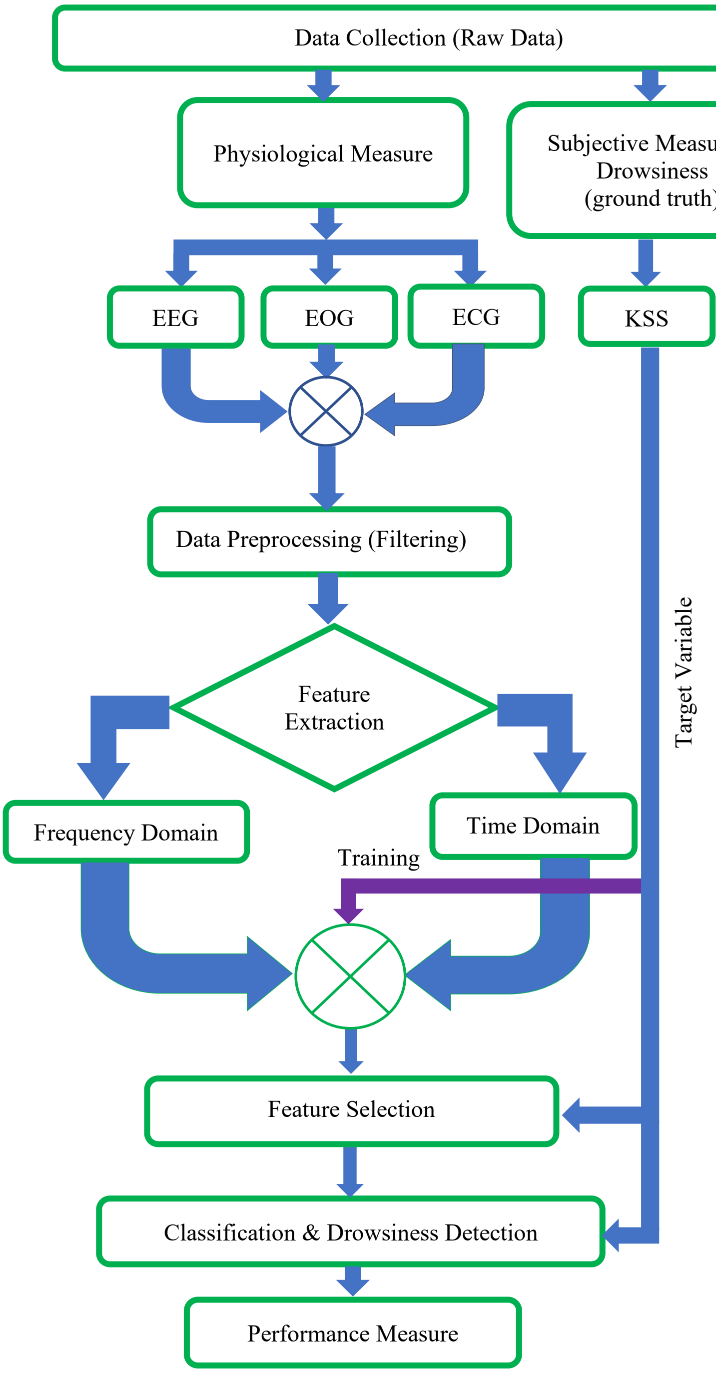


Figure 1: *The proposed model for data analysis in the current study. EEG, electroencephalography; EOG, electrooculography; ECG, electrocardiography; KSS, Karolinska Sleepiness Scale*

2.4.2. Feature Selection

The feature selection procedure was performed on the 22 extracted EEG, EOG, and ECG features, as irrelevant features can reduce the accuracy of classification models (Chandrashekar & Sahin, 2014). Despite the availability of several types of feature selection methods, filter methods succeed in various real-world datasets as they are independent of any learning algorithms (Das, 2001; Suto et al., 2016). After considering the computational complexity and reviewing previous biosignal-based sleep studies (Das, 2001; Suto et al., 2016), this study utilised two feature selection techniques: ANOVA F Test and correlation-based feature selection algorithm. The results from the two methods were aggregated into a final list to ensemble features based on the stability selection technique (Dittman et al., 2013; Haury et al., 2011). According to this technique, an appropriate threshold value is chosen for all the features from different measurements based on a specific threshold. Then the features exceeding the threshold values are given a single point, and the remaining features are assigned zero points, and thus the most important features are selected.

2.4.3. Classification & Drowsiness Level Detection

The five minutes preceding the KSS rating was considered as the KSS level for that specific session and was regarded as the ground truth – this approach is consistent with contemporary research (e.g., Åkerstedt et al., 2010; Barua, Ahmed, Ahlstrom, et al., 2019; Martensson et al., 2019; Persson et al., 2020). Akerstedt et al.'s (2008) crash risk study noted a near exponential relationship with increased crash risk due to drowsiness as calculated with the Sleep/Wake Predictor (SWP) being a biomathematical model of circadian and homeostatic sleep factors. The study results indicated that KSS levels of 7 to 9 are strongly associated with crashing due to drowsiness. Given the exponential relationship, KSS scores of 6 and less correspond to much smaller crash risk estimates. Thus, the current study dichotomised the KSS values into awake and drowsy, using KSS scores of 2 to 6 and 7 to 9, respectively.

There are no uniform or generalised models used for drowsiness classification to date. Considering the distinct computational cost and complexity (Mårtensson et al., 2018; Persson et al., 2020) and based on the current literature, four supervised learning models were applied for classification, namely k-nearest neighbour (KNN), support vector machine (SVM), random forest classifier (RFC), and artificial neural networks (ANN). The hyperparameters of each classifier were set using consecutive iteration process for getting the best accuracy after cross-validation. For the KNN classifiers, a set of the best value of 'k' was used to train the model, and the 'k value' corresponding to the best training accuracy was chosen. For the Gaussian SVM, 'radial basis function (RBF) kernel' was used, 'grid-search' was done for 'C' and 'γ' parameters using the 'grid-search' algorithm, and the best parameter values were chosen for the best accuracy after cross-validation. For the random forest classifier models, a grid-search was done on the 'n_estimate' and 'max-depth' to tune the number of trees in the forest and the maximum depth of the tree, respectively. Number of hidden layers and the number of neurons in hidden layers were tuned for ANNs using the 'grid-search' approach. All the models were implemented, trained, and tested using Python version 3.6.7.

During the classification phase, the four classification models were fitted for each of the three singular biosignals (EEG, EOG, and ECG), as well as for the three dual and one triple combinations of the biosignals (hybrid approach).

2.4.4. Performance Measures and Cross-validation

The machine learning classifiers' performance was assessed on their sensitivity or True positive Rate (TPR), specificity or True Negative Rate (TNR), and accuracy outcomes. All the classification performance were evaluated from the parameters found from the confusion matrix being- true positive (TP), true negative (TN), false positive (FP) and false negative (FN). The definition of the three metrics are given by the equations below (Zhu et al., 2010). The following equations provides the definition of sensitivity, specificity and accuracy, where, true positive (TP) implies the number of cases correctly identified as drowsy, false positive (FP) is the number of cases incorrectly identified as drowsy, true negative (TN) is the number of cases correctly identified as awake and false negative (FN) indicates the number of cases incorrectly identified as awake.

$$Sensitivity\ or\ TPR = \frac{TP}{TP+FN} \times 100\% \quad (1)$$

$$Specificity\ or\ TNR = \frac{TN}{TN+FP} \times 100\% \quad (2)$$

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \times 100\% \quad (3)$$

For the drowsiness detection system, a higher number of false alarms are caused if the system has a high sensitivity with a lower specificity. On the other hand, there is a higher chance of missing drowsy states if the system has higher specificity with relatively lower sensitivity. Therefore, considering the disparity (or absolute difference) between the two measures is an essential factor for the drowsiness detection system (Watling et al., 2021). Moreover, training and evaluating a model's performance on the same data can produce an over-optimistic result (Berrar, 2019; Danjuma, 2015). For all the classification models, 10-fold cross-validation was performed for the whole dataset to overcome this problem, and the mean and standard deviation of the ten experiments were reported as performance measures during the model evaluation.

3. Results

3.1. PVT Outcomes

In total, two PVT metrics were extracted to observe the effect of the task on subjective sleepiness levels (i.e., the KSS levels), being mean reaction time (msec) and lapses. As the

PVT lapses and response times have been validated as being associated with the subjective sleepiness levels, the mean lapses versus KSS scores for the participants were plotted as shown in Figure 2. The graph indicates that as the number of lapses and response time increases, KSS scores also increase. More specifically, the number of lapses gradually increased from KSS score 2-6, but the graph shows a more rapid increase in lapses from KSS 7-9. These findings add confidence to the binary classification of the drowsiness levels based on the KSS scores.

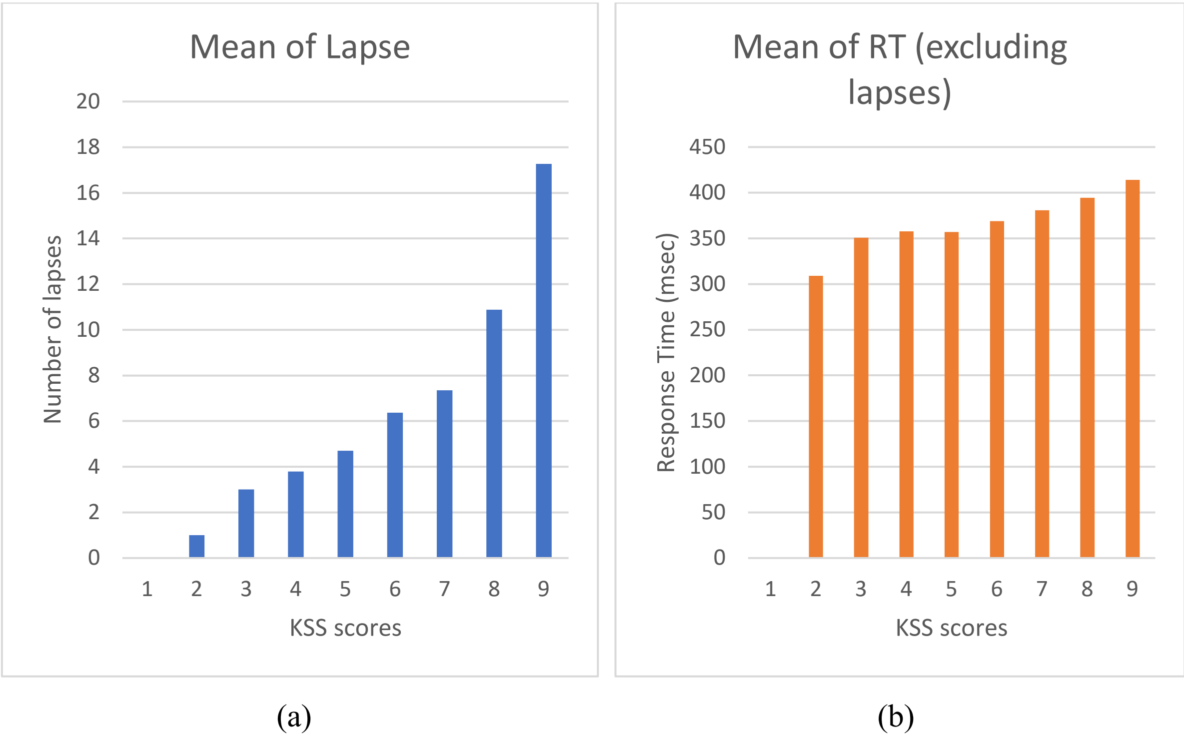
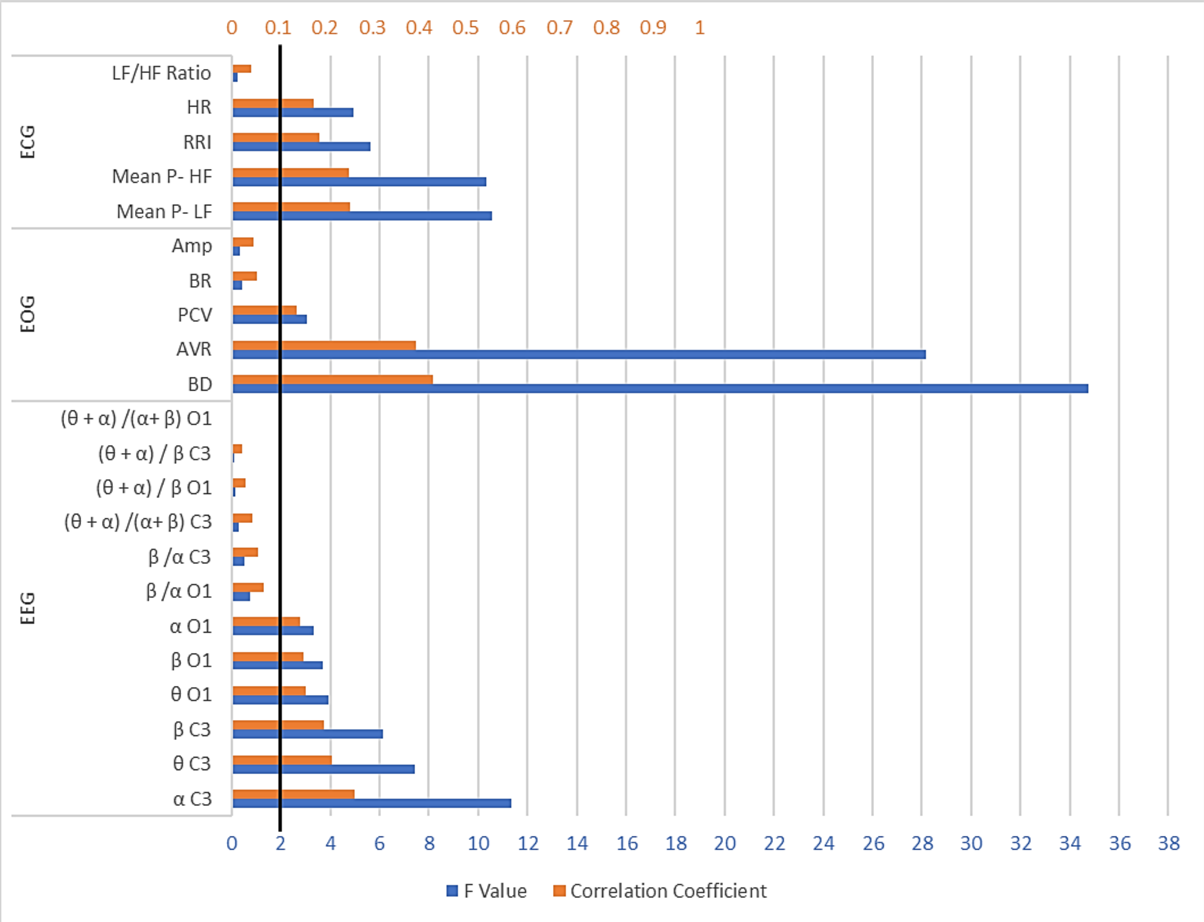


Figure 2: *Psychomotor Vigilance Test (PVT) outcomes (a) lapses and (b) response times (msec) versus subjective sleepiness levels*

3.2. Feature Selection

Stability selection technique (Dittman et al., 2013) was applied in this study (see section 2.4.2. for more details). At first a threshold value of 2.0 was applied as ANOVA F score threshold to select features, which is approximately 5% of the maximum F value obtained from the F based feature ranking. Applying this threshold selects 13 out of 22 features from all the biosignals. In the second step, for correlation coefficient based ranking, it required a threshold of 0.1 (approximately 25% of the maximum value) to include and exclude the same features which has been selected by ANOVA F Test. Given that the F score and correlation coefficients signifies two different measurements, still both satisfy the criteria of including 13 features out 22 having a score which are scored at least 5% of the maximum score of that specific methods. Based on that, a feature was included only if that feature score satisfy the criteria of having a score of 5% of the maximum score for both methods. Based on that inclusion criterion, finally the 13 features were selected for further analysis.



Note: α- C3, Alpha Central; θ-C3, Theta Central; β-C3, Beta Central; α- O1, Alpha Occipital; θ-O1, Theta Occipital; β-O1, Beta Occipital;BD, Blink Duration; AVR, Amplitude Velocity ratio; PCV, Peak Closing Velocity; BR, Blinking Rate; Amp, Amplitude; Mean P-LF, Mean Power at Low Frequency; Mean P-HF, Mean Power at High Frequency.

Table 1: *Features chosen from the physiological signals.*

Feature Rank	EEG	EOG	ECG
1	Alpha Central (α-EEG C3)	Blink Duration (BD)	Mean Power at Low Frequency (Mean P- LF)
2	Theta Central (θ-EEG C3)	Amplitude Velocity Ratio (AVR)	Mean Power at High Frequency (Mean P- HF)
3	Beta Central (β-EEG C3)	Peak Closing Velocity (PCV)	R-R Interval (RRI)
4	Theta Occipital (θ-EEG O1)		Heart Rate (HR)
5	Beta Occipital (β-EEG O1)		
6	Alpha Occipital (α-EEG O1)		

Note: EEG, electroencephalography; EOG, electrooculography; ECG, electrocardiography.

Finally, 13 features were chosen (out of initial 22 features) from EEG, EOG and ECG (see Table 1). The included EEG features were all of the fundamental frequency bands, while the

EOG had the BD, AVR and PCV, and ECG included the Mean P-LF and HF in the frequency domain, and HR and RRI in the time domain.

3.3. Drowsiness Detection

3.3.1. Classification Models

The classification performance (sensitivity, specificity, accuracy; mean ± standard error) from the four models for the singular and hybrid approaches is presented in Table 2. EEG and EOG showed similar performance for most classification models among the singular biosignals performance. For EEG and EOG, a specific pattern is visible, i.e., specificity is higher than the sensitivity, while ECG shows an opposite pattern. Lastly, EOG offers the best accuracy overall (73.9% for ANN), while ECG performs the worst (59.1% for ANN). In the hybrid biosignal based classification models, it is apparent that the triple combination offers the best performance in terms of sensitivity, specificity, and accuracy being 81.3%, 85.7% and 83.5%, respectively, for the ANN model. EOG and ECG's combination also gives some promising results (highest 70.7% accuracy for RFC). Considering all of the classifier models' average performance, as well as the accuracy as the base performance measure, it is evident that EOG outperforms the other singular biosignals. In contrast, the triple combinations outperform the dual combinations among the hybrid approaches.

The accuracy outcomes were separately plotted for all the classifiers against the seven different signal combinations (see Figure 4) to get a better insight into the classification models' accuracy measures and their consistency throughout the singular and hybrid approaches. The figure illustrates that all of the classification models show a similar pattern in terms of performance measures. Considering the overall performance measure in terms of accuracy, the ANN model outperforms the other three classification models (Figure 4). Therefore, the ANN was used as the base model for the rest of the paper to compare the singular and hybrid biosignals' performance for drowsiness detection.

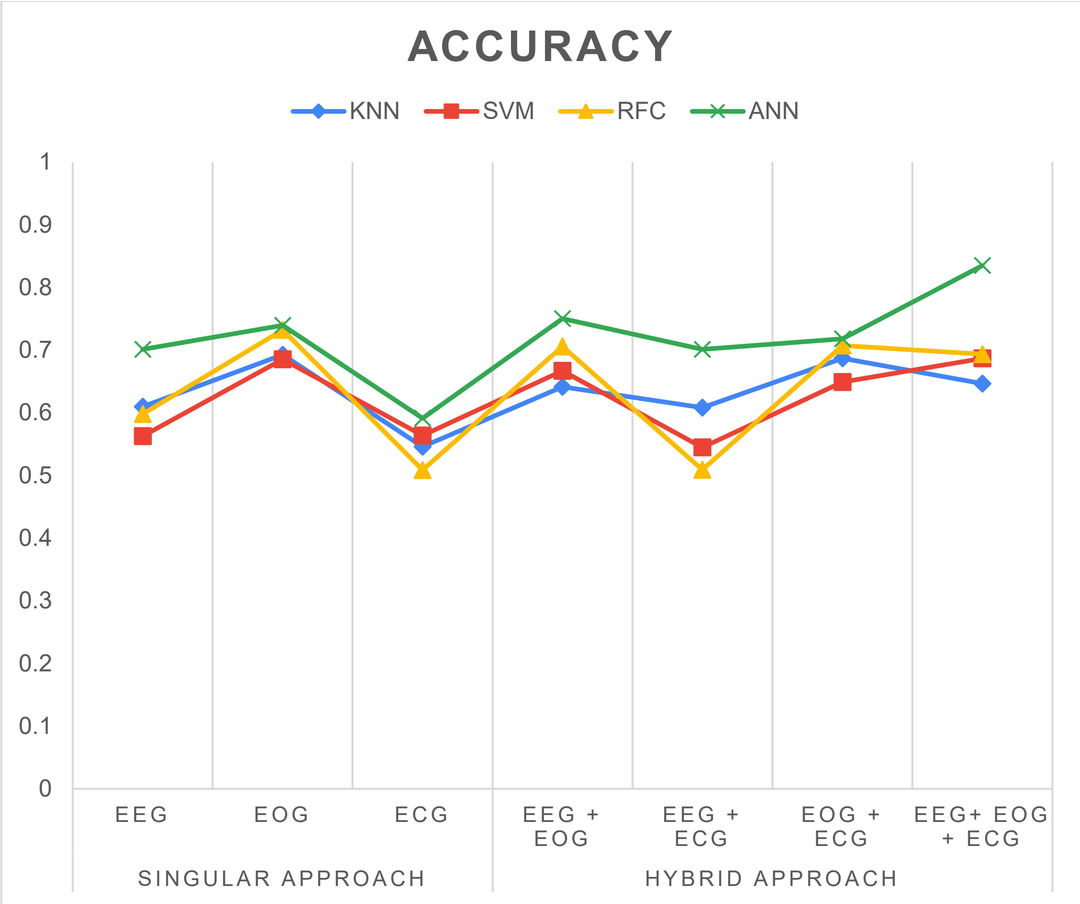


Figure 4: *Plots for each of the individual measures for singular and hybrid approach for four different classifiers. EEG, electroencephalography; EOG, electrooculography; ECG, electrocardiography.*

Table 2: *Classification performance for the four different classification models (test results, mean± standard deviation)*

Classifiers		Singular Approach			Hybrid Approach			
	Measures	EEG	EOG	ECG	EEG + EOG	EEG + ECG	EOG + ECG	EEG+ EOG + ECG
KNN	Sen	49.3 (+/- 16.9)	48.9 (+/- 25.6)	**81.8 (+/- 29.0)**	61.6 (+/- 17.4)	**73.9 (+/- 25.5)**	52.9 (+/- 20.3)	58.9 (+/- 17.3)
	Spec	72.9 (+/- 15.4)	**89.3 (+/- 17.5)**	26.6 (+/- 27.8)	66.3 (+/- 27.4)	47.1 (+/- 28.8)	**84.5 (+/- 15.5)**	70.2 (+/- 22.4)
	Acc	61.0 (+/- 10.6)	**69.3 (+/- 14.1)**	54.6 (+/- 23.0)	64.1 (+/- 13.7)	60.8 (+/- 23.3)	**68.7 (+/- 10.8)**	64.7 (+/- 12.8)
SVM	Sen	36.3 (+/- 21.9)	57.9 (+/- 18.9)	**78.9 (+/- 33.7)**	51.4 (+/- 23.7)	57.9 (+/- 24.0)	**64.1 (+/- 23.5)**	60.5 (+/- 15.9)
	Spec	76.8 (+/- 25.0)	**78.9 (+/- 15.7)**	33.0 (+/- 29.8)	**81.4 (+/- 17.4)**	50.7 (+/- 26.4)	65.2 (+/- 21.5)	76.4 (+/- 16.1)
	Acc	56.3 (+/- 14.7)	**68.5 (+/- 11.2)**	56.4 (+/- 28.0)	66.7 (+/- 9.80)	54.5 (+/- 14.7)	64.9 (+/- 11.3)	**68.7 (+/- 9.0)**
RFC	Sen	60.5 (+/- 16.9)	64.3 (+/- 22.1)	**76.8 (+/- 31.2)**	69.1 (+/- 17.9)	59.3 (+/- 18.9)	63.0 (+/- 21.1)	**70.5 (+/- 16.9)**
	Spec	58.2 (+/- 27.1)	**81.8 (+/- 19.2)**	24.5 (+/- 24.0)	71.4 (+/- 20.8)	42.1 (+/- 27.8)	**78.0 (+/- 18.9)**	67.7 (+/- 20.2)
	Acc	59.8 (+/- 13.1)	**73.2 (+/- 11.2)**	51.0 (+/- 24.1)	70.6 (+/- 13.1)	50.9 (+/- 19.2)	**70.7 (+/- 13.3)**	69.3 (+/- 14.1)
ANN	Sen	62.1 (+/- 18.3)	68.6 (+/- 13.7)	**69.8 (+/- 13.7)**	71.1 (+/- 14.7)	71.1 (+/- 12.3)	67.1 (+/- 16.8)	**81.3 (+/- 16.1)**
	Spec	78.0 (+/- 18.8)	**79.3 (+/- 29.3)**	54.6 (+/- 20.7)	78.9 (+/- 16.4)	69.1 (+/- 25.5)	76.4 (+/- 21.9)	**85.7 (+/- 21.2)**
	Acc	70.1 (+/- 11.3)	**73.9 (+/- 21.5)**	59.1 (+/- 13.8)	75.0 (+/- 12.1)	70.1 (+/- 15.6)	71.8 (+/- 16.5)	**83.5 (+/- 17.3)**
AVERAGE	Sen	52.1 (+/-18.5)	59.9 (+/-20.1)	**76.8 (+/-26.9)**	63.3 (+/-18.4)	65.6 (+/-20.2)	61.8 (+/-20.4)	**67.8 (+/-16.6)**
	Spec	71.5 (+/-21.6)	**82.3 (+/-20.4)**	34.7 (+/-25.6)	74.5 (+/-20.5)	52.3 (+/-27.1)	**76.0 (+/-19.5)**	75.0 (+/-20.0)
	Acc	61.8 (+/-12.4)	**71.2 (+/-14.5)**	55.3 (+/-22.2)	69.1 (+/-12.2)	59.1 (+/-18.2)	69.0 (+/-13.0)	**71.6 (+/-13.3)**

Note: Bolded values represent the largest obtained value. EEG, electroencephalography; EOG, electrooculography; ECG, electrocardiography

3.3.2. Singular Versus Hybrid Approach

Considering ANN as the best performing model, all the singular and hybrid approaches' classification performance is shown in Table 3. Putting accuracy as the performance measure base, EOG outperforms all of the singular biosignals based approaches (accuracy= 73.9%). In contrast, ECG performs the worst (accuracy= 59.1%). It is worth noting that EEG shows an accuracy of 70.1%, which is very close to EOG performance.

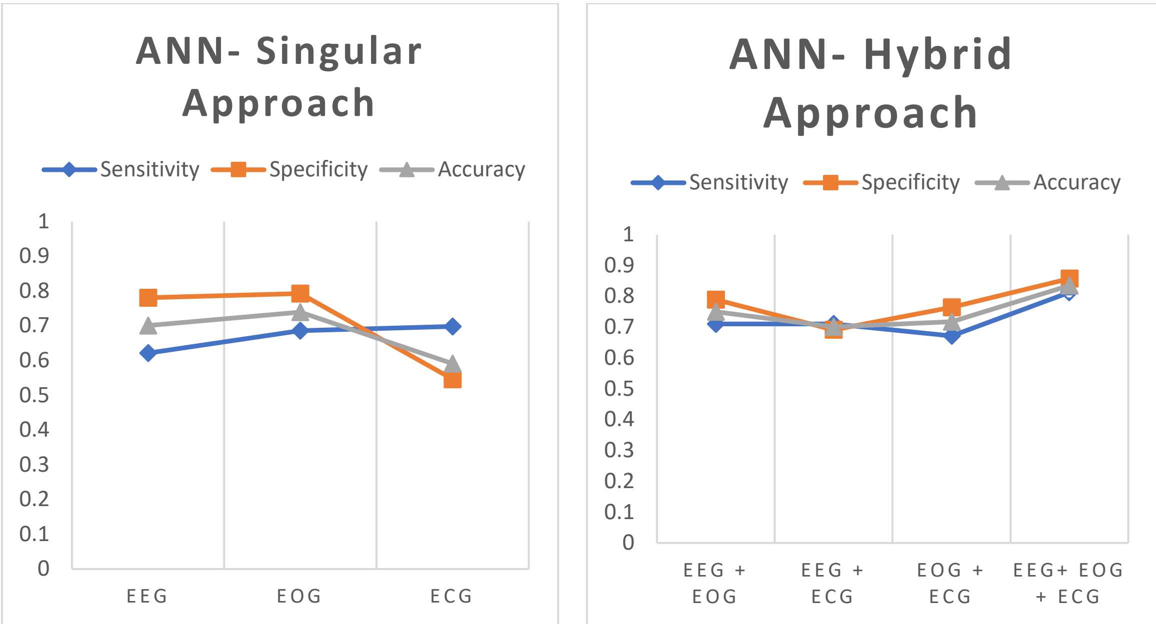
Though ECG solely shows lower accuracy (59.1%) than EOG (73.9%) and EEG (70.1%) among the singular approaches, the combined effect of ECG with EOG and EEG gives better accuracy in the hybrid approaches, being 71.8% and 70.1%, respectively. However, among the two-paired hybrid approaches, the combination of top-performing singular biosignals (EEG + EOG) performs relatively higher (accuracy= 75.0%). Additionally, the accuracy was the highest (83.5%) when the top-performed two-paired approach (EEG + EOG) was further combined with ECG to make a triple combination (EEG + EOG + ECG). The mean performance from singular and hybrid measures also reflects the hybrid approaches' better overall performance compared to the singular approaches.

Table 3: *The classification performance measures for ANN (the bold numbers signify the highest accuracy in singular and hybrid approach)*

Approaches	BioSignals	Number of included features	Sensitivity	Specificity	Absolute difference (sensitivity-specificity)	Accuracy
Singular	EOG	3	68.6%	79.3%	10.7%	**73.9%**
	ECG	4	69.8%	54.6%	15.2%	59.1%
	EEG	6	62.1%	78.0%	15.9%	70.1%
	AVERAGE		66.8%	70.7%	13.9%	67.7%
Hybrid	EOG + ECG	7	67.1%	76.4%	9.3%	71.8%
	EEG + EOG	9	71.1%	78.9%	7.8%	75.0%
	EEG + ECG	10	71.1%	69.1%	2.0%	70.1%
	EEG + EOG + ECG	13	81.3%	85.7%	4.4%	**83.5%**
	AVERAGE		72.7%	77.5%	5.9%	75.1%

Note: EEG, electroencephalography; EOG, electrooculography; ECG, electrocardiography; ANN ANN, artificial neural networks.

Considering the ANN, a plot was performed (see Figure 5 (a)), to observe the absolute difference among the performance metrics (sensitivity, specificity, and accuracy). From the plot of the singular approaches, it is evident that there is a higher disparity in performance between the sensitivity and specificity while using all of the three biosignals individually, being 15.9%, 15.2%, and 10.7% respectively for EEG, ECG, and EOG. From the plots of the hybrid approaches (see Figure 5 (b)), it is evident that the triple combination of EEG + EOG + ECG signals provides the best accuracy (83.5%) with a relatively lower disparity (4.4%). Overall, the findings indicate a higher disparity between sensitivity and specificity in singular approaches, which was further reduced in hybrid strategies.



(a) (b)

Figure 5: *Discrepancy plot of performance metrics for (a) singular (EEG, EOG and ECG) and (b) hybrid approaches for the best performing classifier ANN to observe the discrepancy of the sensitivity, specificity and accuracy measures.*
EEG, electroencephalography; EOG, electrooculography; ECG, electrocardiography; ANN, artificial neural networks.

4. Discussion

4.1. Singular Versus Hybrid Approaches

The research aimed to classify the drowsy and awake states with singular and hybrid combinations of biosignals through comparing the resultant performance. There are two main findings from this study. First, the individual signals show a specific trend or pattern of sensitivity and specificity. Second, the singular approaches (EEG/EOG/ECG) show a larger disparity between sensitivity and specificity, with this disparity minimised by dual or triple combinations of the biosignals, which was useful for improving the overall accuracy.

The results show EOG to be the most effective biosignal after comparing the singular biosignal based models related to the performance measures. The current finding is in line with the study performed by Wilkinson et al. (2013) which achieved good performance outcomes (sensitivity: 77.0-100.0% and specificity: 83.0-85.0%) through the use of an Optalert® device based on the ocular parameters. One potential explanation of the comparatively higher performance of EOG could be the highest correlations of blink duration and peak closing velocity samples with the associated KSS values (see the result section 3.2). It is worth noting that the ECG feature ranking procedure produced features that reached the threshold value; however, these ECG features produced the worst performance outcomes. The mean power at LF and HF (pair-1) and HR and RRI (pair-2) had a higher correlation indices. Thus, adding those four features with two pairs provided less utility to the machine learning models, as each of the two pairs' features are highly correlated. That signifies that incorporating features which are highly inter-correlated does not help in improving the classification performance. Second, there was a specific pattern evident from observing the individual biosignals' sensitivity and specificity index. EEG and EOG show higher specificity than sensitivity, which indicates that these two signals can identify the awake stage better than the drowsy stage. These findings are consistent with the findings from several studies. For example, Poorna et al. (2018) used EEG spectral parameters and achieved 44.4% sensitivity, and 70.6% specificity in classifying drowsy/awake stages using ANN classifier. On the other hand, ECG shows the opposite trend with comparatively lower specificity and a higher sensitivity index, indicating ECG can identify the drowsy stage more correctly than the awake stage. This finding favours the study performed by Maftukhaturrizqoh et al. (2019) (sensitivity=84.8%, specificity=76.9%).

The ultimate goal of this research was to investigate combined features from multiple biosignals to observe the hybrid performance compared to the singular ones. The main findings from the hybrid biosignal based classification are the reduced disparity between the sensitivity and specificity metrics and improved overall performance, in terms of accuracy, in the hybrid models utilising singular biosignals. These findings are consistent with the outcomes from previous studies (Balandong et al., 2018; Khushaba et al., 2013; Ramzan et al., 2019). For instance, the study by Khushaba et al. (2013) combined a total of 30 features obtained from 3 EEG channels, 1 EOG channel and 1 ECG channel features, and achieved 93.0% accuracy using Fuzzy Neighborhood Preserving Analysis (FNPA) method. It is worth noting that significant differences were found with regards to age, gender, types of participants (i.e. normal, partially or fully sleep-deprived), experimental settings (laboratory, driving simulator and real-world driving) and ground truth (subjective/objective measures of sleepiness) (Watling et al., 2021), and thus suffer from a lack of generalisability (Chowdhury et al., 2018).

4.2. Machine Learning Techniques

This study utilised four different supervised machine learning techniques for solving the drowsiness detection problem. As the machine learning models' performance depends on the nature of the data, it is difficult to estimate a predominately generalised model for a specific problem. However, different models work on the different working principle, which could provide a deep insight into its performance. In this study, the ANN classifiers provided the best overall results. That is, with all of the biosignals combined, the sensitivity, specificity, and accuracy outcomes were all above 80.0%. On the other hand, the KNN, SVM, and RFC classifiers outcomes were less than 75.0%. The current findings are consistent with previous research that utilised the same classifiers and found the ANN to achieve the highest overall performance (Larue et al., 2015; Min et al., 2017). This finding is potentially due to the ANN's ability to learn the complex and non-linear relationship, especially from real-world data (Almeida, 2002). This is probably due to the ANN’s ability to effectively map salient information end-to-end through hidden layers by using backpropagation techniques and a predetermined number of epochs, while simultaneously updating weights at each iteration in the forward and backward pass (Setiono & Liu, 1998). In addition to this, the generalisation capacity of neural network with backpropagation also demonstrates reliability (Kotu & Deshpande, 2014). Thus, such characteristics of ANN’s outperform the state-of-the-art methods in terms of overall accuracy, while also maintaining a good trade-off between the sensitivity and the specificity values.

4.3. Practical Applications

The implications of the research findings can be highlighted in three main points. First, it is important to consider using the most important features. Specifically, as the selected 13 features (EEG, EOG, and ECG) have demonstrated utility in reducing the disparity between the sensitivity and specificity, it is highly recommended that they are utilised in future drowsiness detection systems, either through a singular or hybrid approach. Second, it is also important to consider using singular biosignals to improve the overall performance of the hybrid biosignals. Specifically, as the singular physiological signals possess a specific pattern (higher sensitivity with lower specificity and vice-versa) in their performance metrics, their contribution could improve classification accuracy and minimise the disparity between sensitivity and specificity of hybrid biosignal-based detection systems. Finally, utilising hybrid combinations of the biosignals is highly recommended for future research, given that they possess a better profile in terms of performance metrics. The research findings would help in the future implementation of a robust in-vehicle driver drowsiness detection system. Alternatively, if such systems must remain in the lab, those could be used as a ground truth when investigating other metrics that can be obtained from the vehicle. Nonetheless, intrusiveness is a concerning issue surrounding the use of multiple biosignals for drowsiness detection (Sahayadhas et al., 2012). However, this can be potentially overcome by using contactless electrodes, for example.

4.4. Limitations and Future Research Direction

The current study possesses limitations that need to be considered when interpreting the results. First, this is a lab-based study with no driving involved. Further tests should be done, for instance, using driving simulations or real-world driving scenarios. Second, the sample size is small in this study and, as a result, potentially limits the generalisability of the findings. Third, inter-individual differences are a potential issue for generalising drowsiness detection models (Barua, Ahmed, Ahlström, et al., 2019; Naurois et al., 2017). Additionally, there is a significant variation of the physiological signals and cognitive performance within and among the individuals over time. Future consideration could be given to the use of deep learning approaches and the adding of contextual information (such as sleep-awake time and environmental conditions) to the biosignal features for more robust drowsiness detection using biosignals.

5. Conclusion

The findings from the study demonstrate that certain features extracted from the physiological signals are associated with KSS scores. Furthermore, including the selected features from singular physiological signals illustrate a particular pattern in the sensitivity and specificity profile. On the other hand, the hybrid approaches showed a relatively lower disparity between the sensitivity and specificity than the singular approaches. These outcomes demonstrate the utility of singular biosignals in improving the performance measures of the hybrid biosignal-based drowsiness detection system. Future research is needed for real-world driving, the value of the model over time, and inter-individual differences. In conclusion, the physiological signal-based detection system is more accurate and reliable than the existing systems. Therefore, further investigation should be conducted on real-time detection systems and the appropriate market solutions with the proper fusion of the biosignals, all while considering intrusiveness, ergonomics, cost-effectiveness, and user-friendliness.

6. Acknowledgement

This research project was supported by the Australian Government funded Research Training Program Scholarship. The authors would like to sincerely thank Sonali Nandavar for her diligent proofreading of this article.

7. References

Abe, E., Fujiwara, K., Hiraoka, T., Yamakawa, T., & Kano, M. (2014). Development of drowsy driving accident prediction by heart rate variability analysis. *Signal and Information Processing Association Annual Summit and Conference (APSIPA)*, 1-4.

Åkerstedt, T., Anund, A., Axelsson, J., & Kecklund, G. (2014). Subjective sleepiness is a sensitive indicator of insufficient sleep and impaired waking function. *Journal Of Sleep Research, 23*(3), 240-252.

Åkerstedt, T., Connor, J., Gray, A., & Kecklund, G. (2008). Predicting road crashes from a mathematical model of alertness regulation—The Sleep/Wake Predictor. *Accident Analysis & Prevention, 40*(4), 1480-1485.

Åkerstedt, T., & Gillberg, M. J. I. J. o. N. (1990). Subjective and objective sleepiness in the active individual. *International Journal of Neuroscience, 52*(1-2), 29-37.

Åkerstedt, T., Ingre, M., Kecklund, G., Anund, A., Sandberg, D., Wahde, M., Philip, P., & Kronberg, P. (2010). Reaction of sleepiness indicators to partial sleep deprivation, time of day and time on task in a driving simulator--the DROWSI project. *Journal Of Sleep Research, 19*(2), 298-309.

Almeida, J. S. (2002). Predictive non-linear modeling of complex data by artificial neural networks. *Current opinion in biotechnology, 13*(1), 72-76.

Anderson, C., Chang, A.-M., Sullivan, J. P., Ronda, J. M., & Czeisler, C. A. (2013). Assessment of drowsiness based on ocular parameters detected by infrared reflectance oculography. *J Clin Sleep Med, 9*(9), 907-920B.

Anderson, C., & Horne, J. A. (2006). Sleepiness Enhances Distraction During a Monotonous Task. *Sleep, 29*(4), 573-576.

Anderson, C., Wales, A. W. J., & Home, J. A. (2010). PVT lapses differ according to eyes open, closed, or looking away. *Sleep, 33*(2), 197-204.

Anderson, C., Wales, A. W. J., & Horne, J. A. (2010). PVT Lapses Differ According To Eyes Open, Closed, or Looking Away. *Sleep, 33*(2), 197-204.

Anund, A., Fors, C., Hallvig, D., Åkerstedt, T., & Kecklund, G. (2013). Observer Rated Sleepiness and Real Road Driving: An Explorative Study. *Plos ONE, 8*(5), e64782.

Apparies, R. J., Riniolo, T. C., & Porges, S. W. (1998). A psychophysiological investigation of the effects of driving longer-combination vehicles. *Ergonomics, 41*(5), 581-592.

Australian Transport Council. (2011). *National Road Safety Strategy 2011-2020.* https://roadsafety.gov.au/nrss/files/NRSS_2011_2020.pdf

Awais, M., Badruddin, N., & Drieberg, M. (2014). Driver drowsiness detection using EEG power spectrum analysis. *2014 IEEE Region 10 Symposium*, 244-247.

Balandong, R. P., Ahmad, R. F., Saad, M. N. M., & Malik, A. S. (2018). A Review on EEG-Based Automatic Sleepiness Detection Systems for Driver. *IEEE Access, 6*, 22908-22919.

Barbato, G., Ficca, G., Muscettola, G., Fichele, M., Beatrice, M., & Rinaldi, F. (2000). Diurnal variation in spontaneous eye-blink rate. *Psychiatry Res, 93*(2), 145-151.

Barua, S., Ahmed, M. U., Ahlstrom, C., & Begum, S. (2019). Automatic driver sleepiness detection using EEG, EOG and contextual information. *Expert Systems With Applications, 115*, 121-135.

Barua, S., Ahmed, M. U., Ahlström, C., & Begum, S. (2019). Automatic driver sleepiness detection using EEG, EOG and contextual information. *Expert Systems with Applications, 115*, 121-135.

Berrar, D. (2019). Cross-validation. *Encyclopedia of Bioinformatics and Computational Biology, 1*, 542-545.

Bolón-Canedo, V., Remeseiro, B., Alonso-Betanzos, A., & Campilho, A. (2016). Machine learning for medical applications. *European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning 2015*(1), 1-1.

Bonjyotsna, A., & Roy, S. (2014). Correlation of drowsiness with electrocardiogram: A review. *International Journal of Advanced Research in Electrical, Electronics Instrumentation Engineering, 3*(5), 9538-9544.

Chandrashekar, G., & Sahin, F. (2014). A survey on feature selection methods. *Computers & Electrical Engineering, 40*(1), 16-28.

Chen, J., Wang, H., & Hua, C. (2018). Assessment of driver drowsiness using electroencephalogram signals based on multiple functional brain networks. *International journal of psychophysiology, 133*, 120-130.

Chen, J., Wang, H., Wang, Q., & Hua, C. (2019). Exploring the fatigue affecting electroencephalography based functional brain networks during real driving in young males. *Neuropsychologia, 129*, 200-211.

Chowdhury, A., Shankaran, R., Kavakli, M., & Haque, M. M. (2018). Sensor applications and physiological features in drivers' drowsiness detection: A review. *IEEE Sensors Journal, 18*(8), 3055-3067.

Cleveland Medical Devices Inc. (2006). BioCapture User's Guide. 1-44.

Danjuma, K. J. (2015). Performance evaluation of machine learning algorithms in post-operative life expectancy in the lung cancer patients. *arXiv preprint arXiv:1504.04646*.

Das, S. (2001). Filters, wrappers and a boosting-based hybrid for feature selection. *LCML, 1*, 74-81.

Dawson, D., & Reid, K. J. N. (1997). Fatigue, alcohol and performance impairment. *Nature, 388*(6639), 235-237.

Dinges, D. F., & Powell, J. W. (1985). Microcomputer analyses of performance on a portable, simple visual RT task during sustained operations. *Behavior Research Methods, Instruments, & Computers, 17*(6), 652-655.

Dittman, D. J., Khoshgoftaar, T. M., Wald, R., & Napolitano, A. (2013). Classification performance of rank aggregation techniques for ensemble gene selection. *The Twenty-Sixth International FLAIRS Conference*.

Dkhil, M. B., Wali, A., & Alimi, A. M. (2015). Drowsy driver detection by EEG analysis using Fast Fourier Transform. *2015 15th International Conference on Intelligent Systems Design and Applications (ISDA)*, 313-318.

Doudou, M., Bouabdallah, A., & Berge-Cherfaoui, V. (2019). Driver Drowsiness Measurement Technologies: Current Research, Market Solutions, and Challenges. *International Journal of Intelligent Transportation Systems Research*, 1-23.

Eoh, H. J., Chung, M. K., & Kim, S.-H. (2005). Electroencephalographic study of drowsiness in simulated driving with sleep deprivation. *International Journal of Industrial Ergonomics, 35*(4), 307-320.

Galley, N., Schleicher, R., & Galley, L. (2004). Blink parameters as indicators of driver's sleepiness–possibilities and limitations. *Vision in vehicles, 10*, 189-196.

Goovaerts, G., Denissen, A., Milosevic, M., van Boxtel, G. J., & Van Huffel, S. (2014). Advanced EEG Processing for the Detection of Drowsiness in Drivers. *Biosignals*, 205-212.

Gorgoni, M., Ferlazzo, F., Ferrara, M., Moroni, F., D'Atri, A., Fanelli, S., Torriglia, I. G., Lauri, G., Marzano, C., & Rossini, P. M. (2014a). Topographic electroencephalogram changes associated with psychomotor vigilance task performance after sleep deprivation. *Sleep Medicine, 15*(9), 1132-1139.

Gorgoni, M., Ferlazzo, F., Ferrara, M., Moroni, F., D'Atri, A., Fanelli, S., Torriglia, I. G., Lauri, G., Marzano, C., & Rossini, P. M. J. S. m. (2014b). Topographic electroencephalogram changes associated with psychomotor vigilance task performance after sleep deprivation. *Sleep Medicine, 15*(9), 1132-1139.

Guo, M., Li, S., Wang, L., Chai, M., Chen, F., & Wei, Y. (2016). Research on the relationship between reaction ability and mental state for online assessment of driving fatigue. *Int J Environ Res Public Health, 13*(12), 1-15.

Gupta, N., Najeeb, D., Gabrielian, V., & Nahapetian, A. (2017). Mobile ECG-based drowsiness detection. *2017 14th IEEE Annual Consumer Communications & Networking Conference (CCNC)*, 29-32.

Hassan, M. A., Mahmoud, E. A., Abdalla, A. H., & Wedaa, A. M. (2015). A Comparison between Windowing FIR Filters for Extracting the EEG Components. *Journal of Biosensors & Bioelectronics, 6*(4), 1.

Haury, A.-C., Gestraud, P., & Vert, J.-P. (2011). The influence of feature selection methods on accuracy, stability and interpretability of molecular signatures. *PLoS One, 6*(12), e28210.

Hu, S. Y., & Zheng, G. T. (2009). Driver drowsiness detection with eyelid related parameters by Support Vector Machine. *Expert Systems With Applications, 36*(4), 7651-7658.

Ingre, M., Åkerstedt, T., Peters, B., Anund, A., & Kecklund, G. (2006). Subjective sleepiness, simulated driving performance and blink duration: Examining individual differences. *Journal of sleep research, 15*(1), 47-53.

Ji, Q., Zhu, Z., & Lan, P. (2004). Real-time nonintrusive monitoring and prediction of driver fatigue. *IEEE transactions on vehicular technology, 53*(4), 1052-1068.

Johns, M. J. S. (2003). The amplitude-velocity ratio of blinks: a new method for monitoring drowsiness. *Sleep, 26*(SUPPL.).

Kaida, K., Takahashi, M., Åkerstedt, T., Nakata, A., Otsuka, Y., Haratani, T., & Fukasawa, K. (2006). Validation of the Karolinska sleepiness scale against performance and EEG variables. *Clinical Neurophysiology, 117*(7), 1574-1581.

Kecklund, G., Anund, A., Rodling Wahlström, M., Philip, P., & Åkerstedt, T. (2011). Sleepiness and the risk of car crash: a case control study. *Statens väg-och Transportforskningsinstitut*, 1-54.

Khushaba, R., Kodagoda, S., Lal, S., & Dissanayake, G. (2013). Uncorrelated fuzzy neighborhood preserving analysis based feature projection for driver drowsiness recognition. *Fuzzy Sets and Systems, 221*, 90-111.

Khushaba, R. N., Kodagoda, S., Lal, S., & Dissanayake, G. (2011). Driver drowsiness classification using fuzzy wavelet-packet-based feature-extraction algorithm [Article]. *IEEE Transactions on Biomedical Engineering, 58*(1), 121-131, Article 5580017.

Ko, L.-W., Komarov, O., Lai, W.-K., Liang, W.-G., & Jung, T.-P. (2020). Eyeblink recognition improves fatigue prediction from single-channel forehead EEG in a realistic sustained attention task. *Journal of Neural Engineering, 17*(3), 036015.

Komada, Y., Asaoka, S., Abe, T., & Inoue, Y. (2013). Short sleep duration, sleep disorders, and traffic accidents. *IATSS Research, 37*(1), 1-7.

Kotu, V., & Deshpande, B. (2014). *Predictive analytics and data mining: concepts and practice with rapidminer*. Morgan Kaufmann.

Kurt, M. B., Sezgin, N., Akin, M., Kirbas, G., & Bayram, M. (2009). The ANN-based computing of drowsy level. *Expert Systems with Applications, 36*(2), 2534-2542.

Larue, G. S., Rakotonirainy, A., & Pettitt, A. N. (2011). Driving performance impairments due to hypovigilance on monotonous roads. *Accident Analysis & Prevention, 43*(6), 2037-2046.

Larue, G. S., Rakotonirainy, A., & Pettitt, A. N. J. I. P. C. (2015). Predicting reduced driver alertness on monotonous highways. *14*(2), 78-85.

Lee, B.-G., Lee, B.-L., & Chung, W.-Y. (2014). Mobile healthcare for automatic driving sleep-onset detection using wavelet-based EEG and respiration signals. *Sensors, 14*(10), 17915-17936.

Liang, Horrey, W. J., Howard, M. E., Lee, M. L., Anderson, C., Shreeve, M. S., O'Brien, C. S., & Czeisler, C. A. (2019). Prediction of drowsiness events in night shift workers during morning driving [Article]. *Accident Analysis & Prevention, 126*(1), 105-114.

Liang, Y., Horrey, W. J., Howard, M. E., Lee, M. L., Anderson, C., Shreeve, M. S., O'Brien, C. S., & Czeisler, C. A. (2019). Prediction of drowsiness events in night shift workers during morning driving. *Accident Analysis and Prevention, 126*, 105-114.

Liu, Dougherty, E. R., Dy, J. G., Torkkola, K., Tuv, E., Peng, H., Ding, C., Long, F., Berens, M., & Parsons, L. (2005). Evolving feature selection. *IEEE Intelligent systems, 20*(6), 64-76.

Loh, S., Lamond, N., Dorrian, J., Roach, G., & Dawson, D. (2004a). The validity of psychomotor vigilance tasks of less than 10-minute duration. *Behavior Research Methods, 36*(2), 339-346.

Loh, S., Lamond, N., Dorrian, J., Roach, G., & Dawson, D. (2004b). The validity of psychomotor vigilance tasks of less than 10-minute duration. *Behavior Research Methods, Instruments, & Computers, 36*(2), 339-346.

Maftukhaturrizqoh, O., Nuryani, N., & Darmanto, D. (2019). Drowsiness detection using radial basis function network with electrocardiographic RR interval statistical feature. *Journal of Physics: Conference Series, 1153*(1), 012049.

Mahachandra, M., Yassierli, Sutalaksana, I. Z., & Suryadi, K. (2012). Sensitivity of heart rate variability as indicator of driver sleepiness. *2012 Southeast Asian Network of Ergonomics Societies Conference (SEANES)*, 1-6.

Martensson, H., Keelan, O., & Ahlstrom, C. (2019). Driver Sleepiness Classification Based on Physiological Data and Driving Performance From Real Road Driving. *IEEE Transactions on Intelligent Transportation Systems, 20*(2), 421-430.

Mårtensson, H., Keelan, O., & Ahlström, C. (2018). Driver sleepiness classification based on physiological data and driving performance from real road driving. *IEEE Transactions on Intelligent Transportation Systems, 20*(2), 421-430.

Min, J., Wang, P., & Hu, J. (2017). Driver fatigue detection through multiple entropy fusion analysis in an EEG-based system. *PLoS ONE, 12*(12), e0188756.

Nakisa, B. (2019). Emotion classification using advanced machine learning techniques applied to wearable physiological signals data. *PhD Thesis, Queensland University of Technology*

Naurois, C. J., Bourdin, C., Stratulat, A., Diaz, E., & Vercher, J. L. (2017). Detection and prediction of driver drowsiness using artificial neural network models. *Accident Analysis & Prevention, 126*(1), 95-104.

Nguyen, T., Ahn, S., Jang, H., Jun, S. C., & Kim, J. G. (2017). Utilization of a combined EEG/NIRS system to predict driver drowsiness [Article]. *Scientific Reports, 7*, 43933.

Persson, A., Jonasson, H., Fredriksson, I., Wiklund, U., & Ahlström, C. (2020). Heart Rate Variability for Classification of Alert Versus Sleep Deprived Drivers in Real Road Driving Conditions. *IEEE Transactions on Intelligent Transportation Systems*, 1-10.

Poorna, S., Arsha, V., Aparna, P., Gopal, P., & Nair, G. (2018). Drowsiness detection for safe driving using PCA EEG signals. In *Progress in computing, analytics and networking* (pp. 419-428). Springer.

Putilov, A. A., & Donskaya, O. G. (2013). Construction and validation of the EEG analogues of the Karolinska sleepiness scale based on the Karolinska drowsiness test. *Clinical Neurophysiol, 124*(7), 1346-1352.

Ramzan, M., Khan, H. U., Awan, S. M., Ismail, A., Ilyas, M., & Mahmood, A. (2019). A survey on state-of-the-art drowsiness detection techniques. *IEEE Access, 7*, 61904-61919.

Sahayadhas, A., Sundaraj, K., & Murugappan, M. J. S. (2012). Detecting driver drowsiness based on sensors: a review. *Sensors, 12*(12), 16937-16953.

Saleab, M. S., Abd El Ghany, M. A., Toma, R. M., & Hofmann, K. (2016). Real-time sleep detection and warning system to ensure driver's safety based on EEG. *2016 IEEE 19th International Symposium on Design and Diagnostics of Electronic Circuits & Systems (DDECS)*, 1-6.

Sanei, S., & Chambers, J. A. (2008). *EEG Signal Processing*. John Wiley & Sons, Ltd.???

Setiono, R., & Liu, H. (1998). Feature extraction via neural networks. In *Feature Extraction, Construction and Selection* (pp. 191-204). Springer.

Shih, K. H., Hung, H. F., & Lin, B. (2009). Supplier evaluation model for computer auditing and decision-making analysis. *Kybernetes*.

Suto, J., Oniga, S., & Sitar, P. P. (2016). Comparison of wrapper and filter feature selection algorithms on human activity recognition. *2016 6th International Conference on Computers Communications and Control (ICCCC)*, 124-129.

Svensson, U. (2004). *Blink Behaviour Based Drowsiness Detection: Method Development and Validation*. Swedish National Road and Transport Research Institute. https://books.google.com.au/books?id=N6VFGwAACAAJ

Tran, Y., Wijesuriya, N., Tarvainen, M., Karjalainen, P., & Craig, A. (2009). The relationship between spectral changes in heart rate variability and fatigue. *Journal of Psychophysiology, 23*(3), 143-151.

Vicente, J., Laguna, P., Bartra, A., & Bailon, R. (2016). Drowsiness detection using heart rate variability. *Med Biol Eng Comput, 54*(6), 927-937.

Watling, C. N. (2016). The sleep and wake drives: exploring the genetic and psychophysiological aspects of sleepiness, motivation, and performance. *PhD Thesis, Queensland University of Technology*.

Watling, C. N., Hasan, M. M., & Larue, G. S. (2021). Sensitivity and specificity of the driver sleepiness detection methods using physiological signals: A systematic review. *Accident Analysis & Prevention, 150*, 105900.

Watling, C. N., Smith, S. S., & Horswill, M. S. (2014). Stop and revive? The effectiveness of nap and active rest breaks for reducing driver sleepiness. *Psychophysiology*, n/a-n/a.

Watling, C. N., Smith, S. S., & Horswill, M. S. (2016). Psychophysiological Changes Associated With Self-Regulation of Sleepiness and Cessation From a Hazard Perception Task. *Journal of Psychophysiology, 30*(2), 66-75.

Wilkinson, V. E., Jackson, M. L., Westlake, J., Stevens, B., Barnes, M., Swann, P., Rajaratnam, S. M., & Howard, M. E. (2013). The accuracy of eyelid movement parameters for drowsiness detection. *Journal of Clinical Sleep Medicine, 9*(12), 1315-1324.

Zhang, A., Liu, Q., & Bi, J. (2013). Sleepiness Recognition Based on Time-Domain Analysis of ECG and Pulse Signals. In *Advances in Mechanical and Electronic Engineering* (pp. 427-431). Springer.

Zhu, W., Zeng, N., & Wang, N. (2010). Sensitivity, specificity, accuracy, associated confidence interval and ROC analysis with practical SAS implementations. *NESUG proceedings: health care and life sciences, Baltimore, Maryland, 19*, 67.

Zilberg, E., Xu, Z. M., Burton, D., Karrar, M., & Lal, S. (2007). Methodology and initial analysis results for development of non-invasive and hybrid driver drowsiness detection systems. *The 2nd International Conference on Wireless Broadband and Ultra Wideband Communications (AusWireless 2007)*, 16-16.

Zoubek, L., Charbonnier, S., Lesecq, S., Buguet, A., & Chapotot, F. (2007). Feature selection for sleep/wake stages classification using data driven methods. *Biomedical Signal Processing and Control, 2*(3), 171-179.