Challenges of Driver Drowsiness Prediction: The Remaining Steps to Implementation

Emma Perkins, Graduate Student Member, IEEE, Chiranjibi Sitaula, Michael Burke, Member, IEEE, and Faezeh Marzbanrad, Senior Member, IEEE

Abstract—Driver drowsiness has caused a large number of serious injuries and deaths on public roads and incurred billions of taxpayer dollars in costs. Hence, monitoring of drowsiness is critical to reduce this burden on society. This paper surveys the broad range of solutions proposed to address the challenges of driver drowsiness, and identifies the key steps required for successful implementation. Although some commercial products already exist, with vehicle-based methods most commonly implemented by automotive manufacturers, these systems may not have the level of accuracy required to properly predict and monitor drowsiness. State-of-the-art models use physiological, behavioural and vehicle-based methods to detect drowsiness, with hybrid methods emerging as a superior approach. Current setbacks to implementing these methods include late detection, intrusiveness and subject diversity. In particular, physiological monitoring methods such as Electroencephalography (EEG) are intrusive to drivers; while behavioural monitoring is least robust, affected by external factors such as lighting, as well as being subject to privacy concerns. Drowsiness detection models are often developed and validated based on subjective measures, with the Karolinska Sleepiness Scale being the most popular. Subjective and incoherent labelling of drowsiness, lack of on-road data and inconsistent protocols for data collection are among other challenges to be addressed to progress drowsiness detection for reliable on-road use.

Index Terms—Behavioural monitoring, driver physiological monitoring, driver safety, drowsiness detection, fatigue, hybrid models, sensor applications, vehicle-based monitoring.

I. INTRODUCTION

Globally, it is estimated that 1.3 million people die per year as a result of car crashes worldwide, and crashes cost countries approximately 3% of their GDP [1]. Although it has been suggested that in the future Autonomous Driving Systems (ADS) will overcome this problem, it remains the case that the driver is still legally responsible for the vehicle [2]. Furthermore, this extends to take over requests, where the user will need to be fully alert.

Fatigue has been labelled as part of the “fatal five” for driving safety risks, alongside speeding, drugs/alcohol, failure to wear a seat belt and driver distraction. Research has shown and vigorously advertised that the effects of drowsiness when driving is similar to that of driving above the legal alcohol limit [3], [4]. Hence, the monitoring of fatigued driving is required to make driving safer and has emerged as a key priority to reduce road related fatalities and serious injuries.

In Australia, driving whilst fatigued has contributed to 20-30% of road related severe injuries and fatalities, similar to that of drink driving and speeding [5], [6]. Australia has seen an average of 1,209 fatalities on-road within the last 5 years (2016-2020) [7], where road related accidents are estimated to cost $3-4 billion AUD to the economy per year in Victoria alone [8]. In 2015, it was reported that USA had an estimated 5000 people die in crashes influenced by drowsiness, although this is difficult to measure accurately [9]. This was estimated to cost the economy $109 billion USD [9]. Based on reports from the United Kingdom and many European countries, approximately 10 to 30% of crashes are caused by drowsiness [10].

It is clear that driver fatigue remains a pressing concern for road safety. However, despite over a decade of research into detection and prediction measures, this problem is yet to be adequately addressed. Our paper seeks to review the progress made to date, and identify key challenges holding back the implementation of driver drowsiness prediction technologies.

In preparing this paper, 130 studies were surveyed, including 44 Physiological-based studies, 43 Behavioural-based studies, 15 Vehicle-based studies, and 28 Hybrid studies. Of these studies, 69 were published after 2018 and considered to be recent. These studies were found by first searching for keywords listed in Table I. The selection was then expanded to include a greater number of recent studies, with the search restricted to 2018 and later. The search was conducted across Google Scholar, IEEE Xplore, and ScienceDirect.

TABLE I

| Method   | Keywords                           |
|----------|------------------------------------|
| All Studies | Driver, Fatigue, Drowsiness, Drowsy |
| Physiological | Physiological, PPG, ECG, BOG, EEG, EMG, Temperature, Respiration |
| Behavioural | Behavioural, Camera, Yawning       |
| Vehicle-Based | Vehicle, Lane departure, Steering Wheel |
| Hybrid    | Hybrid, Physiological + Behavioural, Behavioural + Vehicle, Physiological + Vehicle |

Manuscript received 17 May 2022; revised 7 November 2022; accepted 22 November 2022. Date of publication 25 November 2022; date of current version 20 March 2023. The work of Emma Perkins was supported by the Australian Government Research Training Program Scholarship. (Corresponding author: Faezeh Marzbanrad.)

The authors are with the Department of Electrical and Computer Systems Engineering, Monash University, Clayton, VIC 3800, Australia (e-mail: emma.perkins1@monash.edu; chiranjibi.sitaula@monash.edu; michael.g.burke@monash.edu; faezeh.marzbanrad@monash.edu).

Color versions of one or more figures in this article are available at https://doi.org/10.1109/TIV.2022.3224690.

Digital Object Identifier 10.1109/TIV.2022.3224690
Existing review papers have thoroughly covered technical aspects such as signal acquisition, feature extraction and detection of driver drowsiness, in three primary categories: physiological, behavioural and vehicle-based monitoring. Sahayadhass et al. [11] outline existing simulator types, causes of crashes, inducement of drowsiness, simulator versus on-road data collection and physiological intrusiveness. Challenges regarding protocols are mentioned, where the environment for acquisition and drowsiness inducement are discussed; however, other issues including drowsiness labelling, early detection and subject diversity are not addressed in this review.

Kaplan et al. [12] highlight ways to detect drowsiness, commercial products available and simulation data collection. Kaplan et al. also expand the review to driver distraction; another cause of deaths on our roads. Within the review, the discussion of physiological methods mention Electrooculography (EOG) and Electrocardiography (ECG); however, only EEG is discussed in depth as this is labelled as the most “promising and feasible method”. Furthermore, limitations of described studies are mentioned, but protocols are not covered, including diversity of subjects and the participant numbers.

DouDou et al. [13] thoroughly examine acquisition systems, fatigue inducement and commercial products; however this differs from the focus of our work, as shown in Table II. Moreover, our paper includes a number of more recent works in the literature. DouDou et al. also look at issues in the literature, including intrusiveness of acquisition, noise in data collection, ease of use and accuracy of different methods. They discuss the variations in fatigue measuring, circadian rhythm and the effect on models as well as the ability to track drowsiness in real-time. However, they do not discuss simulations versus on-road data, predicting drowsiness rather than detecting, and the issues with collected data including subject diversity, number of subjects and accuracy reporting differences; whereas our review addresses further challenges, with a focus on more recent studies.

Ramzan et al. [14] provide a very methodological, systematic review of the literature and accompanies fatigue features with examples from the literature. Classification methods are also expanded upon; however, challenges are not a focus of this review, but rather identifying current studies on drowsiness and what methods are being used.

Hu and Lodewijks [15] provide an in-depth understanding of fatigue versus sleepiness, and how measures can effectively distinguish the two. They also cover the issue of intrusive acquisition, the differences in results between simulator and on-road studies, and inconsistency of protocols. However, less intrusive acquisition systems (such as Photoplethysmography (PPG)) and hybrid studies are not discussed.

Finally, Němcová et al. [16] cover specific features and their link to drowsiness, mass produced drowsiness systems, privacy concerns, cost implications, and the annotation variances in drowsiness labels. Their work also include a number of current challenges in driver drowsiness detection. However, several important issues are not discussed, such as early detection, experimental protocols, number and diversity of participants, and sleep disorders.

In contrast to existing reviews, this paper focuses on feature selection methods and barriers to the implementation of hybrid models, alongside existing challenges. In particular, we focus on prediction vs detection, subject diversity, the number of subjects, accuracy reporting and future possibilities for driver drowsiness, which have not been adequately addressed. The current challenges in drowsiness detection covered by review papers is summarised in Table II. A key challenge identified in this work lies in the distinction between driver drowsiness prediction (prediction of impending drowsiness) and detection (recognising drowsiness at it occurs). The latter is most commonly addressed by current technologies, while the former is often neglected, although of significantly more value from a road safety perspective.

This paper is structured as follows. Initially, we provide a background of the effects of drowsiness and how it is monitored. Second, we define our evaluation metrics to review different drowsiness systems. We then describe existing commercial products for driver drowsiness detection, followed by methods under active research and development to provide an overview of recent literature. Existing challenges are then outlined, with explanations of how these challenges affect models and restrict these from becoming implementable. Future work is also outlined, providing a road-map to more accurate and effective drowsiness technologies and enhanced road safety.

II. BACKGROUND

Drowsiness is the transitional phase between wake and sleep, where a person can be mentally and physically affected [17]. Drowsiness has been shown to depend on the time of day (circadian rhythm) and lack or interruption of sleep [17],

### Table II

| Existing reviews | Features /Modalities | Classification Methods | Simulation vs on-road | Labelling of drowsiness | Prediction vs detection | Subject diversity | Performance discrepancies |
|------------------|----------------------|------------------------|-----------------------|-------------------------|------------------------|-------------------|--------------------------|
| Sahayadhass et al. 2012 [11] | ✓ | ✓ | ✓ | X | X | X | X |
| Kaplan et al. 2015 [12] | ✓ | ✓ | ✓ | X | X | X | X |
| DouDou et al. 2019 [13] | ✓ | ✓ | ✓ | ✓ | ✓ | X | X |
| Ramzan et al. 2019 [14] | ✓ | ✓ | ✓ | X | X | X | X |
| Hu and Lodewijks 2020 [15] | ✓ | ✓ | ✓ | X | X | X | X |
| Němcová et al. 2021 [16] | ✓ | ✓ | ✓ | ✓ | ✓ | X | X |
| Ours | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
Furthermore, increased driving times, warm rooms and monotonous activities have been found to unmask sleepiness rather than cause it [19]. Causes of drowsiness can include work shift schedules (where a decrease in performance has been seen in night shift workers [18]), lifestyle choices, mental health problems, stress and medication or illness [20], [21].

Drowsiness can decrease cognitive performance including vigilance, information processing and decision making [22]. Reaction time is also decreased under the influence of drowsiness, making it unsafe to drive. Visible symptoms can include frequent yawning, eye closure, head tilt, itchy or red eyes, lack of attention, line crossing, a tendency to accelerate, and reduced stopping distances [20].

Reducing the effects of drowsiness can be achieved by taking a nap, while avoiding drowsiness in general can be achieved through regular sleep patterns (more than 7 hours a night) and avoidance of regular daytime napping [23]. Research has shown that caffeine consumption is not beneficial in the long term and can induce sleep in the short term (1–2 hours) [20]. However, slow-release caffeine can improve performance for over 5 hours [24]. Moreover, regular caffeine intake can provide minimal effect on a user, hence, this should not be used as a drowsiness countermeasure.

As drowsiness is difficult to define, subjective measures have been used as a label in most cases for data gathered in studies of fatigue, where subjects have given a number associated with their level of drowsiness. Alternatively, a professional may label data based on visual drowsiness indicators. Numerical drowsiness scores are used for the majority of studies, with most studies using the Karolinska Sleepiness Scale (KSS) [15]. Professional labelling has also been performed on video recordings of subjects, where an Observer Rating of Drowsiness (ORD) is often used. Less common methods are also discussed in this paper.

The main methods explored to monitor drowsiness include physiological, behavioural, and vehicle-based technologies [25]. Physiological-based technologies are mainly based on human physiology-related measurements such as measurements of brain waves, cardiovascular measurements, muscle fatigue, eye movement, and eye closure. The benefit of such technologies is that they are reported to be more accurate than other methods, they can detect drowsiness at earlier stages and have less interference from lighting, environment and road conditions [11], [14]. However, they are limited by their intrusiveness [26], susceptibility to noise [12] and cost of implementation [13]. One common method in the literature includes Electroencephalography (EEG), which provides high accuracy for drowsiness monitoring [13], by monitoring spectral power various bands including the beta (13–30 Hz), alpha (8–13 Hz), theta (4–8 Hz) and delta (0.1–4 Hz) bands [27]. Electrocardiography (ECG) is also common in the literature, where Heart Rate (HR) and Heart Rate Variability (HRV) parameters are often used for drowsiness indication [13]. Electrooculography (EOG) monitors eye features related to drowsiness, but the electrodes used here are often intrusive and can be replaced by behavioural monitoring to reduce intrusiveness [28], [29]. Various classification methods are used in physiological studies including thresholding [29], support vector machine (SVM) [30], k-nearest neighbours (kNN) [31] and more recently, deep learning approaches [32].

Behavioural-based technologies are based on human behaviours that are captured using a camera, where facial features, head inclination, yawning, and eye features are used to measure drowsiness [11]. Such technologies are the second-most accurate and less intrusive than other methods [33]; however, they may be subject to privacy concerns [34]. These systems can fail for people with glasses, moustaches or those with different skin tone [35], [36], [37], where the users face is not adequately detected. Furthermore, lighting (such as at night) and environments also limit behavioural systems, making it harder to detect the face [33], [38], where many systems still do not consider infra-red (IR) cameras which can mitigate the affects. Classification in behavioural studies has evolved from simpler methods where pre-processing, feature extraction and feature selection were used [39], to more advanced techniques largely reliant on deep learning models that learn to identify appropriate features for classification [40]. Before this, other methods including thresholding [37] and SVM [39] models were implemented.

Finally, vehicle-based technologies are mainly based on vehicle status and parameters such as lane departure and steering wheel measurements [41], [42]. These are also dependent on weather and road condition, in particular for lane keeping measurements [11], [20]. Moreover, features from vehicle-based studies can vary due to external factors, such as alcohol consumption [11], [43], [44], [45]. Although vehicle-based methods are not as widely researched, similar classification used above methods have been applied, including SVM [46], thresholding [47] and deep learning methods [48].

Studies have suggested that multi-modal (physiological) and hybrid approaches produce greater accuracy compared to a mono-signal or singular method [28], [30], [49], [50], [51], [52], [53]. Despite this, fewer studies have focused on hybrid models. Furthermore, commercial products mainly focus on behavioural and vehicle-based measures, with the latter being most commonly used in vehicle, as they are most easily implemented by automotive manufacturers. Physiological methods are not widely used commercially, perhaps due to their intrusiveness.

III. EVALUATION METRICS

Throughout this paper we discuss various metrics used to help categorise and summarise the papers surveyed. To clarify our evaluation criteria, we define the following aspects to determine the benefits and downfalls of a system, which were areas of importance and concern in various papers:

- **Performance**: the reported accuracy of correct drowsy and non-drowsy classification. Furthermore, precision, sensitivity and specificity are important factors when reporting performance of a drowsiness system [49], including true/false positive and true/false negative rates, which can also be presented in a confusion matrix.
- **Intrusiveness**: both physical and psychological intrusiveness, which will be defined subsequently.
- Privacy: the ability to identify people or sensitive information using collected data
- Data size: how much data will need to be stored
- System cost: how much the acquisition system will cost to implement in a vehicle
- Computational cost: how much computation is required to detect drowsiness
- Susceptibility to noise: how much noise will be present in the raw data
- Vulnerability to subject diversity: how the model will be affected by differences in ethnicity, age and gender.
- Adaptability: how the system can cope with data loss

Our analysis of the literature showed that definitions of intrusiveness within papers could vary and that the levels of intrusiveness were not clear. Hence, we first introduce a new intrusiveness scale, in order to evaluate systems consistently. Then, based on the frequency of concern in papers of the identified areas, we rank the existing non-hybrid methods according to the dimensions above.

A. Defining Intrusiveness

In many of the papers surveyed the word intrusive was not directly defined, although in some instances, methods were described using phrasing such as less intrusive. In order to clarify types and levels of intrusiveness, we propose a new scaling system as per the following:

- Low: No intrusion and not noticeable by the user. For example, a sensor could be completely immersed within a car seat.
- Medium:
  - Type I: Psychological intrusiveness where a subject is personally monitored by video or images and can be easily identified.
  - Type II: Physical intrusiveness where minimal contact required, for example sensors that are noticeable and in contact but minimalist and lightweight.
- High: Physically intruding on a person, sensors can be felt when movement occurs

Henceforth, our paper uses this proposed intrusiveness scale when referring to levels of intrusiveness.

B. Ranking of Current Methods

Table III provides a ranking of the various methods in the literature and how they perform for each category of interest discussed above, based on the frequency with which a given dimension was mentioned in the literature surveyed. Through reviewing the literature, we have ranked how each method performs in each given category. In Table III, a 1 is used to denote the best performing method, while 3 denotes the worst performing method for a given category. The best performing method was determined by comparing a particular category. For example in terms of privacy concern, physiological data does not provide information on the age, gender or identity. However, behavioural data has large privacy issues as the identity of the user can be determined in videos. Lastly, vehicle data is in between, as it is possible to estimate where a subject lives by tracking the roads the user drives on; however, this is much more difficult to determine. Hence, for privacy concern, physiological data is ranked as 1, vehicle-based as 2 and behavioural as 3. In the case of physical intrusiveness, we ranked both vehicle and behavioural methods as 1, as neither require sensors or data is acquired whilst physically interacting with the body.

C. Computational Requirements

The computing source requirements of driver drowsiness detection depend on the amount of data received, the amount of processing required and the prediction model built. It is noted that the computation time required for training is far greater than estimating and predicting as shown by Wang et al. [54]. Inherently, behavioural models have a large data input as the data is received via video and requires a high rate of acquisition in order to detect subtle changes in the face (such as blinking). Secondly, behavioural models are often built using deep learning approaches, requiring higher computational sources than traditional machine learning methods. Physiological data also requires a high rate of acquisition to detect minor changes and often requires more complex pre-processing. One study explored reducing computational cost across five algorithms using EEG data [55]. Lastly, vehicle-based methods have a lower computational resource requirement due to less data requiring processing (for example line crossing metrics can be denoted by a singular binary occurrence, and velocity as single number). However, vehicle-based methods may also involve some video input that increases computational cost, although not at the same rate as behavioural models. For example, vehicle metrics such as lane deviation, occur over a slower period of time compared to blinking. Moreover, in commercially available systems, the device computes internally, offloading the need for computational power to be provided by the user.

| Metric                      | Physiological |Behavioural |Vehicle-Based |
|-----------------------------|---------------|------------|--------------|
| Performance                 | 1             | 2          | 3            |
| Physical Intrusiveness      | **3**         | 1          | 1            |
| Psychological Intrusiveness | 1             | **3**      | 2            |
| Privacy Concern             | 1             | 3          | 2            |
| Data Size                   | 2             | 3          | 1            |
| System cost                 | 3             | 2          | 1            |
| Computational cost          | 2             | 3          | 1            |
| Susceptibility to noise     | 3             | 1          | **2**        |
| Vulnerability to subject diversity | 2          | 3          | 1            |
| Adaptability                | 1             | **3**      | **2**        |

* This can vary depending on the acquisition system
A number of commercial products have been implemented to counter drowsy driving, many of which use vehicle-based detection methods and are installed by car manufacturers. Other commercial devices include behavioural and physiological components, where the detection systems are optional external purchases (often targeted at truck, mining and logistic companies).

### IV. COMMERCIAL AVAILABLE SYSTEMS

A number of commercial products have been implemented to counter drowsy driving, many of which use vehicle-based detection methods and are installed by car manufacturers. Other commercial devices include behavioural and physiological components, where the detection systems are optional external purchases (often targeted at truck, mining and logistic companies).

#### A. Physiological Commercial Products

Existing physiological products (5) can be relatively expensive, with 3 products identified built around Galvanic Skin Response (GSR), one of which includes heart rate (HR) monitoring. Once product included EEG monitoring with a device named SmartCap, which is targeted at mining settings. Finally, a device that utilises reflectance oculography (with similar features to EOG measures including relative velocity of blinks and relative duration of blinks [56]) is available, in the form of glasses. This product outputs a drowsiness scale, known as the John’s Drowsiness Scale (JDS). These products are listed in Table IV.

1) Behavioural Commercial Products: Commercial products relying on behavioural sensing are the most commonly available product that can be purchased for detecting drowsiness [13]. These make use of cameras focused on the driver and are generally based on eye tracking. The products identified are shown in Table V.

2) Vehicle-Based Commercial Products: Table VI includes current companies, their vehicle-based drowsiness detection products and the methods they use to detect drowsiness.

3) Hybrid Products: We identified two implementations of hybrid methods in cars, including Toyota drowsiness tracking [84], and the Bosch mobility solutions Interior Monitoring System [85]. Both hybrid systems combine vehicle and behavioural approaches.

Current commercial products are largely reliant on vehicle-based methods, which have been shown to be the least accurate for detecting drowsiness, despite being the easiest data to gather and process. These systems are not always defined as drowsiness devices, but are also targeted at distraction detection. They are often labelled in vehicles as lane departure and lane assist products. More reliable and specific devices are required for better accuracy, including the implementation of more complex systems. Moreover, it remains the case that commercial products are focused on detection rather than prediction, which is a major disadvantage of current devices. High false positive rates are also a potential area of concern. Hence, further research and development is required.

### V. METHODS UNDER ACTIVE RESEARCH AND DEVELOPMENT

For the most part the development of drowsiness detection and prediction technologies has followed a standard paradigm. First, labelling of drowsiness has been undertaken using subjective measures, which is then used to develop models in one of the four primary categories: physiological, behavioural, vehicle-based, and hybrid. The development of these models follow the main stages presented in Fig. 1. The output of these models can vary depending on the labelling of drowsiness in the study, where most studies address a binary drowsy/non-drowsy label. A smaller selection of studies include between 3 and 5 drowsiness labels.

Feature selection methods used in the studies include t-tests [29], [30], [51], [52], [86], (multivariate and univariate)
analysis of variance [42], [53], [87], [88], [89], [90], correlation coefficients [28], [31], [90], [91], [92]. In the case of end-to-end approaches, deep learning methods offset this by including feature extraction layers within the architecture [93]. This is particularly true of models relying on fine-tuning [93], where additional layers are appended to a more general pre-trained network to allow training with relatively smaller datasets. Feature selection methods are frequently applied to reduce the number of features, in order to reduce computational and system costs, and increase accuracy.

A summary of the main classification methods in driver drowsiness has been provided in Fig. 2, where the number of studies using the various methods are shown for both before and after 2018. Classification methods in driver drowsiness studies have evolved in the past few years, from centering around SVM, thresholding and various other approaches (including Random Forests (RF)) to more advanced methods including deep learning. Deep learning approaches include artificial neural networks (ANN), convolutional neural networks (CNN) and recurrent neural networks (RNN). Drawing on the representations learned by intermediate layers, deep learning approaches have shown great promise in several areas, including drowsiness detection. In particular, these are most commonly used in behavioural-based technologies, as these are particularly useful for two-dimensional (2D) inputs including image classification tasks [32]. In addition, these approaches are now also emerging in studies including lower dimensional time series inputs typically used in physiological and vehicle-based techniques [48], [94].

Traditional machine learning-based methods such as kNN and SVM are more commonly seen in physiological studies [30], [31] and are considered to have an advantage when it comes to interpretability. Despite this, deep learning methods are emerging as more desirable methods due to the increased accuracy they can provide and ability to work end-to-end with pre-processing, feature extraction, feature selection and classification jointly undertaken by the network [94]. However, deep learning comes at the cost of a 'black box' solution and is often considered less interpretable. To counter this, recent papers have recommended the use of an explainable AI (XAI) system to develop models, in order to better understand how drowsiness is classified with a deep learning model [93], [95], [96]. Qian et al. [95] design a multimodal representation, leveraging environmental data and daily life data, and train a deep learning model for the drowsiness detection. This approach provides a performance of 64.6% unweighted average recall measure. Furthermore, Rajkar et al. [97] employ a CNN for the drowsiness detection using eye closure and yawn data. This provides an average accuracy of 96%. More recently, Cui et al. [98] utilise electroencephalogram signals and design a CNN for drowsiness detection, which results in an average accuracy of 78.35%. Following a similar approach, Nandini et al. [99] employ yawning and blinking features (facial features) to detect drowsiness using a stacked deep CNN model. This work achieves a precision of >95% for the detection. Magan et al. [100] use 2D cropped face images and fine-tune a pre-trained deep learning model, EfficientNetB0, using a transfer learning approach. Their work yields the highest accuracy of 63% on the test data.

A. Drowsiness Labelling

Subjective measures are often used to describe a subject’s state of drowsiness. The most common subjective method used in the literature is the Karolinska Sleepiness Scale (KSS), followed by the Stanford Sleep Scale (SSS) [15]. Other sleep scales seen in the literature include the HFC drowsiness scale, Epworth Sleepiness Scale (ESS), Johns Drowsiness Scale (JDS), Observer rating of drowsiness (ORD), and Subjective Drowsiness...
Rating (SDR). The HFC drowsiness scale, ORD rating and SDR are externally labelled, while the JDS rating is a physiological indication and the remainder of labelling methods are self estimations.

1) Karolinska Sleepiness Scale (KSS): The KSS is the most used subjective scale in the literature, following a 9-point system [101]. Some studies used the KSS to categorise fatigue into two stages (awake and drowsy), others three stages (awake, slight and drowsy), and sometimes 4 stages. Of those studies simplifying to two stages, many left out KSS points to enable a clear difference between awake and drowsy states. The use of KSS in evaluating fatigue is shown in Fig. 3. The decisions behind these splits is not always clear, although for some studies it was based on collected features and how they changes across the reported KSS values [86], [102], [103], [104].

2) Stanford Sleep Scale (SSS): The Stanford Sleep Scale (SSS) [108] has been used similarly to the KSS scale, with one study in particular grouping the ratings to reduce the number of output labels [109]. The SSS is a 7-point scale, where one proposed combination of stages for fatigue labelling is shown in Fig. 4 [109].

3) Epworth Sleepiness Scale (ESS): The ESS [110] is used to measure daytime sleepiness with a questionnaire on the likelihood of dozing during eight common situations [110]. These situations include dozing when watching TV, as a passenger in a car, talking to someone, sitting in traffic etc. This method is useful for an overall view of tendency to sleep in subjects, but cannot be used for detection [103].

4) The Groningen Sleep Quality Scale (GSQS): Similarly to the ESS, the GSQS comprises a pre-study questionnaire that is used to measure the quality of sleep the night before. The GSQS was used in one study to determine the continuation of the experiment with a particular subject, where exceeding the sleep quality threshold of 3 meant the subject could not participate, as that indicated they had intermediate sleep disturbances [111]. The GSQS asks 15 true or false questions, with a maximum score of 14 available.

5) The Chalder Fatigue Scale (CFS/CFQ): The CFQ (adapted from the CFS) is an 11-question survey aiming to measure fatigue levels, with reference to a “usual” state [112]. The former CFS scale comprises 14 questions, where a yes or no answer is required for each [113]. A study finds that the CFS shows fatigue levels similar to that of the more commonly used SSS [111].

6) HFC Drowsiness Scale: The HFC Drowsiness Scale is used to annotate video-based recordings, where trained annotators label video recordings every minute [92].

7) Observer Rating of Drowsiness (ORD): Two different levels of ORD have been used, both for the analysis of video recordings, based on the face and body language of a driver [114]. The first ORD is 100-point rating scale where two studies have grouped the ratings into drowsiness labels. The second ORD is a 5-point scale, with the same classifications as the first ORD but labelled under the numbers 1 to 5. A summary can be found in Fig. 5, where one study grouped moderately, very and extremely drowsy into a single “drowsy” label for classification [115].
8) Subjective Drowsiness Rating (SDR): The SDR was developed for a study that uses a 5-point system, very similar to the 5-point ORD [117]. The fatigue stages used in the study are based on behavioural signs observed in a video recording, where a non-drowsy to extremely drowsy rating is indicated using the labels 0 to 4.

B. Physiological Methods

Approaches to driver drowsiness detection and prediction have been undertaken using both uni- and multi-modal methods, with multi-modal approaches proving more reliable [49]. Acquiring these signals can require a high level of intrusiveness, negatively impacting their potential use in detecting driver drowsiness [26]. Alternatively, some of these signals can be collected using medium to low levels of intrusiveness, with sensors placed in the steering wheel of the vehicle [118], [119], drivers seat [14], [120] or by using wrist worn sensors [26]. Physiological signals have shown to be more accurate when compared to vehicle-based and behavioural methods as they are less prone to interference from lighting, environmental and road conditions [11], [14]. Furthermore, physiological signals can detect drowsiness at earlier stages when compared to behavioural and vehicle methods [11].

Primary physiological measurement methods include Electroencephalography (EEG) [32], [94], [121], Electrocardiography (ECG) [88], [122], [123], Electromyography (EMG) [31], [124] and Electrooculography (EOG) [28], [125], [126]; moreover, some studies have also explored Galvanic Skin Response (GSR) [52], [124], Ballistocardiography (BCG) [127], Seismocardiography (SCG) [127] and Skin Temperature (ST) [53]. Photoplethysmograms (PPG) have also been applied with similar drowsiness correlations to ECG [51], [128], [129]. These signals are typically processed and analysed for use in predicting drowsiness using the process shown in Fig. 1.

A list of features commonly used in physiological drowsiness detection and their relationship to drowsiness can be found in Table VII.

The limitations of physiological systems are as follows:

- Intrusiveness signal collection, where sensors can be placed on the head, face or chest [26], [103], [152]. Nonetheless, alternative methods have been used in some instances. EOG and EEG cannot avoid intrusiveness as the data is collected from the face/head; despite some studies attempting to reduce it [52], [153], [154]. Despite this, EEG-based measures are still the most popular method for detecting drowsiness [15]. Further data capture mechanisms and corresponding levels of intrusiveness are described in more detail below.
- Many studies have not considered various sleep disorders in their studies, with some finding differences in the signals for those that do [136]. This has included differing RRI intervals and respiration in ECG data for people with sleep disorders versus those without [136].
- A limitation exists where almost all studies focus on simulated data, with very few on actual road data, where almost all of the on-road experiments found do not have sleep
The time frame for analysis is shown to exceed multiple EOG recordings require electrodes to be placed. EMG recordings are often captured using surface EMG (sEMG) as this is a non-invasive method of collecting. The steering wheel sensor is not viable when the subject is using deprived subjects [103]. It has also been noted that on-road data needs to be collected to test systems [30].

- The time frame for analysis is shown to exceed multiple minutes in some studies [87], which is not feasible in early detection, where an adverse event will have likely to have already occurred by the time of detection.
- Gender can influence signals, where one study demonstrated this using EEG signals [12].

1) EEG: Studies using EEG signals do not often report the specific electrodes and relevant features but rather the number of electrodes used in acquisition. This includes between 1 and 43 measurement electrodes [30], [87], [132], where the most relevant electrodes have found to be around the parietal [30] and occipital lobes [25], [30], [132]. Traditional EEG has a high level of intrusion, but alternative methods specific to driver drowsiness detection including the use of a headband [153], [154] and in-ear EEG have been used [52]. Nonetheless, EEG signal acquisition faces a trade-off between intrusiveness and accuracy, where systems offering lower levels of intrusion, such as headband EEG recording, provide a lower detection accuracy when compared to conventional methods [154]. This further extends to system cost and computational cost where fewer electrodes results in cheaper acquisition and computation.

2) Heart Rate Monitoring Methods: ECG also has a high level of intrusiveness, but can be collected less intrusively by using electrodes in a steering wheel rather than the chest [119]. However, this requires direct contact with the steering wheel using both hands. A few ECG studies have mentioned using 1-lead [30] and 2-lead ECG [139], but most do not specify this. PPG can be used as an alternative to ECG with medium level intrusiveness and is an easy to collect signal [51], [52]. Furthermore, BCG and SCG do not require skin contact and offer a low level of intrusiveness [127]. These methods can use sensors positioned in inconspicuous locations including the head rest of a driver’s seat [120], [127], [155], the car’s steering wheel [3], [127], [140], the seat belt [127] or on the finger [51]. Furthermore, video monitoring can also be used to detect heart rate and heart rate variability in both day and night conditions, eliminating the need for sensors [156], [157]. However, the alternative methods are more susceptible to noise and data loss. The steering wheel sensor is not viable when the subject is using gloves or in the case of ADS where the user is not required to have their hands on the steering wheel. Furthermore, these methods are more susceptible to movement noise than alternative capture methods.

3) EOG: EOG recordings require electrodes to be placed close to the eyes for a stronger signal [103], [144]. Studies reported using two electrodes for each horizontal and vertical placements [103], [126], whereas others used just two vertical electrodes [125]. This associates EOG recordings with a high level of intrusion; however, studies have suggested this can be replaced with a camera for many of the blink features typically identified using EOG [28], [29]. This shifts the intrusiveness of acquisition to psychological intrusiveness, corresponding to a medium level of type I intrusiveness.

4) EMG: EMG recordings are often captured using surface EMG (sEMG) as this is a non-invasive method of collecting EMG signals. Electrodes have been placed on the legs (anterior tibialis) [136], chin [136], upper arm (on the deltoids) [149], [158], back (on the trapezius) [149], [158], neck (splenius capitis) [158] as well as lower arm [124]. EMG acquisition is associated with high-level of intrusion.

5) GSR: The GSR signal acquisition can be categorised as a medium or a low level intrusiveness as it can be placed in a steering wheel, but contact is still required if used in wearable devices [52], [124], [152]. Studies have also reported collecting GSR from the finger [52], [124].

6) Temperature: Temperature that has been taken externally at the forehead was shown to have the greatest relationship to drowsiness, when compared to body temperature captured at other locations [150]. Other points for detecting skin temperature have been proposed through the steering wheel alongside various other methods, both with medium and low levels of intrusion [119].

7) Breathing: Respiration sensors offer a medium level of intrusion, for example those using a band around the chest [51]. Respiration can also be measured from videos, making the process non-contact and of a low level of intrusion [156], [157]. However, this method can be challenging due to noise and may be difficult to apply on-road. A hybrid method can be used in this case, where the collection methods can be cross-validated between a seat belt, camera, heart-rate derived breathing and seat pressure to determine the true breathing rate of a subject and help counter the effects of noise.

8) Thermal Imaging: Thermal imaging has been proposed to remove all physical intrusiveness, and is categorised as a method with a low level of intrusiveness, where it is a viable method in detecting drowsiness [116].

9) Grip Strength: Grip strength is another feature that can be integrated into the steering wheel, which has been correlated with EMG in predicting muscle fatigue [151]. Grip strength data collection can also offer low level intrusiveness as this can be integrated in the steering wheel.

Many of the features described above can be combined or extracted using other acquisition systems to reduce the number of sensors required to collect relevant drowsiness features. For example, respiration (thermal imaging and respiration sensors) can be derived from ECG or video, reducing the need for a breathing belt. Low and medium intrusive alternatives could also be used instead of traditional methods. PPG and BCG have a medium intrusive level, whereas ECG has a high level of intrusiveness, but can extract some of the same features. These methods can be useful in reducing the cost and intrusiveness of signal acquisition.

C. Behavioural Methods

Behavioural monitoring is becoming more popular as it is not physically intrusive [38]; hence proving more desirable than physiological monitoring. However, behavioural methods are psychologically intrusiveness, due to their association with surveillance, falling under medium type II for intrusiveness. This method has been shown to be more accurate than vehicle monitoring [33] and uses cameras to detect drowsiness...
indicators such as yawning, eye features, facial features and head orientation [11]. Eye feature detection can be used as a non-intrusive alternative to EOG for measuring blink duration, blink frequency and percentage of eyelid closure (PERCLOS). Of the behavioural indicators, two studies reported that head movement/pose features provide the highest correlation with drowsiness [38], [159]; however, eye movement-based measures are the most widely used measures [15].

Behavioural detection processes have also followed the stages in Fig. 1 where the signal is acquired using camera and pre-processing can often involve face and landmark detection. Behavioural feature extraction often relies on features including head, mouth and eyes. The related eye features are similar to those presented under EOG in Table VII, and behavioural features used in studies are presented in Table VIII. Finally, this method is commonly classified as described previously, by both SVM or thresholding methods. However, in more recent works behavioural monitoring tends to involve deep learning approaches, where the image is fed in to a network and trained for the corresponding drowsiness label. This approach is known as an end-to-end method, where the neural network is able to extract and select features before classifying them.

The limitations of behavioural studies include a variety of areas:

- Monitoring can be affected by lighting, especially at night [33], [38]. Infra-red (IR) LED can be used as an alternative; however, this makes the day time monitoring worse [35]. IR illumination has also shown incorrect detection on older people as they have conflicting features such as deeper wrinkles [20].
- Some systems only work at day or night, not both [35].
- Some systems do not work for people with glasses, moustaches or those with different skin tone [35], [36], [37]. Some studies have tackled this problem, including expanding the system to include facial expressions in contrast to the more common method using eye states alone [160].
- Camera movement can occur for a variety of reasons especially outside of simulation settings [33].
- Frame rate can affect the accuracy of the recording [33].
- On-road testing needs to be undertaken as factors such as lighting, vehicle vibration and traffic density may effect models [92]. Accuracy was shown to decrease for models after on-road testing [161].
- Behavioural features such as PERCLOS are said to detect drowsiness too late [144].
- Privacy concerns, where uses may feel uncomfortable having their face monitored [34].

### D. Vehicle-Based Methods

Vehicle based methods are the least common and least accurate method for detecting driver drowsiness, and most suited to roads with clean line makings and good weather conditions [33]. However, these methods are the simplest to implement, and most convenient when it comes to data collection. They use lane position, steering wheel, acceleration pedal, and yaw features to evaluate driver drowsiness [41], [42]. Steering wheel parameters were found to be the most commonly investigated in recent literature despite one study suggesting lane keeping parameters were the most widely used approach [15]. One study showed that steering wheel parameters were more accurate than lane parameters [91]; however, another study suggested that lane variability was better [42]. A summary of the features used in vehicle-based studies can be found in Table IX.

The recording and modelling of vehicle-based studies is simpler than behavioural and physiological methods. The vast majority of studies described little pre-processing other than normalization or standardization and windowing of the data. The
studies are most often classified using SVM and deep learning approaches and also follow the process in Fig. 1.

The limitations of vehicle-based methods, as well as lower accuracy, include the following:

- Vehicle-based methods are weather dependent and road condition dependent, particularly for lane keeping measurements [11], [20].
- It was found that there are some differences between on-road and simulated driving, despite almost all studies undertaken in a simulation. It was found that drowsiness is induced earlier in a simulation than in on-road settings and more research needs to be done on simulations versus on-road, as well as on-road testing to validate models [42].
- Other factors can influence steering variability and lane position, not only fatigue [11], [43], [44], [45].
- Mental fatigue can influence vehicle-based parameters, as well as sleepiness [15].

E. Hybrid Methods

Hybrid methods are relatively new in the literature as the individual methods had to be established before the combined approaches [167]. Thus, most studies of hybrid technologies have reported better results [28], [30], [50], [51], [52], [53], indicating that hybrid approaches are the best method for detecting drowsiness in terms of lowering intrusiveness levels, accuracy, and ability to function with data loss. This is due to a combination of different techniques being able to help overcome the negative components of specific techniques [14].

Hybrid models use the same features as the physiological, behavioural, and vehicle-based methods, with individual models often being developed for each method and combined to develop the hybrid model. Hence, the data is often classified using the individual method, then again with other parameters considered. Additional methods are also sometimes present in hybrid models, such as movement [168] or seat pressure detection [169], [170], [171], driving duration [104], [172] and voice input [173]. Furthermore, other features appeared in hybrid methods that weren’t common across individual method studies including gaze direction and pupil diameter in behavioural methods, vehicle speed, acceleration, and pedal input in vehicle methods. It was found that hybrid studies are able to minimise limitations seen in the other methods, however, some still exist and will be discussed in-depth in the following sections as these are the current limitations in progressing driver drowsiness technology.

VI. EXISTING CHALLENGES

A. Detection Versus Prediction

Behavioural and vehicle-based methods focus on the detection of drowsiness rather than prediction, despite the label “prediction” being used [172]. For behavioural methods, often prolonged eye closure is an indicator of drowsiness, where a microsleep may have already occurred. Furthermore, the commonly used measure, PERCLOS, has been found to detect drowsiness too late [144]. In vehicle-based studies, often lane departure is an indication of drowsiness; however, once a car has departed the lane, it has the potential to crash into oncoming traffic, indicating that this method is also too late to assist in driver safety. Advances in physiological monitoring has allowed for earlier prediction of drowsiness, where some features including changes in the VLF component of ECG can predict drowsiness up to 5 minutes before sleep [136], [137]. We recommend a time to event-based approach for vehicle-based and behavioural methods, where models are built to predict how long until the prolonged eye closure occurs, and focus on early prediction based on other features, rather than the detection of eye closure itself and the same for lane departure monitoring.

In terms of applying prediction in machine learning, various models can be developed with survival analysis [174]. Survival analysis can be used to determine the time until an event will occur, such as severe lane departure or falling asleep on the road. Implementing this technique will allow prediction to be properly implemented in drowsiness monitoring.

B. Feature Trend Discrepancies

Hundreds of features have been explored across physiological, behavioural, and vehicle-based studies. Some studies have differing reports regarding particular features and their correlation with drowsiness. We caution the use of these, as these features could vary depending on acquisition, environment or processing. Certain features may also be inherently different between subjects in different trials. Some of these features, in relation to increased drowsiness, include:

- Breathing rate, where one study noted that breathing rate was assumed to decrease [141] with drowsiness and another used thermal imaging to show this [116]; however, no change was observed in another study [136]. In contrast, one study reported there was an increase in breathing rate, where a breathing belt was used [51].
- The power of the low frequency (LF) component (ECG) was said to increase with drowsiness in [137]; however, other studies found that it decreases [130], [136].
- Blink frequency was said to decrease slightly in both [142], [146]; however, other studies suggested that blink frequency increases [86], [143]. Caffier et al. [142] expand upon other studies and how blink frequency was found to be conflicting for different studies, where the variations could be due to differing circumstances and/or environments for the subjects.

C. Physiological Intrusiveness

Physiological signals are considered be more intrusive, in particular when EEG and EOG signals are involved. The application of hybrid methods can address this, where a camera can be used to detect blinking parameters rather than EOG electrodes. Despite this, some studies still use glasses to monitor eye features [104] and EEG/EOG recordings [105], [168], [169], [171], [175], [176]. This includes the Muse headband, which is less intrusive than traditional EEG, but still more intrusive than other methods [154]. Accuracy reports have still been high however, when all medium level intrusive methods have been used.
D. Subject Diversity

Subject diversity has been lacking between subjects in areas including ethnicity, age, gender and glasses vs non-glasses users. Ethnicity variations can affect the accuracy of behavioural-based models, with most studies using just one ethnicity; however, a further problem exists where ethnicity is often not reported in studies. Research has shown that classifiers often reduce accuracy in darker skin tones, in particular for women, as models are often not diversely trained [177]. Of the available popular datasets (refer to Table X), the NTHU dataset is the most diverse, and well balanced [178]; but, has still been flagged as a potentially problematic when it comes to generalisation for end-to-end models, which require large amounts of training data [93]. This is a particular challenge in machine learning, as highlighted by the seminal gender shades paper [177], that is yet to be adequately addressed in drowsiness literature.

Gender bias was largely present in studies with one study recognising this as biased [169]. Fig. 6 demonstrates the percentages of male subjects in the studies, where 52 out of 126 studies reported the split of female and male subjects. Gender split is particularly important in behavioural and hybrid studies, as facial features can be different across gender.

Variance in ages of subjects is important when developing methods, as models can behave differently between younger and older participants. One hybrid study explored this, where age groups were split to compare those above and below 40 and a difference between age groups was found, where the model worked better for older subjects [138]. In contrast, another behavioural study reported their model to be worse for older subjects [36], demonstrating that a more diverse range of age groups needs to be considered to allow and test for these discrepancies. However, as demonstrated in Fig. 7, the median average age of participants is around 28–29, biasing models towards younger participants.

### TABLE X

| Dataset | Sample size | Modality | Weblink |
|---------|-------------|----------|---------|
| A sustained-attention driving task database [179] | 27 | Physiological | [link](https://figsharc.com/articles/dataset/Multi-channel_EEG_recordings_during_a_sustained-attention_driving_task/64273342) |
| A multimodality drowsiness database (called DROZY) [180] | 14 | Physiological+Behavioural | [link](http://www.drozy.ulg.ac.be/), accessed: 13/10/2022 |
| The National Tsing Hua University (NTHU) database [178] | 18 | Behavioural | [link](http://cv.cs.nthu.edu.tw/php/callforpaper/datasets/DDD/accessed: 13/10/2022) |
| Yawning detection dataset (YawDD) [181] | 29 | Behavioural | [link](https://ieeexplore.ieee.org/open-access/yawdd-yawning-detection-dataset/accessed: 13/10/2022) |
| UTA Real-Life Drowsiness Dataset (UTA-RLDD) [182] | 60 | Behavioural | [link](https://paperswithcode.com/dataset/uta-rldd/accessed: 13/10/2022) |
| Closed Eyes In The Wild (CEW) [183] | 2423 | Behavioural (images) | [link](http://parnec.nuaa.edu.cn/_upload/tpl/02/db/731/template731/pages/xian/CEW_Databases.html/accessed: 13/10/2022) |
| Eye-Chimera [184] | 40 | Behavioural (images) | [link](http://imag.pub.ro/common/staff/cfloreac/cf_research.html/accessed: 13/10/2022) |
| Zhejiang University Eyeblink Database (ZU) [185] | 20 | Behavioural | [link](http://www.cs.zju.edu.cn/gpan/accessed: 13/10/2022) |
| The Second Strategic Highway Research Program (SHRP2) [186] | 3400– | Vehicle+Behavioural | [link](https://insight.shrp2nds.us/accessed: 13/10/2022) |

Fig. 6. Percentage of Male Subjects used in studies.
Fig. 7. Reported average and maximum ages of participants used in studies. (average age reported in 24 studies). The maximum ages used in studies was more often reported (46 studies), where ages up to 80 were used; however, a concerning number of studies had the oldest participants of 40 or less (22 studies), as shown in Fig. 7.

Prescription glasses are worn full-time by 37% of Australians, with 66% of Australians reported wearing prescription glasses in general [187]. Hence, glasses and sunglasses are regularly worn driving and studies have reported reduced accuracy levels in behavioural monitoring for glasses users [35], [36], [37], [40], [188], [189], [190]. Despite this, behavioural and hybrid studies that include behavioural monitoring have specifically been reported to include glasses in only 47% of studies.

E. Number of Subjects

A limited number of subjects was repeatedly seen within the literature, where often below 20 subjects were reported, and a median of less than 15 seen in physiological, vehicle, and hybrid studies (Fig. 8). This relates to the lack of diversity, and models developed with such low numbers of participants cannot be appropriately generalised to the broader public.

F. Labelling of Drowsiness

A variety of drowsiness labels have been used including the previously described subjective measures, EEG recordings, eye closure features, professional labelling, and lane departure occurrences. The different labels can affect what is considered drowsiness and hence produce different results between models. The most common measure for labelling drowsiness was the subjective measure using KSS [15]; however, it has been noted that asking the driver can influence their drowsiness as the measure involves a verbal cue, where the driver is required to think [107]. One study carried out experiments on collecting KSS ratings at times of 5, 10 and 15 minutes, where 10 minutes was found to be optimum [107]. Furthermore, as stated previously, it was found that some studies use feature-based methods to decide where the drowsiness label will split [86], [102], [103], [104], whereas others did not always provide reasoning.

Other methods such as professional labelling (ORD or similar) can vary depending on the professional’s training; however, this measure can reduce the limitations of KSS measures. Lane departure, eye closure, and EEG measures have been used as both predictors and labels, where they do not have 100% accuracy and could be an untrustworthy label. Lane departure and eye closure have also been shown to predict drowsiness too late, meaning these cannot be used as features to create a prediction model. EEG could be a suitable measure and provide consistency if a protocol was developed, as this is a more accurate measure and properly shows the subjects physiological state. However, this could also be influenced by equipment and artefacts, especially on the road. Hence, the labelling of drowsiness is not consistent across studies and remains a barrier to the implementation of driver drowsiness detection and prediction schemes.

G. Performance Reporting

Model performance has been inconsistently reported across studies, making studies difficult to compare. There are a variety of factors that need to be considered when comparing studies, hence it is not possible to compare studies without looking at the full set of parameters. Factors contributing to the accuracy of studies can include:

- The variations in training, testing and validation splits for classification. Of the 80 studies that mention some form of splitting, there was found to be 36 studies that used a singular split and 38 that used some form of cross-validation. 6 studies reported the percentage splits, but not whether cross-validation or the singular split was used. One study demonstrated the issues of inconsistent splitting percentages, where a higher accuracy was presented for a lower amount of testing subjects (80% training, 20% testing) [125]. The most common reported and recommended split was 70% training and 30% testing.
- The level of drowsiness used – detecting slight drowsiness often provided a lower accuracy than severe drowsiness [102], [105], [171].
- The number of drowsiness levels considered. Studies with more than 2 levels of drowsiness often produced lower reported accuracies [102], [105].
- It was observed that studies sometimes made models and collected baselines per subject which reported better accuracies than across subjects [162], [168]. Furthermore,
studies that used cross-subject models should ideally have separate subjects in testing and training; however, they often used portions of each subject’s data in both test and training sets, which does not validate the model for external subjects. This has demonstrated to effect the accuracy in [154]. Furthermore, of the studies, it was found that just 7 of studies used leave-one-subject-out cross-validation; however, 17 studies still reported to separate subjects between testing and training. Leave-one-subject-out cross-validation is particularly valuable in drowsiness detection studies, where the number of subjects is often limited.

- The number of subjects and diversity of subjects as previously discussed can also affect the reported accuracy of a study.
- Almost all studies use different datasets where the data they use or collect is not publicly available [32], [48], [92], [123]
- On-road and simulation studies will produce different results due to the different artefacts present in the car on-road [30].
- The labelling of drowsiness, where different drowsiness labels may produce different results.

The numerical percentage for accuracy is often also expressed differently between studies, where some present the one value under “accuracy”, whereas others include specificity, sensitivity, precision, recall, root mean square error (RMSE) etc. Hence, it is difficult to compare accuracies across subjects, which is a particular concern when many researchers suggest that there are many other factors contributing to the outcome, including acquisition configuration, lighting and environment. We recommend studies clearly outline their protocol, as well as a number of performance metrics including accuracy, sensitivity, specificity and precision.

\section*{H. Time Window}

The time window of feature extraction, batch processing, and duration of the experiment are all important factors when monitoring drowsiness. In some studies, the length of time used for extracting features was shown to exceed multiple minutes, which is not feasible for early detection. Longer window frames can run the risk that an adverse event has already occurred by the time of detection, and possibly differ per user. Furthermore, some studies have collected chunks of data and labelled the entire segment as drowsy or non-drowsy; however, drowsiness can vary over smaller time intervals. Therefore, the collected data may not be a true representation of the instantaneous drowsiness of a subject and can alter models’ perception of drowsiness. Experiment duration should also be considered as some studies have reported links between duration of drive and drowsiness [172].

\section*{I. Data Collection Protocol}

Within the literature, it was found that experiment protocols need improvement and consistency [15]. On top of the previously discussed issues, protocols need to include both day and night studies, in particular when behavioural and vehicle-based data is involved. Drowsiness is largely prominent at night [191], yet only 47% of behavioural, vehicle, and hybrid papers included night scenarios. However, some studies with night monitoring did not work in the daytime [35]. Drowsiness can occur at any time and has often been shown to be present between 14:00 and 16:00 due to the circadian effect, as well as at night [11]. Furthermore, subjects may not maintain protocols which could hinder results.

\section*{J. Simulation Versus on Road}

Within the literature, an absence of on-road studies was markedly present. Of the studies included in this survey, 15 out of 126 had on-road data, often lacking drowsy participants. Earlier studies had more on-road data (11 out of the 15 prior to 2018), perhaps due to the lack of technology available for simulations. The on-road studies often had few participants and minimal drowsy participants (if any), where behavioural studies often asked subjects to “act” drowsy [39], [192]. Differences were reported between simulated and on-road studies, such as the effect of duration and perception of risk, furthering the need for on-road studies to be conducted [11], [161], [193]. However, one study stated that comparable results can be seen in EOG related metrics when the same drive abortion criteria is used for both simulation and on-road drives [103].

\section*{K. Loss of Data}

The loss of data can occur as sensors can detach, weather interfering with vehicle-based recordings or lighting variances hindering behavioural recordings. When using an adaptable hybrid model or a model with multiple sources, data loss does not have to mean ceased drowsiness monitoring. One study explored this, with a system that continued to work when a signal was lost and drowsiness monitoring could continue [194]. Data loss can occur at any time during monitoring, hence the system should continue to work if this is to occur.

\section*{L. Sleep Disorders}

Sleep disorders have been reported to increase the chance of a driver being drowsy as they often affect the quality of sleep [116]. Sleep disorders can also affect the results of drowsiness studies and hence most studies do not allow participants with sleep disorders to take part. However, this area still needs to be explored. For example, 33% of all accidents in Australia have been related to sleep problems [195], and excluding these participants from studies is thus extremely problematic.

One study explored the effects of obstructive sleep apnoea (OSAS), various sleep disorders (VSD) and “normals” in relation to the sleep-wake transition [136]. The study collected EEG data, respiration, EMG and HR data, where the identified differences between subject types are summarised as follows:

- Respiration frequency was different for VSD’s; however, the variability did not change across the three.
- EMG results appeared to be the same across the groups.
- ECG R-R Intervals (RRI) had correlation with drowsiness for the normals and VSD’s but not for OSAS.
• The very low frequency power component of RRI was consistent across all three groups.
• The low frequency power had correlation with drowsiness for normal and VSD groups but not OSAS.

This study demonstrated that differences are seen between subject types and this should be explored when developing devices for on-road use. Almost all other drowsiness studies omit these subjects from trials; however, driver drowsiness products cannot be considered effective until sleep disorder integration and monitoring is included.

M. Public Acceptance

For behavioural monitoring in particular, public acceptance is a challenge for driver drowsiness technologies due to concerns regarding data storage and privacy protection. As camera monitoring is required, concerns of what else the data may be used for may hinder public acceptance of the product. Ways to counter this can include driver deidentification as described in [34]. This can include extracting only relevant components required to determine if the driver is drowsy, such as eye and mouth segments. The rest of the image data can then be blurred, blackened or the components completely extracted from the image to make it harder to identify drivers.

VII. CONCLUSION AND FUTURE DIRECTIONS

Progressing driver drowsiness research requires current challenges to be addressed in order to develop reliable models for commercial use. We discussed how each of these modalities may be affected by noise and data loss/contact loss, ambient conditions, subject diversity etc. It is hoped that multi-modal data may alleviate this, however, this provides further implications such as an increase in the number of sensors. The use of deep learning-based methods (e.g., CNN or RNN) could be useful for the multi-modal data as they have a higher-order feature extraction and fusion strategies in an end-to-end fashion. Light-weight deep learning models could be useful to deploy as a low-cost and real-time option in a vehicle for the drowsiness detection. However, this does not negate the need for data quality issues raised above to be addressed.

Current research is promising in terms of accuracy and foundations; however, further research should be conducted to validate approaches for on-road use. Adoption of models should also be considered, with the objective of prediction rather than detection, before the driver is put at risk.

A large number of challenges were presented in the literature, many of which relate to the data collection for driver drowsiness detection. Hence, for studies that are able to collect their own data, we emphasise that authors should be consistent and address the issues raised insofar as possible, with particular focus on the following:
• Collecting on-road data where possible, with appropriate ethical considerations.
• Recruiting a diverse range and larger number of participants, with generalisability in mind.

• Labelling of drowsiness should be reasoned and if KSS is used, an adequate time frame between KSS samples should be maintained.
• Protocols for data collection needs to be thorough, with both day and night studies conducted for vehicle-based and behavioural-based studies in particular.

Furthermore, analysis of collected data needs to be addressed, with appropriate model validation. Proper window sizes for feature collection should be used in order to detect drowsiness in a timely manner. Prediction rather than detection should be used, in order to notify the driver in time, before an accident or adverse event occurs. Further research can include how far in advance drowsiness can be predicted and what methods are best suited for this task.

Hybrid models have shown to be more accurate, flexible and use the benefits of each individual method. They can allow for data loss and overcome the downfalls of intrusiveness, as a variety of sources can be used to collect data. However, they come at an increased complexity of development and cost. Further research should be undertaken for these approaches, bearing in mind the complexity they carry.

Going forward, it is important for studies of drowsiness detection to also require subjects with sleep disorders to be included in studies, so they too can be monitored effectively on-road. Future work could include further exploring the effects of gender, age and driver experience on drowsiness models also. Once these challenges are addressed, accurate monitoring of drowsiness and hopefully more commercial implementation can be achieved.

REFERENCES

[1] “Road traffic injuries,” World Health Organisation, 2021. [Online]. Available: https://www.who.int/news-room/fact-sheets/detail/road-traffic-injuries
[2] L. Collingwood, “Privacy implications and liability issues of autonomous vehicles,” Inf. Commun. Technol. Law, vol. 26, no. 1, pp. 32–45, 2017.
[3] G. Li and W.-Y. Chung, “Detection of driver drowsiness using wavelet analysis of heart rate variability and a support vector machine classifier,” Sensors, vol. 13, no. 12, pp. 16494–16511, 2013. [Online]. Available: https://www.mdpi.com/1424-8220/13/12/16494
[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,” Occup. Environ. Med., vol. 57, no. 10, pp. 649–655, 2000. [Online]. Available: https://pubmed.ncbi.nlm.nih.gov/10984335https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1739867/
[5] M. D. Mulhall et al., “Sleepiness and driving events in shift workers: The impact of circadian and homeostatic factors,” Sleep, vol. 42, no. 6, 2019, Art. no. zsz074. [Online]. Available: https://doi.org/10.1093/sleep/zsz074
[6] 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
[7] WA Road Fatalities 2020. 2021. [Online]. Available: https://www.rsc.wa.gov.au/RSC/media/Documents/Road%20Data/Statistics/WA-Road-Fatality-Progress-2020.pdf
[8] Inquiry into the Increase in Victoria’s Road Toll, 2020. [Online]. Available: https://www.parliament.vic.gov.au/images/stories/committees/SCEI/Inquiry_into_the_Increase_in_Victorians_Road_Toll/Submissions/S53_-_RACV_Redacted.pdf
[9] “Drivers are falling asleep behind the wheel.” 2015. [Online]. Available: https://www.nsc.org/road-safety/safety-topics/fatigued-driving
[10] “Drowsy driving,” National Sleep Foundation. Accessed: Aug. 4, 2021. [Online]. Available: https://drowsydriving.org/about/facts-and-stats/
[11] A. Sahayadh, K. Sundaraj, and M. Murugappan, “Detecting driver drowsiness based on sensors: A review,” Sensors, vol. 12, no. 12, pp. 16937–16953, 2012. [Online]. Available: https://www.mdpi.com/1424-8209/12/12/16937
[12] S. Kaplan, M. A. Kuvamsan, A. G. Yavuz, and Y. Karalurt, “Driver behavior analysis for safe driving: A survey,” IEEE Trans. Intel. Transp. Syst., vol. 16, no. 6, pp. 3017–3032, Dec. 2015.
[13] M. Doudou, A. Bouabdallah, and V. Berge-Chefrouoi, “Driver drowsiness measurement technologies: Current research, market solutions, and challenges,” Int. J. Intell. Transp. Syst. Res., vol. 18, pp. 297–319, 2019.
[14] M. Ramzan, H. U. Khan, S. M. Awan, A. Ismail, I. Aylas, and A. Mahmood, “A survey on state-of-the-art drowsiness detection techniques,” IEEE Access, vol. 6, pp. 61904–61919, 2019.
[15] X. Hu and G. Lodewijks, “Detecting fatigue in car drivers and aircraft pilots by using non-invasive measures: The value of differentiation of sleepiness and mental fatigue,” J. Saf. Res., vol. 72, pp. 173–187, 2020. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0022437519306735
[16] A. Németh et al., “Multimodal features for detection of driver stress and fatigue,” IEEE Trans. Intell. Transp. Syst., vol. 22, no. 6, pp. 3214–3233, Jun. 2021.
[17] C. Stuts, J. W. Wilkins, and B. V. Vaughn, “Why do people have clumsy driving crashes,” In Proc. Drivers Who Just Did, vol. 202, no. 638, 1999, Art. no. 5944.
[18] T. Akerstedt, B. Peters, A. Anund, and G. Kecklund, “Impaired alertness and performance driving home from the night shift: A driving simulator study,” J. Sleep Res., vol. 14, no. 1, pp. 17–20, 2005. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1365-2869.2004.00437.x
[19] J. M. Lyznicki, T. C. Doege, R. M. Davis, and M. A. Williams, “Sleepiness, driving, and motor vehicle crashes,” JAMA, vol. 279, no. 23, pp. 1908–1913, 1998.
[20] K. Cherney, “How long do caffeine stay in your system?,” vol. 2021, pp. 7821–7826, 2014. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0001457520317206
[21] J. C. Stutts, J. W. Wilkins, and B. V. Vaughn, “Why do people have clumsy driving crashes,” In Proc. Drivers Who Just Did, vol. 202, no. 638, 1999, Art. no. 5944.
[22] T. Akerstedt, B. Peters, A. Anund, and G. Kecklund, “Impaired alertness and performance driving home from the night shift: A driving simulator study,” J. Sleep Res., vol. 14, no. 1, pp. 17–20, 2005. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1365-2869.2004.00437.x
[23] K. Cherney, “How long does caffeine stay in your system?,” vol. 2021, pp. 7821–7826, 2014. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0001457520317206
[24] J. C. Stutts, J. W. Wilkins, and B. V. Vaughn, “Why do people have clumsy driving crashes,” In Proc. Drivers Who Just Did, vol. 202, no. 638, 1999, Art. no. 5944.
[25] T. Akerstedt, B. Peters, A. Anund, and G. Kecklund, “Impaired alertness and performance driving home from the night shift: A driving simulator study,” J. Sleep Res., vol. 14, no. 1, pp. 17–20, 2005. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1365-2869.2004.00437.x
[26] K. Cherney, “How long does caffeine stay in your system?,” vol. 2021, pp. 7821–7826, 2014. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0001457520317206
[27] J. C. Stutts, J. W. Wilkins, and B. V. Vaughn, “Why do people have clumsy driving crashes,” In Proc. Drivers Who Just Did, vol. 202, no. 638, 1999, Art. no. 5944.
[28] T. Akerstedt, B. Peters, A. Anund, and G. Kecklund, “Impaired alertness and performance driving home from the night shift: A driving simulator study,” J. Sleep Res., vol. 14, no. 1, pp. 17–20, 2005. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1365-2869.2004.00437.x
[29] J. C. Stutts, J. W. Wilkins, and B. V. Vaughn, “Why do people have clumsy driving crashes,” In Proc. Drivers Who Just Did, vol. 202, no. 638, 1999, Art. no. 5944.
[52] T. Hwang, M. Kim, S. Hong, and K. S. Park, “Driver drowsiness detection using the in-ear EEG,” in Proc. IEEE 38th Annu. Int. Conf. Eng. Med. Biol. Soc., 2016, pp. 4646–4649.

[53] J. Gielen and J.-M. Aerts, “Feature extraction and evaluation for driver drowsiness detection based on thermoregulation,” Appl. Sci., vol. 9, no. 17, 2019, Art. no. 3555. [Online]. Available: https://www.mdpi.com/2076-3417/9/17/3555

[54] W. Wang, D. Zhao, W. Han, and J. Xi, “A learning-based approach for lane departure warning systems with personalized driver model,” IEEE Trans. Veh. Technol., vol. 67, no. 10, pp. 9145–9157, Oct. 2018.

[55] D. Wu, V. J. Lawhern, S. Gordon, B. J. Lance, and C.-T. Lin, “Driver drowsiness estimation from EEG signals using online weighted adaptation regularization for regression (OwARR),” IEEE Trans. Fuzzy Syst., vol. 25, no. 6, pp. 1522–1535, Dec. 2017.

[56] M. W. Johns, A. Tucker, R. Chapman, K. Crowley, and N. Michael, “Monitoring eye and eyelid movements by infrared reflectance oculography to measure drowsiness in drivers,” Somnologie-Schlafforsch. und Schlafmedizin, vol. 11, no. 4, pp. 234–242, 2007.

[57] “Driver vigilance telemetric control system - vigilton,” 2013. [Online]. Available: http://www.neurocom.ru/en2/product/vigilton.html

[58] “Stopsleep,” 2017. [Online]. Available: https://www.stopsleep.com.au/

[59] “Stopsleep,” 2017. [Online]. Available: https://www.stopsleep.com.au/

[60] “Life by smartcap,” 2016. [Online]. Available: http://www.smartcaptech.com/

[61] “Eagle light.” Accessed: May 21, 2021. [Online]. Available: https://www.optalert.com/explore-products/scientifically-validated-glasses-mining/

[62] “Hxgn mineprotect operator alertness system light vehicle,” 2019. [Online]. Available: https://www.hexagonmining.com/solutions/safety-portfolio/hxgm-mineprotect-operator-alertness-system

[63] “Guardian.” Accessed: May 21, 2021. [Online]. Available: https://www.fraunhofer.de/en/press/newsroom/2010/10/eye-tracker-driver-drowsiness.html

[64] “Infrared led detectors in microsolv,” 2021, vol. 21 May, 2010. [Online]. Available: http://w1.siemens.com.cn/news/en Frontier_technology_en/1748.aspx

[65] “Hxgm mineprotect operator alertness system light vehicle,” 2019. [Online]. Available: https://hexagonmining.com/solutions/safety-portfolio/hxgm-mineprotect-operator-alertness-system

[66] “Saab tech: Keep drivers awake and focused,” vol. 2014, no. 2014. [Online]. Available: https://www.saabplanet.com/saab-tech-keep-drivers-aware-and-focused

[67] “Intercor’s driver alertness detection systemTM,” 2014. [Online]. Available: https://www.newswire.com/news/intercores-driver-alterness-detection-system-now-available

[68] “Drivealer.” Accessed: May 21, 2021. [Online]. Available: https://driverisk.com.au/drivealer/

[69] “Driver fatigue monitoring system driving status detection.” Accessed: May 21, 2021. [Online]. Available: https://stonum.com/products-driver-fatigue-monitoring-system-ad-ai.html

[70] “TS driver fatigue monitor DFMZ.” Accessed: May 21, 2021. [Online]. Available: http://www.transportssupport.co.uk/product/ts-driver-fatigue-monitor-dfmz2/ykbbr6gzaUk

[71] “Fas100-user-guide-may2007.” 2007. [Online]. Available: http://rasice.alan-electronics.de/Archiv/Fahrer-Assistenz-Systeme/FAS%20100/FAS100-user-guide.pdf

[72] “Drowsy driver detection systems sense when you need a break,” vol. 2021, no. 24 May, 2016. [Online]. Available: https://www.cars.com/articles/drowsy-driver-detection-systems-sense-when-you-need-a-break/142068/44/9199

[73] “Driver alert control (DAC),” 2018. [Online]. Available: https://www.volvocars.com/en-th/support/manuals/s60/2016/w17/driver-support/driver-alert-system/driver-alert-control-dac

[74] “Cognex vision helps vehicle “see” the road in 2007 darpa urban challenge,” 2007. [Online]. Available: https://www.cognex.com/company/press-releases/2007/cognex-vision-helps-vehicle-see-the-road-in-2007-darpa-urban-challenge

[75] “A Texas safety system, Detects: Accessed: May 24, 2021. [Online]. Available: https://drivers.lexus.com/lexus-drivers-theme/pdf/LSS%20Quick%20Guide%20Link.pdf

[76] “Subaru eyesight.” Accessed: May 24, 2021. [Online]. Available: https://www.subaru.com/engineering/eyesight.html

[77] “Ford summary of operation,” FMI Report. Accessed: May 24, 2021. [Online]. Available: https://fmiapplications.com/astid.html

[78] “Saab tech: Keep drivers awake and focused,” vol. 2021, no. 24 May, 2008. [Online]. Available: https://www.ietis.com/news/ietis-announces-introduction-industry-first-lane-departure-warning-data-collection-product

[79] “Driver fatigue detection.” Accessed: May 24, 2021. [Online]. Available: https://www.milesoncontinental.co.nz/news/features/driver-fatigue-detection

[80] “Toyota launches is and mirai equipped with “advanced drive” that enables drivers and cars to drive together in Japan,” vol. 2021, no. 24 May, 2021. [Online]. Available: https://global.toyota/en/newsroom/corporate/35063150.html

[81] “Audi driver Fatigue—Overview.” Accessed: May 24, 2021. [Online]. Available: https://www.bosch-mobility-solutions.com/en/solutions/interior-interior-monitoring-systems/

[82] B. Thorslund, “Electrooculogram analysis and development of a system for defining stages of drowsiness,” Statens väg-och transportforskningstit, Linköping, 2004.

[83] M. Simon et al., “EEG alpha spindle measures as indicators of driver fatigue under real traffic conditions,” Clin. Neurophysiol., vol. 122, no. 6, pp. 1168–1178, 2011.

[84] S. Murugan, J. Selvaraj, and A. Sahayadhas, “Detection and analysis of driver fatigue monitoring system with electrocardiogram (ECG),” Phys. Eng. Sci. Med., vol. 43, no. 2, pp. 525–537, 2020. [Online]. Available: https://doi.org/10.1007/s12265-020-00853-8

[85] M. Chai, S.-W. Li, W.-C. Sun, M.-Z. Guo, and M.-Y. Huang, “Drowsiness monitoring based on steering wheel status,” Transp. Res. Part D: Transport Environ., vol. 66, pp. 95–103, 2019. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1361920917306582

[86] M. M. Hasan, C. N. Watling, and G. S. Larue, “Physiological signal-based drowsiness detection using machine learning: Singular and hybrid signal approaches,” J. Saf. Res., vol. 80, pp. 215–225, 2022.

[87] H. Zhang, C. Wu, Z. Huang, X. Yan, and T. Z. Qiu, “Sensitivity of lane position and steering angle measurements to driver fatigue,” Transp. Res. Rec., vol. 2585, no. 1, pp. 67–76, 2016. [Online]. Available: https://doi.org/10.3141/2585-08

[88] M. H. Baccour, F. Driewer, T. Schäck, and E. Kasneci, “Camera-based driver drowsiness state classification using logistic regression models,” in Proc. IEEE Int. Conf. Sport, Cybern., 2020, pp. 1–8.

[89] M. Ngxande, J.-R. Tapamo, and M. Burke, “Bias remediation in driver drowsiness detection systems using generative adversarial networks,” IEEE Access, vol. 8, pp. 55592–55601, 2020.

[90] Z.-K. Gao, Y.-L. Li, Y.-X. Yang, and C. Ma, “A recurrence network-based convolutional neural network for fatigue driving detection from EEG,” Chaos: Interdiscip. J. Nonlinear Sci., vol. 29, no. 11, 2019, Art. no. 113126.

[91] K. Qian, T. Koike, T. Nakamura, B. Schuller, and Y. Yamamoto, “Learning multimodal representations for drowsiness detection,” IEEE Trans. Intell. Transp. Syst., vol. 23, no. 8, pp. 11539–11548, Aug. 2022.

[92] A. Adadi and M. Berrada, “Peeking inside the black-box: A survey on explainable artificial intelligence (XAI),” IEEE Access, vol. 6, pp. 52138–52160, 2018.

[93] A. Rajkar, N. Kulkarni, and A. Raut, “Driver drowsiness detection using deep learning,” in Proc. IEEE Int. Conf. Indi. Inf., 2022, pp. 73–82.

[94] S. Nandhini, V. Venkatasubramanian, and C. Aparna, “Drowsiness detection using facial features, image processing and machine learning,” in Computer Networks, Big Data and IoT. Berlin, Germany: Springer, 2022, pp. 565–574.
A. Joshi, S. Kyal, S. Banerjee, and T. Mishra, “In-the-wild drowsiness detection from facial expressions,” in Proc. IEEE 1st Int. Conf. Comput., Commun., Comput. Technol., 2020, pp. 1794–1800.

H. D. Rosario, J. S. Solaz, N. Rodriguez, and L. M. Bergasa, “Controlled induction and measurement of drowsiness in a driving simulator,” IET Intell. Transport Syst., vol. 4, no. 4, pp. 280–288, 2010.

J. Gwak, A. Hirao, and M. Shino, “An investigation of early detection of driver drowsiness using ensemble machine learning based on hybrid sensing,” Appl. Sci., vol. 10, no. 8, 2020, Art. no. 2890. [Online]. Available: https://www.mdpi.com/2076-3417/10/8/2890

C. J. D. Naurios, C. Bourdin, A. Stratulat, E. Diaz, and J.-L. Vercher, “Detection and prediction of driver drowsiness using artificial neural network models,” Accident Anal. Prevention, vol. 126, pp. 95–104, 2019. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0001457517304347

C. Craye, A. Rashwan, M. S. Kamel, and F. Karray, “A multi-modal driver fatigue and distraction assessment system,” Int. J. Intell. Transp. Syst. Res., vol. 14, no. 3, pp. 173–194, 2016. [Online]. Available: https://doi.org/10.1371/journal.pone.0122243

G. Li, B. Lee, and W. Chung, “Smartwatch-based wearable EEG system for driver drowsiness detection,” IEEE Sensors J., vol. 15, no. 12, pp. 7169–7180, Dec. 2015.

A. Mehreen, S. M. Anwar, M. Haseeb, M. Majid, and M. O. Allah, “A hybrid scheme for driver drowsiness detection using wearable sensors,” IEEE Sensors J., vol. 19, no. 13, pp. 5119–5126, Jul. 2019.

S. Koh et al., “Driver drowsiness detection via PPG biosignals by using multimodal head support,” in Proc. IEEE 4th Int. Conf. Control, Decis. Inf. Technol., 2017, pp. 0383–0388.

R. Favilla, V. C. Zuccalà, and G. Coppini, “Heart rate and heart rate variability from single-channel video and ICA integration of multiple signals,” IEEE J. Biomed. Health Inform., vol. 23, no. 6, pp. 2398–2408, Nov. 2019.

F. Zhao, M. Li, Y. Qian, and J. Z. Tsien, “Remote measurements of heart and respiration rates for telemedicine,” PLoS ONE, vol. 8, no. 10, 2013, Art. no. e71384.

V. Balasubramanian and K. Adalarasu, “EMG-based analysis of change in muscle activity during simulated driving,” J. Bodywork Movement Therapies, vol. 11, no. 2, pp. 151–158, 2007. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1360859207000334

A. Joshi, S. Kyal, S. Banerjee, and T. Mishra, “In-the-wild drowsiness detection from facial expressions,” in Proc. IEEE Int. Conf. Signal Image Process. Appl., 2011, pp. 337–341.

P. Philip et al., “Fatigue, sleep restriction and driving performance,” Accident Anal. Prevention, vol. 37, no. 3, pp. 473–478, 2005.

E. Vural, M. Cetin, A. Ercil, G. Littlewort, M. Bartlett, and J. Movellan, “Drowsy driver detection through face expression analysis,” in Proc. Int. Workshop Hum.-Comput. Interact., 2007, pp. 6–18.

Z. Li, L. Chen, J. Peng, and Y. Wu, “Automatic detection of driver fatigue using driving operation information for transportation safety,” Sensors, vol. 17, no. 6, 2017, Art. no. 1212.

G. Zhenhai, L. DinhDat, H. Hungyu, Y. Ziwen, and W. Xinyu, “Driver drowsiness detection based on time series analysis of steering wheel angular velocity,” in Proc. IEEE 9th Int. Conf. Measuring Technol. Mechatron. Automat., 2017, pp. 99–101.
[187] “The 2020 vision index,” Optometry Australia, South Melbourne, VIC, Australia, Report, 2020. [Online]. Available: https://www.optometry.org.au/wp-content/uploads/GVFL/Vision_Index/2020-Vision-Index-Report-FINAL.pdf

[188] S. Jamshidi, R. Azmi, M. Sharghi, and M. Soryani, “Hierarchical deep neural networks to detect driver drowsiness,” Multimedia Tools Appl., vol. 80, no. 10, pp. 16045–16058, 2021. [Online]. Available: https://doi.org/10.1007/s11042-021-10542-7

[189] R. Jabbar, K. Al-Khalifa, M. Kharbeche, W. Alhajyaseen, M. Jafari, and S. Jiang, “Real-time driver drowsiness detection for android application using deep neural networks techniques,” Procedia Comput. Sci., vol. 130, pp. 400–407, 2018. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1877050918304137

[190] F. Dornaika, J. Reta, I. Arganda-Carreras, and A. Moujahid, “Driver drowsiness detection in facial images,” in Proc. 8th Int. Conf. Image Process. Theory, Tools Appl., 2018, pp. 1–6.

[191] B.-L. Lee, B.-G. Lee, and W.-Y. Chung, “Standalone wearable driver drowsiness detection system in a smartwatch,” IEEE Sensors J., vol. 16, no. 13, pp. 5444–5451, Jul. 2016.

[192] I.-H. Choi, C.-H. Jeong, and Y.-G. Kim, “Tracking a driver’s face against extreme head poses and inference of drowsiness using a hidden Markov model,” Appl. Sci., vol. 6, no. 5, 2016, Art. no. 137. [Online]. Available: https://www.mdpi.com/2076-3417/6/5/137

[193] E. Blana and J. Golias, “Differences between vehicle lateral displacement on the road and in a fixed-base simulator,” Hum. Factors, vol. 55, no. 1, pp. 303–313, 2002.

[194] S. Samiee, S. Azadi, R. Kazemi, A. Nahvi, and A. Eichberger, “Data fusion to develop a driver drowsiness detection system with robustness to signal loss,” Sensors, vol. 14, no. 9, pp. 17832–17847, 2014. [Online]. Available: https://www.mdpi.com/1424-8220/14/9/17832

[195] S. J. Motlagh, M. Shabany, K. S. Haghighi, A. N. Nasrabadi, and S. H. E. Razavi, “Relationship between sleep quality, obstructive sleep apnea and sleepiness during day with related factors in professional drivers,” Acta Medica Iranica, vol. 55, no. 11, pp. 690–695, 2017.

Emma Perkins (Graduate Student Member, IEEE) received the B.Eng. degree in biomedical engineering from RMIT University, Melbourne, VIC, Australia, in 2020. She is currently working toward the Ph.D. degree with the Department of Electrical and Computer Systems Engineering, Monash University, Melbourne, VIC, Australia. She is currently researching and developing a driver drowsiness detection system. Her research interests include physiological monitoring, machine learning, biomedical devices, and computer vision.

Chiranjibi Sitaula received the Ph.D. degree from Deakin University, Burwood, VIC, Australia, in 2021. He is currently a Research Fellow under the Department of Electrical and Computer Systems Engineering, Monash University, Melbourne, VIC, Australia. He was with industry and academia in Nepal for a number of years before joining Deakin University for a Ph.D. degree. He has authored or coauthored research articles in top-tier conferences and journals in the field of deep learning and machine learning. His research interests include computer vision, signal processing, and machine learning.

Michael Burke (Member, IEEE) received the Ph.D. degree in statistical signal processing from the University of Cambridge, Cambridge, U.K., in 2016. He is currently a Lecturer specialising in computer vision, machine learning and control for autonomous robots with the Department of Electrical and Computer Systems Engineering, Monash University, Clayton, VIC, Australia. From 2018 to 2020, he was a Research Associate working on robot learning with the University of Edinburgh, Edinburgh, U.K. Before this, Michael led the Mobile Intelligent Autonomous Systems Group with the Council for Scientific and Industrial Research (CSIR), South Africa.

Faezeh Marzbanrad (Senior Member, IEEE) received the Ph.D. degree from the University of Melbourne, Parkville, VIC, Australia in 2016. She is currently a Lecturer and the Head of Biomedical Signal Processing Lab with the Department of Electrical and Computer Systems Engineering, Monash University, Clayton, VIC, Australia. Her research interests include biomedical signal processing, machine learning, affordable medical technologies, and mobile-health.