Can Steering Wheel Detect Your Driving Fatigue?

Jianchao Lu, Xi Zheng, Member, IEEE, Lihong Tang, Tianyi Zhang, Member, IEEE, Quan Z. Sheng, Member, IEEE, Chen Wang, Jiong Jin, Member, IEEE, Shui Yu, Senior Member, IEEE, Wanlei Zhou, Senior Member, IEEE

Abstract—Automated Driving System (ADS) has attracted increasing attention but the state-of-the-art ADS largely depend on vehicle driving parameters and facial features, which lacks reliability. Approaches using physiological based sensors (e.g., electroencephalogram or electrocardiogram) are either too clumsy to wear or impractical to install. In this paper, we propose a novel driver fatigue detection method by embedding surface electromyography (sEMG) sensors on a steering wheel. Compared with the existing methods, our approach is able to collect bio-signals in a non-intrusive way and detect driver fatigue at an earlier stage. The experimental results show that our approach outperforms existing methods with the weighted average F1 scores of about 90%. We also propose promising future directions to deploy this approach in real-life settings, such as applying multimodal learning using several supplementary sensors.

Index Terms—Fatigue, Road accidents, EMG measurement, Vehicle driving, Feature generation, Biomedical monitoring, Vehicle safety, Steering wheel, Human vehicle interaction

I. INTRODUCTION

The Automated Driving System (ADS) refers to an automated driving mechanism that takes over the vehicle and allows human drivers to leave all responsibilities to the driving system. Several companies have been actively implementing Level 3 ADS projects, meaning that the vehicles can guide themselves automatically under certain conditions [1]. However, driver-less vehicles are still far away from us. Alternatively, ADS is currently adopting a Human in The Loop (HITL) design, where the driver’s anomaly needs to be detected [2].

Detecting anomalous driving behavior is not only important for designing automated driving systems, but also is critical for driving safety. One of the most serious anomalous driving behaviors is fatigue, which can be referred to as a state where the person is neither in sleep nor awake state [3]. Fatigue driving has become one of the major causes of deaths and accidents across the world. According to National Highway Traffic Safety Administration (NHTSA), around 50,000 injuries and 800 deaths were reported from car accidents because of the fatigue driving in the United States alone in 2017 [4]. Similarly, it accounts for 25%-30% of the total road accidents in China [5]. The ever increasing number of accidents caused by fatigue motivates us to propose a practical yet accurate driving fatigue detection system, which could be eventually integrated into the ADS.

Driver fatigue detection (FDD) is the technology that helps prevent accidents caused by the driver’s drowsiness. Several approaches have been proposed for detecting driving fatigue, such as the vehicle behaviour based method. In [6], the authors offer fatigue detection systems by monitoring vehicle movements using various sensors, including cameras, speed sensors, GPS, and accelerometer sensors. However, the shortcoming of such systems is that drivers are typically warned only at a deep fatigue state, which could be too late when the warning occurs. The facial-feature based method is another way to detect fatigue driving. For example, head movement is extracted through the camera to evaluate the drivers’ fatigue state in [7]. But such methods may fail to detect fatigue driving because of the surrounding context such as sunlight or darkness, and wearable devices. Moreover, previous work also uses bio-signals for fatigue driving detection. Electromyography (EMG), electroencephalogram (EEG), electrocardiogram (ECG), and electrooculography (EOG) are typical bio-signals to measure the physiological state of a driver [5]. It is, however, difficult to collect data from such bio-sensors since they require attaching electrodes and wires directly to the driver. This requirement of attaching sensors to the human body while driving is not practical for real-world fatigue detection [8]. Some recent work also explores the possibility of attaching sensors to driving seats for fatigue detection [9]. But in such a scenario, a driver needs to wear thin clothing to facilitate direct contact with the sensors, making such a technique impractical in real life. To overcome these limitations, our system adopts surface electromyography (sEMG) sensors, which are attached to the steering wheel of a vehicle for convenient and practical fatigue detection.

In this paper, we propose a steering-wheel based sensor deployment solution to detect fatigue driving degree. First, we design a sEMG sensor, which is designed with a detachable style. The sensor is connected with a data collection board to collect sensor data in real-time. Second, to process the data collected from our sEMG sensor, we design a Valid sEMG Selection Machine (VsESM) to detect the driving fatigue state. It addresses the problem aroused from the different interval settings of the fatigue detection points. Besides, it also distin-
guishes the valid signal from the raw signal. Compared with existing methods, our approach is able to (1) keep monitoring the drivers’ fatigue state while driving rather than a later warning only at a deep fatigue state, (2) avoid failing to detect fatigue driving because of the surrounding context, and (3) collect the bio-signals in a non-intrusive way. The experiment shows that our approach outperforms existing methods with the weighted average F1 scores of about 90%. In addition, we conduct two real-life qualitative studies. The studies show that there are insights and promising future directions to make our solution a reality. Our primary contributions are three-fold:

1) Novel feature engineering for customized sEMG sensors: We design and build our sensors together with a signal collection device. The sensor can be detached to the steering-wheel which collects the bio-signal in a non-intrusive way.

2) Sound methodology: We propose a novel approach called VsESM to extract valid sEMG signals from noisy data caused by lane changing and resultant hand movement etc. and design two-layer features for the underlying sEMG signals.

3) Extensive studies: We conduct two extensive studies in both lab environment and real-world. 17 experienced drivers are invited to our experiment, and each is asked to drive one and a half hour approximately. The results show useful insights to adapt our approach for real-life usage.

The rest of the paper is organized as follows. Section II discusses the related work of fatigue driving detection systems. Section III walks through the unique characteristics of sEMG sensors. The overview of our proposed method is presented in Section IV. The empirical settings are discussed in Section V. The experimental results are reported in Section VI. Finally, the conclusion and future works are drawn in Section VII.

II. RELATED WORK

A. Vehicle Behavior based Methods

When getting fatigued, a driver’s ability to perceive the surrounding traffic circumstance and judge the driving situation will decrease. This change impacts how a driver controls the vehicle, which can be reflected by the vehicle’s abnormal performance [10]. Therefore, vehicle behaviors collected by various types of sensors inside a vehicle (e.g., accelerometer and the force sensor embedded in the steering wheel) can be used to determine whether the driver is in fatigue driving.

Ma et al. [11] studied the correlation between speed variability and driving fatigue. Their results indicated that the variability of vehicle speed was negatively correlated with drivers’ fatigue state. They argued that drivers with less variability of vehicle speed had more fatigue than those with more variability of vehicle speed did during the experiment. Krajewski et al. [12] proposed a fatigue detection system based on the steering wheel angle. They found that the operation of the steering wheel would be obviously abnormal when the driver felt drowsy or sleepy. More specifically, they indicated that the driver in a fatigue state would take a longer time to correct the steering wheel and the instantaneous correction range could be extremely larger than normal. LEE et al. [13] developed a fatigue detection system based on the grip force on the steering wheel. They analysed the changing pattern of the driving steering wheel grip force and argued that the steering wheel grip force would be gradually decreased while the participants’ fatigue state increased concurrently over time.

Vehicle behavior based methods can easily collect the data from sensors embedded in the car without affecting the driver’s normal driving. However, its detection accuracy is easily affected by a driver’s driving habits, weather, road, and traffic conditions and other external factors. Moreover, the method is only able to detect fatigue driving when the driver is about to lose the control of the vehicle, which is obviously not safe for the driver [14]. Therefore, the fatigue detection result from this method is better to be used as supplementary information for the driver.

B. Facial-feature based Methods

The techniques based on facial features have been extensively adopted in fatigue driving detection, mainly due to their non-disturbance on the driver’s attention in the driving process. Most existing methods attempt to detect drivers’ fatigue facial features, e.g., yawning, blink frequency, gaze direction, eye state, and head position.

Alioua et al. [15] introduced a driving fatigue detection system based on the yawning feature. The authors applied the circular Hough transform (CHT) to detect wide open mouth edge for yawning extraction. The experiment results showed that their system with SVM classifier could achieve a great result (98%) for driving fatigue detection. Ji et al. [16] proposed a real-time driving fatigue monitoring system based on the simultaneous use of various visual cues including the characteristics of the eyelid, gaze, head, and facial expression. A Bayesian network was developed to detect the fatigue state using the multiple visual cues obtained. The experiment result stated that their system could generate a more robust, reliable, and accurate driving fatigue detection system than using a single visual feature. Yang et al. [17] proposed a real-time system to monitor a driver’s nodding movements using RFID sensors. In their experiment, each driver was required to wear a hat where two RFID tags were attached to the back of it. The nodding movements were measured by computing the phase difference between the two RFID readings. In this study, the authors used a long short-term memory (LSTM) autoencoder to learn the nodding features for detecting driving fatigue. The experiment result showed that their proposed system could generate a highly accurate performance in both the lab environment and real-world driving environment.

In most cases, such methods are capable of recognizing fatigue facial features. However, they may easily fail to detect fatigue driving because of the surrounding context such as sunlight or darkness, and wearable devices [18]. Zhang et al. [19] reported that glasses disturbed the detection of the eye state. Moreover, the shape of the eyes in the camera changed significantly during head rotation [20]. Therefore, it becomes difficult to recognize the state of the eyes for any fatigue detection. The current facial-feature based fatigue
driving methods lack precise gaze-estimation algorithms to detect head orientation aligning with eye movement. Also, they fail to differentiate the closed eyes state caused by fatigue or vigorous laughter. Another major weakness of the existing methods is their attempt to recognize expressions from high-resolution facial images that need to be generated from a controlled environment [21]. However, in the real scenario, the surveillance images are often of low resolutions, making it more difficult to recognize expressions in this setting.

C. Physiological based Methods

Physiological based methods are widely used for driving fatigue detection. They are built to effectively evaluate the fatigue symptoms of drivers based on the bio-signals, including electroencephalogram (EEG), electrooculography (EOG), electrocardiogram (ECG), and surface electromyography (sEMG) collected by real-time portable sensors.

Wang et al. [22] argued that the psychological based method was an effective approach to detect driving fatigue. In their study, they developed a real-time driving fatigue detection system based on dry EEG signals. The EEG signals were collected from an EEG headset which each driver was required to wear all the time while driving. Two methods were proposed to detect mental fatigue including power spectrum density (PSD) and sample entropy (SE). Their experiment results indicated that there was an increasing trend observed in the PSD patterns while there was a decreasing trend observed in the sample entropy patterns when the driving fatigue occurred. Ahn et al. [23] stated that, besides EEG, ECG and EOG were another two effective approaches to detect driving fatigue. In their research, they proposed a multiview (EEG/ECG/EOG) based algorithm to detect drivers’ fatigue states. In order to collect data, all the sensors were attached to the drivers’ bodies. Their experiment result showed that their proposed system could generate good discriminative results to determine the fatigue state of each driver. Although physiological based methods have exhibited optimum performance in FDD, it is intrusive in nature [24] and drivers may be simply not willing to drive with many sensors attached to different parts of their bodies. Another main issue with physiological based methods is the difficulty of collecting the bio-signals in a non-contact way (usually electrodes and wires need to be in direct contact with drivers). The drivers are inevitably disturbed by the measuring electrodes. In order to address this issue, in [9], sEMG and ECG are collected from the sensors that were deployed in a cushion on a driver’s seat. Similarly, a European project called HARKEN developed a sensor system built into a safety belt and seat cover of cars, which is able to detect fatigue driving behaviour [25]. The proposed method of reading signals in [9] requires a driver to wear cloth of thickness less than 2 mm. Besides, there are also methods to use sEMG sensors to analyze muscle signals and detect fatigue driving [26]. However, they require sEMG sensors to be directly attached to the human hands, skin, or neck, which may not be practical [24]. In contrast to the existing approaches, we apply sEMG sensors on the steering-wheel to avoid attaching sensors to different parts of the human body.

In summary, there are no practical and reliable solutions for fatigue driving detection. Therefore, we propose the idea of using a sEMG sensor on the steering wheel in order to collect physiological features from bio-sensors for FDD. Our work aims to provide a practical and stable FDD compared to the existing methods. In particular, we use objective quantification of fatigue degree, mainly on the basis of features extracted from the sEMG signals.

III. MOTIVATION AND PROBLEM STATEMENT

A. Motivation

Most of the sEMG-based fatigue driving detection systems are intrusive, and only a few studies are based on non-intrusive detection systems but they usually have specific requirements, such as the thickness of clothes (less than 2 mm) in [9]. In order to detect fatigue driving using non-intrusive and sEMG based detection systems with minimal special requirements, a new method needs to be proposed. Considering the exposed skin on palms and fingertips are the most frequent contact part of the human body with the steering wheel, detecting fatigue driving through the sEMG sensors attached to a steering wheel could be a new feasible way. A similar design could be found in gym and health clubs, where sensors are attached to the handles of fitness equipment. For example, pulse heart rate sensors are attached to the handgrip of treadmills. The limitations of this kind of design are that: (1) the valid sensor reading requires users to keep their hands very still. However, it is not an easy thing to do if a user is running especially at peak intensity; (2) handlebar fixed with sensors is awkward for running posture and is discouraged. People could not run naturally if they have their arms placed in front of them and not swinging arms naturally by their side.

B. Problem Statement

For the steering wheel based sEMG sensor, though it is a lot more intuitive and easier to use and deploy, we have greater problems comparing with the existing deployment (e.g., handlebar in fitness equipment):

1) **Problem 1 (P1):** When drivers change a lane, or make a turn, it is difficult to ask them to hold the steering wheel still or constant contact with the steering wheel. Such resultant hand movement causes sEMG signals distortion or even loss.

2) **Problem 2 (P2):** Since different drivers may have different driving habits, we cannot fix the position of a sEMG sensor on a steering wheel.

These problems derive some particular research challenges of fatigue driving detection as follows:

1) **Challenge 1 (C1):** sEMG signals distortion or loss caused by P1 increases the difficulty of valid sEMG signals acquisition from the raw signal. On the one hand, distorted signals or lost signals (represented as a flat reading) are noise signals which must be removed from the raw signals. On the other hand, as the position of a sEMG sensor is not fixed on a steering wheel (P2), different drivers may have different holding postures,
resulting in various signal patterns of the valid sEMG signals. Figure 1 presents an example regarding how the signal patterns could be various among four participants with different holding postures. Therefore, how to differentiate between the valid sEMG signals and the noise signals, especially the distorted signals, is one of the key challenges in this study.

2) **Challenge 2 (C2):** In order to obtain distinctive fatigue states from the sEMG signals, the interval of the detection points for such fatigue state (Fatigue Detecting Point: FDP) cannot differ greatly. For example, if we set the first FDP at 15:00 and the second FDP at 15:05, it is clearly not appropriate if we set the third FDP at 15:06, but 15:10 will be reasonable. However, sEMG signals distortion or loss caused by F1 may disturb the balance of these intervals. Due to the complex road conditions and unpredictable hand movements, the interval between the occurrence of a valid sEMG signal in \( t_n \) and \( t_{n+1} \) can be significantly different with the one in \( t_{n+1} \) and \( t_{n+2} \) ("1" is a specific point of time for detection). A specific example could be found in Figure 2. There are three consecutive 5-minute window frames in the figure, where each index is 10 seconds data. It means "each window frame has 30 items and each item contains 10 seconds data, which means each window framework is 5 minutes long. We assign a unique index for each item starting from 1 to 30." Only the valid sEMG sample indices are illustrated in the figure (for example, for the first 5-minute window frame, the signal data located in the indices of sub-window "1", "5", "7" are recognized as the valid sEMG samples). Now, if we set the first index of each window frame as the detection point. It is clear to see that the distance of detection points between the second and third window frame is only 2 (i.e., the third detection point only has 20 seconds interval to the second detection point), which is too close to obtain useful information. Therefore, how to choose the distinctive FDP to make the intervals more reasonable is another key challenge in this study.

IV. **Proposed Method**

In this section, we explain the hardware and the system we design in our proposed method. The overall framework of our proposed method is illustrated in Figure 3, which consists of two parts: hardware design and system design.

A. **Hardware Design**

In order to address the problem \( \mathbf{P2} \) in Section III and satisfy different driving habits, the sEMG sensor is designed with a detachable style by using a flexible printed circuit (FPC) board (10 cm × 5 cm). It easily fits the shape of the steering wheel. Therefore, drivers can adjust the position of the sensor based on their own habit as shown in Figure 4. The signal acquisition electrodes on the sEMG sensor include four strips of copper in each FPC. Two strips are used as positive and negative electrodes to collect the signals from sEMG in a non-contact way. Such a design allows getting the sEMG signal using only one FPC (i.e., one hand). Another strip is ECG electrode, which is designed for our future research of leveraging sEMG with ECG signals for fatigue detection. All the strips of copper are vertically arranged in each FPC so that the palms and fingertips can maximally contact with the signal acquisition electrodes.

The data collection board (right side of Figure 4) contains two mini USB ports and a micro USB port. The former is connected to two sEMG sensors for left and right hands and the latter is connected to the computer for transmitting the electric signals. The data collection board is able to collect the sEMG sensor data in real-time.

As compared to the state-of-the-art in detecting driving fatigue, our solution would not disturb the driver’s attention, while traditional approaches either require sEMG sensors to be attached to the human body [24] or to be placed on the cushions of the seats with a thickness requirement on clothing [9]. Therefore, our tailored sensor and device are convenient, practical and suitable for detecting driving fatigue in real scenarios.

B. **System Design**

Our study consists of three main components: (1) signal processing and noise filtering, (2) dynamic fatigue detection point selection, and (3) feature generation. In the first component, we pass the raw sEMG signals to filter the noise using band-pass filters. Even so, there are still some noises left after that (e.g., mechanical noises caused by the big movement of the steering wheel). We thus segment the filtered signals into 5-minute sliding windows and further divide each sliding window into much smaller sub-sliding windows with 50% overlapping. They are then fed into the "valid sEMG sample selection machine" to identify valid sEMG samples. Afterwards, fatigue detection points are selected using our dynamic fatigue detection point selection method. Finally, the valid sEMG samples in the fatigue detection points become the input of the feature generation component, where 28 features will be generated based on our two-layer feature generation method. In the following, we explain each component in detail.

1) **Signal Processing and Filtering:** sEMG signals have a low signal-to-noise ratio and are sensitive to noises. To remove these noises, we firstly use a fourth-order band-pass filter to de-noise the sEMG signals with a lower cutoff frequency of 10 Hz and a higher cutoff frequency of 300 Hz. These two parameters are designed based on the frequency range of sEMG (10 Hz-300 Hz) [27] [28] [29] [9].

Figure 5 presents the pattern of sEMG signals before and after filtering. Clearly, the signals are still noisy. These noises are generated due to the challenge C1. In order to further remove noises and obtain valid sEMG signals, we design a “Valid sEMG Selection Machine (VsESM)”, which is a semi-supervised learning based on PU Learning algorithm [30].

In the first stage, we prepare two sample sets: labelled valid sEMG sample set and unlabelled sample set. All samples are collected based on a two-layer sliding window technique. Generally speaking, driving fatigue is gradually developed as an accumulation process over a period of time, and there is hence no significant change in a relatively short period.
However, if the sliding window is set too large, the fatigue detection system may fail to respond timely. In this study, we record the drivers’ driving conditions every five minutes and thus set a 5-minute interval as the window size. A 5-minute window size may be too long to extract the clean feature under certain scenarios, due to a significant amount of noises inside sEMG signals deployed on a steering wheel, e.g., mechanical noise, signal shift, electricity interference [28] and no sEMG signals at all when drivers’ hands are off the steering wheel (e.g., a sharp turn of the steering wheel or turning in a round-about). Therefore, a dual-layer sliding window with a much smaller sub-window with 50% overlapping is adopted in our approach. The total number of points in one sample is sub-window size * sample frequency data. For example, for a 10 second sub-window, the amount of points in this sub-window is: 10 second sub-window size * 1000Hz = 10,000 points.

And then, both labelled valid sEMG samples and unlabelled samples are used to select valid sEMG samples. We believe that the labelled valid sEMG samples are different to each other, and they should not be simply classified into one cluster. The reason is that the different drivers may have different holding postures, making that the valid sEMG signals are really diversified. Therefore, we first try to separate them into different clusters, so that the samples in each cluster are similar to each other. For unlabelled samples, we try to filter both potential valid sEMG samples and noise samples from them according to the isolation score and their similarity score to the labelled valid sEMG samples. The intuition is that, on one hand, the potential valid sEMG signals should be different to noise samples (i.e., can be easily isolated); on the other hand, they should be similar to some labelled valid sEMG signals. In the second stage, we build a weighted multi-class model (in this study, we use “XGBoost”† as the weighted multi-class model) to distinguish different valid sEMG samples from the noise samples. For the labelled valid sEMG samples, the weights are set to 1, and for the filtered samples, the weights are set according to the confidence of their attached labels (i.e., label “0”: noise, label “1”: valid sEMG samples).

2) Dynamic Fatigue Detection Point Selection: In order to address the challenge C2, we propose a dynamic fatigue detection point selection method to make the detection intervals reasonable. Since each 5-minute window frame may contain many valid sEMG samples and the indices of these samples in the current 5-minute window frame can be significantly different from the ones in its neighbour frame, we use the shortest absolute distance of index of the valid sEMG samples between two consecutive 5-minute window frames to select the fatigue detection point.

\[
\text{absolute distance} = |\text{value}_{n+1} - \text{value}_n| \tag{1}
\]

An example is shown in Figure 7. There are five consecutive 5-minute window frames in the figure, where each frame includes several valid sEMG samples (for example, for the first 5-minute window frame, the data located in the indices of sub-window 2, 4, 13, 20 are the valid sEMG samples). Now, if we set the index “2” in the first 5-min window frame

†https://xgboost.ai/
as the first fatigue detection point, we can find index “3” in the second 5-minute window frame has the shortest absolute distance with the index “2” in the first 5-min window frame. Therefore, we set the index “3” as the fatigue detection point for the second 5-minute window frame, by doing so, the most representative feature could be obtained from the valid sEMG sample of index “3” in this 5-minutes window. We can use the same way to get the rest detection points, i.e., index “10” in the third 5-min window frames, index “30” in the fourth 5-min window frames, index “20” in the fifth 5-min window frames, respectively.

3) Feature Generation: Initially, 14 features are generated from the detection point obtained from the previous section, including the 12 most commonly used features in previous studies: mean, standard deviation (std), median, maximum (max), minimum (min), the difference between maximum and minimum (max_min), the signal magnitude area (sma), skewness, kurtosis, zero crossing (time_over_zero), mean frequency, median frequency; and two particular features: sample entropy and Lempel-Ziv complexity, which have been used by our most related study [9].

In [9], a smooth decreasing tendency on the change of fatigue state is presented using the cushion solution. In order to evaluate whether our solution has a similar tendency with theirs, we draw the tendency chart for all the 14 features and find all of them present different tendency (ups and downs trend) compared with the physiological features tendency illustrated in [9]. One explanation for this difference is that when using the steering wheel based fatigue driving detection method, drivers tend to make small adjustments of hand holding posture when they feel fatigued. It is intuitive as a body’s natural reaction to fatigue. Since a human body movement increases the amplitude of the sEMG signals [31], for this natural fatigue recovery behavior, we can find the corresponding unique sEMG signal patterns, a wave trend (down and up). However, the frequency of adjustments on body sitting posture is much less than the frequency of posture adjustments on hand holding posture, and therefore the feature tendency in [9] is much smoother than ours. As a result, we cannot directly use the features introduced in [9].

In our case, we define a term fatigue state transition (FST), that is, transiting from not feeling fatigued to start feeling fatigued or feeling more fatigued. FST helps us to analyze the drivers’ fatigue behaviours. In order to capture those characteristics of sEMG signals corresponding to FST,
we design the second layer sub-features from these wave trends. We first calculate the slope of every two adjacent data points for a given feature to obtain such “wave” like patterns. However, in some real-life cases, the rising part of the “wave” form (the increase of sEMG feature value) could be caused by changing lane, when a driver needs to turn the steering wheel, or some unconscious little finger movements (e.g., some drivers like to rub steering wheel using fingers unintentionally). It is shown as the red circle of point 1 in Figure 8. To differentiate those from the actual FST-related “wave” form (points 2 and 3), we use the absolute distance between every two adjacent data points, because the bigger the distance, the more likely the FST-related “wave” form is. That is, slope and absolute distance, on top of those four features. Finally, 28 features are generated in our study (i.e., 14 initial features × 2 slope and absolute distance).

V. EMPIRICAL SETTING

In order to evaluate the effectiveness and practicality of our sensors and proposed methodology, we conducted two extensive experiments (E1 and E2) in different settings. Both experiments were done with the sEMG sensors installed on a driving steering wheel and the signals were collected with our data acquisition board. The sensor data were stored in a laptop for driver fatigue detection.

E1 is conducted in a lab environment, as shown in Figure 9b. The ambient temperature was set to 26 °C. We recruited 13 experienced drivers (7 males and 6 females) with an average of 5 years’ driving experience. All participants were healthy adults of 20-30 years of age, who did not feel any fatigue when they started driving. (1) For training sEMG sample set collection, each participant continuously drove for 1.5 hours. The experiments were done in four different time slots from 9 am to 8:30 pm. More specifically, three participants were asked to perform the experiments from 9 am to 10:30 am, three participants from 1 pm to 2:30 pm, another three from 4 pm to 5:30 pm, and finally four participants from 7 pm to 8:30 pm. All the participants were asked to drive freely without any extra constraints on speed and steering movement. (2) For sEMG validation sample set collection,
each participant’s driving was done in a more controlled way than in the collection of training sEMG sample set. They were asked to hold the steering wheel tightly and maintained a speed between 60-80 km/hr. The controlled driving involved fewer sudden movements of the steering wheel. In addition, participants had a tighter contact with the sensors while driving. All the other settings are identical to the training sEMG sample set collection.

We used the car racing game “Need for Speed: Hot Pursuit-free driving model” [32] as the simulation software, and collected the signal data with a sampling frequency of 1000 Hz. In the free driving model, the participants drove freely in a big city map, even though they need to pay more attention to the driving due to the cars they drove are the high-performing sport cars. This setting makes the participants much easier to get fatigue.

For E2, as displayed in Figure 10, we conducted four on-road tests in a major city in Australia. The cars used were a Mercedes GLA200 2016 Model and an Audi Q7 2011 model with one sEMG sensor attached to the left-hand side of their steering wheel, respectively. The reason for not deploying the sensor to the right-hand side is to avoid dangling the cable connecting our board with the sensor, which may generate more noises. Four male drivers with an average of 10 years driving experience were asked to drive in four road tests each. All participants were healthy adults of 30-40 years of age. In all real-world tests, the drivers drove freely and reported their fatigue level every 5 minutes. The first two on-road tests were conducted in the morning from 9 am to 11 am, as shown in Figure 10a. A total of 63.9 km driving was done in about 100 minutes. The second two were conducted at night from 6 pm to 8 pm, as shown in Figure 10b. A total of 54.1 km driving was done in about 95 minutes.

In this study, the results of subjective questionnaire “Swedish Occupational Fatigue Inventory” (SOFI) [9] were used as labels for defining the fatigue state. SOFI intends to measure work-related fatigue by adopting a multidimensional perspective questions, where each question will be ranked from 0 to 10. We asked these questions to participants every 5 minutes and calculated their scores (the higher the score they obtain, the more fatigue the participants they feel). In this study, the classification of NFST and FST is from the SOFI questionnaire. According to the result of the SOFI questionnaire in literature [33], [34], we defined score between 0 and 5 as the NFST and the score between 5 and 10 as FST. Moreover, we recorded participants facial features using the GoPro Hero7 4K Action Cam [35], which features high video stabilisation, and used a computer vision based fatigue detection system [8] to get the participants’ fatigue state as a benchmark to improve the labels reliability.

VI. RESULTS AND DISCUSSION

A. E1 Result

1) Accuracy Analysis with Fixed Sub-window Size: Machine learning algorithms have been widely used in the vehicle-based tasks [36], [37]. In our experiment E1, we applied the following 5 different machine learning algorithms: Support Vector Machine (SVM), Random Forest (RF), Logistic Regression (LR), K-nearest neighbor (KNN) and Naive Bayes (NB). The performance comparison of these models is given in Table I, where a 30 second sub-window size is used for feature generation, as the same configurations with the study [9], and the hyper-parameters of models were tuned using the grid search approach. Table I indicates that our random forest based detection model is capable of distinguishing between NFST and FST, with the weighted average F1 score of 96%. AUC curve is used to evaluate our fatigue detection model. Figure 11 shows the AUC curve of the RF for E1, where AUC score of 0.95 indicates that our RF-based detection model is capable of distinguishing between NFST and FST. Table II shows the Precision and Recall score besides the average F1 score, where the Precision score of 0.97 for NFST and 1.00 for FST, while the Recall score of 0.99 for NFST and 0.94 for FST, respectively. The confusion matrix for E1 are given in Table III, in which 32 FSTs are correctly identified
and 6 FST is not. The reason behind false positives and false negatives is mainly due to noises during the experiment. The participants in E1 were asked to drive freely without any speed and steering movement limitations. With no speed limitation, participants tended to drive fast. In the simulation game, it was very difficult to control the vehicle in high speed, so a large number of big wheel adjustments were performed to control the vehicle. Such significant movements of steering wheel produced significant amount of noises in data, thereby causing misclassifications. Also, with no steering wheel movement limitations, participants did not hold the steering wheel very tight. The hands often moved away from the deployed sEMG sensors, producing high motion artefacts that are part of the transient baseline change caused by the electrode motions due to a subject’s movement. These motion artefacts created more noises for the underlying sEMG sensor data, thus leading to some false negatives. The experimental results of experiment E1 showed promising potential of identifying fatigue from our proposed solution. We intended to check further if the accuracy can be affected by the different sub-window size.

2) Accuracy Analysis with Various Sub-window Size: We also analyzed the general effects of the windowing operation on the fatigue driving detection process. The performance results for diverse window sizes and each specific methodology are depicted in Figure 12. The RF, KNN and SVC models show a trend of first increase and then decrease performance as the size of the window grows. Except the NB, the maximum performances are all obtained in the sub-window size of 30s. The performance of LR increases monotonically throughout the change of diverse window sizes, while the performance of NB firstly increases (from 10s to 40s) and then remains unchanged (from 40s to 60s). Both of the performance on LR and NB are not significantly influenced by the change of the sub-window size, and therefore, we designed a “cut-off” window size at 30s for feature extraction. Based on the experiment result, we observed that the RF model stands out among all evaluated models, which provides the highest performance, with an F1-score of 96%. However, increasing the window size to more than 30s drags down the detection performance more rapidly comparing with the other models. The reason is mainly due to the imbalanced samples between NFST and FST. Imbalanced training samples produce an important deterioration of the classification accuracy, in particular with the performances belonging to the less represented classes [38]. In other words, the change of the sub-window size has a more significant impact (sensitivity) on the detection of FST (small class) than NFST (large class). For the models excluding the RF, they have a high accuracy on NFST but low
accuracy on FST. Therefore, even changing the sub-window size is sensitive to the performance of FST, but because there is not much space to drop the accuracy of FST, the weighted average F1 score does not present a steeply decrease. Conversely, as the tree-based machine learning models work by learning a hierarchy of “if/else” questions and this can force both classes to be addressed, tree-based models usually perform well on imbalanced data. RF, as a tree-based model, has a high accuracy on both FST and NFST, and the change of sub-window size can easily drop the accuracy of FST a lot (from 90% to 36%), and therefore decrease the weighted average F1 score significantly (Figure 13).

3) Feature Analysis: We are also interested in what are the important features for the fatigue driving detection and how these features impact the detection performance.

Feature Importance. In this study, since we used the XGBoost to select the valid sEMG samples and Random Forest to detect the fatigue state transition (FST), in order to avoid the bias, we used another tree based model “lightGBM” to determine the feature importance. The importance of the features we generated is illustrated in Figure 14, by which we can obtain the highly important features and lesser important features for FST detection.

Slop of Empel Ziv Complexity, the slop of median, slop of Skew, slop of kurtosis and absolute distance of sma are the top 5 features which have significant contributions to detect FST, while slop of max, slop of median frequency, slop of max_min, absolute distance of min, absolute distance of std and slop of std are less help to the FST detection. In Figure 15a, the cumulative importance versus the number of features is illustrated.

The vertical line is drawn at threshold of the cumulative importance, in this case is 90%, which indicates that 23 features contribute 90% effort for FST detection. This result give a complement explanation to the fact that low important features do not contribute a lot in the fatigue state detection.

Feature Correlation. In order to understand the reason why these features are low important to the FST, we conducted the correlation analysis using a heatmap. We find most of the lesser important features (excluding slop of median frequency) are collinear features which are highly correlated with one another, as shown in Figure 15b. In machine learning, this lead to decreased generalization performance due to high variance and less model interpretability. In a tree-based model, such as“lightGBM”, the collinear features are seldom or not used to split any nodes, and therefore these features are categorised as lesser important features eventually.

In order to further investigate the effect of the lesser important features, we removes those lesser important features and used the rest features to evaluate the fatigue state detection performance again. This time, only RF model was used due to its insensitivity on imbalanced samples. The weighted average F1 score and confusion matrix after removing the lesser important features under sub-window size of 30s are illustrated in Table IV and Table V. The benefits of lesser important feature removal are reflected by an increase of F1 score from 96% to 98%, and a decrease on false positive (from 1 to 0) and false negative (from 6 to 3).

B. E2 Result

The data collected from real driving were used to test against the trained classification model in E1. The weighted average F1 scores of the four drivers are given in Table VI and
Fig. 14. The importance of the features generated in our system

TABLE IV. F1 score of RF after removing the row important features (sub-window size: 30s)

<table>
<thead>
<tr>
<th>Model</th>
<th>F1 score</th>
<th>Parameters</th>
</tr>
</thead>
<tbody>
<tr>
<td>RF</td>
<td>0.98</td>
<td>bootstrap=True, class_weight=None, criterion=entropy, max_depth=20, max_features=auto, max_leaf_nodes=None, min_impurity_decrease=0.0, min_impurity_split=None, min_samples_leaf=1, min_samples_split=2, min_weight_fraction_leaf=0.0, n_estimators=15, n_jobs=None, oob_score=False, random_state=None, verbose=0, warm_start=False</td>
</tr>
</tbody>
</table>

Figure 16. From Table VI, we can see, except the third driver’s performance, the weighted average F1 scores generated from the feature set which is removed the lesser important features (fs1) are superior to the weighted average F1 scores generated from the whole feature set (fs2). More specifically, using fs1, the weighted average F1 scores for the first to third driver are all above 90%, achieving 92%, 91% and 91%, respectively. For the third driver, although the weighted average F1 score of fs1 (91%) is lower than the weighted average F1 score of fs2 (100%), the result is still acceptable. Moreover, Figure 16 illustrates the NFST and FST detection results of RF using the feature set which is removed the lesser important features, where driver 1 has 10 NFST ground truth and 3 FST ground truth, driver 2 has 11 NFST ground truth and 2 FST ground truth, driver 3 has 11 NFST ground truth and 2 FST ground truth, and driver 4 has 12 NFST ground truth and 1 FST ground truth, respectively. From Figure 16, we can see the F1 score regarding the detection of NFSTs and FSTs are 95% and 80% for the first driver, 96% and 67% for the second driver, 96% and 67% for the third driver, and 92% and 0% for the fourth driver, respectively. The result indicates that, except the fourth driver, most of the NFST and FST points can be detected using our solution.

The weighted average F1 scores generated from the both feature sets for the fourth driver are worse than those for the other three drivers. The reason is mainly due to the additional/unusual driving behaviors of the fourth driver. During the experiment, we noted the first to third drivers drove safely and turned smoothly. However, the fourth driver has been a...
professional Uber driver for more than 2 years. He was used to driving for long hours and did not get tired in the first 100 minutes of driving. This observation also reflected on the camera-based fatigue detection method, by which there was no highly confident result indicating he was under a fatigue state. During his driving, he only reported “he felt a little bit fatigue” at the end of the experiment. We labelled his report as FST, but the sEMG samples we collected at that FDP may not represent a real fatigue driving state corresponding to the FST. Furthermore, the fourth driver also had some unconscious habits to rub fingers with the driving wheel especially the sensors and sometimes holding the bottom part of the steering wheel (the sEMG sensors were deployed on the upper part). All of these unusual driving behaviors of the fourth driver caused noises that led to the fairly low F1 score of the fourth driver in Figure 16. Overall, this real-world driving experiment shows encouraging results and useful insights to adopt our approach for real-life usage.

TABLE VI. Overall F1 score of RF for E2

<table>
<thead>
<tr>
<th>Driver</th>
<th>NFST (feature set with lesser important features removed)</th>
<th>FST (whole feature set)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Driver 1</td>
<td>0.92</td>
<td>0.78</td>
</tr>
<tr>
<td>Driver 2</td>
<td>0.91</td>
<td>0.85</td>
</tr>
<tr>
<td>Driver 3</td>
<td>0.91</td>
<td>1.00</td>
</tr>
<tr>
<td>Driver 4</td>
<td>0.78</td>
<td>0.09</td>
</tr>
</tbody>
</table>

C. Discussion and Insights

**Result Validity.** As per our study, the “wave” forms from sEMG signals are observed whenever participants feel fatigue and intuitively change hand and body postures. The rise and drop of the corresponding sEMG feature values thus are crucial for us to detect FST. Yet, such forms may also be caused by hand movement, diverse noises or other confounding activities similar to the FST-correlated one. However, from our extensive studies in the two experiments, our approach is able to filter out the majority of these false “wave” forms. Since there are much more NFST samples than FST samples in our experiments, we decided to evaluate both NFST and FST instead of just FST and introduce weighted average F1 to show more balanced evaluation results.

Our approach primarily uses “wave-form” to identify FST. However, it is still challenging to identify the correlation among FSTs. In our experimental settings, participants self-reported their fatigue level while driving. But in real life, drivers may take a break in the middle of driving or some energy drink in between, so that the FST detected afterwards might not be necessarily more severe than those detected before.

In E2, when drivers feel very tired after driving long hours, they may be too exhausted or even fall asleep without any reactive hand and body postures (no rise part of the “wave” form). In such cases, our approach may not be able to detect FST. However, before drivers feel such a level of fatigue, they are very likely to have already exhibited multiple “wave” forms for FST, which should have been successfully captured by our system. Thus, our system could effectively provide feedback in the real-life scenario and prevent drivers from entering into such an exhaustion status.

**Insights.** The use of two sEMG sensors deployed on the steering wheel may not be sufficient for all drivers. For example, the fourth driver in E2 has a personal habit of using bottom part of the steering wheel occasionally that impacted sensor data collection, hence, four sEMG sensors may need to be installed quarterly on the wheel. Also, plenty of noises are caused by finger rubbing of the sEMG sensors and the big movement of the wire that connects the data collection board and the sensors. Therefore, it is desirable to design a better hardware for noise-resilient sEMG sensors with a wireless connection to the board. A graphene-based sEMG sensor design solution is a potential research direction due to its mechanical flexibility and ultra-thinness high signal-to-noise ratio (SNR), and efficient signal transmission with the high electrical conductivity [39].

Noises present in the obtained signals often compromised the accuracy of our proposed solution. It is challenging to remove such background noises entirely. This actually points out a promising research direction to use complementary sensors. That is, sEMG sensors can be leveraged with night-vision cameras and other bio-signals sensors, such as ECG devices, to minimize the noise impacts. In this direction, multimodal Deep Learning can be investigated further due to its ability to deal with the differences among complementary yet heterogeneous sensors with varying sampling rates, data types, and data format (discrete and continuous data) [40].

VII. Conclusion

In this paper, we present a novel approach to detect driving fatigue by deploying detachable sEMG sensors on steering wheels, which collects bio-signal in a non-intrusive way compared with other existing solutions. To be more specific, we design and build the detachable styled sEMG sensors and data collection board to collect the bio-signal, and then develop an innovative solution VsEM to extract valid sEMG bio-signal from underlying diverse noises, use multi-layer features to identify driver fatigue state transitions. Moreover, using
comprehensive experiments involving 17 participants on the simulated driving platform and on-road tests, we confirm and verify that our approach is able to detect driver fatigue state transition with an acceptable accuracy. We also found that the deployment of the sensors (e.g. the places where the sensors are detached on the steering wheel) and the design of the sensors (e.g. whether the sensors are connected with the data collection broad wirelessly) can affect the quality of the collected bio-signal. Based on our experimental results, we raise a few important insights to push forward the state-of-the-art in detecting driving fatigue. Our future work will focus on further improving the accuracy of our approach in the real-life scenarios. We are currently investigating the use of multi-view learning [41] to leverage our sEMG sensors with complimentary sensors (e.g., night vision camera and ECG sensors).

VIII. ACKNOWLEDGEMENT
This research was supported by Data61 Collaborative Research Project C020996 and Australian Research Council (ARC) LIEF Project LE180100158.

REFERENCES


Jianchao Lu received the Master of Research with Distinction from Macquarie University, Master of Professional Accounting and Master of Commerce from Deakin University. Now he is a PhD candidate at Macquarie University. Specialized in Human Behaviour Detection in Autonomous Driving and published a paper in IEEE Engineering in Medicine and Biology Society (EMBS), Australasian Computer Science Week (ACSWeek), IEEE International Conference on Web Services (ICWS), International Conference on Mobile and Ubiquitous Systems: Computing, Networking and Services (MobiQuitous).

Xi Zheng received the PhD in Software Engineering from UT Austin, now Director of Intelligent systems research center (isteg.org) at Macquarie University. Specialised in Service Computing, IoT Security and Reliability Analysis. Published more than 80 high quality publications in top journals and conferences (PerCOM, ICSE, IEEE Communications Surveys and Tutorials, IEEE Transactions on Cybernetics, IEEE Transactions on Industrial Informatics, IEEE Transactions on Vehicular Technology, IEEE IoT journal, ACM Transactions on Embedded Computing Systems).

Lihong Tang currently is a PhD student at Department of Computer Science and Software Engineering, Swinburne University of Technology. She received her Bachelor degree with Honours from Deakin University, Australia. She is working on Android malware detection and evolution analysis, and her research interests are android malware analysis, adversarial attacks under the Android context, human-centric research, and explainable machine learning.

Tianyi Zhang is a Postdoctoral Fellow in Computer Science at Harvard University. He obtained his Ph.D. in Computer Science from UCLA and his bachelor's degree in Computer Science from Huazhong University of Science and Technology. His research interests reside primarily in Software Engineering and Human-Computer Interaction, with a specialization in mining software repositories, interactive system design, and software evolution.

Michael (Quan Z.) Sheng is a full Professor and Head of Department of Computing at Macquarie University, Sydney, Australia. His research interests include Internet of Things, service oriented computing, distributed computing, Internet computing, and pervasive computing. Michael holds a PhD degree in computer science from the University of New South Wales (UNSW) and did his post-doc as a research scientist at CSIRO ICT Centre. Prof Michael Sheng is the recipient of AMiner Most Influential Scholar in IoT Award (2007-2017), ARC Future Fellowship in 2014, Chris Wallace Award for Outstanding Research Contribution in 2012, and Microsoft Fellowship in 2003.

Chen Wang received his B.S., M.Eng. and Ph.D. in Computer Science from Nanjing University, China. He is currently a Research Scientist in the Software and Computational Systems Program (SCS), Data61 at CSIRO, which he joined in 2008. Prior to that, he was a research fellow with the Centre for Distributed and High Performance Computing / School of Information Technology at The University of Sydney, a software engineer with Amicas Inc. in the US and a Postdoc with National University of Singapore. His current research interests are on the interpretability, robustness and scalability of data analytics systems.

Jiong Jin received the B.E. degree with First Class Honours in Computer Engineering from Nanyang Technological University, Singapore, in 2006, and the Ph.D. degree in Electrical and Electronic Engineering from the University of Melbourne, Australia, in 2011. From 2011 to 2013, he was a Research Fellow in the Department of Electrical and Electronic Engineering at the University of Melbourne. He is currently a Senior Lecturer at Swinburne University of Technology, Australia. His research interests include network design and optimization, edge computing and distributed systems, robotics and automation, and cyber-physical systems and Internet of Things as well as their applications in smart manufacturing, smart transportation and smart cities.

Shui Yu is a Professor of School of Computer Science, University of Technology Sydney, Australia. Dr. Yu’s research interest includes Big Data, Security and Privacy, Networking, and Mathematical Modelling. He has published three monographs and edited two books, more than 350 technical papers, including top journals and top conferences, such as IEEE TPDS, TC, TIFS, TMC, TKDE, TETC, ToN, and INFOCOM. He is currently serving a number of prestigious editorial boards, including IEEE Communications Surveys and Tutorials (Area Editor), IEEE Communications Magazine, and IEEE Internet of Things Journal. He is a Senior Member of IEEE, a member of AAAS and ACM, and a Distinguished Lecturer of IEEE Communication Society.

Wanlei Zhou has been appointed Vice Rector of the City University of Macau on 12 October 2020. Professor Zhou received a PhD degree in Computer Science from The Australian National University in 1991 and a DSc degree (a higher Doctorate degree) from Deakin University in 2002. Professor Zhou specialized in research on network security and privacy protection, computer network and online teaching. He has been in charge of and completed 13 ARC projects and supervised 30 PhD students in recent years. More than 400 academic papers have been published in a series of international journals and presented at international conferences. Professor Zhou has also participated in the edition of seven academic monographs and published four of his own. He was a moderator of more than 30 international computer conferences.