Learning multimodal representations for drowsiness detection

Kun Qian, Tomoya Koike, Toru Nakamura, Björn W. Schuller, Yoshiharu Yamamoto

Learning Multimodal Representations for Drowsiness Detection

Kun Qian, *Senior Member, IEEE*, Tomoya Koike, *Student Member, IEEE*, Toru Nakamura, *Member, IEEE*, Björn W. Schuller, *Fellow, IEEE*, and Yoshiharu Yamamoto, *Member, IEEE*

Abstract—Drowsiness detection is a crucial step for safe driving. A plethora of efforts has been invested on using pervasive sensor data (e.g., video, physiology) empowered by machine learning to build an automatic drowsiness detection system. Nevertheless, most of the existing methods are based on complicated wearables (e.g., electroencephalogram) or computer vision algorithms (e.g., eye state analysis), which makes the relevant systems hardly applicable in the wild. Furthermore, data based on these methods are insufficient in nature due to limited simulation experiments. In this light, we propose a novel and easily implemented method based on full non-invasive multimodal machine learning analysis for the driver drowsiness detection task. The drowsiness level was estimated by self-reported questionnaire in pre-designed protocols. First, we consider involving environmental data (e.g., temperature, humidity, illuminance, and further more), which can be regarded as complementary information for the human activity data recorded via accelerometers or actigraphs. Second, we demonstrate that the models trained by daily life data can still be efficient to make predictions for the subject performing in a simulator, which may benefit the future data collection methods. Finally, we make a comprehensive study on investigating different machine learning methods including classic 'shallow' models and recent deep models. Experimental results show that, our proposed methods can reach 64.6% unweighted average recall for drowsiness detection in a subject-independent scenario.

Index Terms—Sensors, vehicular technology, drowsiness detection, machine learning, deep learning.

(Kun Qian and Tomoya Koike contributed equally to this work.)
(Corresponding author: Kun Qian.)

Kun Qian is with the Group on Audition for Intelligent Medicine (AIM), Institute of Engineering Medicine, Beijing Institute of Technology, Haidian, Beijing 100081, China (e-mail: qian@bit.edu.cn).

Tomoya Koike and Yoshiharu Yamamoto are with the Educational Physiology Laboratory, Graduate School of Education, The University of Tokyo, Bunkyo-ku, Tokyo 113-0033, Japan (e-mail: tommy@p.u-tokyo.ac.jp; yamamoto@p.u-tokyo.ac.jp).

Toru Nakamura is with the Graduate School of Engineering Science, Osaka University, Suita 565-0871, Japan (e-mail: t-nakamura@sangaku.es.osaka-u.ac.jp).

Björn W. Schuller is with the Group on Language, Audio and Music (GLAM), Imperial College London, London SW7 2AZ, U.K., and also with the Chair of Embedded Intelligence for Health Care and Wellbeing, University of Augsburg, 86159 Augsburg, Germany (e-mail: schuller@ieee.org).

I. Introduction

ACCORDING to a recent report by the U.S. Department of Transportation, drowsy driving killed 697 lives in 2019 [1]. In the past two decades, drowsiness detection has increasingly attracted tremendous works due to its crucial role in a safe driving [2]–[6]. These previous studies can be mainly categorised into three directions: i) *computer vision* (CV) based methods; ii) physiological signals based methods; iii) a fusion of the aforementioned methods.

For the CV based methods, the advantage is their non-intrusive characteristic. Ji *et al.* proposed a real–time monitoring system for detection of driver fatigue [7]. In their study, multiple visual cues including *eyelid movement*, *gaze movement*, *head movement*, and *facial expression* were used to build a Bayesian network (BN) model [7]. Mbouna *et al.* extracted visual features such as *eye index*, *pupil activity*, and *head pose* for driver alertness monitoring by using a support vector machine (SVM) [8]. They indicated that, a combination of eye and head information can achieve the best performance [8]. An Occluded Stacked Hourglass method was proposed in [9], to render the algorithms for the analysis of head, eye, or mouth more robust in the case that any objects or sunlight occlude any related facial parts. Mandal *et al.* used the percentage of eyelid closure (PERCLOS) estimation method to predict the driver's state, i.e., *normal* or *fatigue* [10]. You *et al.* took the individual differences into account, which can reduce the effects by eye sizes of different drivers [11]. Besides the PERCLOS, *speech* and *touch response* were considered to be additional stages for detection of driver drowsiness in [12], which was demonstrated to be more efficient than only using a single stage. Yu *et al.* considered spatio-temporal representation learning from video by using a 3D deep convolutional neural network (3D-DCNN) [13]. Furthermore, more advanced convolutional neural network (CNN) based models were introduced for the driver drowsiness detection task, e.g., multiple CNNs with the usage of kernelised correlation filters (KCF) [14], and a multi-tasking CNN [15]. To overcome the (labelled) *data scarcity* in the training set, sparse representations and generative adversarial networks were used in [16] and [17], respectively. Even though the previous studies showed promising results, the drawbacks are obvious: First, CV based methods may be vulnerable to environmental interference, e.g., head movements, or illumination conditions. Second, the high computational costs of the CV methods, in particular, for CNN based models, may render it inapplicable for real

life driver safety systems which are mainly relying on low energy-consumption and cheap embedded devices.

As an alternative, physiological signals were widely used in driver drowsiness detection. Among most of these studies, electroencephalogram (EEG) and its related brain-computer interface (BCI) technologies were investigated. Lin *et al.* used a combination of independent component analysis (ICA), power-spectrum analysis, correlation evaluations, and linear regression model to estimate the driver's cognitive state when (s)he drives a car in a virtual reality (VR)-based dynamic simulator [18]. In addition, a further study on using a dual-core signal processing unit was proposed in [19], which can implement real-time processing, automatic analysis, and online warning feedback in a real-world operation and living environments. To make their systems more applicable in real-world driving situation; the authors designed a portable, wireless, and real-time processing EEG-BCI system for driver drowsiness detection [20]. A generalised EEG-based self-organising neural fuzzy system to monitor driver's drowsiness was presented in [21], which was demonstrated to be efficient in both subject-dependent and cross-subject scenarios. Further, Li *et al.* developed a smartwatch based wearable EEG device to predict the probability values of the driver drowsiness level instead of discrete class labels [22]. An adaptive Hermite decomposition for extracting more efficient EEG features was given in [23], which also benefited from the efficacy of an extreme learning machine (ELM) for its classification performance. To obtain high-quality EEG signals with convenience and long-term wearing comfort, a method using EEG from non-haring-bearing (NHB) scalp areas was proposed in [24]. Reddy *et al.* introduced a multi-task deep neural network (MTDDN) in a regression scheme for driver drowsiness estimation in [25]. Choi *et al.* made efforts towards instantaneous drowsiness (i. e., lapse or micro-sleep) detection using wired and wireless EEG [26]. Besides the EEG signals, some other physiological signals were used, like the electrocardiogram (ECG) [27], and the respiratory rate variability (RRV) [28]. The limitations of these physiological signal analysis-based methods are: First, the multiple sensors connected to humans may interfere with the driver's driving status (uncomfortable intrusive experience); second, the mostly complicated system architectures may make them inapplicable to be used in real–world scenarios.

The third research direction is to use a fusion of the two aforementioned kinds of signals, i. e., using both visual and physiological signals to detect a driver's drowsiness [29]–[31]. Results demonstrated that, fusion of different kinds of signals can improve the performance achieved by using a single kind of signal. In addition, Sunagawa *et al.* found that, introducing the posture information can contribute to the drowsiness detection, in particular, for the case of weak drowsiness [32]. Nevertheless, there are still some challenges needing to be addressed in the existing studies: First, subject-independency was ignored in many of the previous works, which may lead to over-optimistic results. Second, the often complicated system architectures (both for hardware and software) may make it difficult to be implemented in real driving applications. Third, some other environmental sensor

data were not considered, which may determine and/or help detect the driver's status. Last but not least, driver behaviour data in a driving situation is inconvenient to be collected, which may restrain a further large-scale data analysis. To overcome these challenges, we propose a novel data analysis and model building method in this work. In particular, compared to the method proposed in [32], our approach has several advantages: First, we do not use any sensitive data related to personal privacy, whereas a camera was still used in [32]. Second, the environmental condition was not sufficiently considered in their multimodal information [32]. In this study, we take the environmental data into account. Third, a novel finding and an assumption of this study is that, the human behaviour in daily life could be modelled and used for prediction of the driving drowsiness status. The assumption of this study is that, the human behaviour in daily life can be modelled via the statistical functionals extracted from the *spontaneous physical activity* (SPA) and be used for predicting the drivers' drowsiness status. In our previous studies [33], [34], we found that using wearable sensor data collected from daily life can be efficiently exploited to analyse the human behaviour via machine learning methods. For instance, the SPA data recorded via a watch-type device can be used to detect if the subject is suffering from a major depressive disorder (MDD) [33]. In that study, both 'classic' ML models and a recent DL model were demonstrated to be feasible to make predictions based on representations extracted from the human behaviour data. In addition, a bag-of-behaviour-words (BoBW) approach was introduced to analyse the SPA data in [34]. It showed that more robust representations can be extracted even from the raw sensor data in comparison to human expert designed features. Inspired by these pilot studies, we investigate the feasibility of using AI-enabled methods for automatic detection of driver drowsiness by analysing human behaviour and environment data. A similar research related to this work was proposed by Lee *et al.* [35]. In their study, the motion data collected from a smartwatch, i. e., the accelerometer and the gyroscope, were used to analyse driver behaviour. Eight features extracted from the raw sensor data were employed to train an SVM model for predicting the drowsiness level of the driver. Their encouraging results included a high accuracy (more than 98.0%) on Karolinska sleepiness scale [35], [36]. Nevertheless, this study has several limitations: First, two pre-defined models were used (for the left and right hands, respectively), which may cause complicated applications in the wild. Second, more advanced (ML) methods were not involved, e g., DL models. To this end, we introduce a novel method in this work by combining wearable sensor data and environmental data for analysis of a driver's drowsiness status.

The main contributions of this work can be summarised as: 1) motivated by a pioneer's preliminary work in [37], we for the first time propose a comprehensive study on a multimodal, data-based driver drowsiness detection method, which includes not only human activity data, but also environmental sensor data. We find that the complementary information provided by environmental data can be helpful for the final analysis. 2) we first demonstrate that, using the model trained by daily life data can be directly used for predicting the driver
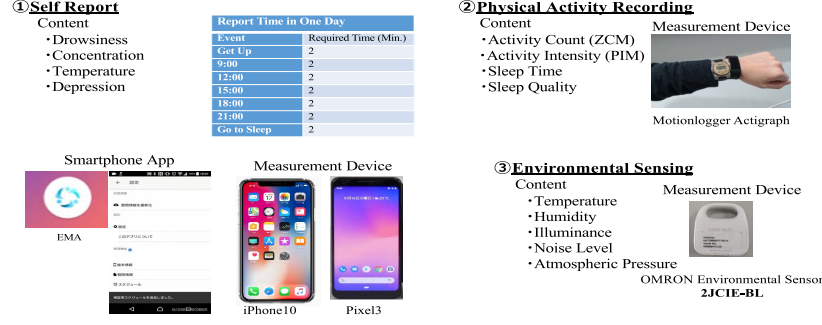
Fig. 1. Scheme of the data collection and annotation of the participants in daily life. The EMA was conducted to annotate the corresponding data by each participant at around 9:00, 12:00, 15:00, 18:00, and 21:00 o'clock during a day.

Fig. 2. Scheme of the data collection and annotation of the participants when driving in a simulator. We give two examples of the instruction course when the subject is driving in the simulator–Top: "Please drive your car following the blue car in front of you. Please be aware of the motorcycles and the people accidentally appearing from the two sides of the road." Bottom: "Please keep driving in the centre lane. There will be regularly traffic jams, but keep driving safely." Simultaneously, the activity data (ZCM and PIM) and the environmental data (Env) are collected the same as in the aforementioned daily life data.

states in simulator experiments. This novel finding can be a good guidance for further data collection and model training work. Third, we validate our proposed method in a rigid subject-independent scenario, which provides good guidance for researchers who share similar interests. Moreover, all the results are based on the averaged values of cross-validated experiments, which can guarantee a more convincing conclusion. 3) we comprehensively investigate the performances between different machine learning (ML) models including the 'classic' shallow models and recent deep learning (DL) models.

The remainder of this article is organised as follows: Firstly, we describe the database and methods used in Section II. Then, the experimental results and discussion are presented in Section III and Section IV, respectively. Finally, we give a conclusion in Section V.

II. DATABASE AND METHODS

In this section, we firstly introduce the database used in this study. Then, we describe the methods in detail accompanied by the evaluation metrics. In general, a novel method is proposed for using SPA data collected from daily life to predict a drivers' drowsiness status (see Fig. 1 and Fig. 2). Moreover, the environmental condition is also taken into account.

A. Data Acquisition and Annotation

1) Subjects: We recruited 24 office workers (female: 3, male: 21, age: 39.3 ± 9.6 years old) in our experiments. Written informed consent was obtained from all participants

after they were fully informed about the purpose; procedures were including risks and benefits of the study. This study was conducted as part of a comprehensive ethical application for ecological momentary assessment (EMA) [38] studies at The University of Tokyo, Japan. It should be noted that, the participants are recruited by volunteers and agreed to participate in the experiment and use of their data for research purposes. Unfortunately, male participants dominated as this was given by the gender balance of the research/development sections from which we could recruit the participants.

2) Daily Life Assessment of Drowsiness: EMA was adopted to obtain momentary drowsiness in daily life using a custom-made smartphone application. The custom app we used has a standard visual analog scale (VAS) [39]. In addition, the EMA studies using VAS for sleepiness/drowsiness have been extensively conducted. The participants were instructed to answer a questionnaire on the current drowsiness level at randomly selected times within ±5 minutes of prescheduled times (9:00, 12:00, 15:00, 18:00, and 21:00 o'clock) during the study period (five consecutive weekdays, from Monday to Friday), by using a visual analogue scale from 0 to 100 displayed on a smartphone screen. They were also asked to complete the EMA questionnaire on drowsiness when they woke up and before they went to bed. If the EMA score was equal to or greater than 60, we classified it into 'drowsiness', otherwise, it was classified into 'wakefulness'. The EMA scores among the participants in their daily life have a distribution of 7.4 ± 25.2 (from 0 to 100 scale). Fig. 1 shows the scheme of daily life assessment of drowsiness in this study.

3) Driving Simulator Protocols: Participants performed a driving task in a driving simulator (DS) once a day. We used two different protocols with 30 minutes length: traffic-jam and jump-out (UC- win/Road; FORUM 8 FORUM8 Co., Ltd., Tokyo, Japan). The traffic-jam protocol is a monotonous task, in which a driver is requested to keep driving safely in the centre lane of a street in a traffic jam, likely making a driver drowsy. This is supported by the mean and the standard deviation of the individually averaged drowsiness levels of 3.85 ± 0.79 for the traffic-jam protocol and 3.04 ± 0.92 for the following jump-out protocol ($p = 0.00005$ by a paired t-test). In the jump-out protocol, a driver follows the vehicle running in front of the own vehicle with keeping proper distances. Similar to the traffic-jam protocol, this is a 'boring' protocol, however, a pedestrian suddenly jumps out onto a street at pre-scheduled timings. Thus, this imposes a driver on staying focused on the task. Participants repeatedly conducted two protocols on a one-day rotation during the experimental period.

During the protocols, a driver orally reported their subjective drowsiness level using a scale from 1 to 5 (from "not sleepy at all" to "very sleepy") at the time when a message requesting a driver to answer the current drowsiness level appeared on the screen of the simulator vehicle. An experimenter manually recorded the level reported by the driver. The drowsiness level was assessed every 5 minutes in the jump-out protocol (5 recordings in total), while the assessment was conducted at the times of 8, 17, and 26 minutes passed from the start time in the traffic-jam protocol (3 recordings in total). The obtained drowsiness levels were also binarised; if the drowsiness level was equal to or greater than 3, it was labelled as "drowsiness"; otherwise, it was labelled as "wakefulness". The value 3 is chosen as a boundary dividing larger/smaller halves of the scale. Fig. 2 presents the simulator data collection processing scheme in this study.

B. Multimodal Data

1) Environmental Sensor Data: Participants were asked to carry a portable environmental sensor (2JCIE-BL; Omron Co. Ltd., Kyoto, Japan) with them at all times during the experimental period, including the duration of the driving simulator protocols, to record their ambient data. Therefore, the sensor was attached to their wrist during daytime and located at their bedside during sleep.

The environmental sensor data, including *temperature* (°C), *humidity* (relative humidity % RH), *illuminance* (lx), atmospheric pressure (hPa), and *noise level* (dB) are recorded as one of the modalities of the environment (Env.). These data were sampled every 5 minutes. Additionally, we add the *elapsed time* from the beginning of the day, i.e., 0:00 AM, into this kind of modality.

2) Physical Activity Data: All participants were also instructed to wear a watch-type activity monitor (actigraph; Ambulatory Monitoring Inc., Ardsley, NY, USA) on the wrist of their respective non-dominant hand throughout the study period, except while bathing, showering, performing rigorous exercises, or any other activity likely to damage the device. This device is equipped with a piezoelectric
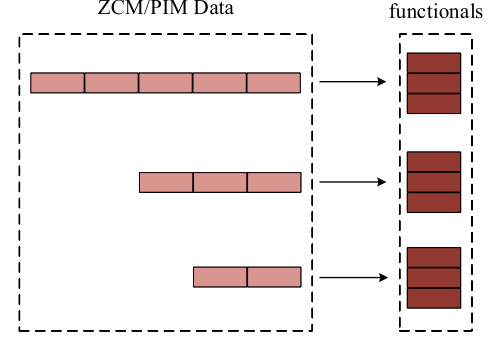

Fig. 3. An overall scheme of the statistical functionals approach. Independent of the length of the instances, statistics, e.g., *max*, *min*, *mean*, etc., are extracted from per one instance of the ZCM/PIM data.

accelerometer capable of detecting small changes in wrist acceleration (up to 0.01 G/rad/s), which enable to register even slight movements of daily life. The reliability of this device as a sleep/awake/vigilance detector can be found in detail on the website.[1] We use Zero Crossing Mode (ZCM) and Proportional Integration Mode (PIM) to obtain information on movement frequency and intensity, respectively. The ZCM counts the number of times the accelerometer signal crosses the zero-level within a specific time interval (referred to as epoch time). Generally, ZCM data are used for sleep studies because of their ability to estimate sleep-wake states with a high accuracy. On the other hand, The PIM calculates the area under the curve of the accelerometer signal. Therefore, PIM data provide a measure of physical activity level or vigour of motion. In this study, we set the epoch time to 1 minute. The activity data during which the participants were not wearing the device were identified and excluded from the analysis.

C. Feature Extraction

To extract high-level representations from the ZCM/PIM data, we use statistical functionals in this study. These statistical functionals, as one of the supera-segmental features [40], can be extracted independent of the instance length, which makes them suitable for static ML models, e.g., a support vector machine (SVM) [41]. Moreover, Dozza *et al.* proposed a method called "chunking" that aims to assure a more consistent and robust calculation of parameters from naturalistic driving data [42]. Motivated by the success in analysing human behaviour data [33], we extract nine statistics per instance of ZCM/PIM data, i.e., *maximum*, *minimum*, *mean*, *range*, *standard deviation*, *skewness*, *kurtosis*, *slope*, and *bias* of linear regression approximation. The details are as follows:

Let $y = x(n)$, $n = 1, 2, \ldots, N$ denote a time series of the ZCM/PIM data in a given time period. Then, the maximum (x_{max}), minimum (x_{min}), mean (μ_x), range (η_x), standard deviation (σ_x), skewness (ς_x), and kurtosis (κ_x) values are defined as:

$$x_{max} = \max\{x(1), x(2), \ldots, x(N)\}, \tag{1a}$$

$$x_{min} = \min\{x(1), x(2), \ldots, x(N)\}, \tag{1b}$$

$$\eta_x = x_{max} - x_{min}, \tag{1c}$$

[1] http://www.ambulatory-monitoring.com/

$$\mu_x = \frac{1}{N} \sum_{n=1}^{N} x(n), \tag{1d}$$

$$\sigma_x = \sqrt{\frac{1}{N-1} \sum_{n=1}^{N} |x(n) - \mu_x|^2}, \tag{1e}$$

$$\varsigma_x = \frac{E(x - \mu_x)^3}{\sigma_x^3}, \tag{1f}$$

$$\kappa_x = \frac{E(x - \mu_x)^4}{\sigma_x^4}, \tag{1g}$$

where $E(\cdot)$ denotes the expected value.

Considering a linear regression estimation, the target is to approximate a line ($\tilde{y} = \alpha n + \beta$) that has the minimised *quadratic error* ($\hat{e}^2$) between the approximated line ($\tilde{y}$) and the actual value series (y). The quadratic error can be written as:

$$\hat{e}^2 = \sum_{n=1}^{N} (y - \tilde{y})^2 = \sum_{n=1}^{N} (x(n) - \alpha n - \beta)^2$$
$$= \sum_{n=1}^{N} (x(n)^2 - 2\alpha n x(n) - 2\beta x(n) + 2\alpha\beta n + \alpha^2 n^2 + \beta^2), \tag{2}$$

where α, and β represent the *slope* and the *bias*, respectively. To minimise the quadratic error, the following differential equations should be solved (cf. [40]):

$$\frac{\partial \hat{e}^2}{\partial \alpha} = \sum_{n=1}^{N} (-2n x(n) + 2\beta n + 2\alpha n^2) = 0, \tag{3a}$$

$$\frac{\partial \hat{e}^2}{\partial \beta} = \sum_{n=1}^{N} (-2 x(n) + 2\alpha n + 2\beta) = 0, \tag{3b}$$

which can be re-written as:

$$-\sum_{n=1}^{N} n x(n) + \beta \sum_{n=1}^{N} n + \alpha \sum_{n=1}^{N} n^2 = 0, \tag{4a}$$

$$-\sum_{n=1}^{N} x(n) + \alpha \sum_{n=1}^{N} n + N\beta = 0. \tag{4b}$$

Finally, the solutions for α and β can be written as:

$$\alpha = \frac{N \sum_{n=1}^{N} n x(n) - \sum_{n=1}^{N} n \sum_{n=1}^{N} x(n)}{N \sum_{n=1}^{N} n^2 - (\sum_{n=1}^{N} n)^2}, \tag{5a}$$

$$\beta = \frac{\sum_{n=1}^{N} x(n) \sum_{n=1}^{N} n^2 - \sum_{n=1}^{N} n \sum_{n=1}^{N} n x(n)}{N \sum_{n=1}^{N} n^2 - (\sum_{n=1}^{N} n)^2}. \tag{5b}$$

These statistics are calculated from ZCM/PIM data with 60 min length prior to each drowsiness recording. As well, mean values of ambient environmental variables during the corresponding period are calculated.

TABLE I
THE GRID-SEARCH STRATEGY FOR OPTIMISING THE HYPER-PARAMETERS OF THE ML/DL MODELS. ALL THE CANDIDATE HYPER-PARAMETERS ARE EMPIRICALLY CHOSEN BASED ON INITIAL EXPERIMENTS

Model	Main Hyper-parameters
k-NN	*k value*: 2, 4, 8;
LDA	*solver*: 'Singular value decomposition', 'Least squares solution', 'Eigenvalue decomposition';
SVM	*kernel*: 'linear'; *C value*: 10^{-3}, 10^{-2}, 10^{-1}, 1;
LightGBM	*learning rate*: 10^{-3}, 10^{-4}; *max number of leaves in one tree*: 6, 16; *max number of depth*: 3, 6;
DNN	*number of hidden layers*: 1, 2, 3; *optimiser*: 'adam'; *number of hidden units*: 50, 50-50, 50-50-50;

D. ML Models

As a feasible investigation, we use a series of popular and mature classifier models in this study. For classic ML models, we use k nearest neighbour (k-NN) [43], linear discriminant analysis (LDA) [44], support vector machine (SVM) [41], and an ensemble learning strategy, i.e., LightGBM [45]. For DL models, we use a multilayer perceptron (MLP) [46] architecture that concatenates multiple feedforward neural networks (FNNs) [47]. The recurrent neurlal networks (RNNs) [48] are suitable to extract high-level information from the time-series data. However, in our initial experiments, the RNNs were not found to be superior to other ML/DL models for this database. We think the reason might be 30-minute length of the SPA data is still short for RNNs to capture sufficient strong representations, which restrains the feasibility of RNNs to be good models for predicting the drowsiness status of the subjects. All the hyper-parameters will be optimised in the validation set by a grid-search strategy (see Table I) and applied to the test set in the following experiments. The reader can refer to details of these selected models in the given references which are not the main direction of this study. We should indicate that, all the models we selected are fully considered as efficient and applicable based on our experience. Instead of a comprehensive comparison between the ML/DL models, this study focuses more on investigating the feasibility of the assumption that the multi-modal data collected in daily life can be modelled and used for predicting a drivers' drowsiness status.

E. Evaluation Metrics

Considering the imbalanced characteristic of the database, we use the unweighted average recall (UAR) as the main evaluation metric in this study, which is thought to be more representative in such case than the weighted average recall (*aka* accuracy) [49]. UAR is defined as:

$$\text{UAR} = \frac{\sum_{i=1}^{N_c} Recall_i}{N_c}, \tag{6}$$

where $Recall_i$ is the *recall* of the i-th class. N_c is the total number of classes. The widely used weighted average recall (WAR), i.e., accuracy is defined as:

$$\text{WAR} = \sum_{i=1}^{N_c} \lambda_i Recall_i,$$

$$\lambda_i = \frac{N_i}{N}, \tag{7}$$

where λ_i is the *weight* for the i-th class, N_i is the total number of instances labelled as 'i-th class', N is the total number of instances.

In addition, due to the fact that that drowsiness detection is a binary classification task, we also calculate sensitivity (Sens.), specificity (Spec.), precision (Prec.), F1 measure (F1), and geometric mean (G-mean) values (combined with WAR) as complementary metrics. These values are defined as:

$$\text{Sens.} = \frac{TP}{TP + FN}, \tag{8a}$$

$$\text{Spec.} = \frac{TN}{TN + FP}, \tag{8b}$$

$$\text{Prec.} = \frac{TP}{TP + FP}, \tag{8c}$$

$$\text{F1} = 2 \cdot \frac{\text{Prec.} \cdot \text{Recall}}{\text{Prec.} + \text{Recall}}, \tag{8d}$$

$$\text{G-mean} = \sqrt{\text{Sens.} \cdot \text{Spec.}}, \tag{8e}$$

where Recall is the same as Sens.; TP, TN, FP, and FN are the number of *true positive* ('Drowsiness' correctly identified as 'Drowsiness'), *true negative* ('Wakefulness' correctly identified as 'Wakefulness'), *false positive* ('Wakefulness' incorrectly identified as 'Drowsiness'), and *false negative* ('Drowsiness' incorrectly identified as 'Wakefulness'), respectively.

When comparing two models' performances, we use as significance test the one-tailed z-test [50]. The results are regarded as significant when the p-value is less than .05.

III. EXPERIMENTAL RESULTS

The experimental setup and results are given in this section. A discussion based on this study will then be conducted in the following section.

A. Setup

In this study, we adopt a subject-independent paradigm for validating our proposed method. Namely, the instances used for training or validating/testing are from different subjects. Firstly, the total 24 subjects are divided into six groups. Then, the hyper-parameters of the classifiers are optimised on train/validation (four vs one) in five groups. Finally, the remaining one group will be used as the test set. Firstly, the total 24 subjects are equally divided into six groups. Among these six groups, one group will be selected as a test set in an independent experiment. In each experiment, in order to optimise the hyper-parameters, the other five groups together are split into train/validation for five-fold cross-validation. The same procedure will be replicated six times by selecting a different test set in each turn.

Group ID	**W**	**D**	Σ
1	206	378	584
2	263	397	660
3	251	396	647
4	281	389	670
5	274	377	651
6	270	378	648
Σ	1 545	2 315	3 860

The final test results are the averaged values of the six independent experiments. The numbers of instances in each label and each test subject group are summarised in Table II. This cross validated subject-independent model evaluation can avoid over-optimistic results and allows a more solid conclusion from the results. The experiments are done by scripts written in Python with the scikit-learn toolkit. We conduct the experiments in four scenarios: First, we use the daily life data trained model to detect the drowsiness in daily life data (D2D). Second, we use the daily life data trained model to detect the drowsiness in simulator data (D2S). Third, we use the simulator data trained model to detect the drowsiness in simulator data (S2S). Finally, we use the simulator data trained model to detect the drowsiness in daily life data (S2D).

B. Results

The experimental results of the aforementioned four scenarios are shown in Table III, Table IV, Table V, and Table VI, respectively. When looking at the D2D scenario (see Table III), the best result (a UAR of 60.3%) is achieved by the LightGBM model using a single modality, i.e., the ZCM data. For this case, introducing more data modalities cannot yield better results. Similarly, for the D2S scenario (see Table IV), using the environmental information only can make the LightGBM classifier reach the best result, i.e., a UAR of 56.1%. For the other two scenarios, i.e., the cases S2S and S2D, the best results are achieved by employing multi-modal analysis of the data. In particular, the Env. + PIM case provides the best result in the S2S scenario (a UAR of 64.6%) while analysing all the modalities reaches the best performance in the S2D scenario (a UAR of 55.7%).

The complementary metrics for measuring the performances of the best models in four scenarios are listed in Fig. 4. In the S2S scenario, the best model can reach the best WAR (68.5%), the best Sens. (75.5%), the best Prec. (75.9%), and the best G-mean (63.0%). When considering the Spec. and the F1-score, the top performance is reached in the D2D scenario at 69.4% and 54.2%, respectively. In summary, all the experimental results demonstrate that the proposed method can be feasible to give an alert for predicting the drivers' drowsiness status in a totally non-invasive way. There are no significant differences between the selected ML/DL models. The environmental sensor data shows its importance to help the models reach best performance in several scenarios. More importantly, we demonstrate that the data collected in daily life can be modelled and used for predicting the driving drowsiness status.

TABLE III

EXPERIMENTAL RESULTS (UARS: IN [%]) FOR USING DAILY LIFE DATA TO PREDICT THE DAILY LIFE DATA. THE RESULTS ARE AVERAGED ACROSS SIX INDEPENDENT EXPERIMENTS. THE BEST RESULTS ON THE TEST SET ARE HIGHLIGHTED IN BOLD FONT WITH A GREY BACKGROUND. CHANCE LEVEL: 50.0%

UAR [%]	k-NN		LDA		SVM		LightGBM		DNN	
	Val.	Test	Val.	Test	Val.	Test	Val.	Test	Val.	Test
Env.	53.8	50.7	48.2	49.3	50.0	50.0	51.3	49.7	53.8	51.2
ZCM	53.9	56.0	52.1	52.3	51.4	53.7	54.6	**60.3**	60.0	51.9
PIM	58.0	53.8	50.1	51.4	50.4	48.8	58.6	57.9	57.4	52.1
Env.+ZCM	59.7	58.1	51.8	50.9	51.1	50.9	50.7	57.6	58.1	53.1
Env.+PIM	54.2	52.4	47.0	49.3	50.2	48.7	56.4	57.2	56.2	53.7
ZCM+PIM	57.7	58.4	54.1	57.4	52.9	53.7	56.0	57.9	57.7	53.3
Env.+ZCM+PIM	58.4	57.6	54.5	54.6	52.3	52.4	53.5	58.1	59.3	53.2

TABLE IV

EXPERIMENTAL RESULTS (UARS: IN [%]) FOR USING DAILY LIFE DATA TO PREDICT THE SIMULATOR DATA. THE RESULTS ARE AVERAGED ACROSS SIX INDEPENDENT EXPERIMENTS. THE BEST RESULTS ON THE TEST SET ARE HIGHLIGHTED IN BOLD FONT WITH A GREY BACKGROUND. CHANCE LEVEL: 50.0%

UAR [%]	k-NN		LDA		SVM		LightGBM		DNN	
	Val.	Test	Val.	Test	Val.	Test	Val.	Test	Val.	Test
Env.	53.8	53.9	48.2	49.7	50.0	50.0	51.3	**56.1**	53.8	51.5
ZCM	53.9	54.2	52.1	52.6	51.4	52.8	54.6	55.0	60.0	50.8
PIM	58.0	54.6	50.1	50.2	50.4	49.6	58.6	47.8	57.4	54.1
Env.+ZCM	59.7	51.1	51.8	50.5	51.1	49.0	50.7	52.5	58.1	48.8
Env.+PIM	54.2	52.5	47.0	49.4	50.2	49.9	56.4	50.4	56.2	51.5
ZCM+PIM	57.7	50.4	54.1	50.3	52.9	51.3	56.0	47.8	57.7	48.1
Env.+ZCM+PIM	58.4	55.8	54.5	49.8	52.3	51.6	53.5	49.4	59.3	54.9

TABLE V

EXPERIMENTAL RESULTS (UARS: IN [%]) FOR USING SIMULATOR DATA TO PREDICT THE SIMULATOR DATA. THE RESULTS ARE AVERAGED ACROSS SIX INDEPENDENT EXPERIMENTS. THE BEST RESULTS ON THE TEST SET ARE HIGHLIGHTED IN BOLD FONT WITH A GREY BACKGROUND. CHANCE LEVEL: 50.0%

UAR [%]	k-NN		LDA		SVM		LightGBM		DNN	
	Val.	Test	Val.	Test	Val.	Test	Val.	Test	Val.	Test
Env.	67.4	55.3	66.9	62.3	66.4	61.0	71.4	64.2	67.6	55.9
ZCM	61.9	51.5	56.4	50.1	50.0	50.0	63.0	48.9	64.4	41.7
PIM	59.1	52.2	54.3	56.8	50.2	49.9	60.4	57.3	67.3	50.1
Env.+ZCM	63.0	52.3	63.5	55.7	62.7	56.1	70.0	61.3	66.2	51.2
Env.+PIM	61.9	60.0	65.6	60.6	66.2	60.1	70.5	**64.6**	70.3	54.0
ZCM+PIM	49.1	52.6	55.4	56.7	53.3	52.5	61.7	53.8	62.2	49.2
Env.+ZCM+PIM	60.3	53.3	61.5	55.4	63.9	55.8	71.2	62.2	66.4	52.0

TABLE VI

EXPERIMENTAL RESULTS (UARS: IN [%]) FOR USING SIMULATOR DATA TO PREDICT THE DAILY LIFE DATA. THE RESULTS ARE AVERAGED ACROSS SIX INDEPENDENT EXPERIMENTS. THE BEST RESULTS ON THE TEST SET ARE HIGHLIGHTED IN BOLD FONT WITH A GREY BACKGROUND. CHANCE LEVEL: 50.0%

UAR [%]	k-NN		LDA		SVM		LightGBM		DNN	
	Val.	Test	Val.	Test	Val.	Test	Val.	Test	Val.	Test
Env.	67.4	50.9	66.9	49.4	66.4	49.2	71.4	49.3	67.6	50.0
ZCM	61.9	54.1	56.4	52.7	50.0	50.0	63.0	52.1	64.4	47.5
PIM	59.1	53.4	54.3	53.5	50.2	50.2	60.4	51.0	67.3	51.5
Env.+ZCM	63.0	53.1	63.5	48.6	62.7	48.2	70.0	48.1	66.2	51.6
Env.+PIM	61.9	52.7	65.6	48.0	66.2	48.8	70.5	48.8	70.3	50.5
ZCM+PIM	49.1	50.5	55.4	47.5	53.3	47.8	61.7	49.6	62.2	50.2
Env.+ZCM+PIM	60.3	**55.7**	61.5	47.8	63.9	46.4	71.2	50.0	66.4	50.3

IV. DISCUSSION

In this section, we firstly present the findings in this study. Then, we indicate the current limitations and perspectives in future work.

A. Current Findings

In this study, it is demonstrated that, directly using a model trained by daily life data can be feasible to predict the subjects' drowsiness status when they are in a simulator, and vice versa. Even though the current results are modest, this finding is still encouraging because we can see the potential of utilising large scale daily life behavioural data to train a model that could be possible for predicting the subjects' drowsiness status when they are driving.

In addition, we can find that, ZCM and PIM are two important representations that can reflect the subjects' behaviour.

TABLE VII

Confusion Matrices (Normalised: In [%]) for of the Best Models in Four Scenarios: D2D: Using Daily Life Data to Predict Daily Life Data; D2S: Using Daily Life Data to Predict Simulator Data; S2S: Using Simulator Data to Predict Simulator Data; and S2D: Using Simulator Data to Predict Daily Life Data. The Results Are Averaged Across Six Independent Experiments. **D**: Drowsiness; **W**: Wakefulness

(a) D2D

Pred ->	D	W
D	66.4	33.6
W	48.1	51.9

(b) D2S

Pred ->	D	W
D	50.6	49.4
W	37.5	62.5

(c) S2S

Pred ->	D	W
D	56.5	43.5
W	25.4	74.6

(d) S2D

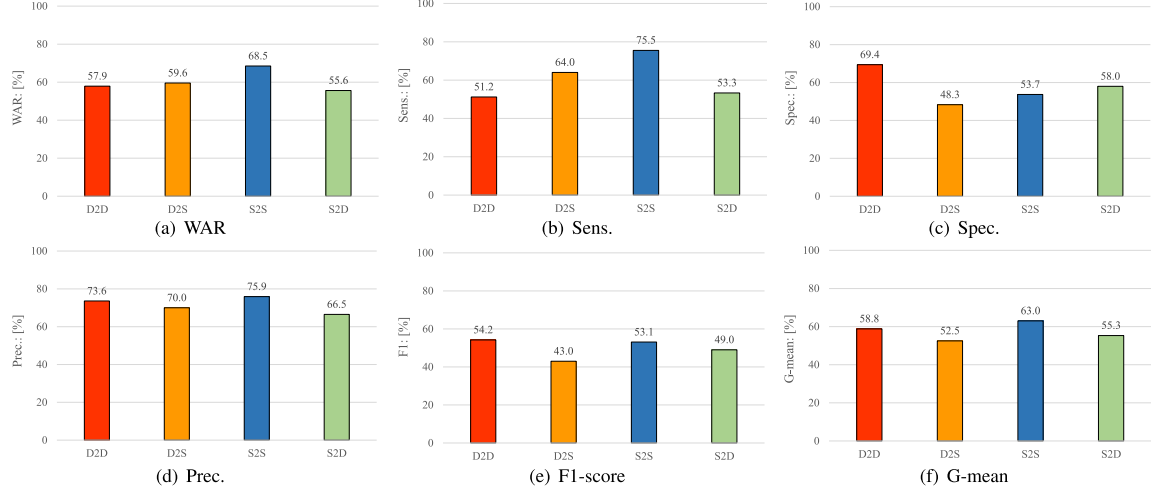| Pred ->	D	W
D	58.9	41.1
W	46.7	53.3

Fig. 4. Complementary metrics (in [%]) achieved by the best models in the four scenarios: D2D: Using daily life data to predict daily life data; D2S: Using daily life data to predict simulator data; S2S: Using simulator data to predict simulator data; and S2D: Using simulator data to predict daily life data. The results are averaged across six independent experiments.

Moreover, the environmental condition cannot be ignored when predicting the subjects' drowsiness status. But it should be noticed that, the data collected from the wearable sensor and environmental condition are not complicated to be calculated. This will be applicable in real-world usage.

We surmise that DNN model performance suffers due to the limited data size for training. Nevertheless, simple classifiers, e.g., k-NN, can be efficient and robust when fed with multimodal representations. As an ensemble learning model, LightGBM has shown its encouraging performance in this study.

We believe that our proposed method was evaluated within a rigorous scheme: First, all the experiments were executed in a subject-independent scenario, which means the training and validation/test sets were collected from different participants. Second, to avoid some biased evaluation due to the data partition strategy, a cross-validation method was used. Third, the main evaluation metric, i.e., UAR is better suited than the widely used WAR (*aka.* accuracy) for the considered imbalanced data set.

B. Limitations and Perspectives

First of all, a fundamental study on differences between the behaviour of daily life vs driving in a simulator is lacking. In future work, one needs to explore this relationship. Also, bigger lifestyle differences (e.g., poor sleep, shift work, sleep apnea) that cause drowsiness should be considered and

investigated in future work. More importantly, the presence of poor sleep or circadian rhythms could be a crucial factor that leading to a high prevalence of drowsiness. One should study deeply about the individual lifestyle and its effect on detection of drowsiness.

Second, the current experimental results need to be improved. Note that, it is difficult to collect a sufficient amount of data in practice. Therefore, one will need to investigate some state-of-the-art techniques that can address the data scarcity challenge – for instance, active learning [51], [52], cooperative learning [53], and/or augmentation by generative adversarial networks (GANs) [54], [55]. We also have to confirm that gender appears a highly relevant factor for predicting a driver's behaviour. It was indicated that, compared to female drivers, male drivers are more likely to exhibit various patterns of aggressive driving [56]. In future work, more female participants should be involved in the experiments, which can improve the models' generalisation. In addition, diverse gender should be integrated into considerations.

Third, the ground-truth of the data used in this study is not sufficiently accurate due to the current small size of the data and the experimental conditions. This limitation may restrain the real applications of our proposed methods. In future work, one should investigate a ternary classification scenario, i.e., *mild*, *moderate*, or *severe* drowsiness level. Moreover, a regression task should also be validated.

Finally, one should consider building an explainable AI (XAI) system [57] to make the models interpretable for further deep study on human behaviours in both daily life and the driving case.

V. Conclusion

A novel non-invasive multimodal ML analysis method was proposed for the driver drowsiness detection task. Environmental sensor data was firstly introduced as complementary information of human activity data for building the prediction model. The experimental results then demonstrated the efficacy of the method. Even though the current results are moderate due to the limited data size, the feasibility has been successfully validated in a rigorous scheme via a cross validated subject-independent scenario. In addition, the model trained by the data collected in subjects' daily life can be used directly to predict the drivers' states in simulator experiments. The best result for UAR was 64.6% in a subject-independent scenario in this study.

References

[1] N. H. T. S. Administration. *Drowsy Driving*. Accessed: Oct. 2020. [Online]. Available: https://www.nhtsa.gov/risky-driving/drowsy-driving

[2] M. I. Chacon-Murguia and C. Prieto-Resendiz, "Detecting driver drowsiness: A survey of system designs and technology," *IEEE Consum. Electron. Mag.*, vol. 4, no. 4, pp. 107–119, Oct. 2015.

[3] S. Kaplan, M. A. Guvensan, A. G. Yavuz, and Y. Karalurt, "Driver behavior analysis for safe driving: A survey," *IEEE Trans. Intell. Transp. Syst.*, vol. 16, no. 6, pp. 3017–3032, Dec. 2015.

[4] C. Bila, F. Sivrikaya, M. A. Khan, and S. Albayrak, "Vehicles of the future: A survey of research on safety issues," *IEEE Trans. Intell. Transp. Syst.*, vol. 18, no. 5, pp. 1046–1065, May 2017.

[5] A. Chowdhury, R. Shankaran, M. Kavakli, and M. M. Haque, "Sensor applications and physiological features in drivers' drowsiness detection: A review," *IEEE Sensors J.*, vol. 18, no. 8, pp. 3055–3067, Apr. 2018.

[6] M. Ramzan, H. U. Khan, S. M. Awan, A. Ismail, M. Ilyas, and A. Mahmood, "A survey on state-of-the-art drowsiness detection techniques," *IEEE Access*, vol. 7, pp. 61904–61919, 2019.

[7] Q. Ji, Z. Zhu, and P. Lan, "Real-time nonintrusive monitoring and prediction of driver fatigue," *IEEE Trans. Veh. Technol.*, vol. 53, no. 4, pp. 1052–1068, Jul. 2004.

[8] R. Oyini Mbouna, S. G. Kong, and M.-G. Chun, "Visual analysis of eye state and head pose for driver alertness monitoring," *IEEE Trans. Intell. Transp. Syst.*, vol. 14, no. 3, pp. 1462–1469, Sep. 2013,

[9] K. Yuen and M. M. Trivedi, "An occluded stacked hourglass approach to facial landmark localization and occlusion estimation," *IEEE Trans. Intell. Vehicles*, vol. 2, no. 4, pp. 321–331, Dec. 2017.

[10] B. Mandal, L. Li, G. S. Wang, and J. Lin, "Towards detection of bus driver fatigue based on robust visual analysis of eye state," *IEEE Trans. Intell. Transp. Syst.*, vol. 18, no. 3, pp. 545–557, Mar. 2017.

[11] F. You, X. Li, Y. Gong, H. Wang, and H. Li, "A real-time driving drowsiness detection algorithm with individual differences consideration," *IEEE Access*, vol. 7, pp. 179396–179408, 2019.

[12] A. Dasgupta, D. Rahman, and A. Routray, "A smartphone-based drowsiness detection and warning system for automotive drivers," *IEEE Trans. Intell. Transp. Syst.*, vol. 20, no. 11, pp. 4045–4054, Nov. 2019.

[13] J. Yu, S. Park, S. Lee, and M. Jeon, "Driver drowsiness detection using condition-adaptive representation learning framework," *IEEE Trans. Intell. Transp. Syst.*, vol. 20, no. 11, pp. 4206–4218, Nov. 2019.

[14] W. Deng and R. Wu, "Real-time driver-drowsiness detection system using facial features," *IEEE Access*, vol. 7, pp. 118727–118738, 2019.

[15] B. K. Savas and Y. Becerikli, "Real time driver fatigue detection system based on multi-task ConNN," *IEEE Access*, vol. 8, pp. 12491–12498, 2020.

[16] C.-Y. Chiou, W.-C. Wang, S.-C. Lu, C.-R. Huang, P.-C. Chung, and Y.-Y. Lai, "Driver monitoring using sparse representation with part-based temporal face descriptors," *IEEE Trans. Intell. Transp. Syst.*, vol. 21, no. 1, pp. 346–361, Jan. 2020.

[17] 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.

[18] C.-T. Lin, R.-C. Wu, S.-F. Liang, W.-H. Chao, Y.-J. Chen, and T.-P. Jung, "EEG-based drowsiness estimation for safety driving using independent component analysis," *IEEE Trans. Circuits Syst. I, Reg. Papers*, vol. 52, no. 12, pp. 2726–2738, Dec. 2005.

[19] C. T. Lin *et al.*, "Development of wireless brain computer interface with embedded multitask scheduling and its application on real-time driver's drowsiness detection and warning," *IEEE Trans. Biomed. Eng.*, vol. 55, no. 5, pp. 1582–1591, May 2008.

[20] C.-T. Lin, C.-J. Chang, B.-S. Lin, S.-H. Hung, C.-F. Chao, and I.-J. Wang, "A real-time wireless brain–computer interface system for drowsiness detection," *IEEE Trans. Biomed. Circuits Syst.*, vol. 4, no. 4, pp. 214–222, Aug. 2010.

[21] F.-C. Lin, L.-W. Ko, C.-H. Chuang, T.-P. Su, and C.-T. Lin, "Generalized EEG-based drowsiness prediction system by using a self-organizing neural fuzzy system," *IEEE Trans. Circuits Syst. I, Reg. Papers*, vol. 59, no. 9, pp. 2044–2055, Sep. 2012.

[22] G. Li, B.-L. Lee, and W.-Y. Chung, "Smartwatch-based wearable EEG system for driver drowsiness detection," *IEEE Sensors J.*, vol. 15, no. 12, pp. 7169–7180, Dec. 2015.

[23] S. Taran and V. Bajaj, "Drowsiness detection using adaptive Hermite decomposition and extreme learning machine for electroencephalogram signals," *IEEE Sensors J.*, vol. 18, no. 21, pp. 8855–8862, Nov. 2018.

[24] C.-S. Wei, Y.-T. Wang, C.-T. Lin, and T.-P. Jung, "Toward drowsiness detection using non-hair-bearing eeg-based brain-computer interfaces," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 26, no. 2, pp. 400–406, Feb. 2018.

[25] T. K. Reddy, V. Arora, S. Kumar, L. Behera, Y.-K. Wang, and C.-T. Lin, "Electroencephalogram based reaction time prediction with differential phase synchrony representations using co-operative multi-task deep neural networks," *IEEE Trans. Emerg. Topics Comput. Intell.*, vol. 3, no. 5, pp. 369–379, Oct. 2019.

[26] H.-S. Choi *et al.*, "Learning-based instantaneous drowsiness detection using wired and wireless electroencephalography," *IEEE Access*, vol. 7, pp. 146390–146402, 2019.

[27] K. T. Chui, K. F. Tsang, H. R. Chi, B. W. K. Ling, and C. K. Wu, "An accurate ECG-based transportation safety drowsiness detection scheme," *IEEE Trans. Ind. Informat.*, vol. 12, no. 4, pp. 1438–1452, Aug. 2016.

[28] F. Guede-Fernández, M. Fernández-Chimeno, J. Ramos-Castro, and M. A. García-González, "Driver drowsiness detection based on respiratory signal analysis," *IEEE Access*, vol. 7, pp. 81826–81838, 2019.

[29] B.-G. Lee and W.-Y. Chung, "Driver alertness monitoring using fusion of facial features and bio-signals," *IEEE Sensors J.*, vol. 12, no. 7, pp. 2416–2422, Jul. 2012.

[30] A. Picot, S. Charbonnier, and A. Caplier, "On-line detection of drowsiness using brain and visual information," *IEEE Trans. Syst., Man, Cybern., A, Syst. Humans*, vol. 42, no. 3, pp. 764–775, May 2012.

[31] C. Zhang, X. Wu, X. Zheng, and S. Yu, "Driver drowsiness detection using multi-channel second order blind identifications," *IEEE Access*, vol. 7, pp. 11829–11843, 2019.

[32] M. Sunagawa, S.-I. Shikii, W. Nakai, M. Mochizuki, K. Kusukame, and H. Kitajima, "Comprehensive drowsiness level detection model combining multimodal information," *IEEE Sensors J.*, vol. 20, no. 7, pp. 3709–3717, Apr. 2020.

[33] K. Qian *et al.*, "Teaching machines to know your depressive state: On physical activity in health and major depressive disorder," in *Proc. 41st Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC)*, Berlin, Germany, Jul. 2019, pp. 3592–3595.

[34] K. Qian *et al.*, "Automatic detection of major depressive disorder via a bag-of-behaviour-words approach," in *Proc. 3rd Int. Symp. Image Comput. Digit. Med. (ISICDM)*, Xi'an, China, 2019, pp. 71–75.

[35] 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.

[36] A. Sahayadhas, K. Sundaraj, and M. Murugappan, "Detecting driver drowsiness based on sensors: A review," *Sensors*, vol. 12, no. 12, pp. 16937–16953, Dec. 2012.

[37] S. Kim *et al.*, "Detecting driver drowsiness based fusion multi-sensors method," in *Proc. CUTE & CSA*, Kuala Lumpur, Malaysia, 2018, pp. 459–464.

[38] S. Shiffman, A. A. Stone, and M. R. Hufford, "Ecological momentary assessment," *Annu. Rev. Clin. Psychol.*, vol. 4, no. 1, pp. 1–32, 2008.

[39] L. Klimek *et al.*, "Visual analogue scales (VAS): Measuring instruments for the documentation of symptoms and therapy monitoring in cases of allergic rhinitis in everyday health care," *Allergo J. Int.*, vol. 26, no. 1, pp. 16–24, Feb. 2017.

[40] F. Eyben, *Real-Time Speech Music Classification by Large Audio Feature Space Extraction*. Cham, Switzerland: Springer, 2015.

[41] C. Cortes and V. Vapnik, "Support-vector networks," *Mach. Learn.*, vol. 20, no. 3, pp. 273–297, 1995.

[42] M. Dozza, J. Bärgman, and J. D. Lee, "Chunking: A procedure to improve naturalistic data analysis," *Accident Anal. Prevention*, vol. 58, pp. 309–317, Sep. 2013.

[43] T. Cover and P. Hart, "Nearest neighbor pattern classification," *IEEE Trans. Inf. Theory*, vol. 13, no. 1, pp. 21–27, Jan. 1967.

[44] R. A. Fisher, "The use of multiple measurements in taxonomic problems," *Ann. Eugenics*, vol. 7, no. 2, pp. 179–188, 1936.

[45] G. Ke *et al.*, "LightGBM: A highly efficient gradient boosting decision tree," in *Proc. NeurIPS*, Long Beach, CA, USA, 2017, pp. 3146–3154.

[46] C. M. Bishop, *Pattern Recognition and Machine Learning*. New York, NY, USA: Springer, 2006.

[47] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, May 2015.

[48] M. Schuster and K. K. Paliwal, "Bidirectional recurrent neural networks," *IEEE Trans. Signal Process.*, vol. 45, no. 11, pp. 2673–2681, Nov. 1997.

[49] B. W. Schuller, S. Steidl, and A. Batliner, "The INTERSPEECH 2009 emotion challenge," in *Proc. INTERSPEECH*, Brighton, U.K., 2009, pp. 312–315.

[50] T. G. Dietterich, "Approximate statistical tests for comparing supervised classification learning algorithms," *Neural Comput.*, vol. 10, no. 7, pp. 1895–1923, 1998.

[51] B. Settles, "Active learning," *Synth. Lectures Artif. Intell. Mach. Learn.*, vol. 6, no. 1, pp. 1–114, 2012.

[52] K. Qian, Z. Zhang, A. Baird, and B. Schuller, "Active learning for bird sound classification via a kernel-based extreme learning machine," *J. Acoust. Soc. Amer.*, vol. 142, no. 4, pp. 1796–1804, Oct. 2017.

[53] Z. Zixing, E. Coutinho, D. Jun, and B. Schuller, "Cooperative learning and its application to emotion recognition from speech," *IEEE/ACM Trans. Audio, Speech, Lang. Process.*, vol. 23, no. 1, pp. 115–126, Jan. 2015.

[54] I. Goodfellow *et al.*, "Generative adversarial nets," in *Proc. NIPS*, Montreal, QC, Canada, 2014, pp. 2672–2680.

[55] Z. Zhang, J. Han, K. Qian, C. Janott, Y. Guo, and B. Schuller, "Snore-GANs: Improving automatic snore sound classification with synthesized data," *IEEE J. Biomed. Health Informat.*, vol. 24, no. 1, pp. 300–310, Jan. 2020.

[56] G. Fountas, S. S. Pantangi, K. F. Hulme, and P. C. Anastasopoulos, "The effects of driver fatigue, gender, and distracted driving on perceived and observed aggressive driving behavior: A correlated grouped random parameters bivariate probit approach," *Analytic Methods Accident Res.*, vol. 22, Jun. 2019, Art. no. 100091.

[57] 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.

Tomoya Koike (Student Member, IEEE) received the B.Sc. degree from Kobe University, Kobe, Japan, in 2020. He has working experience with AI related projects on research and development. He is currently at The University of Tokyo, Tokyo, Japan. He is the main author of the open source toolkit DEEPSELF. He has successfully published academic papers in a series of prestigious conferences, such as EMBC and INTERSPEECH. His research interests include machine learning, deep learning, and healthcare applications.

Toru Nakamura (Member, IEEE) received the M.E. and Ph.D. degrees in engineering from the Graduate School of Engineering Science, Osaka University, Osaka, Japan, in 2002 and 2005, respectively.

He is currently a Project Professor at the Center for Industry-University Collaboration, Graduate School of Engineering Science, Osaka University. His current research interests include nonlinear biodynamics, biomedical signal processing, and health informatics.

Björn W. Schuller (Fellow, IEEE) received the Diploma and Ph.D. degrees in electrical engineering and information technology and the Habilitation degree in electrical engineering and information technology (signal processing and machine intelligence) from Technische Universität München (TUM), Germany, in 1999, 2006, and 2012, respectively. He received the Adjunct Teaching Professorship in the subject area of signal processing and machine intelligence from TUM in 2012. He is currently a tenured Full Professor heading the Chair of Embedded Intelligence for Health Care and Wellbeing, University of Augsburg, Germany, and a Professor of artificial intelligence heading GLAM, Department of Computing, Imperial College London, London, U.K. He (co)authored five books and more than 1,000 publications in peer-reviewed books, journals, and conference proceedings leading to more than 39 000 citations (H-index 89). He is the President Emeritus of AAAC. He is a fellow of the Golden Core Awardee of the IEEE Computer Society, ISCA, and BCS. He is a Senior Member of ACM. He is the Field Chief Editor of *Frontiers in Digital Health* and the former Editor-in-Chief of the IEEE TRANSACTIONS ON AFFECTIVE COMPUTING.

Kun Qian (Senior Member, IEEE) received the Ph.D. degree in electrical engineering and information technology from Technische Universität München (TUM), Germany, in 2018, for his study on automatic general audio signal classification. Since 2021, he has been appointed as a (Full) Professor with a title of "Teli Young Fellow" at Beijing Institute of Technology, China. He has a strong collaboration connection to prestigious universities in Germany, U.K., Japan, Singapore, and the USA. He (co)authored more than 70 publications in peer-reviewed journals and conference proceedings having received more than 1 000 citations (H-index 19). He serves as an Associate Editor for the IEEE TRANSACTIONS ON AFFECTIVE COMPUTING, *Frontiers in Digital Health*, and *BIO Integration*.

Yoshiharu Yamamoto (Member, IEEE) received the B.Sc., M.Sc., and Ph.D. degrees in education from The University of Tokyo, Tokyo, Japan, in 1984, 1986, and 1990, respectively. Since 2000, he has been a Professor with the Graduate School of Education, The University of Tokyo, where he is currently teaching and researching physiological bases of health sciences and education. He (co)authored more than 230 publications in peer-reviewed books, journals, and conference proceedings leading to more than 11 000 citations (H-index 55). His research interests include biomedical signal processing, nonlinear and statistical biodynamics, and health informatics. He is an Associate Editor of the IEEE TRANSACTIONS ON BIOMEDICAL ENGINEERING. He is an Editorial Board Member of the *Technology* and *Biomedical Physics and Engineering Express*. He is also the President of the Healthcare IoT Consortium, Japan.