Towards Reliable Driver Drowsiness Detection Leveraging Wearables

Yetong Cao, *Student Member, IEEE,* Fan Li, *Member, IEEE,* Xiaochen Liu, Song Yang, *Member, IEEE,* and Yu Wang, *Fellow, IEEE*

***Abstract*—Driver drowsiness is a significant factor in road crashes. Unfortunately, the state-of-the-art methods for driver drowsiness detection have major drawbacks of requiring expensive hardware, being inconvenient to use, being restricted to limited scenarios, and have insufficient accuracy. To overcome these drawbacks, we propose FDWatch, a novel and accurate drowsiness detection system that exploits the low-cost Photoplethysmogram (PPG) sensor and motion sensor integrated into wrist-worn devices. We design a set of novel algorithms to extract multiple drowsiness-related indicators (including yawning frequency, the rate of yawning interval shortening, heart rate variability features, steering wheel turning angles and speeds, and circadian-rhythm-related alertness level) and jointly consider them to assess drowsiness. In particular, we subvert the traditional understanding of PPG and demonstrate that commodity PPG sensors can be utilized to detect yawning behavior. The core of FDWatch is based on Dempster-Shafer evidence theory. It considers different indicators as evidence describing the state of the driver from different angles. To make the extracted indicators applicable to Dempster-Shafer evidence theory, we employ backpropagation neural networks to obtain the basic probability assignment. Moreover, we propose a similarity-distance-based method to handle evidence conflicts. Extensive experiments with real-road driving data show that FDWatch can accurately detect driver drowsiness with a missing alarm rate of 3.57% and a false alarm rate of 3.68%.**

***Index Terms*—Driver drowsiness detection, wearables, information fusion, Dempster-Shafer evidence theory, PPG.**

I. Introduction

With the rapid growth of car ownership, our daily lives are becoming more convenient, but with it comes the increase in the number of traffic accident deaths. In addition to a large number of cars, some people's indifference to traffic safety is also a major cause. Among them, drowsy driving, as one of the causes of accidents that are difficult to detect, is related to approximately 20% of global traffic accidents [1]. Therefore, it is urgent to develop an automatic drowsiness warning system to improve road safety.

Recent clinical practice validates that some physiology features (e.g., heart rate variability) convey information about the autonomic nervous system and can be utilized to detect drowsiness [2]. Therefore, some methods leverage the electroencephalogram (EEG), electrocardiogram (ECG), and electrooculogram (EOG) to estimate drowsiness-related features and evaluate the driver drowsiness stage [3]–[5]. However, the user is usually tethered to the electrodes in a wired manner or needs to wear bulky devices, which is inconvenient and not preferred in daily life.

To enable convenient drowsiness detection in daily life, a variety of prior work has looked at characterizing notable driver behaviors instead of physiology features to detect driver drowsiness. Manufacturers (e.g., Mercedes-Benz, Nissan, Hyundai, and Volvo) have been updating their vehicles to include infrared cameras and other dedicated hardware. With these vehicles, drowsiness can be detected by characterizing facial expressions (e.g., eye blinking frequency, eye closure duration, percentage of eye closure, and yawning) [6], [7] and driving operations (e.g., steering wheel movements and lane change movements) [8]–[10]. However, the deployment cost of these methods is high, which is difficult to be adopted widely, especially in developing countries and areas.

Under the recent trend of mobile sensing, researchers have exploited the cameras integrated into smartphones to assess drowsiness-related facial expression features [11]–[13]. CarSafe [14] utilizes the smartphone's rear-facing camera to detect driver face direction, eye state, following distance, and lane trajectory. However, smartphone cameras are sensitive to ambient light intensity. Moreover, they require users to have the line of sight to the cameras without covering their faces (e.g., with a mask), which limits their scenarios, especially in the current COVID-19 pandemic. D^3-Guard [15] uses acoustic signals sent by the smartphone audio sensor to detect nodding, operating steering wheel, and yawning of the driver. However, its usefulness is limited to the scenario that no other passengers are in the vehicle. The smartwatches and fitness bands also exhibit the trend to automate driver drowsiness detection. The system proposed in [16], [17] monitor the drowsiness indicators, including steering motions and heart rate variability features using the PPG sensor and accelerometer integrated into wrist-worn smartwatches. However, changes in different indicators are generated gradually and not synchronously. These studies do not take conflicts between indicators into account, which leads to insufficient accuracy in real-world situations.

The above limitations motivate us to design and implement a low-cost, convenient, and accurate driver drowsiness detection system, called FDWatch. It is not restricted to any ambient light intensity and works well when other passengers are

Y. Cao, F. Li, X. Liu and S. Yang are with School of Computer Science, Beijing Institute of Technology, Beijing, 100089, R.P.China.
E-mail: {yetongcao, fli, liuxiaochen, S.Yang}@bit.edu.cn
Y. Wang is with Department of Computer and Information Sciences, Temple University, Philadelphia, Pennsylvania 19122, USA.
E-mail: wangyu@temple.edu
F. Li is the corresponding author.

Manuscript received MONTH DAY, YEAR; revised MONTH DAY, YEAR.

in the vehicle. The design is inspired by our observations, including (1) yawning lead to light reflection intensity changes of the blood, which can be profiled using the PPG sensor available in commercial wrist-worn devices; (2) steering wheel operation patterns such as turning angles and speeds lead to unique wrist motion patterns, which can be characterized by analyzing data of the motion sensor available in commercial wrist-worn devices. Moreover, many other drowsiness-related indicators such as heart rate variability features and circadian rhythm can be extracted by sensors integrated into wrist-worn devices. Therefore, FDWatch detects yawning, unsafe steering wheel operations, heart rate variability features, and circadian rhythm with a wrist-worn smartwatch. Then, FDWatch jointly considers the above factors and determines whether the driver is under drowsiness.

We encountered the following three key technical challenges in designing FDWatch: 1) **How to distinguish yawning actions from other interference actions based on PPG data?** Based on the fact that yawning brings in more oxygen during inhale and removes more carbon dioxide during exhale than usual breath, we associate the two-wavelength PPG with blood oxygen content to detect yawning actions. However, many factors such as hand movements, deep breaths, and blood pressure dynamics could interfere with the PPG data collected at the wrist. To address these, we first extract blood oxygen saturation descriptors from the two-wavelength PPG signal to segment the possible yawning frames. Then, we design a transformation method that effectively improves the discrimination of these actions. Finally, we select actions that satisfies certain conditions as yawning actions. 2) **How to infer fine-grained steering wheel operation patterns using a single wrist-worn device?** Fine-grained indicators like steering wheel turning angle and speed provide more information for drowsiness detection in real driving environments. However, wrist-worn devices have limited sensing modalities and only provide partial knowledge of hand motions. It is challenging to monitor fine-grained steering wheel operations using only the wrist-worn device. To facilitate single-sensor-based detection, we leverage the great power of machine learning. We select 19 motion signal features by feature importance feedback from a random forest classifier. Moreover, we carefully compare several highly used machine learning methods. The random forest classifier outperforms other classifiers and is adopted to detect steering wheel turning angles and speeds. 3) **How to achieve accurate driver drowsiness detection with the extracted indicators?** FDWatch detects drowsiness by jointly consider multiple indicators that change gradually and not synchronously. However, the relationship between these indicators is unclear and rarely studied, which brings many difficulties to handle the contradictions when they imply different mental states of drivers. Therefore, we firstly employ a backpropagation (BP) neural network to estimate the reliability of each indicator. Then, we calculate the similarity distance between indicators and use Dempster-Shafer evidence theory (DST) to fuse these indicators and finally determine whether a driver is under drowsiness.

Extensive experiments with ten participants show that FDWatch can realize reliable drowsiness detection in real driving environments. Our contributions are summarized as follows:

- We present the design of FDWatch, a wrist-worn device-based driver drowsiness detection system. To the best of our knowledge, our work is the first attempt to enable accurate, low-cost, unconstrained, and convenient driver drowsiness assessment from wearables. It addresses the asynchrony change problem of multiple drowsiness-related indicators.
- We design several novel algorithms to extract drowsiness-related indicators. Specifically, we detect yawning actions by associating the two-wavelength PPG signal with blood oxygen content. Besides, we design a machine learning method to detect steering wheel turning angles and speeds using motion sensor data collected at the wrist. Moreover, we explore several heart rate variability features and the alertness level related to circadian rhythm.
- We design an efficient method to overcome the contradictions of multiple indicators. We first employ the BP neural network to estimate the reliability of each indicator. Then, we use a similarity distance-based conflict resolve method together with DST to handle evidence conflicts.
- We thoroughly evaluate FDWatch's performance with real-driving data. The results show that FDWatch can effectively detect drowsy driving with a missing alarm rate lower than 3.57% and a false alarm rate lower than 3.68%.

The rest of the paper is organized as follows. Section II surveys the related work. Section III introduces the preliminary of drowsy driving and system overview. Section IV presents the extraction method of various indicators, followed by Section V explaining the necessity of information fusion and how to implement it. Then Section VI introduces the implementation and evaluation results. Finally, Section VII summarizes the paper.

II. Related Work

Drowsy driving is one of the most significant threats to driving safety. Therefore, drowsy driving detection has drawn significant attention in recent years. The systems proposed in [3]–[5] monitor physiological signs (i.e., electroencephalogram (EEG), electrocardiogram (ECG), and electrooculogram (EOG)) changes to infer driver drowsiness. However, these monitors often cause inconvenience and discomfort as they are bulky and use multiple body-attached sensors. Besides, Baur *et al.* [18] propose an automatic sleepiness classification scheme using features from electroencephalogram (EEG), eye blink duration, and driving conditions. Karuppusamy *et al.* [19] process image data, ECG data, gyroscope data from the head, and use a neural network to detect driver drowsiness. Despite being uncomfortable and inconvenient to use, these methods require multiple additional hardware with high costs and a complicated setup, which makes them difficult to popularize.

Recently, emerging smartphone-based sensing provides some promising solutions. Some prior contributions [11]–[13] have been made to assess drowsiness-related facial expression features. CarSafe [14] tracks eye state, following distance, and lane trajectories using the smartphone's rear-facing camera

to alert drowsy drivers. However, smartphone cameras are sensitive to ambient light intensity, which restricts their use to limited scenarios. Moreover, these systems require users to have the line of sight to the cameras and incur privacy leakage issues. D^3-Guard [15], [20] detects nodding, yawning, and abnormal turning using acoustic signals sent by the smartphone. However, it is built on the assumption that there are no other passengers in the vehicle that could interfere with the acoustic signal.

The popularity of wearables has made them promising tools for preventing car accidents. Therefore, there have been active research efforts in detecting driver drowsiness using wearables. Zhang *et al.* [17] detect driver drowsiness by extracting steering wheel usage and monitoring the driver's heart rate. Similarly, Lee *et al.* [21] estimate driver vigilance level by deriving steering angles {-90, 0, 90} and respiration rate. However, the heart rate and respiration rate are susceptible to many factors other than drowsiness. Thus they only provide partial information about the driver's status. Moreover, these approaches regard all hand movements as steering wheel operations, which makes the detection of drowsiness often inaccurate in practice.

Compared to the above approaches, FDWatch has the following advantages. FDWatch makes good use of sensors integrated into wrist-worn devices with low computational cost and does not require additional devices or modifications. Users can assess their drowsiness by wearing the smartwatch, which is comfortable and convenient. Besides, it is not affected by ambient light conditions or other passengers. Moreover, it effectively addresses the conflict of multiple indicators (i.e., some indicators imply users are drowsy while others imply users are not drowsy). Although the drowsiness-induced indicator changes do not occur synchronously, the system can accurately detect driver drowsiness.

III. Preliminaries and System Overview

In this section, we first present the basics of driver drowsiness and associated signs. Then we introduce the overview of FDWatch.

A. Driver Drowsiness and Associated Signs

A drowsy driver is in the intermediate state between wakefulness and sleep, usually accompanied by performance and psychophysiological changes. The drivers may experience impairment of visual perception, inability to maintain visually focused attention, impairment of higher cognitive functions, and of volition [22]. Many countermeasures to derive driver drowsiness have been proposed. In this work, we consider the combination of physiological signs, vehicle operation dynamics, and circadian rhythm as follows:

Yawning Event: Yawning event is a psychically or reflexly excited inspiration, which consists of three distinct phases: (1) inhale deeply with mouth open slowly and widely; (2) a short period of apnea; and (3) exhale with a quickly closed mouth [23]. We take the frequency of yawning and the rate of yawning interval shortening as drowsiness indicators. As drowsiness deepens, the frequency of yawnings increases.

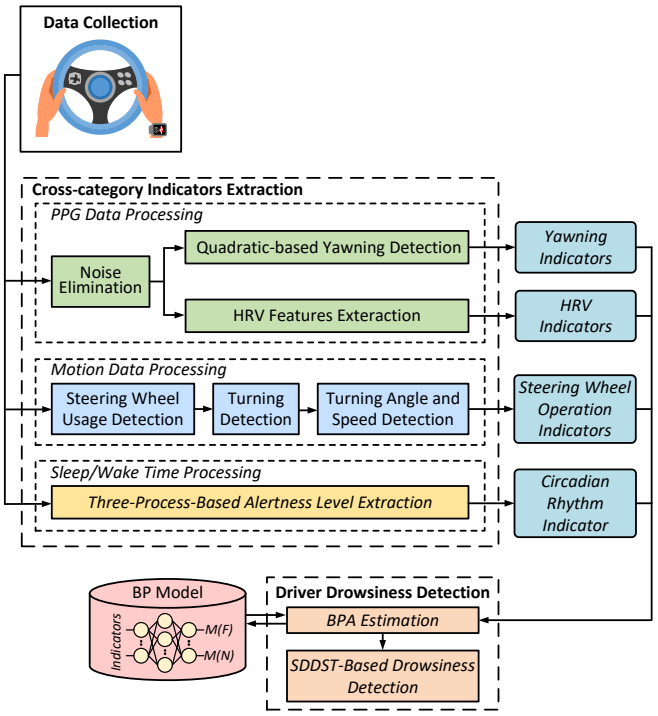


Fig. 1. Framework of FDWatch.

Heart Rate Variability (HRV): HRV is the physiological phenomenon of the beat-to-beat temporal variation of the heart. HRV features contain information about the autonomic nervous system, which conveys information about the alertness of a person. Under the influence of drowsiness, the heart slows down and beats less regularly. That is, the heart rate decreases and other HRV features increase [24].

Steering Wheel Operation: The transition from the normal state to the drowsiness state is always accompanied by decreased attention and reduced ability to operate vehicles. Even on a straight road, alter drivers will adjust the steering wheel steadily, slowly, and frequently at small angles. However, drowsy drivers suffer from increased reaction times and will suddenly realize the direction deviation and quickly turn the steering wheel at a large angle [15]. We take the steering turning angle and steering turning speed as steering wheel operation indicators. The increase in turning angle and speed are signs of drowsiness.

Circadian Rhythm: The regulation of sleep is processed by the homeostatic physiology of the circadian rhythm, i.e., the sleep/wake cycle. This cycle affects human vigilance and drowsiness [25]. If a driver has insufficient sleep before getting behind the wheel, he or she will experience stronger drowsiness than usual. We derive alertness level from circadian rhythm as a sign of drowsiness.

B. Overview

The basic idea of FDWatch is to extract four types of drowsiness indicators from the driver's wrist motion data and PPG data. Then utilize an information fusion method to estimate the driver drowsiness.

Fig. 1 illustrates the framework of FDWatch. In **Data Collection**, the accelerometer, gyroscope, and PPG sensor of the wearable device continuously collect the driver's motion data and PPG data at a low cost. **Cross-category Indicators Extraction** and **Driver Drowsiness Detection** are the key components of FDWatch. In **Cross-category Indicators Extraction**, *PPG Data Processing* first eliminates noise in the PPG signals. Then it detects yawning events based on the unique pattern of blood oxygen saturation descriptors deduced from PPG signals and extracts HRV features from each cardiac cycle to obtain yawning indicator and HRV indicator. *Motion Data Processing* first conducts driving scenario detection to distinguish steering wheel operations from other daily activities. Then, it segments steering wheel turning actions using turning detection. By extracting several significant features and developing an RF-based classifier, FDWatch continuously detects the steering wheel turning angles and speeds. *Sleep/Wake Time Processing* reads the driver's sleep/wake cycle from the wearable device and models the driver alertness level using a three process method.

All indicators serve as pieces of evidence for estimating driver drowsiness. In **Driver Drowsiness Detection**, FDWatch performs information fusion based on Dempster-Shafer evidence theory (DST) and Similarity Distance-based Dempster-Shafer evidence theory (SDDST) based method to detect driver drowsiness. To obtain the basic probability assignment (BPA) of each evidence and resolve the contradiction between evidence, we utilize the power of the backpropagation (BP) neural network. Even if some of the indicators contain errors, the use of multiple indicators and the information fusion method can guarantee the detection of drowsy driving.

IV. Cross-Category Indicators Extraction

A. PPG Data Processing

1) Noise Elimination: We first filter the PPG signals in [0.5 *Hz*, 5 *Hz*], which corresponds to the physiological range of heartbeat signals. Then, we segment heartbeat in the PPG data based on the fact that the boundary point of each heartbeat manifests in the form of the maximum peak of PPG waveform. Specifically, we make use of the fact that the first derivatives of boundary points are "cross-zero" points (first derivatives cross from positive to negative). Furthermore, the hand motion (e.g., fast steering wheel operations) could distort the PPG signals, which is not easy to remove due to its characteristics of frequency overlap. To obtain a reliable PPG signal, we use the percentage change [24] to eliminate outlier heartbeat. An outlier heartbeat is defined as having an interbeat interval deviating more than 30% from the mean of the four previous accepted intervals.

2) Quadratic-based Yawning Detection: Yawning brings in more oxygen during inhale and removes more carbon dioxide during exhale than the usual breath. When a subject inhales oxygen, the hemoglobin carrying oxygen (oxyhemoglobin) causes the blood to be bright red. Different contents of oxyhemoglobin cause light absorption changes of blood that can be captured with the PPG sensor. Therefore, we continuously profile the changing pattern of blood light absorption to detect yawning actions.

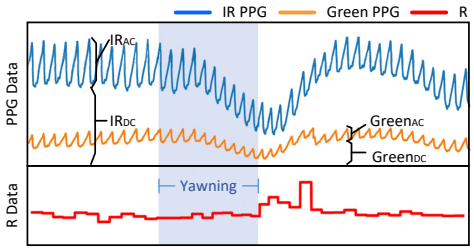


Fig. 2. Examples of Green PPG signal, IR PPG signal and the blood oxygen saturation descriptor R during yawning.

PPG is a simple and inexpensive optical technique that detects the blood volume by illuminating a skin/tissue and measuring light absorption. PPG signal has two main components, including alternating component (AC) and direct current (DC) component. AC is the pulsatile component of the synchronous changes in the blood volume with each heartbeat. The AC component is superimposed onto a large DC component related to the tissues and the average blood volume [26]. A typical PPG sensor employs green and infrared light sources and photodiode chips to detect volumetric changes in blood in the peripheral circulation. Different contents of oxyhemoglobin in the blood show different absorption capacities for the green light and infrared light [27]. Such absorption capacity changes affect the amplitude of AC and DC components. Therefore, we use ratio-of-ratios R to profile the absorption changes by the following equation:

$$R = \frac{Green_{AC}/Green_{DC}}{IR_{AC}/IR_{DC}}, \tag{1}$$

where $Green_{AC}$, $Green_{DC}$, IR_{AC}, and IR_{DC} denote the AC components and DC components of Green PPG and IR PPG, respectively. Previous works [28], [29] have validate the close correlation between blood oxygen content and the derived R. Fig. 2 shows the collected PPG signals during yawning and R derived from yawning PPG signals. It can be seen that R changes immediately after yawning, suggesting that it is feasible to detect yawning based on R results.

However, we find that PPG signals are also sensitive to many other factors such as hand movement, breathing variations, deep breathing, coughing, talking, and blood pressure dynamics. Coughing could cause a fast, small change in R, while talking and breath-holding can cause slow and small changes in R. These activities can be filtered using several thresholds. Moreover, blood pressure dynamics, especially the low arterial pressure, could affect peripheral arterial pulsations. Despite the PPG amplitude decreases with poor arterial pulsations, R is stable within normal daily blood pressure fluctuation [30]. However, deep breathing and slow steering rotating can generate R that shares similar changing duration and fluctuation patterns with yawning. Therefore, we focus on distinguish yawning from them.

Our goal is to find a lightweight transform function that can transform the R values of yawning and other interference actions into significantly different values. We adopt the least square regression method to extensively explore plenty of formulas. Because the quadratic equation of R is computationally efficient and can provide better-discriminating capability

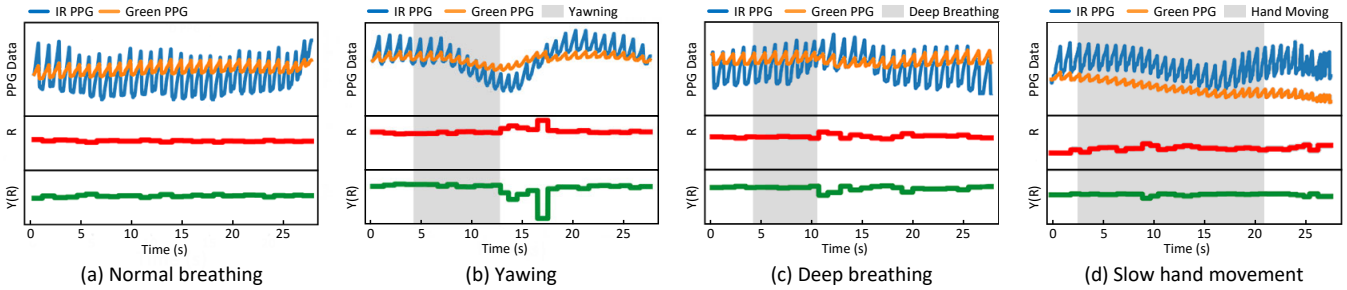


Fig. 3. PPG signals, R and $Y(R)$ during normal breathing, yawning, deep breathing, and slowly operating the steering wheel.

between yawning and other actions, we use the quadratic equation of R as the transform function. The transform function is initially defined as:

$$Y(R) = \alpha R^2 + \beta R + \gamma, \tag{2}$$

where α, β, and γ are sensor-specific and should be learned through maximizing the average difference between $Y(R)$ of yawning and other actions. In our case, we set $Y(R) = -45.060R^2 + 30.354R - 4$ based on our experimental study.

Fig. 3 (a-d) shows the results of $Y(R)$ and R during normal breathing, yawning, deep breathing, and slowly operating the steering wheel. Note that infrared (IR) PPG and green PPG signals are aligned to provide a clear display. R increases with deep breathing, operating the steering wheel, and yawning. In contrast, $Y(R)$ shows significant changes only in yawning, and this change lasts for multiple cardiac cycles regardless of other interference.

Fig. 3 (b) shows a typical $Y(R)$ sample during yawning. $Y(R)$ decreases first, then reduces substantially, and then goes back to its original value. After analyzing the yawning PPG data of ten volunteers, we find that this general threshold can be used to detect yawnings. If $Y(R)$ deviating three times from the standard deviation of 4 previous epochs, FDWatch starts to check whether its fluctuation patterns satisfy the criteria. After a yawning event is detected, we term the time when $Y(R)$ begins to change as the time of the yawning event. Then, the average yawning frequency and the average difference between adjacent yawning intervals (rate of yawning interval shortening) are used as indicators to detect driver drowsiness. Specifically, it receives an accuracy of 97.53%, a recall rate of 96.40%, a precision of 89.59%, and an F1 score of 95.87%, which are described in Section VI.

3) HRV Features Extraction: The peak-valley data of PPG signals validate the activities of the autonomic nervous system, which makes the waveform of PPG signal worth studying for drowsiness detection. We extracted HRV features from PPG signals as HRV indicators. Researches have been tracking HRV features for decades because they can be good indicators of several health-related issues. HRV is concerned with the analysis of the intervals between heartbeats. We extensively survey features and manually select those that are robust against differences in emotion and stress. Table I summarizes the selected features.

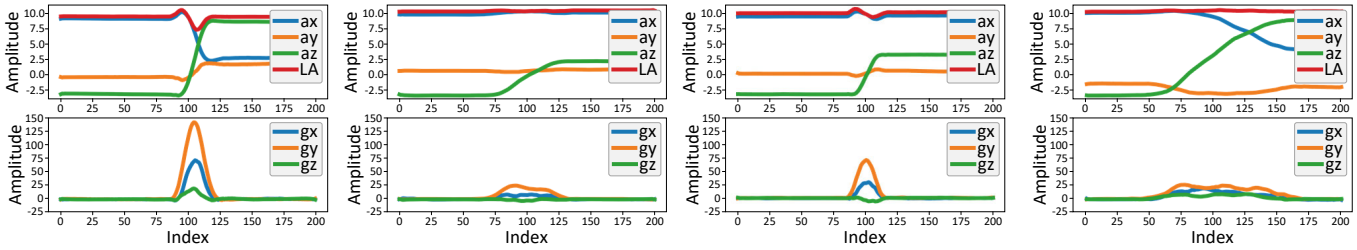
TABLE I
SELECTED HRV FEATURES.

Category	Features
Time Domain	3min respiratory rate minimum, 3min heart rate maximum, standard deviation of all IBIs, number of adjacent IBI with absolute difference more than 50 ms, average of the absolute difference of adjacent IBIs, median absolute deviation, mean of the standard deviations of IBIs in all 5-minute segments, the average of the absolute difference of every two normalized IBIs
Frequency Domain	the power in the low frequency band (LF) the power in the high frequency band (HF) the ratio of power in the LF to HF bands the total power in the band 0-0.4 Hz normalized LF median 3min normalized LF median

B. Motion Data Processing

We estimate the speed and angle of the steering wheel rotations and steering wheel usage frequency as indicators to monitor driver drowsiness. To avoid running expensive estimation algorithms in other scenarios (e.g., eating, cooking, exercising), we first distinguish the driving scenario from other non-driving scenarios. The driving scenario always starts with the driver drops his/her arms on the steering wheel and ends with the driver lifts his/her hands from the steering wheel. We segment the steering wheel operating period from dropping the hands to lifting the hands. For each period, we profile the motion sensor reading collected by wrist-worn wearables and estimate steering wheel operation indicators. This solution, only relying on motion sensors (i.e., a triaxial accelerometer and a triaxial gyroscope) data collected at the wrist, provides cost-effective and energy-efficient service.

1) Steering Wheel Usage Detection: When the driver drops or lifts his/her arms, the accelerations at wrists present unique patterns [31]. When the hand is dropped or lifted, there will be a speed-up process and a slowdown process. We compare the Linear Acceleration ($LA = \sqrt{(a_x)^2 + (a_y)^2 + (a_z)^2}$) with the gravitational acceleration (g), where a_x, a_y, a_z are the accelerations in the X, Y, Z axis, respectively. Hand dropping has the same direction as gravity, causing the LA to be higher than the gravitational acceleration at first and then lower, and hand lifting is the opposite. Therefore, hand dropping is defined as LA below g at first and then over g. Hand lifting is defined as LA over g firstly, then less than g. Dropping behavior triggers the detection of wrist gestures.



(a) Fast large angle rotation (b) Slow large angle rotation (c) Fast small angle rotation (d) Slow small angle rotation

Fig. 4. The accelerations and angular velocities at different rotation speeds and angles. a_x, a_y, a_z denote the acceleration of the X, Y, Z axis, LA denotes the combined acceleration, and g_x, g_y, g_z denote the angular velocity of the X, Y, Z axis, respectively. Data are filtered into $0.1Hz$-$2Hz$ to remove noise caused by vehicle vibrations.

The false pairing of dropping and lifting will lead to incorrect detection of steering wheel usage. To monitor the start of each possible steering wheel usage, we detect each gesture after the hand lifting and dropping. The duration of holding the steering wheel is limited to greater than I_{hold} to prevent false recognition. In Section VI, we discuss the effect of I_{hold} on driver drowsiness detection accuracy.

2) Turn Detection: When drivers make a turn, their hands will rotate clockwise or counterclockwise, causing changes in acceleration data and angular velocity data. We apply the sliding window method to analyze these data and learn more information from them. The left and right local minimum of the collected data is taken as the candidate beginning and the ending of a turn [31]. After segment a steering turning event, there is still a gap to obtain the turning angle and speed. The wrist-worn device's coordinate system rotates as the driver turning the steering wheel. As a result, the angular velocity data we collected is not directly proportional to the angular velocity of steering wheel rotations.

Depending on the turning speed and angle, we characterize the steering wheel rotations into fast-large-angle rotations, slow-large-angle rotations, fast-small-angle rotations, and slow-small-angle rotations. After consulting with 20 drivers, we define a turn higher than 30° as a large angle and a turn higher than 20° per second as a fast rotation. Fig. 4 (a-d) illustrate the angular velocity data and acceleration data in four cases, respectively. The driving data are collected at [20 *km/h*, 40 *km/h*]. The data are filtered in [0.1 *Hz*, 2 *Hz*] range, where most of the noise caused by vehicle vibration is removed. We can observe that the angle and speed of steering wheel rotations are related to the geometric features of data collected by the accelerometer and gyroscope. The results motivate us to use a statistic-based method to distinguish these four conditions.

3) Turning Angle and Speed Detection: We firstly use a feature extraction tool tsfresh [32] to extract several candidate features. Then we apply a Random Forest (RF) classifier to rank these features for steering wheel turning angle and speed detection. To overcome the impact of vehicle speed variations, we also rank candidate features for detecting the vehicle speed and remove those that contribute most to vehicle speed detection. Finally, we select 19 best-performing features to describe the shape of the signal, including *standard deviation*, *maximum*, *kurtosis*, *autocorrelation*, *length*, *time reversal asymmetry statistic*, *absolute energy*, *time series complexity*, *the percentage of values that are lower than* $0.5*$*maximum vale*, *median*, *skewness*, *mean*, *the percentage of non-unique data points*, *the sum of all data points that are present in the time series more than once*, *the number of values that are higher than the mean value*, *the spectral centroid of the absolute fourier transform spectrum*, *the length of the longest consecutive subsequence that is bigger than the mean value*, *sum*, and *the number of values that are lower than the mean value*. These features contain hand motion information such as displacement, velocity, acceleration, etc. They can provide wonderful discriminating capability between different turning angles and speeds.

Then, we compared several commonly used classifiers. RF-based classifier outperforms all the other tested classifiers and achieves the highest accuracy of 94.30%. Thus, we set RF to be our default solution. Section VI shows the details of the comparison.

C. Sleep/Wake Time Processing

Human vigilance is regulated by their circadian rhythm. Insufficient sleep could make the driver sleepy. According to [33], an estimation of $200,000$ crashes each year occur early in the morning (between 2:00 AM and 6:00 AM) when drivers fail in fighting off the urge to sleep. Based on the close correlation between drowsiness and circadian rhythm, we use the sleep and awake time (commonly available in commercial smartwatches and fitness trackers) to help extract drowsiness indicators.

1) Three-Process-Based Alertness Level Extraction: We adopt the three-process model [34] to predict the alertness of a driver. The three-process model includes three parameters (shown in Fig. 5): (1) Process C represents circadian influences and has a sinusoidal form with an afternoon peak. (2) Process S is an exponential function of time since awakening. It is high on awakening, drops rapidly initially, and gradually approaches a lower asymptote. At sleep onset, S is revised (termed as S') and increases rapidly, but subsequently levels off toward an upper asymptote. (3) Process W describes sleep inertia, but it is not part of the present modeling since no recent awakenings are involved. $S = (Sa - L)e^{-0.0353ta} + L$, where Sa is the value of S at awaking, L is the lower asymptote (2.4), ta is the time since awakening. $S' = U - (U - Sr)e^{-0.0381tf}$,

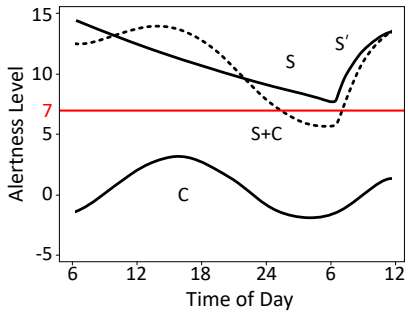


Fig. 5. The components of the three-process model [34].

where Sr is the value of S at retiring, U is the upper asymptote (14.3), tf is the time since falling asleep. $C = 2.5cos(td - p)\pi/12$, where p represents acrophase, and td is the time of day. The predicted vigilance is denoted as $C + S$, which is the arithmetic sum of the C and S (W presently excluded). Alertness levels on the three process method range from -5 to 15. With a score of 7 and higher generally considered alert (in the initial study establishing the three process model, normal controls had an average score above 7).

V. Driver Drowsiness Detection

A. Necessity of Information Fusion

FDWatch is a hybrid system that combines various drowsiness indicators, including circadian rhythm provided by wrist-worn devices and three categories of drowsiness indicators extracted from PPG signals, acceleration data, and angular velocity data. The features we use are summarized as follows.

- *Yawning Indicators (Behavior-Based)*: the average frequency of yawning event and the rate of yawning interval shortening.
- *HRV Indicators (Physiological-Based)*: HRV features such as heart rate, deviation, approximate entropy, etc.
- *Steering Wheel Operation Indicators (Vehicle-Based)*: the frequencies of steering wheel operations at different angles and speeds.
- *Circadian Rhythm indicator* (if not available, it can be ignored): alertness level derived from the time of a day, the time awake, and prior sleep.

The transformation from a normal state to drowsiness is always accompanied by the changes in these indicators. However, these changes are gradually generated and asynchronously. For example, when a driver operates the steering wheel at a low frequency, he/she is considered to be driving under drowsiness from the vehicle-based perspective. However, the driver might not have yawned. He/she is considered driving normally from a behavior-based perspective. Our basic idea is to use various indicators to make up for the limitation caused by using single indicators. To this end, we adopt the information fusion method to resolve evidence conflicts and provide reliable service when some error occurs.

To handle the incomplete knowledge of each indicator, we adopt the DST to jointly consider various indicators and detect driver drowsiness. It has advantages in expressing uncertainty, measurement, and combination [35]. DST allows us to assign a probability mass to various possible conclusions without more specific information. DST has been widely used in multi-sensor and information fusion-based systems to expand the range in space and time for specific sensing tasks, such as fire detection [36] and object recognition [37].

B. Fundamental of DST

The reliability-oriented approach to DST is based on a scenario that contains the system with all hypotheses, pieces of evidence, and data sources. The hypotheses account for all the possible states. Evidence presents symptoms or events that may occur within the system. Each piece of evidence is lead to a unique hypothesis or a set of hypotheses. Data sources are experts that provide information.

The definitions of the DST are described as follows:

- BPA: The BPA function is a mass function (denoted as M) to describe the degree of belief in evidence in the framework Θ. $M(\cdot)$ satisfies $M(\phi) = 0$, $\sum_i M(A_i) = 1$, $A_i \subseteq \Theta$ under the map $2^\Theta \rightarrow [0, 1]$.
- Combination Rules: The new probability $M(A)$ can be derived by combining all pieces of evidence:

$$M(A) = \frac{\sum_{A_1 \bigcap ... \bigcap A_n = A} M_1(A_1)M_2(A_2)...M_n(A_n)}{\sum_{A_1 \bigcap ... \bigcap A_n \neq \phi} M_1(A_1)M_2(A_2)...M_n(A_n)}.$$

- Evidential Interval: The interval $[Bel(A), Pls(A)]$ describes the true range of belief. The lower bound is called Bel, which is the belief function denoting the total belief of a set and all its subsets, $Bel : 2^\Theta \rightarrow [0, 1]$ and $Bel(A) = \sum_{B \subseteq A} M(B)$. The upper bound is called Pls, which is the plausibility function indicating the degree to which the evidence obtained cannot be rejected, $Pls(A) = 1 - Bel(\bar{A})$.
- Final Decision: The proposition with the maximum belief function or plausibility function is the result of the combination.

When detecting driving drowsiness, we consider two hypotheses, the driver is under drowsiness (F), and the driver is driving normally (N). $2^\Theta = \{\phi, \{F\}, \{N\}, \{F, N\}\}$. Each indicator provides a piece of evidence. Depending on the content of the evidence, each hypothesis will be assigned a BPA mass. For example, evidence of the increased frequency of yawning and increased rate of shortening of yawning intervals suggests that BPA for drowsiness is higher than 0.5 ($0.5 < M_1(F) \leqslant 1$) and BPA for normal is less than 0.5 ($0(0 < M_1(N)) \leqslant 0.5$); evidence of alertness level derived from circadian rhythm under 7 suggests that BPA for drowsiness is higher than 0.5 ($0.5 < M_2(F) < 1$) and BPA for normal is less than 0.5 ($0 < (M_2(N)) \leqslant 0.5$); Similarly, steering wheel operation and HRV indicators estimate BPA for drowsiness and normal state $M_3(F)$,$M_3(N)$,$M_4(F)$,$M_4(N)$. We should follow the DST combination rules and get new probability mass for the two hypotheses ($M(F)$ and $M(N)$) and their evidential interval ($[Bel(F), Pls(F)]$ and $[Bel(N), Pls(N)]$). If drowsiness state F has a higher belief function or plausibility function, we will alert the driver of the risk of falling asleep.

C. BPA Estimation

Traditionally, BPA could be assessed by experts through experience, but in our system, BPA is obtained automatically by the power of BP neural networks. The BP-based network is a unilateral spread network that has gotten a wide range of applications in the field of information fusion. We design three BP networks for drowsiness indicators, including yawning indicators, HRV indicators, and steering wheel operation indicators. Each BP network consists of an input layer, a hidden layer, and an output layer. The number of nodes in the input layer of each network is the same as the number of indicators of each indicator group. Specifically, to process yawning indicators, we build two nodes in the input layer, which input yawning frequency, the rate of yawning interval shortening, respectively. Because we extract 14 HRV features from PPG signals, we build 14 input nodes in the BP network for processing HRV indicators. As for processing steering wheel operation indicators, we set four nodes in the input layer, which represents the frequency of four types of steering wheel operations. There are two nodes in the output layer of each network, one representing the probability of drowsiness and the other representing the probability of a normal state. The number of nodes in the hidden layers is determined by the formula: $K = \lceil \sqrt{I + J} \rceil$, where $\lceil \cdot \rceil$ is the ceiling operation and I, J are the number of nodes in the input layer and the output layer.

BP networks obtain BPA, which describes the probability that the implied user state of an indicator is the same as the real user state. We train the BP networks with paired indicators and the ground truth state of users (drowsiness state or normal state). The parameters of BP networks are updated literally until the number of matches between the predicted state and the actual state converges. Since we aim to strike a balance between accuracy and computation efficiency, the input indicators are extracted every 300 s with 270 s overlap, include time averaging.

Unlike these indicators, the alertness levels extracted from circadian rhythm do not adopt the BP neural network to obtain BPA mass. The predicted result is mapped to [0, 1] by a modified sigmoid function to extract the BPA. An alertness level greater than 7 results in a high BPA of the normal state and a low BPA of drowsiness state, and an alertness level less than 7 has the opposite effect.

D. SDDST-Based Drowsiness Detection

Classical DST holds that every piece of evidence has the same importance and does not assign weights to the evidence. This makes DST get wrong results when some evidence contradicts seriously, which we will encounter when using various indicators to detect driver drowsiness. We utilize the similarity distance to handle the conflict between evidence, which is called SDDST. The basic idea is to increase the weight of evidence with a high similarity distance and decrease the weight of evidence with a low similarity distance. Using a simple Euclidean distance to model the distance between BPAs is not appropriate. Instead, we adopt the distance proposed in [38] to give equivalent weight to all the subset of Θ. The distance between two BPAs M_1 and M_2 to be of form:

$$d(M_1, M_2) = \sqrt{\frac{1}{2}(\vec{M_1} - \vec{M_2})^T \underline{D}(\vec{M_1} - \vec{M_2})}, \quad (3)$$

where $\vec{M_1}$ and $\vec{M_2}$ are the BPA vectors with $\vec{M_i} = \{M_i(A_i)\}, A_i \in \Theta$ and $\underline{D}$ is a $2^N \times 2^N$ positively defined matrix whose elements are

$$D(A, B) = \frac{|A \bigcap B|}{|A \bigcup B|}.$$

FDWatch obtains four evidence and their BPA mass. We calculate the distance between them using Equ. 3 and present the distance in a matrix:

$$\begin{bmatrix} 0 & d_{1,2} & d_{1,3} & d_{1,4} \\ d_{2,1} & 0 & d_{2,3} & d_{2,4} \\ d_{3,1} & d_{3,2} & 0 & d_{3,4} \\ d_{4,1} & d_{4,2} & d_{4,3} & 0 \end{bmatrix},$$

where $d_{i,j}$ denotes the distance between BPA i and j of evidence i and j. We model the similarity distance by $S(M_i, M_j)$ which is defined as $S(M_i, M_j) = 1 - d(M_i, M_j)$. Thus the similarity distance matrix can be presented as:

$$\begin{bmatrix} 1 & S(M_1, M_2) & S(M_1, M_3) & S(M_1, M_4) \\ S(M_2, M_1) & 1 & S(M_2, M_3) & S(M_2, M_2) \\ S(M_3, M_1) & S(M_3, M_2) & 1 & S(M_3, M_4) \\ S(M_4, M_1) & S(M_4, M_2) & S(M_4, M_3) & 1 \end{bmatrix}.$$

We use the similarity distance between M_i and the other BPA to model the support level of M_i: $Sup(M_i) = \sum_{j=i, j \neq i}^{4} S(M_i, M_j)$.

When there is a significant conflict between the evidence i and other evidence, M_i derived from the evidence i will have a lower similarity distance with other BPA, thus will have a lower support level. A BPA with a high support level also means that the conflict between this evidence and other evidence is relatively small. We assign the normalized result of support level to each evidence as weight, which is in the form of:

$$Wei(M_i) = \frac{Sup(M_i)}{\sum Sup(M_i)}. \quad (4)$$

Compared with the classical DST combination rules, the method based on similarity distance expedites the convergence speed as well as improves the accuracy of drowsy driving detection.

VI. System Implementation and Evaluation

In this section, we first study the effectiveness of key algorithms. Then we evaluate the overall performance of FDwatch. Moreover, we study the impact of many factors.

A. System Implementation

FDWatch leverages a wrist-worn device to achieve reliable driver drowsiness detection. We implement our system on a prototype equipped with an MPU-6050 6-axis motion sensor (including a 3-axis accelerometer and a 3-axis gyroscope) and a MAX30105 PPG sensor, which is an alternative to a smartwatch. The prototype is paired with a laptop via WiFi. The

prototype continuously collects the acceleration data, angular velocity data, and PPG data at the wrist with the sampling rate of 100 Hz. The 100 Hz sampling rate allows FDWatch to be implemented on most commercial devices. Fig. 6 shows the prototype and its three-axis coordinate system. When worn on the wrist, the prototype's Y-axis is parallel to the forearm. We develop the random forest classifier and BP neural network use Sklearn. Specifically, we build a personalized model for each participant. Besides, we use 75% data for training and the rest 25% data for testing.

B. Data Collection

We recruit 10 participants (including 6 males and 4 females, aged 23 to 50, with 6 months to 20 years driving experience) to collect driving data. The study is conducted with the approval of our institute's Institutional Review Board. All the participants were in good health and did not drink or take drugs within 24 hours before the collection.

There is no state-of-art dataset that simultaneously monitors PPG and motion signals. Therefore, we collect the dataset to implement and evaluate FDWatch. We ask participants to wear the FDWatch prototype during driving. To accommodate sensor position variations, we suggest participants wear the prototype according to their habits. Fig. 7 shows the experimental setup. The tests are conducted in real driving environments. We ask each participant to drive on any empty road 1 to 3 times under the speed limitation of 40 *km/h*. Each trip lasts 40 minutes to 80 minutes. We assign a co-pilot to sit on the passenger side on each trip to ensure road safety and record road conditions. In order to get complete yawning and steering, we segment sensory readings with a length of 30 seconds. Overall, we collect over $6,100$ pieces of 30-second driving records from 26 trips, including 308 drowsy driving records. Circadian rhythm, such as the time of day, the time awake, and prior sleep can be obtained from commercial wearables. While in the experimental stage, we ask the participants to keep a sleep/weak diary to collect such information.

Meanwhile, we place a motion sensor on the steering wheel to obtain the ground truth for steer wheel turning angles and speeds. Besides, we set a camera to record participants' facial expressions during driving, which are then rated by volunteers to obtain ground truth for driver drowsiness. To guarantee the ground truth of driver drowsiness, we recruit three volunteers to rate each 30-second video segment separately and repeatedly. Specifically, each video record is rated as drowsiness stated or normal state. Every volunteer rate the videos three times to ensure the rate of quality. The second rate session was given right after the first session, and the third rate session was given one week after the first and second rate sessions. Video records were rated in chronological order for the first and third sessions, while video records were given in random order for the second session. When the rate results from all volunteers are the same, the 30-second driving data is labeled. Otherwise, the video will be re-rated until all volunteers agree on the final result.

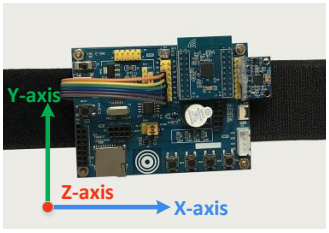


Fig. 6. Prototype.

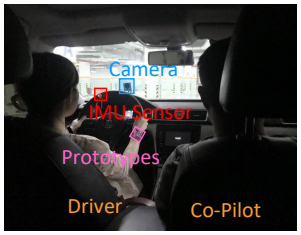


Fig. 7. Experiment setup.

C. Evaluation Methodology

To evaluate the performance of indicators extraction methods, we use the following metrics:

- *Accuracy*: The ratio of samples that are correctly classified among all the samples.
- *Recall*: The ratio of samples that are correctly recognized as label A among all the samples with label A.
- *Precision*: The ratio of samples that are correctly recognized as label A among samples classified as label A.
- *F1 Score*: The weighted average of recall and precision, F1 score= $2(recall * precision)/(recall + precision)$.

When evaluating the performance of drowsy driving detection, the above metrics can be misleading measures for the imbalanced data. Hence, we evaluate FDWatch utilize the following methods:

- *False Alarm Rate*: The ratio of samples that are mistaken as drowsiness state among all the normal state samples.
- *Missing Alarm Rate*: The ratio of samples that are not classified as drowsiness in the samples that truly belong to drowsiness.

D. Key Algorithm Study

1) Effectiveness of Yawning Detection: To evaluate the effectiveness of yawning detection, we ask ten participants to collect PPG data and perform yawning and other actions (including deep breathing, slowly operating the steering wheel, normal breathing, coughing, and talking). In total, the dataset contains $1,500$ actions records, including 250 for each action. At the same time, we record their facial expressions with a camera as ground truth. Of the 250 yawning, 241 are correctly identified, and the other 9 are misidentified as other actions. Besides, of the $1,250$ other actions, $1,222$ are correctly identified, and only 28 are misidentified as yawning. Overall, yawning detection achieves an accuracy of 97.53%, a recall rate of 96.40%, a precision of 89.59%, and an F1 score of 95.87%. The result demonstrates the proposed method is effective, which provides the basis of FDWatch.

2) Effectiveness of Steering Wheel Operation Indicators Extraction: Steering Wheel Usage Detection: The data used to evaluate the performance of steering wheel usage detection comes from two datasets: (1) the collected driving data as described in Section VI-B, which includes 500 steering wheel-holding periods; (2) a new self-collected dataset that includes $3,500$ samples of 7 gestures, such as making phone calls, holding the bottle and adjusting the seat. We conduct experiments with I_{hold} ranging from 0 s to 5 s. Fig. 8 shows the experiment

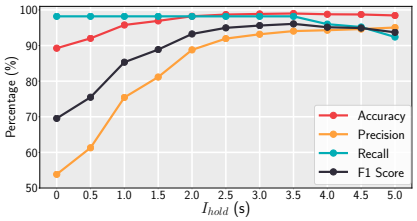


Fig. 8. Performance of the detection of steering wheel usage with different values of the time threshold I_{hold}.

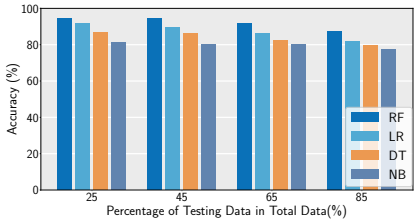


Fig. 9. Comparison of recognition performance of 4 classifiers under different percentage of testing data in total data.

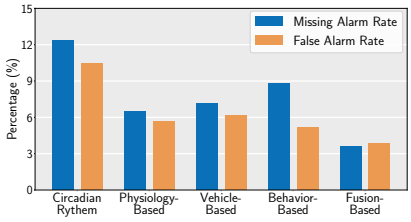


Fig. 10. Performance of drowsinessd driving detection based on various indicators and FDWatch's information fusion method.

results under different values of I_{hold}. When I_{hold} is set to 0 s, many hand motions are misidentified as operating the steering wheel, resulting in 53.84% precision. With the increase of I_{hold}, recognition performance is improved. When I_{hold} is 3.5 s, the precision, recall, and F1 score achieve 94.06%, 98.20%, 96.08%, respectively. When I_{hold} exceeds 3.5 s, the performance of FDWatch decreases as I_{hold} continues to grow because some steering wheel usages are incorrectly refused. Therefore, we use the default value of I_{hold} to be 3.5 s.

Turing Angles and Speeds Detection: We use the machine learning technique to facilitate the classification of steering wheel turning angles and speeds. While there are many options for classification methods, we study the performance of several highly used machine learning methods, including logistic regression (LR), decision tree (DT), and Naive Bayes (NB). The data used to build machine learning algorithms are from our driving data, which includes 21 fast-large-angle steering, 32 fast-small-angle steering, 55 slow-large-angle steering, and 93 slow-small-angle steering. To address the unbalanced samples, we add simulated samples so that each type of steering operation has 100 samples. We conduct experiments with various percentages of training and testing data. Fig. 9 shows the comparison results with different percentages of testing data in total data. With the increase in the percentage of testing data in total data, the performance of all classifiers decreases slightly, and the RF-based classifier has the highest accuracy. When the proportion of testing data is set to 25%, the RF-based classifier reaches the highest accuracy of 94.30%. Thus, we adopt an RF-based classifier for turning angle and speed detection.

E. Performance of Driver Drowsiness Detection

1) Overall Performance: To understand FDWatch's performance in detecting driver drowsiness, we present the intermediate results from the perspective of each indicator and the final results of FDWatch. The intermediate results of each indicator are recorded based on their BPA mass. When a sample's drowsiness BPA generated by evidence is higher than its normal BPA, it means the sample is regarded as drowsy driving from the evidence's perspective. Frequencies of yawning and steering are calculated in 20 minutes. Fig. 10 shows the drowsy driving detection performance based on various indicators and FDWatch. The false alarm rate and missing alarm rate of the intermediate results of each indicator are different. The intermediate result based on yawning indicators has a low false alarm rate of 5.12% and a high missing alarm rate of 8.77% since sometimes drivers do not yawn when they are tired. If we only consider the user's yawning frequency, only a few samples in the normal state are classified as the drowsiness state. However, some samples in the drowsiness state are predicted as normal. The other indicators have high false alarm rates and high missing alarm rates. The missing alarm rates of the intermediate results of circadian rhythm, HRV indicators, and steering wheel operation indicators are 12.34%, 6.49%, and 7.14%. The false alarm rates are 10.48%, 5.67%, and 6.18%, respectively. Reliable drowsy driving detection cannot be guaranteed only based on any category of indicators. FDWatch improves drowsiness detection performance by jointly considering all the indicators. It achieves a 3.57% missing alarm rate and a 3.68% false alarm rate. These demonstrate that FDWatch's information fusion strategy is effective and can reliably detect driver drowsiness. Besides, the detection time of 90% samples is less than 4 seconds, which proves the time efficiency.

2) Performance Under Errors: It is critical to investigate how FDWatch performs when some indicators are wrong. The real-road data we collected can not provide enough error samples, so we modify the estimated BPAs to validate the performance of FDWatch under indicator errors. Specifically, we randomly selected 5%, 10%, 15%, and 20% samples from all the collect data, then randomly selected one of their four indicators to exchange the estimated drowsiness BPA with the normal BPA. Modifying BPA simulates that this indicator makes a wrong estimate of whether the user is in a drowsiness state or normal state. In this way, the information fusion will be carried out on the wrong evidence. Table II shows the missing alarm rates (MA) and false alarm rates (FA) of drowsy driving detection with different percentages of error data in total data. When using 5% error data (307 samples), the performance of the system is almost the same as no error data. Its missing alarm rate is 3.57%, and the false alarm rate is 3.88%. With the increase of error data, both the missing alarm rate and the false alarm rate increase. The reason is that the indicators do not change synchronously. The modified BPA and other indicators' estimated BPA will lead to different similarity matrices in data fusion, which may lead to the wrong drowsiness detection result. In practice, there are errors in the testing data. But, as we discussed in Section VI-E1, the error rates of the intermediate results of all indicators are not more than 10%, which yields 5.52% missing alarm rate and 6.50% missing alarm rate. This suggests that FDWatch is error-tolerant and has advantages in providing reliable drowsy driving detection.

TABLE II
SYSTEM PERFORMANCE UNDER ERRORS

Set Up	MA (%)	FA (%)
0% error data	3.57	3.68
5% error data	3.57	3.88
10% error data	5.52	6.50
15% error data	8.44	8.58
20% error data	10.06	11.70

3) Impact of Indicators: Different indicators contribute differently to the system. To understand the impact of each kind of indicator on system performance, we conduct experiments with different indicator combinations. Table III shows the missing alarm rates (MA) and false alarm rates (FA) of drowsy driving detection without certain indicators. Note that we do not consider the combination of two categories of indicators, because the proposed algorithm will give them the same weight, and the conflict will not be handled. Compared with the fusion of all indicators, when any kind of indicator is excluded, the system performance decreases. Different combinations of indicators perform differently. When steering wheel operation indicators or HRV indicators are excluded, the performance of FDWatch declined the most. The missing alarm rates are 7.14% and 7.74%, and the false alarm rates are 5.21% and 4.47%, suggesting that steering wheel operation indicators and HRV indicators play a vital role in drowsy driving detection. In addition, it is worth further study that one possible reason for the impact of steering wheel operation indicators on the system is that they are extracted from hand motion data, while other indicators are extracted from PPG data.

TABLE III
SYSTEM PERFORMANCE WITHOUT CERTAIN EVIDENCES

Set Up	MA (%)	FA (%)
Without Circadian Rhythm Indicator	4.87	4.19
Without Yawning Indicators	4.22	4.85
Without Steering Wheel Operation Indicators	7.14	5.21
Without HRV indicators	7.74	4.47

4) Impact of Hand Dominance: Drivers tend to use their dominant hands when operating the steering wheel. This results in the different performance of extracting steering wheel operation indicators between the dominant hands and non-dominant hands and further affecting the overall performance of the system. Fig. 11 shows the validation results of drowsiness detection based on data acquired from dominant hands and non-dominant hands. The drowsiness detection results using data collected from the dominant hands are better than those collected from the non-dominant hands. The non-dominant hand operates the steering wheel-less frequently. Hence some samples in the normal driving state are mistaken as drowsiness. The intermediate result based on non-dominate hand's data has a false alarm rate of 6.72% and a missing alarm rate of 8.12% while the dominant hands' data have a false alarm rate of 5.65% and a missing alarm rate of 6.17%. After information fusion, the drowsiness detection performance is improved, but the recognition results of the dominant hands are still better than that of the non-dominant hands. When using the data of the dominant hands, the false alarm rate is lower than 3.57%, and the missing alarm rate is lower than 3.49%. And the non-dominant hand's group yields a 3.57% false alarm rate and a 4.03% missing alarm rate. Based on the comparison results, we recommend that drivers wear the FDWatch prototype on their dominant hand for more reliable safety services.

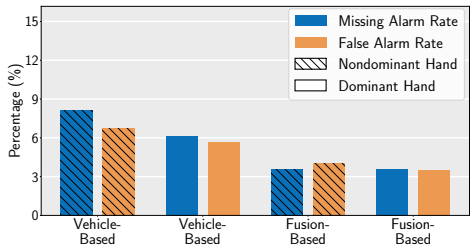


Fig. 11. Impact of hand dominance.

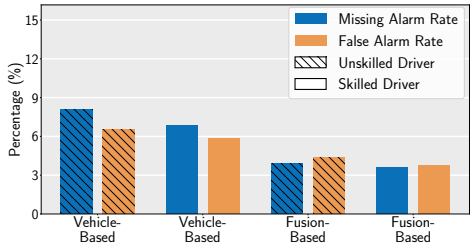


Fig. 12. Impact of driving experience.

5) Impact of Driving Experience: In addition to the way the driver wears the device, the driving experience is also one of the factors affecting the driver's vehicle operations. Hence it will affect drowsy driving detection. Among the ten participants, four have less than two years of driving experience, and six have more than two years of driving experience. We divide them into the unskilled driver group and the skilled driver group and compare the experimental results. Fig. 12 shows the driver drowsiness detection results of skilled drivers and unskilled drivers. The steering wheel operations of skilled drivers are more in line with the pattern of smooth adjustments of the normal state and sudden adjustments of the drowsiness state, which results in a better performance of steering wheel operation indicators of skilled drivers. The intermediate results using unskilled drivers' data achieve a false alarm rate of 6.51% and a missing alarm rate of 8.12%. The false alarm rate and the missing alarm rate of the skilled driver group are 5.89% and 6.82%, respectively. After being processed by the DST method and conflict resolution technique, the system performances based on skilled and unskilled drivers are improved. Still, the results of skilled drivers surpass that of unskilled drivers. Their false alarm rates are 3.78% and 4.38%, respectively. Their missing alarm rates are 3.57% and 3.90%, respectively.

VII. CONCLUSION

In this paper, we exploit the opportunity of realizing accurate, low-cost, unconstrained, and convenient drowsy driving

detection using wrist-worn devices. We propose FDWatch, an information fusion-based system leveraging motion sensors and PPG sensors, which are commonly implemented in modern wearables. Given that drowsy driving is a process that involves many aspects (physiological, behavioral, vehicle, and circadian), FDWatch jointly considers drowsiness-related indicators, including yawning frequency, the rate of yawning interval shortening, frequencies of steering wheel operations at different angles and speeds, HRV features, and alertness level. We design several novel algorithms to extract accurate indicators. Furthermore, FDWatch takes advantage of the BP neural network to estimate the probability assignment from each indicators' prospect. Then, it utilizes DST to fuse all indicators and leverages a distance-based method to resolve the conflict between indicators. We conduct extensive experiments with real-road driving data, and the performance of FDWatch is very promising. It achieves a 3.57% missing alarm rate and a 3.68% false alarm rate.

Further research will be focusing on studying the relationship between drowsiness level and various drowsiness indicators in order to build a more superior drowsiness level detection model. Moreover, we are planning to collect more driving data in real driving environments, and validate the system performance with a larger user group and various readily available wrist-worn devices.

ACKNOWLEDGMENT

The work of Fan Li is partially supported by the National Natural Science Foundation of China (NSFC) under Grant No.62072040, 61772077, and Natural Science Foundation of Beijing Municipality under Grant No.4192051. The work of Song Yang is partially supported by NSFC under Grant No.61802018 and Beijing Institute of Technology Research Fund Program for Young Scholars.

REFERENCES

[1] G. Zhang, K. K.W. Yau, X. Zhang, and Y. Li, "Traffic accidents involving fatigue driving and their extent of casualties," *Accident Analysis and Prevention*, vol. 87, pp. 34–42, Feb 2016.

[2] M. Awais, N. Badruddin, and M. Drieberg, "A hybrid approach to detect driver drowsiness utilizing physiological signals to improve system performance and wearability," *Sensors*, vol. 17, no. 9, p. 1991, 2017.

[3] M. Ben Dkhil, A. Wali, and A. M. Alimi, "Drowsy driver detection by eeg analysis using fast fourier transform," in *2015 15th International Conference on Intelligent Systems Design and Applications (ISDA)*, 2015, pp. 313–318.

[4] R. N. Khushaba, S. Kodagoda, S. Lal, and G. Dissanayake, "Driver drowsiness classification using fuzzy wavelet-packet-based feature-extraction algorithm," *IEEE Transactions on Biomedical Engineering*, vol. 58, no. 1, pp. 121–131, Jan 2011.

[5] C. Lin, C. Chang, B. Lin, S. Hung, C. Chao, and I. Wang, "A real-time wireless brain-computer interface system for drowsiness detection," *IEEE Transactions on Biomedical Circuits and Systems*, vol. 4, no. 4, pp. 214–222, Aug 2010.

[6] W. Tipprasert, T. Charoenpong, C. Chianrabutra, and C. Sukjamsri, "A method of driver's eyes closure and yawning detection for drowsiness analysis by infrared camera," in *2019 First International Symposium on Instrumentation, Control, Artificial Intelligence, and Robotics (ICA-SYMP)*, 2019, pp. 61–64.

[7] S. Murugan, J. Selvaraj, and A. Sahayadhas, "Driver hypovigilance detection for safe driving using infrared camera," in *2020 International Conference on Inventive Computation Technologies (ICICT)*, 2020, pp. 413–418.

[8] Volvo driver alert control and lane departure warning. [Online]. Available: http://www.zercustoms.com/news/Volvo-Driver-Alert-Control-and-Lane-Departure-Warning.html

[9] A. D. McDonald, C. Schwarz, J. D. Lee, and T. L. Brown, "Real-time detection of drowsiness related lane departures using steering wheel angle," *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, vol. 56, no. 1, pp. 2201–2205, 2012.

[10] J. Krajewski, D. Sommer, U. Trutschel, D. Edwards, and M. Golz, "Steering wheel behavior based estimation of fatigue," in *International Driving Symposium on Human Factors in Driver Assessment, Training and Vehicle Design*, Oct 2017, pp. 118–124.

[11] Y. Qiao, K. Zeng, L. Xu, and X. Yin, "A smartphone-based driver fatigue detection using fusion of multiple real-time facial features," in *2016 13th IEEE Annual Consumer Communications Networking Conference*, Jan 2016, pp. 230–235.

[12] I. Chatterjee and S. Roy, "Smartphone-based drowsiness detection system for drivers in real-time," in *2019 IEEE International Conference on Advanced Networks and Telecommunications Systems (ANTS)*, 2019, pp. 1–6.

[13] R. Jabbar, K. Al-Khalifa, M. Kharbeche, W. Alhajyaseen, M. Jafari, and S. Jiang, "Real-time driver drowsiness detection for android application using deep neural networks techniques," *Procedia Comput. Sci.*, vol. 130, no. C, pp. 400–407, May 2018.

[14] C. You, N. D. Lane, F. Chen, R. Wang, Z. Chen, T. J. Bao, M. Montes-de Oca, Y. Cheng, M. Lin, L. Torresani, and A. T. Campbell, "Carsafe app: Alerting drowsy and distracted drivers using dual cameras on smartphones," in *Proceeding of the 11th Annual International Conference on Mobile Systems, Applications, and Services*, 2013, pp. 13–26.

[15] Y. Xie, F. Li, Y. Wu, S. Yang, and Y. Wang, "D^3-Guard: Acoustic-based drowsy driving detection using smartphones," in *IEEE Conference on Computer Communications*, Apr 2019, pp. 1225–1233.

[16] L. B. Leng, L. B. Giin, and W. Chung, "Wearable driver drowsiness detection system based on biomedical and motion sensors," *IEEE Sensors*, pp. 1–4, 2015.

[17] S. Zhang, H. He, Z. Wang, M. Gao, and J. Mao, "Low-power listen based driver drowsiness detection system using smartwatch." *Cloud Computing and Security*, vol. 11067, pp. 453–464, 2018.

[18] S. Barua, M. U. Ahmed, C. Ahlstrom, and S. Begum, "Automatic driver sleepiness detection using EEG, EOG and contextual information," *Expert Systems with Applications*, vol. 115, pp. 121–135, 2019.

[19] N. Senniappan Karuppusamy and B.-Y. Kang, "Multimodal system to detect driver fatigue using eeg, gyroscope and image processing," *IEEE Access*, vol. PP, pp. 1–1, 07 2020.

[20] Y. Xie, F. Li, Y. Wu, S. Yang, and Y. Wang, "Real-time detection for drowsy driving via acoustic sensing on smartphones," *IEEE Transactions on Mobile Computing*, pp. 1–14, 2020.

[21] B. Lee, B. Lee, and W. Chung, "Wristband-type driver vigilance monitoring system using smartwatch," *IEEE Sensors*, vol. 15, no. 10, pp. 5624–5633, Oct 2015.

[22] N. Lamond and D. Dawson, "Quantifying the performance impairment associated with fatigue." *Journal of sleep research*, vol. 8, no. 4, 1999.

[23] J. Barbizet, "Yawning," *Journal of Neurology, Neurosurgery and Psychiatry*, vol. 21, no. 3, pp. 203–209, 1958.

[24] F. Forcolin, R. Buendia, S. Candefjord, J. Karlsson, B. A. Sjoqvist, and A. Anund, "Comparison of outlier heartbeat identification and spectral transformation strategies for deriving heart rate variability indices for drivers at different stages of sleepiness," *Traffic Injury Prevention*, vol. 19, 2018.

[25] W. Harris, *Fatigue, Circadian Rhythm, and Truck Accidents*. Springer, 1977, pp. 133–146.

[26] M. R. Ambekar and S. Prabhu, "A novel algorithm to obtain respiratory rate from the ppg signal," *International Journal of Computer Applications*, vol. 126, no. 15, pp. 9–12, 2015.

[27] T. Tamura, "Current progress of photoplethysmography and spo 2 for health monitoring," *Biomedical engineering letters*, vol. 9, no. 1, pp. 21–36, 2019.

[28] V. Rybynok, J. May, K. Budidha, and P. Kyriacou, "Design and development of a novel multi-channel photoplethysmographic research system," in *2013 IEEE Point-of-Care Healthcare Technologies (PHT)*, 2013, pp. 267–270.

[29] J. G. Webster, *Design of pulse oximeters*. CRC Press, 1997.

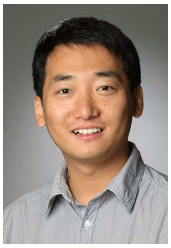
[30] P. Kyriacou, K. Shafqat, and S. Pal, "Pilot investigation of photoplethysmographic signals and blood oxygen saturation values during blood pressure cuff-induced hypoperfusion," *Measurement*, vol. 42, no. 7, pp. 1001–1005, 2009.

[31] C. Karatas, L. Liu, H. Li, J. Liu, Y. Wang, S. Tan, J. Yang, Y. Chen, M. Gruteser, and R. Martin, "Leveraging wearables for steering and driver tracking," in *IEEE Conference on Computer Communications*, Apr 2016, pp. 1–9.
[32] (2016) tsfresh toolkit. [Online]. Available: http://tsfresh. readthedocs.io/en/latest/
[33] R. Grace and S. Steward, "Drowsy driver monitor and warning system," 2001.
[34] T. Akerstedt and S. Folkard, "The three-process model of alertness and its extension to performance, sleep latency, and sleep length," *Chronobiology International*, vol. 14, no. 2, pp. 115–123, 1997.
[35] N. Blaylock and J. Allen, "Chapter 1 - hierarchical goal recognition," in *Plan, Activity, and Intent Recognition*, G. Sukthankar, C. Geib, H. H. Bui, D. V. Pynadath, and R. P. Goldman, Eds., Boston, 2014, pp. 3 – 32.
[36] Q. Ding, P. Zhenghong, T. Liu, and Q. Tong, "Multi-sensor building fire alarm system with information fusion technology based on d-s evidence theory," *Algorithms*, vol. 7, 10 2014.
[37] A. Dempster and W. F. Chiu, "Dempster-shafer models for object recognition and classification," *International Journal of Intelligent Systems*, vol. 21, no. 3, pp. 283–297, 2006.
[38] A. L. Jousselme, D. Grenier, and E. Bosse, "A new distance between two bodies of evidence," *Information Fusion*, vol. 2, pp. 91–101, 06 2001.

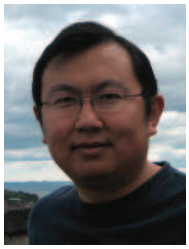
Yetong Cao received the BEng degree in computer science and technology from Shandong University, Shandong, China, in 2017. She is now working toward the PhD degree in the School of Computer Science, Beijing Institute of Technology. Her research interests include mobile computing and smart health.

Fan Li received the PhD degree in computer science from the University of North Carolina at Charlotte in 2008, MEng degree in electrical engineering from the University of Delaware in 2004, MEng and BEng degrees in communications and information system from Huazhong University of Science and Technology, China in 2001 and 1998, respectively. She is currently a professor at School of Computer Science in Beijing Institute of Technology, China. Her current research focuses on wireless networks, ad hoc and sensor networks, and mobile computing. Her papers won Best Paper Awards from IEEE MASS (2013), IEEE IPCCC (2013), ACM MobiHoc (2014), and Tsinghua Science and Technology (2015). She is a member of ACM and IEEE.

Xiaochen Liu received the BEng degree in Internet of Things from China University of Petroleum, Shandong, China, in 2020. She is now working toward the MEng degree in the School of Computer Science and Technology, Beijing Institute of Technology. Her research interests include computer networks and the IoT.



Song Yang received the Ph.D. degree from Delft University of Technology, The Netherlands, in 2015. From August 2015 to July 2017, he worked as postdoc researcher for the EU FP7 Marie Curie Actions CleanSky Project in Gesellschaft für wissenschaftliche Datenverarbeitung mbH Göttingen (GWDG), Göttingen, Germany. He is currently an associate professor at School of Computer Science in Beijing Institute of Technology, China. His research interests focus on data communication networks, cloud/edge computing and network function virtualization.



Yu Wang is currently a Professor in the Department of Computer and Information Sciences at Temple University. He holds a Ph.D. from Illinois Institute of Technology, an MEng and a BEng from Tsinghua University, all in Computer Science. His research interest includes wireless networks, smart sensing, and mobile computing. He has published over 200 papers in peer reviewed journals and conferences. He has served as general chair, program chair, program committee member, etc. for many international conferences (such as IEEE IPCCC, ACM MobiHoc, IEEE INFOCOM, IEEE GLOBECOM, IEEE ICC), and served as Editorial Board Member for several international journals, including IEEE Transactions on Parallel and Distributed Systems. He is a recipient of Ralph E. Powe Junior Faculty Enhancement Awards from Oak Ridge Associated Universities (2006), Outstanding Faculty Research Award from College of Computing and Informatics at the University of North Carolina at Charlotte (2008), Fellow of IEEE (2018), and ACM Distinguished Member (2020).