

DetectDUI: An In-Car Detection System for Drink Driving and BACs

Yanjiao Chen^{ID}, Senior Member, IEEE, Meng Xue, Jian Zhang^{ID}, Runmin Ou^{ID},
Qian Zhang^{ID}, Fellow, IEEE, and Peng Kuang

Abstract—As one of the biggest contributors to road accidents and fatalities, drink driving is worthy of significant research attention. However, most existing systems on detecting or preventing drink driving either require special hardware or require much effort from the user, making these systems inapplicable to continuous drink driving monitoring in a real driving environment. In this paper, we present *DetectDUI*, a contactless, non-invasive, real-time system that yields a relatively highly accurate drink driving monitoring by combining vital signs (heart rate and respiration rate) extracted from in-car WiFi system and driver's psychomotor coordination through steering wheel operations. The framework consists of a series of signal processing algorithms for extracting clean and informative vital signs and psychomotor coordination, and integrate the two data streams using a self-attention convolutional neural network (i.e., C-Attention). In safe laboratory experiments with 15 participants, *DetectDUI* achieves drink driving detection accuracy of 96.6% and BAC predictions with an average mean error of $2 \sim 5\text{mg/dl}$. These promising results provide a highly encouraging case for continued development.

Index Terms—Drink driving detection, mobile sensing, inertial measurement unit.

I. INTRODUCTION

IN 2018, The US suffered 10,511 deaths from drunk driving crashes [1]. The WHO reports that, in high-income countries, as many as 20% of fatally injured drivers have excess alcohol in their blood.¹ COVID-related deaths may dwarf these numbers, but it is important not to lose our pre-pandemic perspective. Tens of thousands of deaths per year due to drink driving is a staggering loss of life. At an estimated cost of \$44 billion² for just one year in the US alone [1],

Manuscript received January 18, 2021; revised July 6, 2021 and August 25, 2021; accepted November 2, 2021; approved by IEEE/ACM TRANSACTIONS ON NETWORKING Editor N. Zhang. Date of publication November 24, 2021; date of current version April 18, 2022. The work of Yanjiao Chen was supported in part by the National Natural Science Foundation of China under Grant 61972296 and in part by Wuhan Application Advanced Funding under Grant 2019010701011419. The work of Qian Zhang was supported in part by the RGC under Contract CERG 16204418, Grant 16203719, and Grant R8015, and in part by the Guang-dong Natural Science Foundation under Grant 2017A030312008. (Corresponding author: Yanjiao Chen.)

Yanjiao Chen, Meng Xue, Jian Zhang, Runmin Ou, and Peng Kuang are with the School of Computer Science, Wuhan University, Wuhan 430072, China (e-mail: chenyanjiao@whu.edu.cn; xuem.yun@gmail.com; jzhang@whu.edu.cn; ormmroor@gmail.com; pengkuang@whu.edu.cn).

Qian Zhang is with the Department of Computer Science and Engineering, Hong Kong University of Science and Technology, Hong Kong, China (e-mail: qianzh@cse.ust.hk).

This article has supplementary downloadable material available at <https://doi.org/10.1109/TNET.2021.3125950>, provided by the authors.

Digital Object Identifier 10.1109/TNET.2021.3125950

¹Road safety-Alcohol https://www.who.int/violence_injury_prevention/publications/road_traffic/world_report/alcohol_en.pdf

²Estimates are from 2010 – the last year for which data was available.

the economic impact is not insignificant either. According to the newly-enacted Halt Act (H.R. 2138) [2] and Ride Act (S.B. 1331) [3], drunk-driving prevention technology will be a safety standard for all new cars in the future. There is an urgent demand for in-vehicle drunk-driving detection systems to help prevent drunk-related accidents.

Most traditional methods for detecting drunk drivers need to interrupt the driving process. To administer a breathalyzer, the police must hail the driver to pull over, giving the driver time to implement means of avoiding detection. Blood tests are invasive, and require the driver to stop. Similarly, urine tests and pupil measuring tests require special operations and expert examiners. While it is desirable to detect whether the driver is drunk before driving and prevent potential risks, it is possible that alcohol consumption takes time to take effect and the driver may consume alcohol during driving. Therefore, the most reliable way is to have a continuous monitoring of drunkenness during driving without interfering the driving process.

In this regard, there are many existing studies on leveraging sensing technologies to determine inebriation levels and Blood Alcohol Content (BAC). You *et al.* [4] devised a transdermal sensing wristband with an accompanying smartphone app connected via Bluetooth that calculates and tracks a person's BAC, while Jung *et al.* [5] developed a smartphone attachment that performs a colorimetric analysis on saliva. Nonetheless, these systems require extra devices, some of which are expensive. Researchers have found others ways to detect drunkenness with only smartphones. Kao *et al.* [6], for example, devised an app that measures a person's walking patterns with an inference model that learns to detect abnormal gaits associated with inebriation. Bae *et al.* [7] use the sensor data a person's phone collects to train a machine learning model to distinguish between drinking and non-drinking episodes. Similarly, Markakis *et al.* [8] also focus on changes in a person's unique patterns of coordination to determine inebriation. These solutions discern psychomotor and cognition skills under the influence of alcohol. However, they require users to perform certain activities that interfere with the driving process.

Our solution is a passive continuous drunk-driving detection system called *DetectDUI*. *DetectDUI* measures a person's vital signs through WiFi signals and their psychomotor coordination through steering wheel operations. However, the complicated driving conditions make it challenging to extract clear vital signs and it is essential to find a non-disturbing way of measuring psychomotor coordination. We manage the interference through a multi-step process. To eliminate reflections

from other passengers and from car interiors (e.g., seats, windows), we leverage power delay profile to separate the direct reflection from the chest of the driver from multipath interference. WiFi signals are carried by multiple subcarriers. Different subcarriers have different sensitivities to subtle chest motions. To take full advantage of the diverse information from all subcarriers and avoid interference, we adopt principal component analysis (PCA) to sift noises and preserve the first principal component. Due to bumpy driving conditions, the received signals contain sudden changes with increased amplitude. We remove sudden changes and preserve only signals during relatively stable driving periods, which show a clear cyclic pattern that corresponds to breathing cycles, but the heartbeat pattern is drowned due to its much weaker amplitude. To address this problem, we propose a novel adaptive variational mode decomposition (AVMD) method to separate the mixed signal into multiple modes, and then keep the modes that relate to breathing and heartbeat respectively. Previous works usually measure psychomotor coordination of a person using interactive games or operations with a smartphone or computer, which is not applicable to the driving environment. We find a natural way to gauge the psychomotor coordination of the driver by monitoring their steering wheel operations. In particular, we use IMU to record the acceleration and gyroscope data during operation. In this way, we obtain a continuous monitoring of psychomotor coordination of the driver without interfering with their driving. Integrating the vital signs and the psychomotor signals is done with neural networks and an attention mechanism. Random Forest (RF) is then used to predict concrete BAC values. Trials with 15 volunteers show that *DetectDUI* was able to detect a drunken driver with 96.6% accuracy. Further, it was able to predict a person's BAC within a mean absolute error (MAE) of 0.002% to 0.005%.

DetectDUI can be supported by in-car IMU and WiFi systems. The lightweight learning-based detection model can be deployed locally, with data collected to fine-tune the model locally without privacy leakage.

In summary, the contributions of this research include:

- As far as we are concerned, *DetectDUI* is the first contactless method of detecting drink driving, including measuring the driver's BAC that can be administered while driving.
- We have proposed a series of signal processing algorithms for extracting human vital signs from WiFi signals given chest motions with high levels of accuracy.
- We have proposed to use C-Attention to combine the information of vital signs and psychomotor coordination to reach a well-round drink driving prediction.
- Extensive experiments on 15 individuals show *DetectDUI* is able to distinguish normal driving from drink driving in real-time with a 96.6%-accurate estimation and the driver's BAC to within an MAE of 0.002% to 0.005%.

II. MOTIVATIONS

In this section, we briefly introduce the impact of alcohol on the vital signs and psychomotor coordination of human. We conduct feasibility study to demonstrate that the change

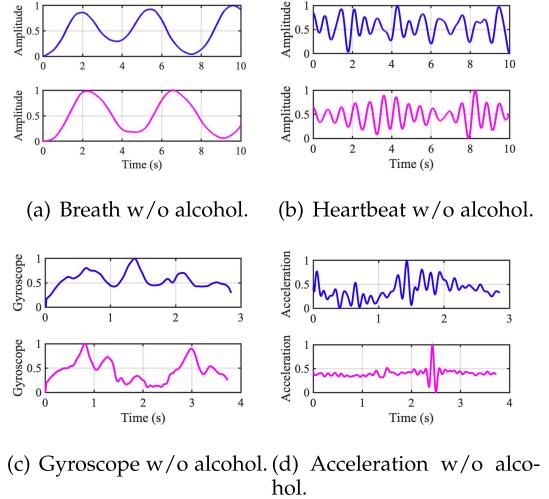


Fig. 1. Alcohol's impact on vital signs and psychomotor coordination. The top graph of each subfigure plots the signals when the person is sober, and the bottom graph of each subfigure plots the signals when the person is under the impact of alcohol.

in vital signs and psychomotor coordination can be leveraged for drink driving detection to enhance driving safety.

Alcohol is the major component in alcoholic beverages, which are one of the most popular drinks in history [9]. However, alcohol has various influences on the human body [10]. The National Advisory Council on Alcohol Abuse and Alcoholism reports that alcohol can disrupt brain functions, resulting in mood swing, behavior change, inability to think clearly and move with coordination [8], [11]. A small amount of alcohol can accelerate the respiratory rate [12], and increase heart rate and blood pressure [13]–[15]. A large amount of alcohol will decrease the respiratory rate, pulse and blood pressure [16].

To validate the feasibility of detecting drink driving and predicting BAC based on vital signs and psychomotor coordination, we collect these two kinds of signals under and without the influence of alcohol. The signals of breathing and heartbeat are extracted from WiFi signals (a detailed process is presented in Section IV) and the psychomotor coordination is monitored by the in-car inertial measurement unit (IMU) [8] including the gyroscope and the acceleration signals.

Fig. 1(a) shows the result of alcohol's effect on breathing. It can be observed that the frequency of respiration is higher when the person is sober. Fig. 1(b) shows the result of alcohol's effect on heartbeat. It demonstrates that the frequency of heartbeat is relatively higher when the person drinks. These results confirm that the alcohol varies vital signs of humans [13], [16]. Fig. 1(d) and Fig. 1(c) display the acceleration and the gyroscope signals as a person conducts the same steering wheel operation when he is sober and drunk. There is a clear discrepancy in the signals, especially for the acceleration signal. After drinking, the acceleration signal is more bursty, indicating that the operation is more abrupt. In comparison, when sober, the person tends to steer the wheel in a more smooth manner. This indicates that alcohol indeed disrupts psychomotor coordination of human [11].

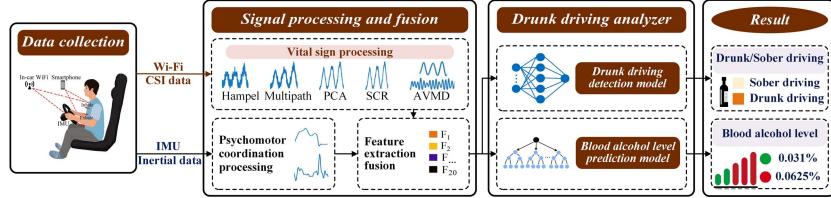


Fig. 2. The system architecture of *DetectDUI*.

III. BACKGROUND & RELATED WORK

A. Drunkenness Detection

Hardware-Based Detection: First used in the United Kingdom in the 1970s [17], breathalyzers are the world's most commonly used tools for testing inebriated drivers. Over its years of usage, researchers have connected breathalyzers, as well as other types of breath alcohol sensors, to smartphones via Bluetooth to improve BAC tracking, especially for self-monitoring by drivers themselves. Example systems include: BACtrack Mobile Pro [18], Breathmeter [19]. One major disadvantage of breathalyzers is that the results are highly susceptible to the oral environment [20] and certain diseases (e.g., diabetes, liver and kidney diseases [20]), which may lead to false detection. Alternatives to breathalyzers include SCRAM, a transdermal sensor that measures the wearer's BAC through their sweat every 30 minutes [21]. The same kind of system is available in a tight wristband that fits closely to the skin [4]. However, SCRAM-based systems require a close contact between the skin and the sensor. Any space or anything between the skin and the sensor will affect the detection accuracy. Moreover, these systems require users to purchase extra devices or sensors, which may be expensive.

Camera-Based Detection: Camera-based drunk driving systems have also been developed [22], [23]. In [22], facial landmarks and motions are recognized in images to detect whether the driver is drunk driving or not. In [23], an audio-visual database is utilized to realize bimodal intoxication detection. However, camera-based approaches are sensitive to lighting conditions and there is potential risk of privacy violation [24].

Behavior-Based Detection: The side effects of alcohol consumption include arrhythmia [14], slowed respiratory rates [15], impaired psychomotor performance [8], and unsteady gait [6]. This abnormality in vital signs and behaviors can be leveraged to detect whether the user is under the influence of alcohol. Bae *et al.* [7] developed a smartphone-based system to track the drinking episodes of users based on built-in sensors (e.g., accelerometer) and the smartphone status (e.g., battery and network usage). Leveraging alcohol's influence on motor coordination and cognition, Markakis *et al.* [8] designed five human-computer interactions to detect BACs (such as swiping or touching the screen in particular ways), akin to the finger-to-nose DUI tests. However, these works require users to interact with their phones (swipe the phone or engage in games), which interrupts the driving task and cannot offer a continuous drunk driving detection.

B. Contactless Vital Sign Monitoring

Vital signs are crucial indicators of physical and mental health. In recent years, extensive studies have been undertaken on contactless vital sign monitoring using wireless signals. For example, Yang *et al.* [25] do the same using the received signal strength (RSS) of millimeter-wave signals. All these systems, however, depend on sophisticated hardware, e.g., radar.

For this reason, some researchers have explored acoustic signals, which can be easily captured by smart devices such as smartphones, in vital sign monitoring [26]. In this sphere, BreathListener [26] recovers breathing patterns based on the energy spectrum density of chest motions. Unfortunately, the sensing range of acoustic signals is small (approximately 30cm), which limits the applications of these systems to scenarios such as in-car heart rate monitoring.

With the advancement of WiFi technologies, some are looking to wireless local area networks (WLAN) not just for their use as a communication tool but also for sensing human activities. Analyzing the characteristics of received signals, WiFi signals can capture the chest motions caused by breathing and heartbeats. Some notable products include UbiBreathe [27], which leverages the RSS of off-the-shelf WiFi devices to monitor breathing rates. Liu *et al.* [28] use the channel state information (CSI) of WiFi signals to track breathing and heart rates during sleep. Zhang et al [29] tracks breathing status using CSI. Compared to acoustic signals, WiFi signals have a broader sensing range.

In this paper, we develop a WiFi-based drunk driving detection system, which contactless vital sign monitoring, which is more suitable for the in-car environment. Compared with existing works on drunkenness detection, our system does not require extra devices, is privacy-preserving and can perform continuous detection without interfering the driving process. WiFi is being deployed in many automobiles, such as Chevrolet Cruze, Chrysler 200, Dodge Dart, Audi A3, Ford Escape, and Mercedes-Benz GLS. Different from existing works on WiFi-based breath or heartbeat monitoring, we have developed a series of signal processing techniques to remove interference in the received signals to cater to the complicated in-car environment.

IV. DetectDUI: DETAILED DESIGN

DetectDUI detects drink driving and predicts BAC through a driver's vital signs and psychomotor coordination. Fig. 2 shows the architecture of *DetectDUI*. In *DetectDUI*, vital signs are tracked through a WiFi sensing system. The transmitter is a hotspot on a mobile phone, and the receiver is the on-board unit (OBU) of the car. Psychomotor coordination is

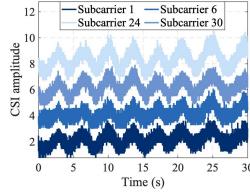


Fig. 3. The CSI amplitudes of various subcarriers.

measured by an IMU attached to the car's steering wheel. The system comprises two main modules: a signal processing and fusion module, which prepares, cleans, processes, and fuses the two data streams, and a drink driving analyzer, which is responsible for extracting features, determining whether the driver is drunk, and producing BAC estimation. If the system determines a driver to be drunk, it issues an alert through the smartphone. Future real-world version of the system could be designed to adopt other actions in the case of emergency, e.g., taking over from the driver and placing the car in autonomous driving mode.

A. Extracting Vital Signs

Due to the complicated driving conditions, the in-car wireless transmission environment is noisy and unstable. The chest motions caused by breathing and heartbeat are subtle and easily contaminated by the interference. To obtain accurate vital signs from the received signals, we have carefully designed five steps to remove the interference and extract clear breathing and heartbeat patterns.

To begin with, we use a Hampel filter to smooth the signal by removing outliers with abnormally high or low amplitudes. Then, we leverage power delay profile to eliminate multipath effects since the in-car environment is compact and contains many reflection paths that interfere the direct reflection from the human body. After that, we adopt principal component analysis to distill and integrate important information of subcarriers. In the fourth step, we eliminate sudden changes due to bumpy driving conditions, which have an overlapping frequency range with vital signs and cannot be removed by previous three steps. Finally, we leverage adaptive variational mode decomposition to separate and recover patterns of breathing and heartbeat. These five steps work together to purify the noisy received signals and help restore accurate vital signs.

1) Collecting CSI Data: Vital signs are derived from the minute chest motions caused by breathing and heartbeats recorded by an in-car WLAN as CSI signals. While RSS only provides the average power of WiFi signals, CSI is capable of capturing more information, e.g., fading and scattering in each narrowband subcarrier channel. We leverage the amplitude and phase of these subcarriers to extract subtle chest movements.

Fig. 3 displays the CSI amplitude of four subcarriers, i.e., subcarriers 1, 6, 24 and 30, between a transmitter-receiver antenna pair. Despite having different amplitudes, these subcarriers are similar in general profiles. A clear periodic pattern can be observed, which reflects the breathing cycles. Nonetheless, the heartbeat cycles are hard to differentiate, due to much weaker amplitudes. The heartbeat cycles tend to be obscured

by the breathing cycles. To extract both the breathing and the heartbeats, the data needs to be further processed.

2) Smoothing CSI Data: As shown in Fig. 3, the CSI signals exhibit many glitches that result from factors like environmental noise and hardware imperfections. To remove these artifacts, we first adopt Hampel filter, which smooths the signals while preserving important information associated with chest motions. Applied to each subcarrier separately, the Hampel filter removes outliers that have an abnormally high or low amplitude. A data point h_i is considered to be an outlier if the following condition is satisfied,

$$|h_i - \text{median}(h_i, K)| > \delta \times \text{std}(h_i, K), \quad (1)$$

where $\text{median}(h_i, K)$ and $\text{std}(h_i, K)$ are the median and standard deviation of K neighboring samples of h_i , and δ is a constant parameter. A detected outlier h_i is replaced by $\text{median}(h_i, K)$. K is the length of local windows.

Fig. 26 in Appendix VII-A shows the subcarrier with the largest variance before and after Hampel filter. We can observe that many glitches around 10s have been removed by the Hampel filter. Compared to the raw signal, the filtered signal is smoother and gives a better representation of the periodic inhale/exhale chest motions of breathing.

3) Eliminating Multipath Reflections: Ideally, we need only the direct reflection from the chest of the user for vital sign extraction. Unfortunately, the signals will also be reflected by other parts of the driver (e.g., head, arm), other passengers and car interiors (e.g., seat). Based on the fact that direct reflection has the shortest path, shortest delay and strongest power, we use power delay profile (PDP) [30] to differentiate direct reflection from the chest and other multipath reflections.

CSI characterizes the frequency response of each subcarrier under orthogonal frequency-division multiplexing (OFDM), a technique adopted to improve spectrum utilization by commercial WiFi devices. The calibrated signals from the step above are first divided into time domains via an inverse fast Fourier transformation. The power PDP is calculated for each time interval as

$$\Delta T = N/(F * C_n * 2), \quad (2)$$

where $N = 64$ is the total number of subcarriers, $F = 40\text{MHz}$ is the bandwidth of the WiFi signal, and $C_n = 60$ is the number of collected subcarriers. In [31], a delay of over 500ns is deemed as multipath reflections in the indoor environment. Since the in-car environment is narrower than the indoor environment, we find that a delay threshold of 300ns is suitable. Any paths with a delay over 300ns are deemed to be multipath reflections and removed.

Fig. 27 in Appendix VII-A displays the subcarrier that has been processed by the Hampel filter before and after multipath elimination. We can observe that before multipath removal, the signal is a mixture of multipath reflections, which blur the chest movement pattern. Specifically, the peaks and valleys are hazy before multipath elimination and becomes sharp after multipath elimination. The signal after multipath removal carries more detailed information regarding the chest movement pattern.

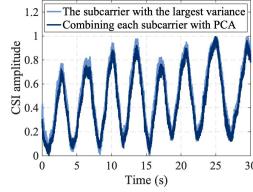


Fig. 4. The first principal component of all subcarriers.

4) Principal Component Analysis: As shown in Fig. 3, the variations in each subcarrier is caused by chest motions. Thus, the information carried by subcarriers are highly correlated and contains redundancy. In addition, subcarriers with different frequencies (i.e., different ID numbers) have different sensitivities to the subtle chest motions. Some subcarriers are sensitive to chest motions and bear rich information of vital signs, while some subcarriers are insensitive to chest motions and contains mostly noises. Therefore, we need to integrate information of subcarriers in a way that keeps useful information and removes redundancy and interference.

In existing works on vital sign monitoring, Liu *et al.* [28] selects a single subcarrier with the largest variance and ignores all other subcarriers, which leads to potential information loss. Instead, we leverage principal component analysis (PCA) [32], an effective way of integrating and denoising highly-correlated data.

To identify the principal components, we compute

$$P = D * \text{eig}(\text{cov}(D)), \quad (3)$$

where D is a $\text{len} * 60$ matrix that contains data of all subcarriers, len is the number of samples in each subcarriers, 60 is the number of collected subcarriers, $\text{cov}(D)$ is the covariance matrix of D , and $\text{eig}(\text{cov}(D))$ is the eigenvector of $\text{cov}(D)$. Each column of P represents a principal component, and we select the first principal component with the largest variance. The selected principal component is normalized since the amplitude of signals can become very large after PCA.

Fig. 4 compares the subcarrier with the largest variance (after multipath elimination) and the first principal component of PCA. It is clear that the latter contains less interference and exhibits clearer breathing cycles. Specifically, compared to the subcarrier with the largest variance, PCA removes the fake peak at 6.2s, and recovers the true valley at 27.3s. PCA merges all the useful information of all subcarriers between the two transmitter-receiver antenna pairs (30 subcarriers of each antenna pair and there are two antenna pairs). Keeping the first principal component and discarding others helps eliminate uncorrelated noises that is difficult to remove using traditional low-pass filters.

5) Removing Sudden Changes: Different from indoor environments, in-car environments provide poorer conditions for wireless transmissions. Interruptions due to poor driving conditions, interference from bumps in the road, and other surrounding disturbances are inevitable. These interruptions usually have large amplitudes but low frequencies that overlap with the frequency of chest motions, thus cannot be removed by the previous steps. Our strategy is to divide the signal

into time windows and eliminate window that contains sudden unexpected changes, leaving only periods of relative stable signals. This is done with Algorithm 1.

Algorithm 1 Sudden Change Removal (SCR)

Input: CSI data $R(t)$, window size L , energy threshold E_T , detection time T_S .

Output: Time-serials data without sudden changes $B(t)$.

```

1:  $I = 0$ . // Index of the current window
2: while The detection module is active do
3:    $I = I + 1$ .
4:   Compute FFT for the signals  $R(t)$  in this window.
5:   Compute the energy of signals in this window  $E_I$ .
6:   if  $E_I > E_T$  then
7:     Delete this window from  $R(t)$ .
8:   else
9:     Store the data in this window to  $B(t)$ .
10:    if  $I > T_S/L$  then
11:      Output  $B(t)$ .
12:    end if
13:  end if
14: end while

```

We first transform the signal to the frequency domain using fast Fourier transform. The change in signals is measured in terms of signal energy and is considered sudden if the total energy E_I in a window of size L exceeds a certain threshold E_T . The energy in a window is calculated as

$$E_I = \frac{1}{N} \sum_{i \in \Omega_I} (R_{fft,i})^2, \quad (4)$$

where Ω_I is the set of data points in the current window after fast Fourier transform, R_i is the i -th data point, and N is the total number of points in the window.

According to the average energy, *DetectDUI* utilizes the sliding window method to remove sudden changes in signals. *DetectDUI* calculates the average energy within a window and compares it with the energy threshold. The window size L and the threshold E_T are adjusted according to the received data. In our experiments, the threshold E_T is set as twice the average energy of the received signal so far. The window size is initiated as 2s. If the average energy is greater than the threshold for four consecutive windows, the window size will increase by 1s. After that, if the average energy is half of the threshold, the window size will return to the initial size.

Fig. 5 shows the signal before and after applying Algorithm 1. We can observe that the sudden change from 80s to 150s has been removed.

After sudden change removal, we apply a Butterworth low-pass filter to remove high-frequency interference. Fig. 6 shows an example of the signal after Butterworth filter.

6) Adaptive Variational Mode Decomposition: Until now, we have got a relatively clean signal after intensive interference removal thanks to the previous four steps, as shown in Fig. 6. However, even though the signal demonstrates clear cycles that corresponds to breath patterns, the information of

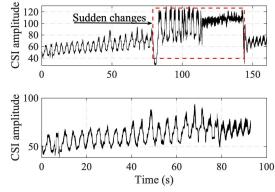


Fig. 5. The red box highlights a sudden change in the signal. The bottom subgraph demonstrates the signal after sudden change removal.

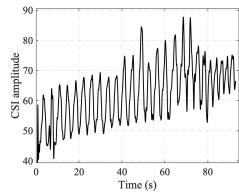


Fig. 6. The signals after Butterworth low-pass filter.

heartbeat is imperceptible due to its much weaker amplitude. Furthermore, there will be noises from the in-car environment in the processed signal [26]. The frequency of heartbeat, breathing, the noises from the in-car environment, thus the vital signs cannot be separated using filters. The challenge now is how to separate the noises introduced by the in-car environment from the vital signs.

A potential solution to this problem is to decompose the mixed signal into different modes of spectral bands, and Empirical Mode Decomposition (EMD) has been widely used in existing works on similar problems [33]. Nonetheless, the recursive nature of EMD makes it suffer from noises. To address this difficulty, we resort to a more advanced signal decomposition approach, variational mode decomposition (VMD) [34], which is non-recursive and outperforms many other mode decomposition methods.

However, the number of decomposed modes affect each mode's central frequency and narrow bandwidth, resulting in reduced detection accuracy of breathing and heartbeat. To solve this problem, we design a novel adaptive variational mode decomposition (AVMD) approach to adaptively determine the numbers of decomposed modes. First, AVMD decomposes the time-series signal $B(t)$ into K sub-signals (modes) via an intrinsic mode function (IMF) [34],

$$B(t) = \sum_{k=1}^K u_k(t), \quad (5)$$

where $u_k(t)$ is the k -th mode.

Similar to the conventional VMD algorithm, we need three steps to estimate the bandwidth of each IMF. First, the analytic signal for each mode is computed. Then, the modes are mixed with an exponential to shift their frequencies to the baseband. Finally, the bandwidth of each mode is estimated via the squared L^2 norm of the gradient. The constrained variational

problem can be expressed as

$$\begin{aligned} \min_{u_k, w_k} \quad & \sum_{k=1}^K \left\| \frac{\partial[(\delta(t) + \frac{j}{\pi t}) * u_k(t)] * e^{-jw_k t}}{\partial t} \right\|_2^2 \\ \text{s.t.} \quad & B(t) = \sum_{k=1}^K u_k(t). \end{aligned} \quad (6)$$

where $j^2 = -1$, w_k is the center frequency of mode $u_k(t)$, and $\delta(t)$ is a Dirac delta function.

The constrained variational problem in (6) can be solved using alternate direction method of multipliers ADMM [34].

Algorithm 2 Adaptive Variational Mode Decomposition

Input: $\hat{u}_k^1, \hat{w}_k^1, n \leftarrow 0$, curvature threshold $C_T \leftarrow 0.3$, mode's number $K \leftarrow 1$, curvature $I_C \leftarrow 0$, mean instantaneous frequency $mif(K)$.

Output: \hat{u}_k^n, \hat{w}_k^n

1: **while** $n \leftarrow n + 1$ **do**
2: **while** $\|I_C\| < C_T$ **do**
3: update \hat{u}_k for all $\omega \geq 0$:

$$\hat{u}_k^{n+1}(\omega) \leftarrow \frac{\hat{b}(\omega) - \sum_{i \neq k} \hat{u}_i(\omega) + \frac{\hat{\lambda}(\omega)}{2}}{1 + 2\alpha(\omega - \omega_k)^2}. \quad (7)$$

4: update \hat{w}_k :
5: $\hat{w}_k^{n+1} \leftarrow (\int_0^\infty \omega |\hat{u}_k(\omega)|^2 d\omega) / (\int_0^\infty |\hat{u}_k(\omega)|^2 d\omega)$.
6: Compute the Hilbert transform for \hat{u}_k .
7: Compute the instantaneous frequency $uif(k)$ for \hat{u}_k based on Hilbert transform.
8: Compute the mean instantaneous frequency $mif(k)$ for \hat{u}_k based on the instantaneous frequency $uif(k)$.
9: Compute the curvature I_C based on $mif(k)$.
9: $k \leftarrow k + 1$.
10: **end while**
11: Execute dual ascent for all $\omega \geq 0$:
12: $\hat{\lambda}^{n+1}(\omega) \leftarrow \hat{\lambda}^n(\omega) + \beta(\hat{b}(\omega) - \sum_k \hat{u}_k^{n+1}(\omega))$.
13: convergence:
14: **if** $\frac{\sum_n^N \|\hat{u}_N^{k+1} - \hat{u}_N^k\|^2}{\|\hat{u}\|_2^2}$ **then**
14: break
15: **end if**
16: **end while**

The complete AVMD algorithm is described in Algorithm 2. Lines 3 and 4 update the modes and their center frequencies. In Equation (7), $\hat{b}(\omega)$ is the frequency-domain presentation of the filtered signal. λ is a quadratic penalty term [34]. α is the variance of white noise. The Hilbert transform for \hat{u}_k is calculated in line 5, based on which the instantaneous frequency $uif(k)$ is computed in line 6 and the mean instantaneous frequency $mif(k)$ is computed in line 7. The curvatures I_C is derived according to $mif(k)$ in line 8. The penalty term λ is updated in line 11 based on an update rate β , and the convergence condition is given in line 12.

After decomposition, we can determine which IMF is the breathing signal and which IMF is the heartbeat signal by comparing the bandwidth of the IMF and the frequency range

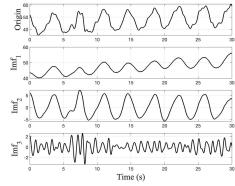
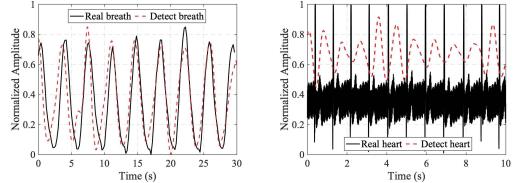


Fig. 7. The signals after AVMD decomposition.



(a) Detected vs real breath
(b) Heart beat pattern for
patterns.

Fig. 8. Vital signs detection.

of breathing/heartbeat. The breathing rate of a person varies from 0.16Hz to 0.6Hz [26], and the heartbeat rate ranges from 1Hz to 2Hz [35]. Fig. 7 shows the decomposition results of our proposed AVMD on a 30s-long signal. AVMD has divided the signal into three modes, i.e., IMF_1 , IMF_2 , and IMF_3 , each corresponding to a certain frequency. For instance, IMF_1 has a frequency of 0.2667Hz, IMF_2 has a frequency of 0.2667Hz, and IMF_3 has a frequency of 1.067Hz. Although the first mode and the second mode have the same frequency, the second mode has the similar pattern with breathing as is shown in Fig. 8(a). AVMD adaptively chooses the optimal number of modes as shown in Algorithm 2. Specifically, when the number of modes is 1, the curvature I_C is zero according to line 8, and then, the number of modes is increased by 1. When the number of modes is 2, the curvature I_C is -0.227 . When the number of modes is 3, the max absolute value of curvature I_C is -0.136 . When the number of modes is 4, the max absolute value of curvature I_C is 0.664 , which is larger than the curvature threshold C_T . Therefore, the number of modes is chosen to be 3. In comparison, traditional VMD has a pre-set number of modes. If the number of modes k is set to 2, the decomposed modes will not contain the pattern of heartbeat. If the number of modes k is set to 4, the frequency of the third mode and the forth mode are 1.033Hz and 1.333Hz , which are far away from the ground-truth heart rate of 1.066Hz . If the number of modes is more than 4, the performance of decomposition is worse. By adjusting the number of modes automatically, our proposed AVMD can realize an ideal decomposition.

We verify that the extracted breathing and heartbeat signals are valid by comparing them with the ground truth. Fig. 8(a) and Fig. 7.3 in Appendix VII-C shows the breathing signal extracted by *DetectDUI* and the ground-truth breathing signal measured by a *Neulog236* breathing belt at the same time. The *Neulog236* breathing belt has a sampling rate of 100Hz. Fig. 8(b) and Fig. 7.3 in Appendix VII-C displays the heartbeat signals extracted by *DetectDUI* and measured by a commercial ECG sensor at the same time. In both cases, there is a close match between the extracted signals and the ground truth, which confirms that our vital sign extraction is quite

accurate, contributing to a precise drink driving detection and BAC prediction.

B. Measuring Psychomotor Coordination

Although vital signs are important indicators of drink driving, we further measure the psychomotor coordination of drivers to improve prediction accuracy. Existing behavior-based drink driving detection methods mostly adopt smart devices to passively monitor user movements or actively require user interactions. Nevertheless, these methods are not applicable to the driving condition since it is unlawful for drivers to play with smartphones while driving. Instead, we have designed a novel way to monitor psychomotor coordination by monitoring the driver's operations of the steering wheel during driving using IMU deployed on the steering wheel. Many car manufactures have already deployed IMU in their cars to monitor the status of the car but not the human. In the future, we expect that IMU will be deployed on the steering wheel to help track the status of the driver. In this way, we realize a continuous non-interrupting monitoring in a user-friendly way.

The IMU on the steering wheel captures the angular velocity $G(t)$ and acceleration $A(t)$ of the driver's operations. Compared with WiFi signals that are easily interfered during transmission, IMU data is relatively stable. We process the IMU data in two steps.

First, we clean the data using a low-pass Butterworth filter to eliminate high-frequency noise. Since most human body movements are within the range of $0 \sim 20\text{Hz}$ [36], we set the cutoff frequency as 40Hz .

Then, we segment the data so that each data segment corresponds to a single steering wheel operation. The key to data segmentation is to pinpoint the start and the end of each operation. We leverage the gyroscope signal rather than the acceleration signal for data segmentation due to the following reason. If the driver turns the steering wheel slowly and smoothly, the acceleration signal shows very small fluctuations. But as the direction of the steering wheel alters, the gyroscope signal will exhibit clear changes. When there is no steering wheel operation, the car is going straight, thus the z -axis of the gyroscope is zero. If the driver turns the steering wheel to the left or the right, the angular velocity of the car changes, which is reflected by the z -axis data of the gyroscope. In our experiment, we decide that an operation starts if the z -axis reading of the gyroscope is greater than 0.01rad/s and the operation ends when the reading returns to zero. The acceleration data is synchronized with the gyroscope data so that we segment the acceleration data accordingly.

As shown in Fig. 9, the vertical dotted lines illustrates the demarcation of each operation. According to the gyroscope, *DetectDUI* extracts 12 steering wheel operations, based on which the acceleration signal is segmented. Note that the acceleration signal itself is much less informative in showing the start and the end of each operation.

C. Extracting Features

After signal pre-processing, we extract the most informative features from the signals for drink driving detection.

TABLE I
PEARSON, SPEARMAN AND KENDALL CORRELATION COEFFICIENTS

		Pearson	Spearman	Kendall			Pearson	Spearman	Kendall
Vital	Heart rate	0.651	0.662	0.563	Vital	Breath rate	0.291	0.246	0.221
	Skewness	0.049	0.117	0.077		Skewness	0.115	0.158	0.107
	CV	0.171	0.166	0.111		CV	0.169	0.179	0.126
	Energy	-0.030	-0.025	-0.014		Energy	0.099	0.089	0.061
	Max amp.	0.063	0.057	0.037		Max amp.	0.062	0.067	0.046
	Mean amp.	-0.122	-0.112	-0.082		Mean amp.	0.046	0.016	0.012
	Median amp.	-0.104	-0.092	-0.064		Median amp.	0.015	0.008	0.007
	Min amp.	-0.122	-0.118	-0.082		Min amp.	0.048	0.080	0.056
	Std deviation	0.159	0.157	0.105		Std deviation	0.218	0.214	0.143
	Variance	0.139	0.157	0.106		Variance	0.203	0.214	0.143

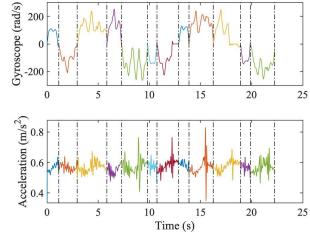


Fig. 9. The gyroscope and acceleration data from steering operations.

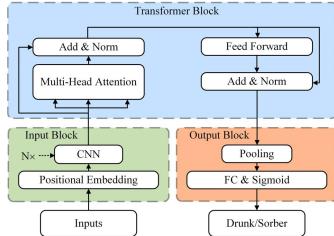


Fig. 10. The C-attention architecture.

To account for both vital signs and psychomotor coordination, we compute the following features.

1) *Vital Signs*: Breath rate and heartbeat rate.

2) *Psychomotor Coordination*:

- Time-domain features, including statistics (mean, median, min, max, standard deviation, variance), skewness (the third moment), and coefficient of variation (the ratio of the standard deviation to the mean multiplied by 100).
- Frequency-domain features, including energy and the square sum of the frequency spectrum related to steering operations.

To confirm that the extracted features are indicative factors of drink driving, we conduct a comprehensive correlation analysis on the relationship between individual features and BAC. Pearson's correlation analysis is used to analyze linear correlation, with ± 1 representing strong positive/negative linear relationship while 0 representing no linear relationship. Spearman's rank correlation coefficient is used to characterize monotonic relationship, with ± 1 indicating monotonically increasing/decreasing relationship, and 0 indicating no monotonic relationship. Kendall rank correlation is used to test similarity, with ± 1 representing consistent positive/negative rank correlation, and 0 means no such relationship.

As shown in Table I, vital sign features have strong positive correlations with BAC with the highest correlation being between the heart rate and BAC, which means that alcohol consumption significantly increases the breathing rate and the heart rate of drivers. The energy, mean amplitude, median

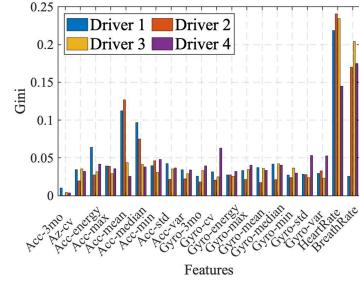


Fig. 11. Gini coefficient for different features of four participants.

amplitude and min amplitude are negatively correlated with BAC, but the max amplitude, standard deviation and variance are positively correlated with BAC. This indicates that the energy level of the driver is reduced due to the alcohol, yet the operations of the driver are more sudden and more rough. All of the gyroscope features are positively correlated with the BAC. This shows that the driver turns the steering wheel at larger angles after drinking, which is an indicator of rash behaviors.

We further analyzed the influence of each feature on the model in terms of Gini coefficient, which reflects the purity of the features, calculated as

$$Gini(D) = \sum_{k=1}^K \sum_{k' \neq k} p_k p_{k'} = 1 - \sum_{k=1}^K p_k^2, \quad (8)$$

where D donates the dataset, p_k is the proportion of samples that belong to class k , and K is the number of classes.

As shown in Fig. 11, vital sign features have larger Gini coefficients than the psychomotor coordination features, which also conforms with the correlation analysis results in Table I. This further confirms the important role that vital signs play in detecting drunkenness. Our extensive effort to purify the vital sign signals for accurately extracting breathing and heartbeat rates enables us to reach high detection accuracy for drink driving and BAC.

D. Detecting Drink Driving

Given the carefully extracted vital signs and psychomotor coordination features, we need a well-performed classifier to determine whether the driver is sober or drunk. As discussed in Section IV-C, different features have different contributions in predicting drink driving and BAC. The problem is how to combine information from vital signs and psychomotor coordination effectively for drink driving detection. The detection model should also be lightweight to be deployed

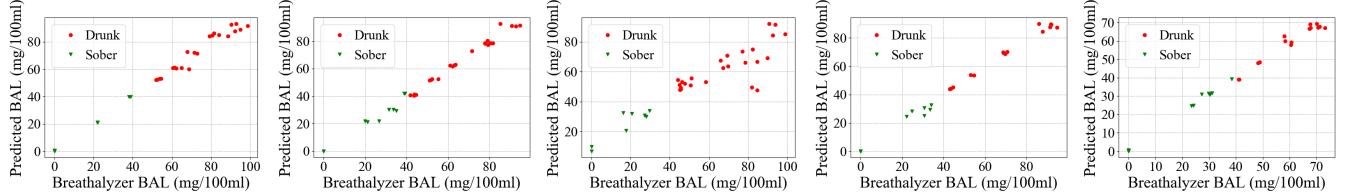


Fig. 12. The performance of BAC prediction of the 15 participants. The rest participants are shown in Appendix VII-B. The green triangle indicates a sober state, and the red circle indicates a drunk state.

in an in-car system. To solve this problem, we propose a C-Attention network with attention mechanism [37]. The self-attention mechanism in the C-Attention network explores the relationship between any two features, resulting in a holistic integration of the vital signs and psychomotor coordination information. The self-attention mechanism of C-Attention also requires less computational resources than traditional CNN. Furthermore, to improve the predictive power of the model without increasing model complexity, the C-Attention network introduces a stack of N one-dimensional convolution layers. In our experiments in Section V-D, we have compared the C-Attention network with commonly-used machine learning algorithms and demonstrated that C-Attention network has the highest prediction accuracy.

Fig. 10 shows the architecture of the drink driving detection model. The model is made up of three blocks: the input block, the transformer block, and the output block.

Input Block: The input block comprises the positional embedding layer and a stack of N one-dimensional convolution layers. The positional embedding layer contains two trainable sublayers: one transfers the features to tokens, and the other assigns indexes to the tokens [37]. With position embedding layers, C-Attention network takes position information of input features into consideration and explores the relationship between features. To optimize the embedding vector derived from the position embedding layer, C-Attention network adds a stack of N one-dimensional convolution layers. In our implementation, we set $N = 3$.

Transformer Block: The transformer block consists of two sub-layers: a multi-head self-attention block and a feed-forward network. After each sub-layer, there is a dropout layer, and the dropout rate is 0.1.

The output of the multi-head self-attention block is

$$\text{MultiHead}(Q, K, V) = \text{Concat}(\text{head}_1, \dots, \text{head}_n)W^O,$$

where

$$\text{head}_i = \text{Attention}(QW_i^Q, KW_i^K, VW_i^V),$$

where $W_i^Q \in \mathbf{R}^{d_{model} \times d_k}$, $W_i^K \in \mathbf{R}^{d_{model} \times d_k}$, $W_i^V \in \mathbf{R}^{d_{model} \times d_v}$ and $W^O \in \mathbf{R}^{hd_v \times d_{model}}$. n is the number of parallel heads. In our implementation, we set $n = 4$, and $d_k = d_v = d_{model}/n = 4$.

At each of the two sub-layers, we utilize a residual connection [38] in the transformer block, followed by a normalization layer, presented as “Add & Norm” in Fig. 10.

Output Block: The output block contains a pooling layer and a fully connected layer. After each layer, there is a dropout layer, and the dropout rate is 0.2. The output block produces

a binary value predicting the state of the driver, i.e., sober or drunk, derived from a self-attention matrix by a sigmoid function as

$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V, \quad (9)$$

where Q is the result of a set of queries calculated by the attention function, K and V are the keys and values, $*^T$ is matrix transformation, and $\frac{1}{\sqrt{d_k}}$ is a scaling factor.

The multi-head attention mechanism in the C-Attention model assigns different weights to different elements of the input so that features of different importance can be leveraged to reach an accurate prediction result. Moreover, the depth of the C-Attention network is relatively small, thus its computational and storage costs are low, making it suitable for the on-board computing system of automobiles. In the experiments, we compare the performance of C-Attention model and other commonly used classification models, and confirm its superiority in achieving higher prediction accuracy.

In our implementation of the C-Attention model, the input feature map size is $1 * 20$. The output feature map size of the positional embedding is $20 * 16$, and the output feature map sizes of the 3 convolution layers is $20 * 64$, $20 * 32$, and $20 * 16$, respectively. The output feature map size of the transformer block is $20 * 16$. The output feature map size of the fully-connected layer in the output block is $1 * 20$. A sigmoid function is applied to the output of the fully-connected layer to yield the final prediction result.

E. Predicting BAC Level

To further investigate how drunk the driver is, we conduct a more precise BAC level prediction. Deep learning networks, though being applied to many other scenarios, are not suitable for our application for two main reasons. First, the feature space and the data size are much smaller than those of image processing or speech recognition. Deep learning networks usually suffer from overfitting problems in the case of small training dataset and feature space. Second, deep learning networks usually have prohibitively high overheads, and not resource-effective for the in-car system. Therefore, we resort to lightweight learning algorithms for BAC prediction.

We compare the performance of various lightweight regression models, including Multilayer Perceptron (MLP), Support Vector Machine (SVM), and Random Forest (RF). The results in Section V show that Random Forest (RF) [39] has the best performance. Using the decision tree as the base learner, RF adopts a random choosing mechanism. Different from the traditional decision tree that selects the optimal attribute



Fig. 13. The simulation scenario.

to partition the current node, RF randomly selects a subset of attributes and then chooses an optimal attribute from this subset to divide the node of a base learner. In this way, RF is highly resistant to overfitting.

V. EVALUATION

A. Experiment Settings

We conduct most experiments in a laboratory environment for safety concerns. Fig. 13 displays the simulation scenario. The driving simulator was run on a computer with a steering wheel controller and an attached IMU. A Lenovo G200 laptop equipped with Intel 5300 WiFi card is used to simulate the on-board unit (OBU). A mobile phone hotspot (HuaweiMate8) attached to a Lenovo Y200 laptop served as a proxy for an in-car WiFi system to form a WLAN. The vital signs of a participant were collected via the WiFi, and the psychomotor coordination data of a participant was gathered by the IMU. Each participant wore an NUL236 breathing belt and ECG sensors to collect ground-truth breathing and heart rate data. Ground-truth BAC values were taken using a breathalyzer.

There are 15 adult volunteers, including five females and ten males. In keeping with the NIAAA's stipulation that the legal drinking age in the US is 21, all participants were aged from 21 to 26, proven by an identification card. Each participant was required to confirm that they do not have a family history of excessive drinking, are not pregnant, and do not take medications that mix with alcohol. All experiments were conducted between 9-11 am or 3-5 pm.

The National Advisory Council on Alcohol Abuse and Alcoholism (NIAAA)³ stipulates that a sober person's BAC should not exceed 0.04%. U.S. Department of Health & Human Services stipulates that a binge-drinking BAC level is 0.08% or above.⁴ Therefore, we use 0.04% as the threshold to differentiate sober and drunk, and 0.08% as the upper bound to limit alcohol intake in our experiments.

Prior to the experiments, all participants were verified to be sober with a breathalyzer and advised to stop the experiment if they felt sick at any time. They were also at liberty to stop the experiment for any reason. No participant was allowed to leave the laboratory until their BAC was below 0.02% to ensure their safety.

To begin with, each volunteer was provided with a standard drink (12oz. of regular beer with 5% alcohol, 5oz. of wine with 12% alcohol, or 1.5oz. of 80-proof distilled spirits with 40% alcohol) [40], [41]. To reach a predefined BAC of 0.08%,

we estimated the drinking capability of each participant based on their weight, the alcohol's proof [42], and the prescribed frequency [43]. As a safety precaution, we only gave the participant one standard drink at a time followed by a 20-minutes rest and a breathalyzer test. Usually, a female needs about four standard drinks and a male needs about five standard drinks to reach BAC of 0.08%. Each drink loosely raises a person's BAC by around 0.02% and, typically, it takes 15–20 minutes for the human body to absorb the alcohol after drinking. Before each breathalyzer, the participant rinsed their mouth with water to avoid interference. If their BAC was lower than 0.02%, they were given another standard drink. If the participant's BAC was higher than 0.08%, they stopped drinking.

Once the participant's BAC had reached 0.08%, we started the experiments. Each volunteer was instructed to sit on a chair with their hands on the steering wheel controller while performing the driving simulation. To simulate the real driving experiences and to increase the diversity of driving conditions, we set four driving terrains/circuits in the simulator. Each circuit contains about 10 intersections and participants can change lanes about twice in each circuit. A participant randomly selects one circuit at each experiment. The laptop was placed in front of the simulator, and the mobile phone was next to the laptop. The WiFi and the IMU data was recorded continuously. We collect the data samples every 2.5 minutes as an epoch. The WiFi signal and the IMU data within the 2.5 minutes are used for predicting the current BAC level. At the end of each epoch, the volunteer was asked to blow into the breathalyzer to collect ground-truth BAC level. The entire experiment lasted for around 240 minutes for each participant, with around 96 data samples collected over the period. For each volunteer, we train the drunkenness detection and BAC prediction model using his/her own data. We use 70% of a person's data for training and the remaining 30% for testing. In Section V-F, we evaluate the performance of complementing the training dataset of a new user with other users' data.

We use accuracy, precision, recall, and F1-score to evaluate the performance of drink driving detection. Let TP , TN , FP , and FN denote the number of true positive, true negative, false positive, and false negative samples, respectively. Accuracy is computed as the ratio of correctly identified samples to all samples, i.e., $(TP + TN)/(TP + TN + FP + FN)$. Precision is the ratio of correctly identified positive samples to all identified positive samples, i.e., $TP/(TP+FP)$. Recall is defined as the ratio of correctly identified positive samples to all real positive samples, i.e., $TP/(TP + FN)$. F1-score strikes a balance between precision (PR) and recall (RR), calculated as $2(PR*RR)/(PR+RR)$. We use Mean Absolute Error (MAE) to evaluate the performance of BAC prediction. MAE is computed as the average absolute value of the difference between the predicted and the ground-truth BAC values.

B. Overall Performance

One customized classification/regression model is trained and tested for each participant, following the practice of

³<https://www.niaaa.nih.gov/Resources/ResearchResources/job22.htm>

⁴<https://www.cdc.gov/alcohol/fact-sheets/binge-drinking.htm>

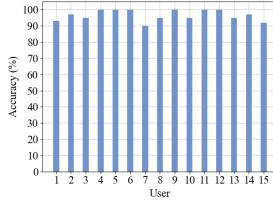


Fig. 14. The accuracy of drink driving detection.

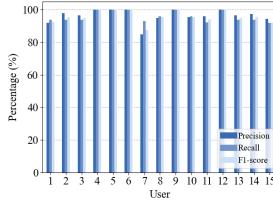


Fig. 15. The precision, recall and F1-scores of drink driving detection.

existing works [6]–[8]. Given the collected data of a particular volunteer, we use 70% for training and 30% for testing. The training is conducted with 10-fold cross-validation [44].

Vital Sign Detection: Fig. 30 in Appendix VII-D shows the CDF of errors of extracted breath rate. *DetectDUI* achieves a median error of 0.08 respirations per minute (rmp) for the participant, and 90% of the errors are smaller than 0.2 rmp. In contrast, in [45], around 80% of the errors are smaller than 0.2 rmp. The results for extracted heart rates are demonstrated in Fig. 31 in Appendix VII-D. We can observe that 90% of the errors are less than 3 beats per minute (bmp), and 75% of the errors are below 2 bmp. As a comparison, in [28], 90% of estimation errors are no more than 4 bpm. These results confirm that our WiFi-based contactless system can extract reliable vital signs, which is essential in accurate drink driving detection.

Drink Driving Detection: The performance of drink driving detection is shown in Fig. 14. *DetectDUI* yielded an average accuracy of 96.6%. Among the 15 participants, participant 7 has the lowest accuracy of 90%. We observed that this participant is in an ecstatic mood, which may have affected the prediction accuracy. We evaluate the influence of mood on the performance of *DetectDUI* later in this section. Fig. 15 shows the precision, recall, and F1-scores for drink driving detection. The lowest precision is above 85%, and the average precision is 96.43%. The recall is no less than 92%, with an average of 95.95%. The F1-scores is higher than 87.5%, with an average of 95.87%. As a comparison, the recall of Mariakakis et al [8] is 93.9% and the accuracy of Bae et al. [7] is 96.6% in detecting three categories of inebriation. Both [8] and [7] require users to interact with smartphones, which is not suitable for drivers. These results verify *DetectDUI* as an effective tool for predicting the drunkenness in drivers.

BAC Prediction: Fig. 12 shows the regression results for BAC predictions. The x-axis is the ground-truth BAC measured by the breathalyzer, and the y-axis is the predicted BAC by *DetectDUI*. The 45-degree diagonal line is a perfect match between the true and the predicted BAC levels. We can observe that most predicted BAC fall close to the true BAC detected by the breathalyzer. Fig. 16 presents the MAE of BAC prediction.

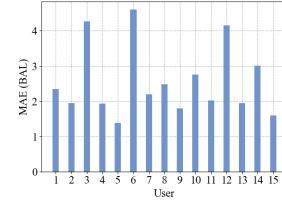


Fig. 16. The MAEs of BAC prediction.

TABLE II
ABLATION STUDY

		Acc.	Precision	Recall	F1-score
Driver 1	psychomotor	72%	78%	67%	72%
	vital signs	93%	100%	88%	93%
	ours	97%	98%	94%	96%
Driver 2	psychomotor	79%	80%	84%	87%
	vital signs	100%	100%	100%	100%
	ours	100%	96%	100%	98%
Driver 3	psychomotor	72%	64%	98%	78%
	vital signs	91%	90%	88%	93%
	ours	95%	97%	94%	95%

The average MAE is 2.66 mg/dl , i.e., less than 5 mg/dl . As a comparison, the absolute mean error of Mariakakis et al [8] is $5 \text{ mg/dl} \pm 7 \text{ mg/dl}$. This confirms that *DetectDUI* can provide an accurate BAC prediction.

C. Ablation Study

We conduct an ablation study to investigate whether the combination of vital signs and psychomotor coordination indeed reach a more accurate prediction with five participants. As shown in Table II, the prediction accuracy based on both vital signs and psychomotor coordination is the highest. With only the psychomotor coordination, the prediction accuracy is less than 80%. The predictive power of vital signs is higher than that psychomotor coordination, but for some users, e.g., driver 4 and 5 whose results are shown in Appendix VII-F, the prediction accuracy based only on vital signs is less than 90%.

D. Performance of Different Machine Learning Algorithms

We use C-Attention, MLP, SVM and RF for classifying whether the driver is sober or drunk. And we use MLP, SVM and RF as regression models to predict BAC. We have introduced C-Attention network in Section IV-E.

MLP is a deep feed-forward network with an input layer, a hidden layer, and an output layer. The layers are fully connected, and each node in the network, except for the input layer, is a neuron with an activation function. We used the default Relu activation function.

SVM is a learning model with a learning kernel that calculates an optimized segmentation hyperplane. We use support vector classification to generate the classifier, a support vector regressor to produce the regressor, and a linear learning kernel. For all three models, we use standard cross-entropy for the loss function.

The Random Forest model had 1000 estimators. We use two layers for classification with L-BFGS as the solvers and three layers for regression with stochastic gradient descent. Based

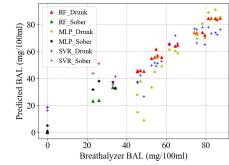


Fig. 17. The regression results of different machine learning algorithms.

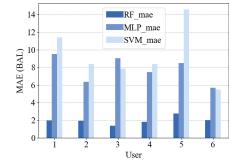


Fig. 18. The comparison of different machine learning algorithms for BAC prediction.

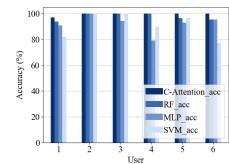


Fig. 19. The comparison of different machine learning algorithms for drink driving detection.

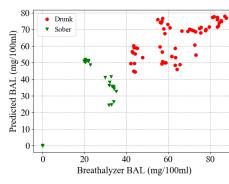


Fig. 20. BAC prediction with cold-start.

on the validation performance, we set the size of the hidden layers to (300, 20) and (6000, 400, 20).

The results of drink driving detection are presented in Fig. 19. C-Attention delivers the highest accuracy. This indicates that the attention mechanism we adopt can focus on the features that are most indicative of drink driving, resulting in the best performance in detecting drink driving than other classifiers. The results of BAC prediction are presented in Fig. 17 and Fig. 18. As shown in Fig. 17, all methods have good performance when the participants are drunk. However, when the BAC of participants is lower than the drink driving separation level of 0.04%, MLP and SVR cannot predict the BAC correctly. Fig. 17 depicts the MAE of each method. It is clear that RF has the best prediction performance.

To validate that C-attention network is suitable to be deployed in an in-car detection system, we evaluate the memory consumption and computational cost of the model. We have implemented the C-attention network using Python-3.7.10 and Tensorflow-2.2 on a MacBook Pro with Intel Core i5 CPU at 2.4GHz and 8G memory. The memory consumption and computational cost of C-attention network are 3.7M and 10.51ms. The computational cost of C-attention network is relatively low because the transformer block has no convolutional layer. The transformer block leverages the

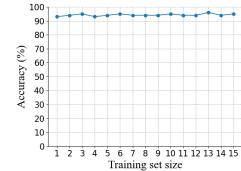


Fig. 21. BAC prediction with different training set size.

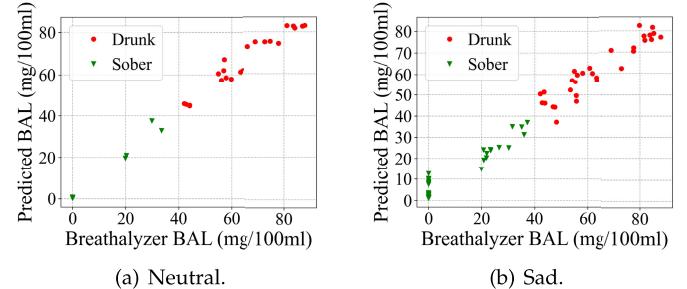


Fig. 22. BAC prediction results with sad participants.

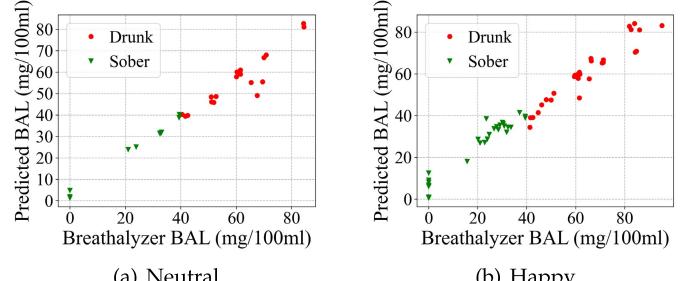


Fig. 23. BAC prediction results with happy participants.

multi-head attention mechanism to reach high accuracy with low computational costs.

E. Impact of Emotions

Our experience with Participant No.7 indicated that emotions may influence a driver's vital signs to the extent that it impacts *DetectDUI*'s performance. Sadness interferes with the cardiovascular adjustments normally associated with exercise, while happiness increases the human heart rate. To evaluate the impact of emotions on *DetectDUI*, we have conducted a set of experiments with the same participants under the influence of different emotions, namely, happiness and sadness.

The experimental set up was exactly the same except that the participants were under the influence of different emotions. Each participant was first tested in a neutral state, then incited to sadness with an abundance of video, music, and pictures and tested again. After that, in a separate session after a certain period of time, the participant is tested again in a neutral state, followed by a test after being exposed to happy materials. We conduct the two sessions separately to ensure the stability of emotions.

The prediction results are shown in Fig. 22 and Fig. 23. The classification accuracy is 96.55% (MAE 2.75mg/100ml) in the first neutral test, 93.10% (MAE 3.48mg/100ml) in the second neutral test, 94.83% (MAE 4.5mg/100ml) in the sad test, and 91.38% (MAE 4.55mg/100ml) in the happy test. It is clear that emotions affects the breathing and heartbeat

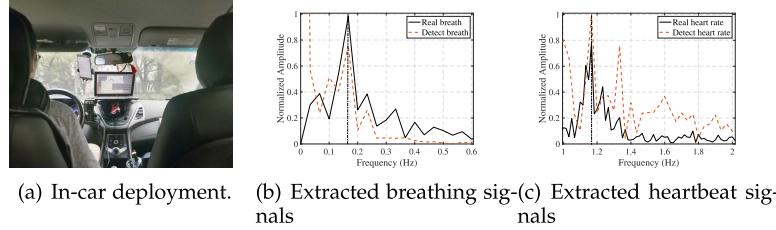


Fig. 24. The deployment of *DetectDUI* on cars and the performance of Wi-Fi signal processing techniques.

rates of a person, thus interferes with the prediction capability of *DetectDUI*, but the influence of emotions on *DetectDUI* is a slight one. The performance degradation due to the sad emotion is larger than the happy emotion, which indicates that sad emotion is more influential than happy emotion on vital signs. Although different emotional states of drivers can affect the performance of *DetectDUI*, the performance degradation is limited.

F. Impact of Training Set

As a final test, we investigated how *DetectDUI* performs with different training sets. In the first set of experiments, we use one participant's data to train a model and then tested it with data collected for all the other participants. This results in an average accuracy of 83% for drink driving detection. The average precision, recall, and F1-scores are 89%, 83%, and 85%, respectively. Moreover, as Fig. 20 shows, the MAE for BAC estimation rises to 13.47. *DetectDUI* shows a certain level of robustness to the cold-start problem. We may train a default prediction model based on massively collected historical user data, preinstalled with each new system. For a new user of *DetectDUI*, the preinstalled model would help to shorten individual training period before the model can accumulate enough personal data to build a high-quality personalized model. In the second set of experiments, we mix the data of several participants, and conduct 10-fold validation. The results, reported in Fig. 21, show a small performance gain as the amount of training data increases. It suggests that the prediction accuracy will increase with more volunteers in the training set.

To study the impact of imbalanced number of positive and negative samples of the training data, we have conducted experiments with data from two different volunteers. For each volunteer, we have collected more negative samples (i.e., sober) with extra experiments. For each volunteer, we have trained two prediction models: one with 90 positive samples and 90 negative samples, and the other with 90 positive samples and 180 negative samples. The experiment results are as follows. For the first volunteer, the detection accuracy, precision, recall, F1-score and false alarm rate of the first model is 100%, 100%, 100%, 100%, and 0% respectively. The detection accuracy, precision, recall, F1-score and false alarm rate of the second model are the same as the first model. For the second volunteer, the detection accuracy, precision, recall, F1-score and false alarm rate of the first model are 96%, 98%, 94%, 96%, and 8.69% respectively. The detection accuracy, precision, recall, F1-score and false alarm rate of the second model are 95%, 95%, 95%, 95%, and 3.92% respectively.

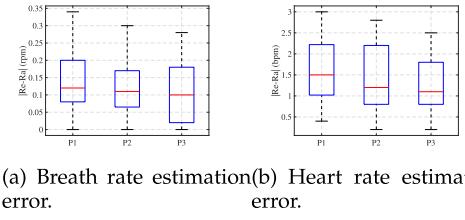


Fig. 25. Vital signs estimation error when the passenger sits at different seats.

We can see that the detection performance indeed drops. Data imbalance is a common problem in machine learning tasks, e.g., wake-up word detection, user authentication, and so on. There are various existing techniques to mitigate this problem, e.g., data augmentation [46].

G. Experiments on Real Roads

To verify the effectiveness of the proposed signal processing algorithms for WiFi signals, we deploy *DetectDUI* on real cars as shown in Fig. 24(a). The WiFi signals are collected by a Lenovo Y200 laptop placed above the dashboard of the car. The privacy leakage is limited since no camera is present and wireless signals are present everywhere. There are many factors that may impact the WiFi signals in the real environments, e.g., noises, glitches caused by bumpy roads, interference from passengers, and so on. In our experiments, each volunteer drives for 1 hour, thus he/she may experience all kinds of road conditions. We also evaluated the case where there are passengers at different seats of the car.

We have recruited five volunteers to drive the car on real roads, each driving for around 1 hour. To obey the laws and guarantee safety, we only conducted tests with sober drivers to evaluate the proposed signal processing techniques.

Fig. 32 in Appendix VII-E show the effect of AVMD performed on real driving traces. Fig. 24(b) and Fig. 24(c) display the extracted breathing and heartbeat signals of real driving traces compared with a Neulog236 breathing belt and a commercial ECG sensor at the same time. We can observe that when deployed in cars and tested on real roads, the proposed signal processing techniques are able to extract good breathing and heartbeat signals, which lead to reliable drunkenness detection.

To validate the performance of the proposed algorithm in eliminating multipath effect from other passengers, we asked another volunteer to sit at the co-pilot seat (P1), the seat behind the driver (P2) and the seat behind the co-pilot (P3) while each volunteer is driving. The volunteer sat at each seat for 20 minutes during driving.

Fig. 25 displays the estimation errors of the vital signs of breath rate and heart rate. It can be seen that DetectDUI has the best performance under P2 and P3. When the passenger sits at the seat behind the copilot (P3), the estimation errors of breath rate and heart rate are the lowest since the passenger has minimal impact on the DetectDUI.

VI. CONCLUSION AND DISCUSSION

In this paper, we presented *DetectDUI*, a non-intrusive, contactless, and continuous system of measuring and monitoring the side effects of alcohol on drivers. To develop *DetectDUI* to this stage, we have overcome two main challenges. The first is to eliminate interference in the WiFi signals caused by the motions of a moving vehicle. This problem was solved with a series of signal processing algorithms. The second is determining which specific features of alcohol's side effects best reflect driving under the influence of alcohol. We have addressed this challenge with a C-Attention network. The results of extensive experiments confirm that *DetectDUI* provides highly accurate drink driving detection and BAC prediction.

Apart from drinking alcohols, other factors may also affect vital signs and psychomotor coordination, e.g., catching a cold or other respiratory diseases. Respiratory diseases will change breathing patterns, which are expected to be different from the breathing patterns of drinking. However, it is difficult to collect training samples to help differentiate the breathing patterns under the two conditions. In the future, we intend to refine our drink driving detection model by considering other impact factors.

REFERENCES

- [1] *Alcohol Impaired Driving: 2018 Data (Traffic Safety Facts. Report No. DOT HS 812 864)*. Accessed: Sep. 2019. [Online]. Available: <https://crashstats.nhtsa.dot.gov/Api/Public/ViewPublication/812864>
- [2] *H.R. 2138: Honoring Abbas Family Legacy to Terminate Drunk Driving Act of 2021*. Accessed: Mar. 2021. [Online]. Available: <https://www.govtrack.us/congress/bills/117/hr2138>
- [3] *S.1331—RIDE Act of 2021*. Accessed: 2021. [Online]. Available: <https://www.congress.gov/bill/117th-congress/senate-bill/1331/text>
- [4] C.-W. You *et al.*, “Enabling personal alcohol tracking using transdermal sensing wristbands: Benefits and challenges,” in *Proc. 21st Int. Conf. Hum.-Comput. Interact. Mobile Devices Services*, Oct. 2019, pp. 1–6.
- [5] Y. Jung, J. Kim, O. Awofeso, H. Kim, F. Regnier, and E. Bae, “Smartphone-based colorimetric analysis for detection of saliva alcohol concentration,” *Appl. Opt.*, vol. 54, no. 31, pp. 9183–9189, 2015.
- [6] H.-L. Kao, B.-J. Ho, A. C. Lin, and H.-H. Chu, “Phone-based gait analysis to detect alcohol usage,” in *Proc. ACM Conf. Ubiquitous Comput. (UbiComp)*, 2012, pp. 661–662.
- [7] S. Bae *et al.*, “Detecting drinking episodes in young adults using smartphone-based sensors,” *Proc. ACM Interact., Mobile, Wearable Ubiquitous Technol.*, vol. 1, no. 2, pp. 1–36, Jun. 2017.
- [8] A. Mariakakis, S. Parsi, S. N. Patel, and J. O. Wobbrock, “Drunk user interfaces: Determining blood alcohol level through everyday smartphone tasks,” in *Proc. CHI Conf. Hum. Factors Comput. Syst.*, Apr. 2018, pp. 1–13.
- [9] *Alcohol*. Accessed: 2021. [Online]. Available: <https://en.wikipedia.org/wiki/Alcohol>
- [10] *Alcohol’s Effects on the Body*. Accessed: 2021. [Online]. Available: <https://www.niaaa.nih.gov/alcohols-effects-health/alcohols-effects-body>
- [11] *Hangovers*. Accessed: 2021. [Online]. Available: <https://www.niaaa.nih.gov/publications/brochures-and-fact-sheets/hangovers>
- [12] *What Alcohol Really Does To Your Body*. Accessed: 2005. [Online]. Available: <https://www.independent.co.uk/life-style/health-and-families/health-news/what-alcohol-really-does-your-body-5349097.html>
- [13] *Alcohol and Cardiovascular Disease*. Accessed: 2020. [Online]. Available: <https://alcoholthinkagain.com.au/alcohol-your-health/alcohol-and-long-term-health/alcohol-and-cardiovascular-disease/>
- [14] C. Wu, K. Tsang, H. Chi, and F. Hung, “A precise drunk driving detection using weighted kernel based on electrocardiogram,” *Sensors*, vol. 16, no. 5, p. 659, May 2016.
- [15] S. Brunner *et al.*, “Alcohol consumption, sinus tachycardia, and cardiac arrhythmias at the munich oktoberfest: Results from the munich beer related electrocardiogram workup study (MunichBREW),” *Eur. Heart J.*, vol. 38, no. 27, pp. 2100–2106, Jul. 2017.
- [16] *First Aider’s Guide to Alcohol by Rick Curtis, Director, Outdoor Action*. [Online]. Available: <https://www.princeton.edu/~oa/safety/alcohol.shtml>
- [17] *Drink Driving UK Law History*. [Online]. Available: https://www.drinkdriving.org/drink_driving_information_uklawhistory.php
- [18] *Breathmeter*. [Online]. Available: <https://www.breathometer.com/>
- [19] *BACtrack Mobile Pro*. [Online]. Available: <https://www.bactrack.com/products/bactrack-mobile-smartphone-breathalyzer>
- [20] *Breathalyzer*. [Online]. Available: <https://en.wikipedia.org/wiki/Breathalyzer>
- [21] *SCRAM CAM Textron*. [Online]. Available: <https://adlmonitoring.com/cam/>
- [22] I. Chatterjee *et al.*, “Driving fitness detection: A holistic approach for prevention of drowsy and drunk driving using computer vision techniques,” in *Proc. South-Eastern Eur. Design Autom., Comput. Eng., Comput. Netw. Soc. Media Conf. (SEEDA CECNSM)*, 2018, pp. 1–6.
- [23] V. Mehta, S. S. Katta, D. P. Yadav, and A. Dhall, “DIF: Dataset of perceived intoxicated faces for drunk person identification,” in *Proc. Int. Conf. Multimodal Interact.*, Oct. 2019, pp. 367–374.
- [24] S. Szymkowski. (2021). *Tesla Interior Cameras Threaten Driver Privacy, Consumer Reports Says*. [Online]. Available: <https://www.cnet.com/roadshow/news/tesla-interior-cameras-driver-privacy-consumer-reports/>
- [25] Z. Yang, P. H. Pathak, Y. Zeng, X. Liran, and P. Mohapatra, “Monitoring vital signs using millimeter wave,” in *Proc. 17th ACM Int. Symp. Mobile Ad Hoc Netw. Comput.*, Jul. 2016, pp. 211–220.
- [26] X. Xu, J. Yu, Y. Chen, Y. Zhu, L. Kong, and M. Li, “BreathListener: Fine-grained breathing monitoring in driving environments utilizing acoustic signals,” in *Proc. 17th Annu. Int. Conf. Mobile Syst., Appl., Services*, Jun. 2019, pp. 54–66.
- [27] H. Abdelnasser, K. A. Harras, and M. Youssef, “UbiBreathe: A ubiquitous non-invasive WiFi-based breathing estimator,” in *Proc. 16th ACM Int. Symp. Mobile Ad Hoc Netw. Comput.*, Jun. 2015, pp. 277–286.
- [28] J. Liu, Y. Wang, Y. Chen, J. Yang, X. Chen, and J. Cheng, “Tracking vital signs during sleep leveraging off-the-shelf WiFi,” in *Proc. 16th ACM Int. Symp. Mobile Ad Hoc Netw. Comput.*, Jun. 2015, pp. 267–276.
- [29] D. Zhang, Y. Hu, Y. Chen, and B. Zeng, “BreathTrack: Tracking indoor human breath status via commodity WiFi,” *IEEE Internet Things J.*, vol. 6, no. 2, pp. 3899–3911, Apr. 2019.
- [30] S. Sen, J. Lee, K.-H. Kim, and P. Congdon, “Avoiding multipath to revive inbuilding WiFi localization,” in *Proc. 11th Annu. Int. Conf. Mobile Syst., Appl., Services (MobiSys)*, 2013, pp. 249–262.
- [31] Y. Y. Jin, W.-S. Soh, and W.-C. Wong, “Indoor localization with channel impulse response based fingerprint and nonparametric regression,” *IEEE Trans. Wireless Commun.*, vol. 9, no. 3, pp. 1120–1127, Mar. 2010.
- [32] A. Ilin and T. Raiko, “Practical approaches to principal component analysis in the presence of missing values,” *J. Mach. Learn. Res.*, vol. 11, no. 1, pp. 1957–2000, 2010.
- [33] S. Ari and G. Saha, “Classification of heart sounds using empirical mode decomposition based features,” *Int. J. Med. Eng. Inform.*, vol. 1, no. 1, pp. 91–108, 2008.
- [34] K. Dragomiretskiy and D. Zosso, “Variational mode decomposition,” *IEEE Trans. Signal Process.*, vol. 62, no. 3, pp. 531–544, Feb. 2013.
- [35] T. Zheng, Z. Chen, C. Cai, J. Luo, and X. Zhang, “V2iFi: In-vehicle vital sign monitoring via compact RF sensing,” *Proc. ACM Interact., Mobile, Wearable Ubiquitous Technol.*, vol. 4, no. 2, pp. 1–27, Jun. 2020.
- [36] H. Zeng, “Bio-inspired inertial sensors for human body motion measurement,” Ph.D. dissertation, Ohio State Univ., Columbus, OH, USA, 2012.
- [37] A. Vaswani *et al.*, “Attention is all you need,” in *Proc. Adv. Neural Inf. Process. Syst.*, 2017, pp. 5998–6008.
- [38] K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for image recognition,” in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2016, pp. 770–778.
- [39] L. Breiman, “Random forests,” *Mach. Learn.*, vol. 45, no. 1, pp. 5–32, 2001.
- [40] L. Blincoe, T. R. Miller, E. Zaloshnja, and B. A. Lawrence. *NIAAA NIH, Drinking Levels Defined | National Institute on Alcohol Abuse and Alcoholism (NIAAA)*. Accessed: 2015. [Online]. Available: <https://www.ncbi.nlm.nih.gov/pubmed>

- [41] *Drinking Levels Defined*. Accessed: 2021. [Online]. Available: <https://www.niaaa.nih.gov/alcohol-health/overview-alcohol-consumption/moderate-binge-drinking>
- [42] *Calculate Your Blood Alcohol Content (BAC)*. Accessed: 2021. [Online]. Available: <http://www.clevelandclinic.org/health/interactive/alcohol-calculator.asp>
- [43] G. J. Schaefer and R. P. Michael, "Effects of an alcohol diet on locomotor activity and the acquisition of brain self-stimulation in rats," *Alcohol*, vol. 8, no. 1, pp. 71–75.
- [44] J. Han, M. Kamber, and J. Pei, *Data Mining: Concepts and Techniques*. Burlington, MA, USA: Morgan Kaufmann, Jan. 2006, pp. 585–631.
- [45] W. Jia, H. Peng, N. Ruan, Z. Tang, and W. Zhao, "WiFind: Driver fatigue detection with fine-grained Wi-Fi signal features," *IEEE Trans. Big Data*, vol. 6, no. 2, pp. 269–282, Jun. 2020.
- [46] J. M. Johnson and T. M. Khoshgoftaar, "Survey on deep learning with class imbalance," *J. Big Data*, vol. 6, no. 1, pp. 1–54, Dec. 2019.



Yanjiao Chen (Senior Member, IEEE) received the B.E. degree in electronic engineering from Tsinghua University in 2010 and the Ph.D. degree in computer science and engineering from The Hong Kong University of Science and Technology in 2015. She is currently a Professor with Wuhan University, Wuhan, China. Her research interests include spectrum management for femtocell networks, network economics, network security, and quality of experience (QoE) of multimedia delivery/distribution.



Meng Xue received the M.S. degree from the School of Electronic Information, Wuhan University, in 2019. He is currently pursuing the Ph.D. degree with the School of Computer Science, Wuhan University. His research interests include the Internet of Things and smart sensing.



Jian Zhang received the B.S., M.S., and Ph.D. degrees in computer architecture from Wuhan University, China, in 1998, 2001, and 2006, respectively. He is currently a Professor with the School of Computer Science, Wuhan University. His research interests include cloud computing, the Internet of Things, and big data.



Runmin Ou is currently pursuing the bachelor's degree with the School of Computer Science, Wuhan University. His research interests include mobile sensing, the Internet of Things, and wearable devices.



Qian Zhang (Fellow, IEEE) received the B.S., M.S., and Ph.D. degrees in computer science from Wuhan University, China, in 1994, 1996, and 1999, respectively. In July 1999, she was at Microsoft Research Asia, Beijing, where she was the Research Manager of the Wireless and Networking Group. In September 2005, she joined the Department of Computer Science and Engineering, The Hong Kong University of Science and Technology, where she is currently a Full Professor. She has published about 300 refereed papers in international leading journals and key conferences in the areas of wireless/internet multimedia networking, wireless communications and networking, wireless sensor networks, and overlay networking. She is a Fellow of IEEE for "contribution to the mobility and spectrum management of wireless networks and mobile communications." She has received MIT Technology Review (MIT TR100) World's Top Young Innovator Award and the Best Asia-Pacific (AP) Young Researcher Award elected by the IEEE Communication Society in 2004.



Peng Kuang is currently pursuing the B.A. degree with the School of Computer Science, Wuhan University. His research interest includes smart sensing.