

# Smart Driver Drowsiness Detection Model Based on Analytic Hierarchy Process

Ting-Cheng Chang,<sup>1</sup> Min-Hao Wu,<sup>1\*</sup> Phan-Zhu Kim,<sup>2</sup> and Ming-Hui Yu<sup>1</sup>

<sup>1</sup>College of Information Engineering, Guangzhou Panyu Polytechnic, Guangzhou, Guangdong 511483, China

<sup>2</sup>College School of Mathematics and Information Engineering, Longyan University, Longyan, Fujian 364012, China

(Received July 20, 2020; accepted November 16, 2020)

**Keywords:** drowsiness detection, electrocardiogram, ECG, heart rate, new features

This paper proposes a smart driver drowsiness detection (SDDD) model for vehicles. The SDDD monitors a driver's heart rate variability (HRV) through electrocardiography (ECG) in real time to detect driver drowsiness. The SDDD processes the data of HRV and ECG to obtain a set of parameters with time-domain analysis, frequency-domain analysis, detrended fluctuation analysis, approximate entropy, and sample entropy. In the process, a machine learning algorithm analyzes the parameters to detect driver drowsiness. The SDDD optimizes critical features with the analytic hierarchy process (AHP), which uses a feature extraction method through an iterative procedure. It is found that the SDDD in this study detects the level of driver drowsiness with higher sensitivity than previous models.

## 1. Introduction

Drowsy driving is dangerous as it frequently causes traffic accidents. Centers for Disease Control and Prevention (CDC) of the U.S.A. reported in 2013 that 1 in 25 adult drivers have fallen asleep while driving and that drowsy driving caused 72000 crashes, 44000 injuries, and 800 deaths in the U.S.A. in 2013, although these numbers are underestimates and up to 6000 fatal crashes may be caused each year.<sup>(1)</sup> There have been many studies to develop an efficient technology to detect drowsy driving based on a driver's heart rate (HR), respiration rate, posture, rapid eye movements, and head movement. Among the various indicators, HR variability (HRV) is known to be the most useful for detecting drowsiness.<sup>(2)</sup> HRV is measured by electrocardiography (ECG) and is used to monitor a driver's physiological and psychological states. ECG is a non-invasive and painless process used to detect the electrical activities of a human's heart. Electrical activities originate from muscle movements of the heart, which generate pulsating electrical waves. ECG requires sophisticated sensors with low noise and high amplification. Advanced sensor technology has enabled more reliable and accurate and faster monitoring of HRV.

HRV devices can detect a driver's states, but not the intensity of the states or the driver's physical condition and its changes. To obtain such information, HRV data need to be analyzed in detail and related with the driver's physical condition.<sup>(3)</sup> In addition, chronological data must

---

\*Corresponding author: e-mail: 18250922163@163.com  
<https://doi.org/10.18494/SAM.2021.3034>

be stored for a certain time period. Then, the necessary information on the physical condition and its changes with time can be extracted by either linear or nonlinear HRV analysis.

There have been many studies on HRV analysis, which mainly focused on spectral modification. Thus, most of the studies performed linear analysis such as time-domain or frequency-domain analysis.<sup>(4)</sup> However, linear analysis will not be appropriate to obtain nonlinear dynamics from HRV data when the data are obtained from a complex system. Thus, nonlinear analyses including abnormal statistical analysis and analysis with the filtering of noise have been suggested to remove the drawbacks of linear analysis.<sup>(5,6)</sup> Nonlinear analysis methods that have been proposed are detrended fluctuation analysis (DFA), wave analysis,<sup>(7)</sup> approximate entropy,<sup>(8)</sup> and sample entropy,<sup>(9)</sup> all of which yield better results than linear analysis for the HRV data of a complex system.

On the basis of the previous analysis methods of HRV data, a new method of analyzing HRV data to detect drowsy driving is proposed in this study. To obtain the best results from the analysis of HRV data, all of the above linear and nonlinear analysis methods are combined to define features and indicators for a new model of smart driver drowsiness detection (SDDD). The result of the new method is compared with that of several combinations of analysis methods. Improved driver drowsiness detection by the new method will minimize traffic accidents caused by drowsy driving.

## 2. Methods

### 2.1 Establishment of SDDD model

The construction process of the SDDD model in this study has the following steps (Fig. 1).

- (1) Preparation for signal detection
- (2) Signal detection of different states of body movements: blinking, closing, opening eyes, and vertical, horizontal, and rotating movements of head
- (3) Calculation of the intersection area of two normal distributions of two indicators
- (4) Calculation of the weights of HRV indicators based on analytic hierarchy process (AHP)
- (5) Machine learning training and model construction

### 2.2 Driving experiment

Twenty-one people (11 males and 10 females) participated in the driving experiment. Their average age was 24.1 ( $\pm 3.41$ ). All of them had held driver's licenses for at least six months. Participants of the experiment were randomly divided into two groups, groups A and B, who carried out simulated drives in the morning (02:00–04:30) and early evening (18:00–20:30), respectively. The times of the day for the simulated drives were chosen on the basis of research<sup>(10)</sup> in which the desire to sleep turned out to be strongest between 02:00 and 04:00 and between 14:00 and 16:30. In the simulated drives, participants rested for 10 min, drove for 90 min, then rested for 30 min again. Physiological parameters (low-frequency and high-frequency power in ECG) were measured for 5 min every 10 min during the experiments. The data were then used to obtain indicators as discussed later.

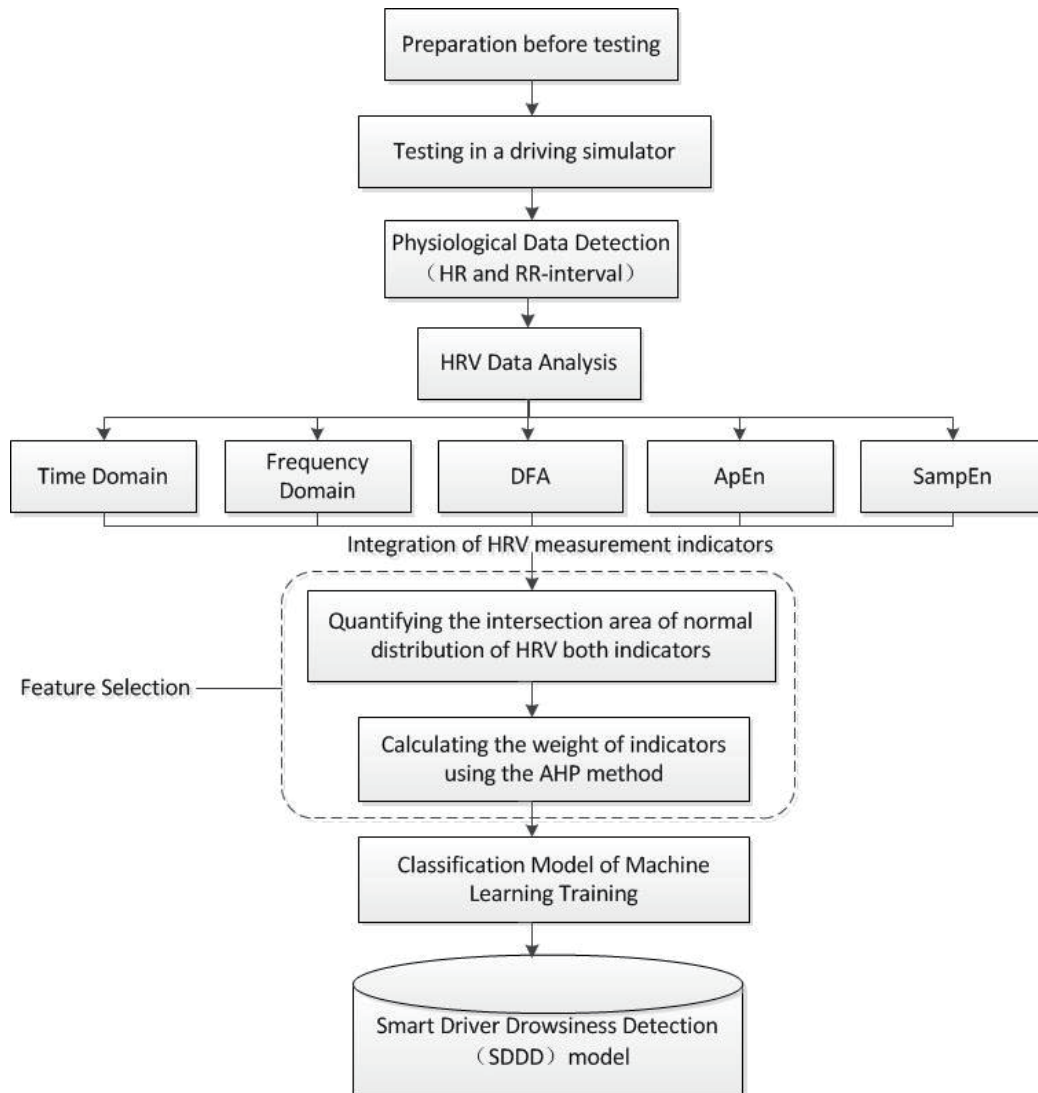


Fig. 1. Architecture used to implement SDDD model.

### 2.3 Experimental devices

The measuring devices for each participant included a chest-worn belt for measuring the HR, a smart mobile device, and an HRV analysis platform (a computer with the SDDD model) (Fig. 2).

- (1) The chest-worn belt had a conductive rubber band to measure the HR. Signals of amplified potential difference generated by the heart were transmitted to a receiving device (the smart mobile device).
- (2) The smart mobile device with a low-power Bluetooth module (4.0 or higher) received the signals of the HR and the R-R intervals of the HR from the belt. The R-R interval is the elapsed time between two successive R-waves. The R-wave is the first upward deflection after a P-wave (atrial depolarization) and part of the QRS complex, which is the main spike

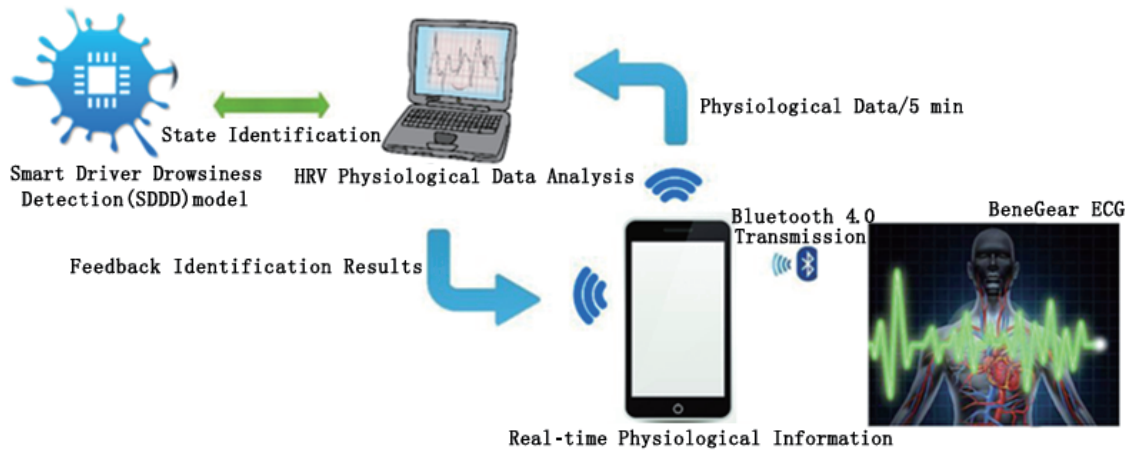


Fig. 2. (Color online) Structure of devices used with SDDD model.

on an ECG line.<sup>(11)</sup> A wireless network transmits the signals to the platform for HRV analysis every 5 min.

- (3) The HRV analysis platform analyzes the signals using the SDDD model in this study.
- (4) The SDDD model identifies a driver's drowsiness.

## 2.4 HRV data analysis

To propose an SDDD model with an AHP to improve the accuracy of detecting drowsy driving, the following different analyses of the signals were carried out.

- Experiment 1: time-domain and nonlinear analyses with feature extraction
- Experiment 2: frequency-domain and nonlinear analyses without feature extraction
- Experiment 3: time-domain, frequency-domain, and nonlinear analyses without feature extraction
- Experiment 4: time-domain, frequency-domain, and nonlinear analyses with feature extraction (the proposed method in this study)

Detection and prediction of driver drowsiness require appropriate physiological data. For accurate prediction, feature extraction was used to select high-sensitivity indicators among 18 indicators for linear and nonlinear analyses as follows: seven indicators for time-domain analysis, seven for frequency-domain analysis, two for DFA, one for the approximate entropy method, and one for the sample entropy method. To select the most sensitive indicators among them, feature extraction was used to determine the degree of mutual dependence of two indicators with the normal distribution, and then an AHP was used to select one for the SDDD model. As a result, 14 indicators were selected to improve the accuracy in detecting driver drowsiness and reduce the probability of misjudgment (Table 1).

*Frequency-domain analysis* calculates the degree of variation of the R-R interval with time series analysis and uses the HR (in beats per minute, BPM) and R-R interval (in ms). Indicators for frequency-domain analysis are average HR ( $\overline{HR}$ ), average HR interval ( $\overline{RR}$ ), standard deviation of HRs (SDHR), and standard deviation of HR intervals (SDNN).

Table 1  
Selected parameters in this study (from MIT-BIH database).

Analysis	Parameter	Description
Time-domain analysis	$\overline{HR}$	Average heart rate
	RR	Average heart rate interval
	SDHR	Standard deviation of heart rates
	SDNN	Standard deviation of heart rate intervals
	R_MSSD	Root mean square of sum of squared differences in heart rate intervals
	NN50	Number of gaps exceeding 50 ms for all pairs of adjacent heart rates
	pNN50	NN50 divided by total number of periods of heart rates
Frequency-domain analysis	TP	Total power (variation in normal heart rate)
	HF	High-frequency power (indicator of parasympathetic nerve activity)
	LF	Low-frequency power (indicator of sympathetic nerve activity)
	VLF	Number of variations in normal heartbeat intervals in very low frequency range
	LF/LH	Indicator of sympathetic regulation
	nHF	Indicator of sympathetic nerve activity
	nLF	Indicator of parasympathetic nerve activity

*Time-domain analysis* uses the root mean square of the sum of squared differences in HR intervals (R\_MSSD), the number of gaps that exceed 50 ms for HRs (NN50), and NN50 divided by the total number of periods of HRs (pNN50). R-MSSD, NN50, and pNN50 are all short-term variability indicators and are mutually correlated. Thus, they are used to evaluate the high-frequency variation of HRV.

*Detrended fluctuation analysis* fragments physiological data to a fixed length and calculates the average and standard deviation of the data in each segment to obtain the trend fluctuations of the data accordingly. The advantage is that the trend fluctuation of the entire segment data is obtained, but the result is easily affected by extreme values.

*Approximate entropy analysis* is a quantified sequence complexity method<sup>(9)</sup> and used to analyze the data of disordered sequences. A data dimension value and a threshold value are set first, and collected physiological data are cut by the data dimension value to obtain several fragments. Then, the proportion of the number of data in each segment above the threshold and the data disorder are calculated as the amount that is algorithmically calculated and is more convenient than measuring bias directly. The amount of information obtained for different data dimensions is defined as ApEn, whose value is between 0 and 1. When an ApEn value is smaller, the disorder of data is less obvious, and the degree of data variation in the sequence is smaller. This analysis is not easily affected by extreme values, but the threshold setting varies.

*Sample entropy analysis* is similar to approximate entropy analysis.<sup>(12)</sup> This analysis calculates the disorder of a system, that is, the degree of data disorder in the system. By using the preset data dimension and threshold value, physiological data are segmented to obtain several fragments. Then the threshold value is used to calculate the degree of data disorder in each segment. Sample entropy analysis mainly considers the proportion of data disorder in different data dimensions. Data disorder decreases when the data dimension is increased. The main advantage is that the proportion of data disorder is obtained, which it is not affected by extreme values. However, this analysis is not applicable with zero data disorder.

Each analysis has its own applicable data distributions. Therefore, all analyses are performed on the physiological data in this study to establish an SDDD model that considers the data characteristics obtained by all analyses.

Indicators from the nonlinear analyses are thought to represent specific body movements. HR is related to the duration of blinking, VLF to the duration of closing eyes, LF to the ratio of the duration of closed to open eyes, HF to the head position on the  $x$  and  $y$  axes, TP to the head position on the  $z$  axis, and nHF and nLF to the rotation of the head along the  $x$  and  $y$  axes, respectively.

## 2.5 Integration of indicators

Calculating the intersection area of normal distributions of two indicators requires a test to ensure that the indicators are normally distributed. From experiments, it is found that at the significance level  $\alpha = 0.05$ , the degree of freedom is  $df = 10$ , the chi-square critical value is  $\chi^2 = 307.182$ , and the verification of the HR data distribution and normal distribution before driving is at the significance level  $\alpha = 0.05$ , degree of freedom  $df = 10$ , and chi-square statistical value 0.982, which is less than the chi-square critical value. Therefore, the HR distribution before exercise conforms to the normal distribution. The chi-square test result of the HR data distribution shows that the data distribution of each category conforms to the normal distribution. Therefore, we can calculate the intersection area based on the normal distribution of all HRV measurement indicators in each category. The distribution of HRV before driving has a chi-square value of 0.982 at a significance level of  $\alpha = 0.05$  and  $df = 10$ . Therefore, the HRV data before driving are normally distributed.

## 2.6 Feature selection

### 2.6.1 Calculation of intersection area

Figures 3 and 4 show the overlapping areas (intersection areas) of the two normal distributions for the two groups. The intersection area of group A is smaller than that of group

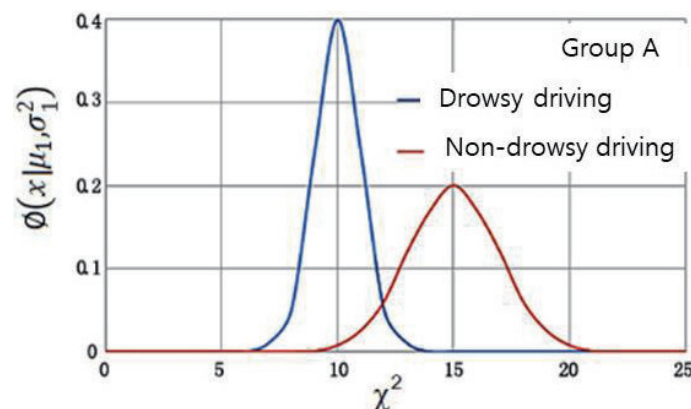


Fig. 3. (Color online) Normal distribution of group A for drowsiness and non-drowsiness.

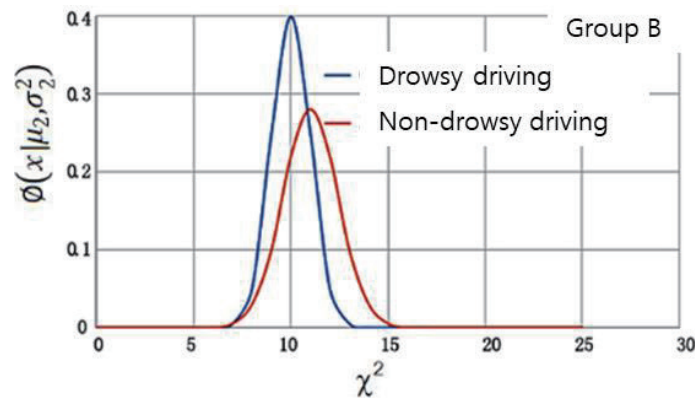


Fig. 4. (Color online) Normal distribution of group B for drowsiness and non-drowsiness.

B. Group A performed simulated drives in the early morning and seemed to have a clearer difference in the indicators between drowsy and non-drowsy driving.

The integrated indicators of the drowsy and non-drowsy driving of groups A and B respectively have normal distributions  $\Phi(x | \mu_1, \sigma_1^2)$  with standard deviation  $\sigma_1$  and mean  $\mu_1$  [Eq. (1)], and  $\Phi(x | \mu_2, \sigma_2^2)$  with standard deviation  $\sigma_2$  and mean  $\mu_2$  [Eq. (2)]. Therefore, the intersection areas of the two normal distributions can be calculated by solving  $\Phi(x | \mu_1, \sigma_1^2) = \Phi(x | \mu_2, \sigma_2^2)$  as in Eq. (3). With a defined  $x$ , the two intersection areas of groups A and B are calculated as  $p_1$  [Eq. (4)] and  $p_2$  [Eq. (5)], respectively. With the cumulative distribution function and the error function of the normal distribution respectively defined as Eqs. (6) and (8), the intersection areas of groups A and B are calculated using Eq. (9).

$$\Phi(x | \mu_1, \sigma_1^2) = \frac{1}{\sigma_1 \sqrt{2\pi}} e^{-\frac{(x-\mu_1)^2}{2\sigma_1^2}} \quad (1)$$

$$\Phi(x | \mu_2, \sigma_2^2) = \frac{1}{\sigma_2 \sqrt{2\pi}} e^{-\frac{(x-\mu_2)^2}{2\sigma_2^2}} \quad (2)$$

$$\begin{aligned} \Phi(x | \mu_1, \sigma_1^2) &= \Phi(x | \mu_2, \sigma_2^2) \\ \Rightarrow \frac{1}{\sigma_1 \sqrt{2\pi}} e^{-\frac{(x-\mu_1)^2}{2\sigma_1^2}} &= \frac{1}{\sigma_2 \sqrt{2\pi}} e^{-\frac{(x-\mu_2)^2}{2\sigma_2^2}} \\ \Rightarrow x &= \frac{\mu_2 \sigma_1^2 - \mu_1 \sigma_2^2 \pm \sqrt{2\sigma_1^4 \sigma_2^2 \ln\left(\frac{\sigma_1}{\sigma_2}\right) + \sigma_1^2 \mu_2^2 \sigma_2^2 + \mu_1^2 \sigma_1^2 \sigma_2^2 - 2\sigma_1^2 \sigma_2^4 \ln\left(\frac{\sigma_1}{\sigma_2}\right) - 2\mu_1 \mu_2 \sigma_1^2 \sigma_2^2}}{(\sigma_1^2 - \sigma_2^2)} \end{aligned} \quad (3)$$

$$p_1 = \frac{\mu_2\sigma_1^2 - \mu_1\sigma_2^2 - \sqrt{2\sigma_1^4\sigma_2^2 \ln\left(\frac{\sigma_1}{\sigma_2}\right) + \sigma_1^2\mu_2^2\sigma_2^2 + \mu_1^2\sigma_1^2\sigma_2^2 - 2\sigma_1^2\sigma_2^4 \ln\left(\frac{\sigma_1}{\sigma_2}\right) - 2\mu_1\mu_2\sigma_1^2\sigma_2^2}}{(\sigma_1^2 - \sigma_2^2)} \quad (4)$$

$$p_2 = \frac{\mu_2\sigma_1^2 - \mu_1\sigma_2^2 + \sqrt{2\sigma_1^4\sigma_2^2 \ln\left(\frac{\sigma_1}{\sigma_2}\right) + \sigma_1^2\mu_2^2\sigma_2^2 + \mu_1^2\sigma_1^2\sigma_2^2 - 2\sigma_1^2\sigma_2^4 \ln\left(\frac{\sigma_1}{\sigma_2}\right) - 2\mu_1\mu_2\sigma_1^2\sigma_2^2}}{(\sigma_1^2 - \sigma_2^2)} \quad (5)$$

$$\Phi(x, p | \mu, \sigma^2) = \int_{x=-\infty}^p \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} = \frac{1}{2} \left[ 1 + \operatorname{erf}\left(\frac{p-\mu}{\sigma\sqrt{2}}\right) \right] \quad (6)$$

$$\operatorname{erf}(z) \approx 1 - t(z), \quad (7)$$

where  $t(z) = 1 / \left( \sum_{i=0}^6 a_i z^i \right)^{16}$ ,  $a_0 = 1$ ,  $a_1 = 0.0705230784$ ,  $a_2 = 0.0422820123$ ,  $a_3 = 0.0092705272$ ,  $a_4 = 0.0001520143$ ,  $a_5 = 0.0002765672$ , and  $a_6 = 0.0000430638$ ,

$$A = \left[ \Phi(x, p_2 | \mu_1, \sigma_1^2) \right] + \left[ \Phi(x, p_1 | \mu_2, \sigma_2^2) - \Phi(x, p_2 | \mu_2, \sigma_2^2) \right] + \left[ 1 - \Phi(x, p_1 | \mu_1, \sigma_1^2) \right] \quad (8)$$

If  $\sigma_1 < \sigma_2$ ,

$$A = \left[ \Phi(x, p_1 | \mu_2, \sigma_2^2) \right] + \left[ \Phi(x, p_2 | \mu_1, \sigma_1^2) - \Phi(x, p_1 | \mu_1, \sigma_1^2) \right] + \left[ 1 - \Phi(x, p_2 | \mu_2, \sigma_2^2) \right]. \quad (9)$$

If  $\sigma_1 > \sigma_2$ ,

$$\begin{aligned} A &= 1 + \frac{1}{2} \left[ \operatorname{erf}\left(\frac{p_2 - \mu_1}{\sigma_1\sqrt{2}}\right) + \operatorname{erf}\left(\frac{p_1 - \mu_2}{\sigma_2\sqrt{2}}\right) - \operatorname{erf}\left(\frac{p_2 - \mu_2}{\sigma_2\sqrt{2}}\right) - \operatorname{erf}\left(\frac{p_1 - \mu_1}{\sigma_1\sqrt{2}}\right) \right] \\ &\approx 1 + \frac{1}{2} \left[ 1 - t\left(\frac{p_2 - \mu_1}{\sigma_1\sqrt{2}}\right) + 1 - t\left(\frac{p_1 - \mu_2}{\sigma_2\sqrt{2}}\right) - 1 + t\left(\frac{p_2 - \mu_2}{\sigma_2\sqrt{2}}\right) - 1 + t\left(\frac{p_1 - \mu_1}{\sigma_1\sqrt{2}}\right) \right] \\ &\approx 1 + \frac{1}{2} \left[ t\left(\frac{p_1 - \mu_1}{\sigma_1\sqrt{2}}\right) + t\left(\frac{p_2 - \mu_2}{\sigma_2\sqrt{2}}\right) - t\left(\frac{p_1 - \mu_2}{\sigma_2\sqrt{2}}\right) - t\left(\frac{p_2 - \mu_1}{\sigma_1\sqrt{2}}\right) \right]. \end{aligned} \quad (10)$$



**2.6.2 Calculation of weight of integrated HRV indicator with AHP**

The intersection areas of groups A and B are the input values of the AHP of this study. Then, the weight of each integrated HRV indicator is obtained according to the AHP. Weights  $\omega_1, \omega_2, \dots, \omega_n$  are calculated using Eq. (11) with the elements  $(a_1, a_2, \dots, a_n)$  of a set of size  $n$ . The  $w_n$  is defined as  $w_n = \sum_{v=1}^n a_{i,v} \times 1 / \sum_{u=1}^n a_{u,v}$ . Then, the matrix for Eq. (11) is

$$M = \begin{bmatrix} a_{1,1} & \dots & a_{1,j} & \dots & a_{1,n} \\ \vdots & & \vdots & & \vdots \\ a_{i,1} & \dots & a_{i,j} & \dots & a_{i,n} \\ \vdots & & \vdots & & \vdots \\ a_{n,1} & \dots & a_{n,j} & \dots & a_{n,n} \end{bmatrix} \cong \begin{bmatrix} \frac{w_1}{w_1} & \dots & \frac{w_1}{w_j} & \dots & \frac{w_1}{w_n} \\ w_1 & & w_j & & w_n \\ \vdots & & \vdots & & \vdots \\ \frac{w_i}{w_1} & \dots & \frac{w_i}{w_j} & \dots & \frac{w_i}{w_n} \\ w_1 & & w_j & & w_n \\ \vdots & & \vdots & & \vdots \\ \frac{w_n}{w_1} & \dots & \frac{w_n}{w_j} & \dots & \frac{w_n}{w_n} \\ w_1 & & w_j & & w_n \end{bmatrix} = W, \tag{11}$$

where  $a_{i,j} > 0, a_{i,j} = 1/a_{j,i}$ , and  $w_i = \sum_{v=1}^n a_{i,v} \times 1 / \sum_{u=1}^n a_{u,v}$ .

Calculation results are divided into  $n$  categories. Therefore,  $m$  features of the indicators are derived from  $q [q = C_2^n = n(n - 1)/2]$  groups of every two categories. For the  $j$ th feature, the intersection area of the  $k$ th group is defined as  $A_{(k,j)}$  [Eq. (10)]. The size of the  $i$ th category is defined as  $G_i$ , and the weight of each category (matrix  $M_G$ ) can be derived from Eq. (12). The weight of each feature in the  $k$ th group can be calculated using Eq. (13). For the final decision, the final weight matrix  $W_F$  is calculated from Eq. (14) based on the weight of  $m$  features of each group. Then, the critical value of the  $k$ th feature is evaluated as  $\omega_k$ . Finally, the feature with the highest value is taken as the critical feature to identify the driver drowsiness.

$$M_G = \begin{bmatrix} \frac{(g_1 + g_2)}{(g_1 + g_2)} & \dots & \frac{(g_1 + g_2)}{(g_i + g_j)} & \dots & \frac{(g_1 + g_2)}{(g_{n-1} + g_n)} \\ \vdots & & \vdots & & \vdots \\ \frac{(g_i + g_j)}{(g_1 + g_2)} & \dots & \frac{(g_i + g_j)}{(g_i + g_j)} & \dots & \frac{(g_i + g_j)}{(g_{n-1} + g_n)} \\ \vdots & & \vdots & & \vdots \\ \frac{(g_{n-1} + g_n)}{(g_1 + g_2)} & \dots & \frac{(g_{n-1} + g_n)}{(g_i + g_j)} & \dots & \frac{(g_{n-1} + g_n)}{(g_{n-1} + g_n)} \end{bmatrix} \cong \begin{bmatrix} \frac{w_{1,G}}{w_{1,G}} & \dots & \frac{w_{1,G}}{w_{j,G}} & \dots & \frac{w_{1,G}}{w_{q,G}} \\ w_{1,G} & & w_{j,G} & & w_{q,G} \\ \vdots & & \vdots & & \vdots \\ \frac{w_{i,G}}{w_{1,G}} & \dots & \frac{w_{i,G}}{w_{j,G}} & \dots & \frac{w_{i,G}}{w_{q,G}} \\ w_{1,G} & & w_{j,G} & & w_{q,G} \\ \vdots & & \vdots & & \vdots \\ \frac{w_{q,G}}{w_{1,G}} & \dots & \frac{w_{q,G}}{w_{j,G}} & \dots & \frac{w_{q,G}}{w_{q,G}} \\ w_{1,G} & & w_{j,G} & & w_{q,G} \end{bmatrix}, \tag{12}$$

where  $r_{1,G} = (g_1 + g_2), r_{2,G} = (g_1 + g_3), \dots, r_{q,G} = (g_{n-1} + g_n), a_{i,j,G} = (r_{i,G} / r_{j,G}) > 0, a_{i,j,G} = 1/a_{j,i,G}$ , and  $w_{i,G} = \sum_{v=1}^q a_{i,v,G} \times 1 / \sum_{u=1}^q a_{u,v,G}$ .

$$M_k = \begin{bmatrix} \frac{(1-A_{k,1})}{(1-A_{k,1})} & \dots & \frac{(1-A_{k,1})}{(1-A_{k,i})} & \dots & \frac{(1-A_{k,1})}{(1-A_{k,m})} \\ \vdots & & \vdots & & \vdots \\ \frac{(1-A_{k,m})}{(1-A_{k,1})} & \dots & \frac{(1-A_{k,m})}{(1-A_{k,i})} & \dots & \frac{(1-A_{k,m})}{(1-A_{k,m})} \\ \vdots & & \vdots & & \vdots \\ \frac{(1-A_{k,m})}{(1-A_{k,1})} & \dots & \frac{(1-A_{k,m})}{(1-A_{k,i})} & \dots & \frac{(1-A_{k,m})}{(1-A_{k,m})} \end{bmatrix} = \begin{bmatrix} \frac{r_{1,k}}{r_{1,k}} & \dots & \frac{r_{1,k}}{r_{i,k}} & \dots & \frac{r_{1,k}}{r_{m,k}} \\ r_{1,k} & & r_{i,k} & & r_{m,k} \\ \vdots & & \vdots & & \vdots \\ \frac{r_{i,k}}{r_{1,k}} & \dots & \frac{r_{i,k}}{r_{i,k}} & \dots & \frac{r_{i,k}}{r_{m,k}} \\ r_{1,k} & & r_{i,k} & & r_{m,k} \\ \vdots & & \vdots & & \vdots \\ \frac{r_{m,k}}{r_{1,k}} & \dots & \frac{r_{m,k}}{r_{i,k}} & \dots & \frac{r_{m,k}}{r_{m,k}} \\ r_{1,k} & & r_{i,k} & & r_{m,k} \end{bmatrix} \tag{13}$$

$$\cong \begin{bmatrix} \frac{w_{1,k}}{w_{1,k}} & \dots & \frac{w_{1,k}}{w_{i,k}} & \dots & \frac{w_{1,k}}{w_{m,k}} \\ w_{1,k} & & w_{i,k} & & w_{m,k} \\ \vdots & & \vdots & & \vdots \\ \frac{w_{i,k}}{w_{1,k}} & \dots & \frac{w_{i,k}}{w_{i,k}} & \dots & \frac{w_{i,k}}{w_{m,k}} \\ w_{1,k} & & w_{i,k} & & w_{m,k} \\ \vdots & & \vdots & & \vdots \\ \frac{w_{m,k}}{w_{1,k}} & \dots & \frac{w_{m,k}}{w_{i,k}} & \dots & \frac{w_{m,k}}{w_{m,k}} \\ w_{1,k} & & w_{i,k} & & w_{m,k} \end{bmatrix} = W_k,$$

where  $r_{1,k} = (1 - A_{k,1}), r_{2,k} = (1 - A_{k,2}), \dots, r_{m,k} = (1 - A_{k,m}), a_{i,j,G} = (r_{i,G} / r_{j,G}) > 0, a_{i,j,k} = 1/a_{j,i,k}$ , and  $w_{i,k} = \sum_{v=1}^m a_{i,v,k} \times 1 / \sum_{u=1}^m a_{u,v,k}$ .

$$W_k = \begin{bmatrix} \sum_{v=1}^q \frac{w_{1,G}}{w_{v,G}} & \dots & \sum_{v=1}^q \frac{w_{i,G}}{w_{v,G}} & \dots & \sum_{v=1}^q \frac{w_{q,G}}{w_{v,G}} \\ \sum_{v=1}^m \frac{w_{1,i}}{w_{v,i}} & \dots & \sum_{v=1}^m \frac{w_{i,i}}{w_{v,i}} & \dots & \sum_{v=1}^m \frac{w_{m,i}}{w_{v,i}} \\ \vdots & & \vdots & & \vdots \\ \sum_{v=1}^m \frac{w_{1,q}}{w_{v,q}} & \dots & \sum_{v=1}^m \frac{w_{i,q}}{w_{v,q}} & \dots & \sum_{v=1}^m \frac{w_{m,q}}{w_{v,q}} \end{bmatrix}, \tag{14}$$

$$= [\omega_1 \cdots \omega_k \cdots \omega_m],$$

where  $\omega_k = \sum_{u=1}^q \left( \sum_{v=1}^q (w_{u,G} / w_{v,G}) \times \sum_{v=1}^m (w_{k,u} / w_{v,u}) \right)$ .

## 2.7 Model construction for machine learning

Algorithms for training in the machine learning in this study include the k nearest neighbors (kNN), support vector machine (SVM), naive Bayes classifier (NBC), neural network, and decision tree. We analyzed and compared the results of using these algorithms.

## 3. Results and Discussion

In this study, parameters such as the sympathetic activity index (LF), parasympathetic activity index (HF), and balance index (LF/HF) were calculated using each algorithm. As shown in Table 2, as the driving time increases, the low-frequency component (LF) gradually increases, indicating that the participants were under stress from driving. The human body tends to maintain constant operation of its physiology, so the sympathetic process is activated. This results in a gradual increase in LF/HF (sympathetic/parasympathetic balance index). After 90 min of simulated driving, the sympathetic nervous system of the participants was deactivated gradually. The balance index began to fall as the parasympathetic nerves were activated.

Changes in the HR of the human body are mainly regulated by sympathetic nerves under stressful conditions such as stress or exercise. Groups A and B showed different states of the sympathetic nerve at the beginning of driving. Group A carried out simulated driving very early in the morning (02:00–04:30). Their physical conditions were in a fresh state, so LF in their HRV data was generally lower than that in group B before driving. After driving, the sympathetic nerves were activated to maintain the normal operation of bodily functions, which increased LF.

In group B, who carried out the experiment between 18:00 and 20:30, the sympathetic nerve function began to decline after 60 min of driving as fatigue had already accumulated due to daily activities. Therefore, the autonomic nervous system stopped the sympathetic nerves from working to rest the body after the simulated driving. After driving, both groups rested and then the sympathetic nerve function was slowly deactivated.

Table 2  
Physiological parameters.

Driving time	LF (Hz)	HF (Hz)	LF/HF
-10 (before driving)			
10	59.26 ± 8.41	40.81 ± 5.91	1.44 ± 0.41
20	69.25 ± 4.89	30.85 ± 2.39	2.15 ± 0.32
30	73.21 ± 6.33	26.89 ± 3.82	2.65 ± 0.59
40	72.55 ± 6.25	27.55 ± 3.75	2.57 ± 0.60
50	74.16 ± 7.61	25.94 ± 5.11	2.88 ± 0.96
60	72.67 ± 8.08	27.43 ± 5.58	2.67 ± 0.81
70	71.59 ± 9.83	28.51 ± 7.33	2.55 ± 0.76
80	74.95 ± 9.34	25.15 ± 6.84	3.03 ± 0.87
90	75.01 ± 7.77	25.06 ± 5.27	3.03 ± 1.01
+10 (after driving)	76.74 ± 8.05	23.36 ± 2.95	3.18 ± 0.69
+20 (after driving)	75.50 ± 9.90	24.60 ± 3.80	3.02 ± 0.83
+30 (after driving)	72.66 ± 7.40	27.44 ± 4.90	2.66 ± 0.91

Table 3  
Summary of the accuracy of experimental results.

Experiment	Classifier						
	KNN-1 (%)	KNN-5 (%)	SVM (%)	Neural network (%)	Naive Bayes (%)	Decision tree (%)	Average (%)
Time-domain analysis	66.66	66.66	83.33	—	—	—	72.23
Experiment 1	87.16	83.78	97.30	93.92	93.92	97.30	92.23
Experiment 2	75.68	74.32	94.59	94.59	93.24	97.30	88.29
Experiment 3	85.81	84.46	95.27	94.59	93.92	97.30	91.89
Experiment 4 (proposed method)	97.97	97.30	97.97	98.65	98.65	97.30	97.97

The independent sample  $t$  test of the HRV data showed that LF was significantly different before and after driving. There was no significant difference during driving ( $p > 0.05$ ).

The experiments on driving drowsiness state identification were performed with different HRV analyses to propose the SDDD model in this study. Among them, experiments 1 to 3 were carried out with a single HRV analysis. In the proposed method, we adopted the state identification of the HRV measurement index with the SDDD model. According to the experimental results, the SDDD model had the highest recognition accuracy and the average accuracy was 97.97%. The average efficiencies of experiments 1 to 3 were 92.23%, 88.29%, and 91.89%, respectively. Table 3 shows the average accuracy for each classification algorithm and experiment. The single HRV analysis had recognition accuracies of 66.66% for KNN-1, 66.66% for KNN-5, and 83.33% for SVM. The overall average efficiency was 72.23%. Therefore, the SDDD model improved the average recognition accuracy by 25.74% compared with the result of single HRV analysis.

#### 4. Conclusions

Our SDDD model measures the degree of overlap between two normal distributions, i.e., the intersection area under the distributions. Then, it uses the AHP to measure the intersection area to distinguish the importance of each HRV measurement feature. This method does not require conventional principal component analysis, multivariate analysis, or multi-objective decision-making to measure the importance of each feature.

The results of the SDDD model provided accurate information on the state of the driver and discriminated drowsiness efficiently with the machine learning algorithm. The SDDD model showed the highest identification accuracy in the experimental results and improved the accuracy by 25.74%. The overall average accuracy of the method in the original literature was low using a single time-domain analysis. Therefore, the proposed method improved the identification accuracy by 25.74% compared with the original method in the literature, which makes the overall efficiency of measuring HR, RR, and VLF more important for distinguishing the sleepiness of the driver. Further research is required to process unusual distributions of features and to predict drowsiness and exhaustion with an improved SDDD model. Our new SDDD model can include more information such as road types, weather conditions, and the speed of vehicles, which affect drivers significantly.

## Acknowledgments

This study was supported by the Big Data Technology Research Center of Guangdong Smart Vocational Education under grant No. 3001/180113491 and Panyu Polytechnic Innovation Team under grant No. 2020CXTD003.

## References

- 1 Centers for Disease Central and Prevention, Department of Health and Human Services: <https://www.cdc.gov/features/dsdrowsydriving/index.html> (accessed March 2020).
- 2 G. Attarodi, S. M. Nikooei, N. J. Dabanloo, O. Pourmasoumi, and A. Tareh: Proc. 45th Computing in Cardiology Conf. 2018 (IEEE, Maastricht, 2018) 1–4. <https://doi.org/10.22489/CinC.2018.014>
- 3 M. Baumert, M. Javorka, A. Seeck, R. Faber, P. Sanders, and A. Voss: *Comput Biol Med.* **42** (2012) 347. <https://doi.org/10.1016/j.compbiomed.2011.03.019>.
- 4 T. Penzel, J. W. Kantelhardt, L. Grote, J. Peter, and H. Bunde: *IEEE Trans. Biomed. Eng.* **50** (2003) 1143. <https://doi.org/10.1109/TBME.2003.817636>
- 5 J. S. Richman and J. R. Moorman: *Am. J. Physiol. Heart Circ. Physiol.* **278** (2000) 2039. <https://doi.org/10.1152/ajpheart.2000.278.6.H2039>
- 6 M. P. Tarvainen, J. P. Niskanen, J. A. Lipponena, P. O. Ranta-aho, and P. A. Karjalainen: *Comput. Methods Programs Biomed.* **113** (2000) 210. <https://doi.org/10.1016/j.cmpb.2013.07.024>
- 7 R. Hedelin, U. Wiklund, P. Bjerle, and K. Henriksson-Larsén: *Med. Sci. Sports Exercise* **32** (2000) 1531.
- 8 T. Penzel, J.W. Kantelhardt, L. Grote, J. Peter, and H. Bunde: *IEEE Trans. Biomed. Eng.* **50** (2003) 1143. <https://doi.org/10.1109/TBME.2003.817636>
- 9 S. M. Pincus, I. M. Gladstone, and R. A. Ehrenkranz: *J. Clin. Monit. Comput.* **7** (1991) 335. <https://doi.org/10.1007/BF01619355>
- 10 J. Horne and L. Reyner: *Occup Environ. Med.* **56** (1999) 289. <http://doi.org/10.1136/oem.56.5.289>.
- 11 P. A. Lanfranchi and K. S. Virend: K. Somers: *Principles and Practice of Sleep Medicine*, 5th ed., M. H. Kryger, T. Roth, W. C. Dement Eds. (Elsevier, Amsterdam, 2011) Chap. 20. <https://doi.org/10.1016/B978-1-4160-6645-3.00020-7>.
- 12 M. P. Tarvainen, J.-P. Niskanen, J. A. Lipponen, P. O. Ranta-Aho, and P. A. Karjalainen: *Comput. Methods Programs Biomed.* **113** (2014) 210. <https://doi.org/10.1016/j.cmpb.2013.07.024>