International Journal of Medical Informatics 119 (2018) 39–46

Contents lists available at ScienceDirect

International Journal of Medical Informatics

journal homepage: www.elsevier.com/locate/ijmedinf

Detection of mental fatigue state with wearable ECG devices

Shitong Huang[a], Jia Li[b], Pengzhu Zhang[a], Weiqiang Zhang[a,*]

[a] Shanghai Jiaotong University, Antai College of Economics & Management, 1954 Huashan Rd., Shanghai 200030, China
[b] East China University of Science and Technology, School of Business, 130 Meilong Rd., Shanghai 200237, China

ARTICLE INFO

Keywords:
Mental fatigue
Wearable devices
ECG
HRV
Feature selection
Machine learning

ABSTRACT

Overwork-related disorders, such as cerebrovascular/cardiovascular diseases (CCVD) and mental disorders due to overwork, are a major occupational and public health issue worldwide, particularly in East Asian countries. Since wearable smart devices are inexpensive, convenient, popular and widely available today, we were interested in investigating the possibility of using wearable smart electrocardiogram (ECG) devices to detect the mental fatigue state. In total, 35 healthy participants were recruited from a public university in East China. Throughout the entire experiment, each participant wore a wearable device that was further linked to a smartphone to upload the data based on Bluetooth transmission. To manipulate the fatigue state, each participant was asked to finish a quiz, which lasted for approximately 80 min, with 30 logical referential and computing problems and 25 memory tests. Eight heart rate variability (HRV) indicators namely NN.mean (mean of normal to normal interval), rMSSD (root mean square of successive differences), PNN50 (the proportion of NN50 divided by total number of NNs), TP (total spectral power), HF (high frequency from 0.15 Hz to 0.4 Hz), LF (low frequency from 0.04 Hz to 0.15 Hz), VLF (very low frequency from 0.0033 Hz to 0.04 Hz) and the LF/HF ratio were collected at intervals of 5 min throughout the entire experiment. After the feature selection was performed, six indicators remained for further analysis, which were the NN.mean, rMSSD, PNN50, TP, LF, and VLF. Four algorithms, support vector machine (SVM), K-nearest neighbor (KNN), naïve Bayes (NB), and logistic regression (LR), were used to build classifiers that automatically detected the fatigue state. The best performance was achieved by KNN, which had a CV accuracy of 75.5%. The NN.mean, PNN50, TP and LF were the most important HRV indicators for mental fatigue detection. KNN performed the best among the four algorithms and had an average CV accuracy of 65.37% for all of the possible feature combinations.

1. Introduction

Overwork-related disorders, such as cerebrovascular/cardiovascular diseases (CCVD) and mental disorders due to overwork, are a major occupational and public health issue worldwide, particularly in East Asian countries [1]. Japan's work culture is so intense that people in the 1970s invented a word, “karoshi,” which translates to "death by overwork." One example of an employee’s death determined to be karoshi was 31-year-old journalist Miwa Sado [2]. She reportedly logged 159 h of overtime in one month at the news network NHK before dying of heart failure in July 2013. In Japan, the government estimates that 200 people die from overwork every year because of heart attacks or cerebral hemorrhages due to long hours spent at the workplace [3]. However, this estimation does not include deaths from mental depression or suicides. If these deaths were included, the number of work-related deaths would dramatically increase. From January 2010 to March 2015, 368 suicides in Japan, from 352 men and 16 women, were deemed as being karoshi [4]. Overwork is also a serious problem in China. According to the China Youth Daily, approximately 600,000 Chinese people each year die from working too hard [5]. In April 2015, China Radio International reported a toll of 1600 deaths from overwork every day in China [5].

Since overwork is a subjective feeling that varies among people, it is very difficult to measure overwork by simply counting working hours. Therefore, mental fatigue is a better way for detecting potential overwork. Mental fatigue is a subjective feeling of mental tiredness. It is a transient decrease in maximal cognitive performance resulting from prolonged periods of cognitive activity (long working hours, shift work, stressful work, anxiety, etc.) [6]. Research proves that stroke and death by karoshi have a strong association with mental fatigue caused by overwork [7]. In addition, intensive work increases the risks for cardiovascular diseases [8], diabetes [9] and cancer [10]. In addition to inducing damage to human health, mental fatigue also has a variety of effects that impair memory, judgement, decision-making and emotion

* Corresponding author.
E-mail addresses: shitong_huang@sjtu.edu.cn (S. Huang), jiali@ecust.edu.cn (J. Li), pzzhang@sjtu.edu.cn (P. Zhang), z.grand@163.com (W. Zhang).

https://doi.org/10.1016/j.ijmedinf.2018.08.010
Received 17 May 2018; Received in revised form 1 August 2018; Accepted 19 August 2018
1386-5056/

management [11]. Routinely working long hours leads to stress and strain, which, in turn, can lead to higher accident levels, greater absenteeism, and reduced productivity [12,13]. Therefore, analyzing wearable devices that can monitor a worker's mental fatigue in a real-time manner and prompt the user to take a rest or leave the office is highly imperative.

However, mental fatigue is elusive and difficult to measure in practice. Extant measures for mental fatigue can be divided into two categories: subjective self-report measures and objective performance measures. Subjective self-report measures require subjects to evaluate their level of mental fatigue typically by a questionnaire [14–16]. Some scales simply involve questions about the participant's perceptions of experienced fatigue or sleepiness at the moment, such as in the Stanford Sleepiness Scale (SSS), Chalder Fatigue Scale (CFS) and Fatigue Severity Scale (FSS). Meanwhile, other scales assess the participant's fatigue level by setting detailed scenarios, such as in the Epworth Sleepiness Scale (ESS) [17] and Specific Fatigue Scale (SFS) [18]. Objective performance measures design many mental tasks to assess the subject's performance of brain function. Some tasks measure the subject's reaction time, memory and decision-making performance, such as in the Psychomotor Vigilance Task (PVT) [19]. Meanwhile, other tasks assess the subject's maintenance of wakefulness and resistance of sleepiness, such as in the Multiple Sleep Latency Test (MSLT) and Maintenance of Wakefulness Test (MWT) [20].

The two measures mentioned above are intrusive in nature because the users must stop their work at hand to finish the questionnaires or mental tasks. Therefore, they cannot be used to monitor mental fatigue without interrupting normal daily life. The equipment approach allows mental fatigue to be measured while the daily work is still going on. For example, the electroencephalograph (EEG) is the most widely used equipment for measuring mental fatigue [21,22]. Some other researchers have proposed a variety of EEG-based algorithms that detect fatigue based on a spectrum analysis [23–25]. Four frequency components obtained from the original EEG signal have been proven to be useful for detecting a subject's brain state, namely, delta (δ) ($\pm$ 0 Hz to 4 Hz), theta (θ) (4–8 Hz), alpha (α) (8–13 Hz), and beta (β) (13–20 Hz) [25]. For a driving fatigue state detection, a group of Australian researchers developed an EEG-based driver-fatigue countermeasure system to monitor driver fatigue [23,24,26]. However, the devices used for EEG-based fatigue detection are usually heavy and large, which is inconvenient for applying to daily life, especially when used in an office space or at home. Since real-time monitoring is very important for helping the users to remain in a healthy state, a convenient wearable device that can ubiquitously monitor the mental fatigue condition is highly desirable [27].

A recent trend in health information technology is the growing popularity of wearable smart devices, such as smart bracelets and wearable ECGs, which makes real-time and distant health monitoring and management possible. According to Gartner's investigation, in 2016, a total of 265.9 M wearable devices were sold. The global market for wearable electronic devices is forecasted to be worth more than $50 billion in 2021 [28]. Therefore, wearable smart devices are becoming increasingly more widely available. We are interested in investigating the possibility of using wearable smart devices to measure mental fatigue.

A number of smart sensors used to continuously obtain physiological parameters, such as an electrocardiogram (ECG), heart rate and blood pressure, with Bluetooth wireless transmission for health monitoring have been developed [29–32]. Among all of these wearable devices, a wearable ECG is a promising one for real-time mental fatigue monitoring. The device provides a relatively easy way to obtain ECG signals compared with that for complex EEG devices. Since the connection between the autonomic nervous system (ANS) and heart rhythms was discovered a long time ago [33], it is possible to measure the mental fatigue status with ECG signals. Therefore, the research question of this study can be interpreted as follows:

Fig. 1. Photograph of the portable ECG equipment 'LaPatch'.

RQ: Can mental fatigue be detected by wearable ECG smart devices? If so, how and with what effect can this be achieved?

To answer this research question, an experiment was designed and executed in this study to test the possibility of measuring mental fatigue with a wearable ECG device. In total, 35 subjects were recruited from a public university in East China. An experiment was carried out to collect self-reported mental fatigue and ECG data.

2. Materials and methods

2.1. The devices

The wearable ECG device used in this study is a portable single-channel electrocardiogram equipment called "LaPatch" and is shown in Fig. 1. This device uses ADS1292R (developed by Texas Instruments) as the core chip to accurately acquire the ECG and multiple respiration states. Bluetooth is used to transmit data from the wearable ECG device to the smartphone [34].

2.2. Experimental design

In total, 35 healthy participants without heart disease were recruited from a public university in East China. We didn't recruit the subjects who have overwork-related problems for matching the issue of mental fatigue. The main reason is that overwork is a transient state which changes over time. A subject who feels tired someday may feel energetic for the next day. Thus it is very difficulty to hire the real overworked subjects who happens to be in the fatigue state before the experiment starts. In contrast, we choose another approach that manipulates the healthy subject's fatigue status with a quiz. Before the quiz, most subjects should be in a non-fatigue state. Then they were asked to finish a quiz. After the quiz, most subjects should be in a fatigue state. The data were collected before and after the quiz. In this way, we collected samples both in fatigue state and non-fatigue state.

Each subject was assigned a unique number to match them with their questionnaires and devices. They had a mean age of 23 $\pm$ 4 years and a male to female ratio of 1:1.3. Throughout the entirety of the experiment, each participant wore a wearable device, as mentioned in the previous section, which was further linked to a smartphone for uploading the data based on real-time Bluetooth transmission. Before the experiment started, each participant was asked to finish a questionnaire containing 14 items (the Chalder Fatigue Scale) [35] to report their fatigue state. The items used for the fatigue scale are shown in

Table 1
Summary of the fatigue self-report assessment.

Item	Describe	scale
1	Do you feel have problems with tiredness now?	1–5
2	Do you need to rest more now?	1–5
3	Do you feel sleepy or drowsy now?	1–5
4	Do you have problems starting things now?	1–5
5	Do you start things without difficulty but get weak as you go on now?	1–5
6	Are you lacking in energy now?	1–5
7	Do you have less strength in your muscles now?	1–5
8	Do you feel weak now?	1–5
9	Do you have difficulty concentrating now?	1–5
10	Do you have problems thinking clearly now?	1–5
11	Do you make slips of tongue when speaking now?	1–5
12	Do you find it more difficult to find the correct word now?	1–5
13	How is your memory now?	1–5
14	Have you lost interest in the things you used to do now?	1–5

Table 1. After that, the participant was asked to complete a quiz. The quiz includes 55 questions. The question type covers spatial imagination, computation, reasoning and memory. The example for each type is provided in Appendix. After finishing the quiz, the participant was asked to fill out the fatigue scale again. To better motivate the participants, the subjects were rewarded with an extra cash bonus of 40 Chinese Yuan (approximately $6). Institutional review board approval was obtained for this study.

2.3. Data preprocessing

The original data (ECG signals) were collected by the wearable device by recording the electrical activity of the heart over a period of time using electrodes placed on the skin. The heart rate variability (HRV) variables were extracted from the original data for further analysis. HRV is a set of systematic evaluation indicators obtained from the ECG [36]. HRV measures the variation in the time interval between heartbeats, which indicates the sympathetic and parasympathetic nervous system (PSNS) regulation, so it has been previously used to examine mental workload [37].

Both time domain and frequency domain HRV indicators [38,39] were extracted and used in this study. The time domain indicators include the mean of normal to normal interval (NN.mean), standard deviation of NN interval (SDNN), the standard deviation of the average NN intervals calculated over short periods, usually 5 min (SDANN), root mean square of successive differences (rMSSD) and the proportion of NN50 divided by total number of NNs (PNN50). The time domain HRV measures are shown in Table 2.

The frequency domain indicators include the total spectral power (TP), high frequency from 0.15 Hz to 0.4 Hz (HF), low frequency from 0.04 Hz to 0.15 Hz (LF), very low frequency from 0.0033 Hz to 0.04 Hz (VLF) and the LF/HF ratio. The TP is a measure of overall autonomic activity [40]. HF is primarily modulated by pneumogastric nerve activities [41]. LF is sensitive to the dual regulation of cardiac

Table 2
Time domain HRV measures.

Measure	Formula	Unit
NN.mean	$\frac{\sum_{i=1}^{N}(NN_i)}{N}$	ms
SDNN	$sqrt\left(\frac{\sum_{i=1}^{N}(NN_i - mNN)^2}{N-1}\right)$	ms
SDANN	$sqrt\left(\frac{\sum_{i=1}^{N}(\overline{NN_i} - \overline{NN_{5min}})^2}{N-1}\right)$	ms
rMSSD	$sqrt\left(mean\left((NN_{i+1} - NN_i)^2\right)\right)$	ms
PNN50	$\frac{count(\mid NN_{i+1} - NN_i \mid) > 50ms}{N-1} \times 100\%$	/

Table 3
Frequency domain HRV measures.

Measure	Frequency Band
TP	0.003 Hz–0.4 Hz
HF	0.15 Hz–0.4 Hz
LF	0.04 Hz–0.15 Hz
VLF	0.003 Hz–0.04 Hz
LF/HF	The balance Ratio of LF to HF

sympathetic and parasympathetic nerve activities [42,43]. VLF reflects the regulation of body temperature [44–46], vasomotor tension and the humoral system [47,48]. The LF/HF ratio indicates the balance of the sympathetic and parasympathetic nervous systems [41]. The frequency domain HRV measures are shown in Table 3.

The ECG data were collected every 0.004 s (sample rate: 250 Hz). Since the recording should be at least 10 times the wavelength of the lowest frequency bound of interest, it takes approximately 1 min for the HF components of the HRV (i.e., a lowest bound of 0.15 Hz is a cycle of 6.6 s, and thus, 10 cycles requires ~60 s) and more than 4 min for the LF component (with a lower bound of 0.04 Hz). If the time interval was 5 min, the SDNN and SDANN had the same value, and we decided to keep the SDNN. Our later correlation analysis after the data collection indicated that the rMSSD is highly related to the SDNN, with a Pearson's coefficient of greater than 0.8. Since multicollinearity has a severe influence on the prediction performance, the SDNN was excluded from further analysis. Therefore, eight HRV variables (i.e., the NN.mean, rMSSD, PNN50, TP, HF, LF, VLF and LF/HF) measured at 5-minute intervals were used for further analysis. The variables were averages among different timestamp for further analysis.

After excluding samples with outliers or missing data, 29 out of 35 subjects were qualified for further analysis. Since each subject were measured twice (before and after the quiz), we collected 58 samples in total. The experiment lasted 60 ~80 min (quiz completion time lasted for 54 ± 8 min). The ECG data 10 min before the quiz started were used to calculate the HRV indicators corresponding to the before-quiz fatigue state. The ECG data during the last 10 min before the quiz ended were used to calculate the HRV indicators corresponding to the after-quiz fatigue state. The self-reported fatigue state was measured by a binary variable, with 0 indicating non-fatigue and 1 indicating fatigue. Since each participant was asked to answer 14 questions using a Likert scale ranging from 1 (no feel) to 5 (very severe), we used the optimum cutoff of 3/4 as suggested in [35]. Therefore, the samples with a fatigue score above 50 were labeled as being fatigue, while the rest were labeled as being non-fatigue. Therefore, the data set used for model training and testing consists of 58 samples, and each sample contains 9 variables (8 HRV variables and 1 fatigue state variable).

3. Results

3.1. Feature selection

In this section, we selected the most salient HRV indicators that best distinguished the fatigue state and non-fatigue state. The HRV indicators for the fatigue and non-fatigue states are shown in Table 4. There are three reasons for this feature selection. First, we made our model simpler and easier to interpret. Second, we could reduce the variance of the model and, therefore, overfitting. Finally, we could reduce the computational cost (and time) of training a model. The process of identifying the most relevant features is called the "feature selection."

The random forest method was used to identify the most salient indicators. A random forest is basically an ensemble of decision trees. Each tree is built from a random subset of the training dataset. In each decision tree model, a random subset of the available variables is used to choose how the dataset is partitioned at each node. Random forests

Table 4
The HRV indicators for the fatigue and non-fatigue states.

Indicator	Fatigue		Non-fatigue	
	Mean	Var	Mean	Var
NN.Mean	0.794	0.013	0.839	0.012
PNN50	0.222	0.014	0.306	0.019
rMSSD	0.266	0.052	0.254	0.051
TP	2.641	19.291	1.828	11.372
LF	0.771	1.590	0.521	0.865
HF	1.679	9.356	1.168	5.426
VLF	0.191	0.043	0.138	0.017
LF/HF	1.860	1.171	1.521	0.603

Table 5
The MDA and MDG scores of the HRV indicators.

Indicator	MDA score*	MDG score*
NN.Mean	3.744	4.717
PNN50	0.468	4.00
rMSSD	1.92	3.658
TP	2.073	3.097
LF	6.039	3.611
HF	-1.210	2.916
VLF	3.471	3.730
LF/HF	−2.867	2.695

* MDA score means the evaluation of random forest under the mean decrease accuracy, and MDG score means the evaluation of random forest under the mean decrease Gini.

also provide a natural way of assessing the importance of input variables (predictors). This process is achieved by removing one variable at a time and assessing whether or not the out-of-bag error changes. If it does, the variable is important for distinguishing the fatigue state from the non-fatigue state.

Two methods were used in the importance evaluation, the mean decrease accuracy (MDA) and mean decrease Gini (MDG) [49]. The MDA measures the importance of a variable by valuing its contribution to increasing the mean squared error. The larger the MDA value of a variable, the more important is that variable. The MDG evaluates the contribution to increasing the node purity. The larger the MDG value of a variable, the purer is that variable. Therefore, the importance of a variable can be measured by the MDA and MDG. A higher MDA score or MDG score indicates that a variable has a more important feature.

The MDA and MDG scores for all of the HRV indicators are shown in Table 5 and are further visualized in Fig. 2. As shown in Table 4, the MDA scores indicate that the VLF, LF, TP, NN.mean, PNN50 and rMSSD are valuable indicators because they positively contribute to the prediction accuracy. However, the MDA scores suggest that HF and LF/HF should be removed from the feature set because they negatively contribute to the prediction accuracy. These results are also consistent with the results suggested by the MDG scores because HF and the LF/HF contribute the least in terms of the MDG scores. Therefore, six indicators, namely, the NN.mean, PNN50, rMSSD, TP, LF and VLF, remained for further analysis.

3.2. Machine learning performance

Four different machine learning algorithms, support vector machine (SVM), K-nearest neighbor (KNN), naïve Bayes (NB), and logistic regression (LR), were used to build classifiers that automatically detect a fatigue state based on HRV indicators. For SVM, the radial basis kernel function was used with optimal parameters cost (c) searching from 0.01 to 10^4 and gamma (G) searching from 10^{-3} to 4. For KNN, the algorithm automatically searched and determined the optimal parameter K from 3 to 9. For LR, a probability of > 0.5 was considered as being indicative of fatigue.

The prediction performance was evaluated by comparing it to human self-reported mental fatigue. To increase the robustness of the evaluation, the performance was evaluated with the 5-fold cross-validation (CV) technique. The performance of classifiers was measured by the CV accuracy and area under the ROC curve (AUC). The CV accuracy is a ratio of the correctly predicted observations to the total observations. The AUC value is equivalent to the probability that a randomly chosen positive example is ranked higher than a randomly chosen negative example. A reliable and valid AUC estimate can be interpreted as the probability that the classifier will assign a higher score to a randomly chosen positive example than to a randomly chosen negative example. In summary, the higher the CV accuracy and AUC are, the higher is the performance of the classifier. The classification results are shown in Table 6.

First, we were interested in the most important feature or feature combination that best predicted the fatigue and non-fatigue states. We tries all the possible combinations of 6 indicators (1 indicator, 2 indicators, 3 indicators, 4 indicators, 5 indicators and 6 indicators). The performance of any 1–6 indicator(s) are summarized as follows.

For a single variable, the PNN50 was the most salient feature since it achieved an average CV accuracy of 64.9% on the three algorithms (SVM, NB and LR). KNN was excluded from the single variable results because it requires at least two variables.

For the two variable combination, NN.mean + PNN50 and NN.mean + TP were the two most important features. NN.mean + PNN50 achieved the best average CV accuracy, which was 66.8%, while

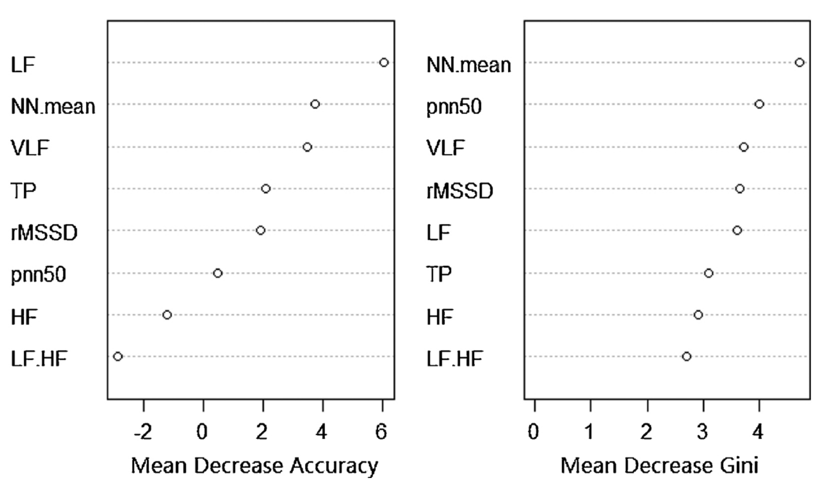

Fig. 2. Evaluation of the importance of the HRV indicators.

Table 6
Summary of the classification results.

HRV Indicators		*SVM CV Accuracy*	*NB CV Accuracy*	*KNN CV Accuracy*	*LR CV Accuracy*
One indicator	NN.mean	59.1%	60.0%	–	**66.1%**
	PNN50	**63.4%**	**65.6%**	–	65.6%
	VLF	53.8%	41.7%	–	50.7%
	TP	51.6%	46.8%	–	37.1%
	LF	51.0%	46.8%	–	37.1%
	rMSSD	41.0%	40.5%	–	40.5%
	Average	**53.32%**	**50.23%**	–	**49.52%**
Two indicators	NN.mean + PNN50	**71.6%**	**64.3%**	68.8%(k = 9)	**62.6%**
	NN.mean + TP	51.5%	49.9%	**73.5%(k = 3)**	61.1%
	NN.mean + VLF	50.9%	41.6%	64.8%(k = 9)	55.9%
	NN.mean + rMSSD	55.1%	51.1%	62.0%(k = 2)	60.6%
	NN.mean + LF	46.8%	49.9%	57.8%(k = 5)	61.1%
	rMSSD + VLF	55.5%	40.2%	63.8%(k = 9)	54.2%
	rMSSD + LF	65.5%	44.6%	68.4%(k = 5)	59.0%
	rMSSD + TP	49.4%	44.0%	65.1%(k = 1)	59.2%
	PNN50 + rMSSD	63.5%	63.2%	59.4%(k = 8)	61.9%
	PNN50 + TP	62.0%	56.5%	69.7% (k = 4)	61.9%
	PNN50 + LF	61.7%	56.5%	58.4%(k = 6)	61.9%
	PNN50 + VLF	58.8%	46.0%	60.6%(k = 6)	61.9%
	LF + VLF	58.4%	42.4%	65.3%(k = 5)	44.2%
	TP + LF	49.2%	46.8%	63.6%(k = 3)	47.9%
	TP + VLF	51.8%	44.0%	63.7%(k = 2)	50.3%
	Average	**56.78%**	**49.40%**	**64.33%**	**57.58%**
Threeindicators	NN.mean + PNN50 + TP	**64.1%**	53.7%	67.2%(k = 5)	58.8%
	NN.mean + PNN50 + LF	**64.1%**	53.7%	59.4%(k = 5)	57.2%
	NN.mean + TP + VLF	63.5%	45.5%	73.5%(k = 3)	58.6%
	NN.mean + TP + LF	51.7%	49.9%	**75.5%(k = 3)**	**65.7%**
	NN.mean + PNN50 + VLF	63.3%	51.4%	62.8%(k = 9)	63.1%
	NN.mean + PNN50 + rMSSD	60.2%	**61.7%**	58.4%(k = 9)	58.8%
	PNN50 + TP + LF	59.4%	51.1%	69.7%(k = 4)	63.8%
	rMSSD + TP + VLF	58.9%	42.4%	62.2%(k = 2)	55.7%
	rMSSD + LF + VLF	58.9%	44.0%	64.3%(k = 5)	54.1%
	PNN50 + rMSSD + VLF	58.1%	41.6%	56.3%(k = 9)	61.0%
	NN.mean + rMSSD + TP	56.2%	51.6%	**71.5%(k = 3)**	62.7%
	rMSSD + TP + LF	55.1%	44.6%	63.6%(k = 3)	55.1%
	PNN50 + LF + VLF	55.0%	48.7%	62.5%(k = 3)	60.1%
	PNN50 + rMSSD + LF	54.8%	48.3%	58.6%(k = 4)	63.4%
	PNN50 + TP + VLF	54.4%	47.0%	66.4%(k = 4)	60.1%
	PNN50 + rMSSD + TP	53.7%	52.8%	69.7%(k = 4)	63.4%
	Average	**58.21%**	**49.25%**	**65.10%**	**60.10%**
Four indicators	NN.mean + PNN50 + LF + VLF	**64.1%**	48.3%	64.3%(k = 6)	62.4%
	NN.mean + PNN50 + rMSSD + LF	63.1%	52.6%	59.4%(k = 5)	66.3%
	NN.mean + PNN50 + TP + LF	62.8%	52.8%	67.3%(k = 5)	65.5%
	NN.mean + PNN50 + rMSSD + TP	62.3%	**54.8%**	68.8%(k = 5)	64.7%
	NN.mean + PNN50 + rMSSD + VLF	59.7%	52.8%	59.2%(k = 8)	63.0%
	NN.mean + PNN50 + TP + VLF	59.7%	48.3%	65.6%(k = 5)	62.4%
	NN.mean + TP + LF + VLF	58.6%	45.5%	73.1%(k = 3)	64.5%
	PNN50 + TP + LF + VLF	57.9%	45.4%	63.1%(k = 4)	63.8%
	NN.mean + rMSSD + TP + LF	57.4%	46.0%	**74.5%(k = 3)**	**66.5%**
	PNN50 + rMSSD + LF + VLF	56.7%	45.5%	63.9%(k = 3)	64.5%
	rMSSD + TP + LF + VLF	53.8%	42.4%	62.9%(k = 3)	50.1%
	NN.mean + rMSSD + TP + VLF	53.4%	45.5%	71.4%(k = 3)	66.1%
	NN.mean + rMSSD + LF + VLF	52.6%	43.8%	58.4%(k = 4)	56.7%
	PNN50 + rMSSD + TP + LF	50.8%	47.7%	69.7%(k = 4)	62.5%
	PNN50 + rMSSD + TP + VLF	48.2%	45.5%	68.0%(k = 4)	63.2%
	Average	**57.41%**	**47.79%**	**65.97%**	**62.81%**
Five indicators	NN.mean + PNN50 + rMSSD + TP + LF	**64.6%**	49.9%	68.8%(k = 5)	65.5%
	NN.mean + PNN50 + TP + LF + VLF	59.7%	47.0%	67.3%(k = 5)	69.9%
	NN.mean + PNN50 + rMSSD + LF + VLF	57.4%	43.2%	61.6%(k = 5)	65.2%
	PNN50 + rMSSD + TP + LF + VLF	56.1%	45.5%	63.1%(k = 4)	60.1%
	NN.mean + PNN50 + rMSSD + TP + VLF	55.8%	43.2%	65.6%(k = 5)	65.2%
	NN.mean + rMSSD + TP + LF + VLF	51.0%	45.4%	**73.1%(k = 3)**	66.1%
	Average	**57.43%**	**45.70%**	**66.58%**	**65.33%**
Six indicators	NN.mean + PNN50 + rMSSD + TP + LF + VLF	59.1%	60.0%	68.8%(k = 5)	66.1%
Total Average		***57.08%***	***48.84%***	***65.37%***	***59.71%***

The bold and underlined values mean the best performance of HRV indicators' combination under different number of indicators in a given algorithm.

NN.mean + TP achieved the best CV accuracy, which was 73.5%, with KNN (k = 3).

For the three variable combination, NN.mean + PNN50 + TP and NN.mean + TP + LF were the two most important features. NN.mean + PNN50 + TP achieved an average CV accuracy of 61.0%, while NN.mean + TP + LF achieved the best CV accuracy, which was

Table 7
The classification performances of the four algorithms.

Algorithm	SVM	NB	KNN	LR
AUC	0.68	0.64	0.74	0.65

75.5%, with KNN (k = 3).

For the four variable combination, NN.mean + PNN50 + rMSSD + TP and NN.mean + rMSSD + TP + LF were the two most important features. NN.mean + PNN50 + rMSSD + TP achieved an average CV accuracy of 62.7%, while NN.mean + rMSSD + TP + LF achieved the best CV accuracy, which was 74.5%, with KNN (k = 3).

For the five variable combination, NN.mean + PNN50 + rMSSD + TP + LF and NN.mean + rMSSD + TP + LF + VLF were the two most important features. NN.mean + PNN50 + rMSSD + TP + LF achieved an average CV accuracy of 62.2%, while NN.mean + rMSSD + TP + LF + VLF achieved the best CV accuracy, which was 73.1%, with KNN (k = 3). Lastly, the full combination achieved an average CV accuracy of 63.5%.

As it can be observed, the performance first increases and then decreases as the number of indicators increases. Therefore, the optimal number of indicators is two or three. NN.mean + PNN50 achieved the best average CV accuracy, which was 66.8%, and NN.mean + TP + LF achieved the best CV accuracy, which was 75.5% with KNN (k = 3). Using only one indicator provides insufficient information, which leads to an inferior performance. In contrast, too many (e.g., four, five or six) indicators provide too much noise, which also leads to an inferior performance.

Second, we were interested in the best machine learning algorithm for the fatigue state prediction task. As shown in Table 5, KNN performed best out of all four of the algorithms and achieved an average CV accuracy of 65.37%. LR and SVM achieved average CV accuracies of 59.17% and 57.08%, respectively. NB performed the worst among the four algorithms and had an average CV accuracy 48.84%.

Table 7 shows the AUC of the four algorithms, and this is further visualized in Fig. 3. Since the AUC is the area under the ROC curve, the higher that the AUC is, the larger is the area under the ROC curve. As it is shown in Table 7 and Fig. 3, KNN achieved the best performance and had an AUC of 0.74, which was followed by SVM (0.68) and LR (0.74). Similar to the results suggested by those of the CV accuracy, the NB performed the worst among the four algorithms, with an AUC of 0.64. This performance is compatible with other recent studies with HRV analysis. For example, Nizami [50] developed a heart disease classification model based on HRV analysis with the best average CV accuracy of 69.9%. In this research, the best average CV accuracy was 74.5%. In another recent study, Pradhapan et al. [51] evaluated the predictive capacity of HRV on rest and recovery state. They achieved the best AUC of 0.7. In this study, the best performance of KNN model had an AUC of 0.74.

4. Discussion

4.1. Principal results

There are several major findings of this study. First, this study demonstrates that the user's mental fatigue state can be detected with a reasonable accuracy by a convenient wearable ECG device. The best performance achieved in this study was a CV accuracy of 75.5%. Therefore, there is great potential to apply inexpensive wearable ECG devices to mental fatigue monitoring and overwork reminding.

Second, our results indicate that the NN.mean, PNN50, TP and LF are the key HRV indicators for mental fatigue detection. Including too few indicators or too many indicators may hurt the classification performance. Three or four indicators is actually a very small number, meaning that the ECG data can be collected and processed very quickly, which is important for real-time monitoring based on the low-cost chip.

Third, our results compared four classification algorithms, and KNN significantly outperformed the other three algorithms. KNN is robust to noisy data and effective when the amount of training data is large. In addition, KNN is actually a simple algorithm that has a low computational cost. Therefore, KNN is a promising approach for mental fatigue state detection when its industrial application is considered in the future.

4.2. Comparison with prior work

Many prior studies have investigated the possibility of detecting fatigue with ECG signals. For time domain indicators, Zhang et al. [52] applied the SVM algorithm to estimate visual fatigue. Instead of using the HRV, they extracted the R, Q, S waves from the original ECG signals and applied SVM to detect visual display fatigue. The result showed that the average heart rate (AHR) and standard deviation of heart rate (SDHR) increased after the visual fatigue test. Song et al. [53]

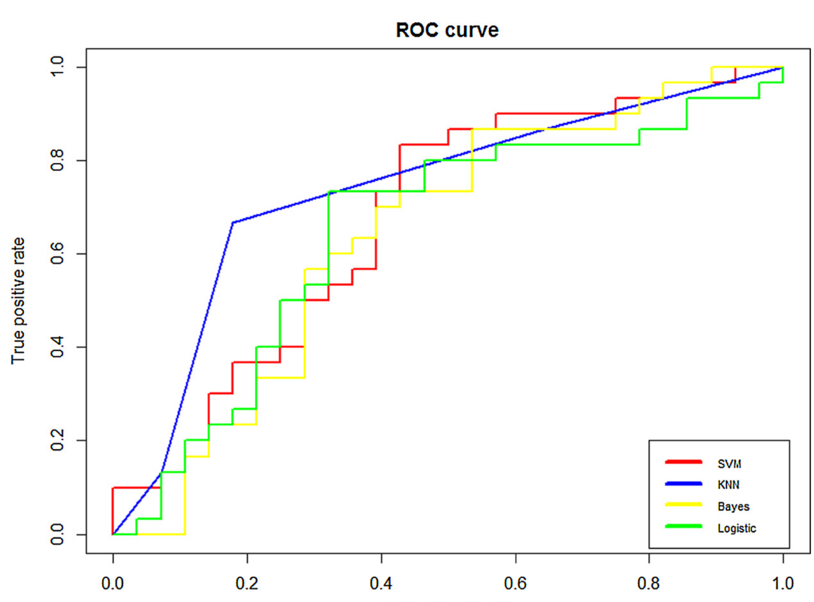

Fig. 3. The ROC curve of the four algorithms.

conducted sports experiments on nine long-distance runners and tested the relationship between the HRV and exercise-induced fatigue. Their result showed that the SDNN, rMSSD, and PNN50 decreased after exercise. However, those works are not closely related to mental fatigue due to overwork. In this study, we found that the rMSSD was positively associated with the mental fatigue state and that the PNN50 and NN.mean were negatively associated with the mental fatigue state.

As for frequency domain indicators, prior studies have explored the detection of driving fatigue and mental stress with ECG signals. Some studies have shown that a significant drop in the LF/HF during a driving test is related to a drowsy state [53–55], while some other studies suggest that the LF/HF may increase with a greater mental workload and higher stress [56,57]. Therefore, there is a conflict in the literature. The results from this study show that the LF, VLF and LF/HF were positively related to mental fatigue.

There are also studies that focus on combinations of indicators. For example, Zhang et al. demonstrated that with the feature combination, the accuracy rate of the visual fatigue classification can increase by 10% [52]. Another study also suggested that the combination of multiple linear and nonlinear features led to an increased classification accuracy for the detection of heart disease [58]. Similar to these prior studies, our study suggests that no single HRV indicator works best and that the best performance was achieved by using 2–3 HRV indicators (e.g., a combination of the NN.mean, PNN50, TP and LF). Therefore, investigating indicator combinations is a promising research direction for increasing the accuracy of mental fatigue detection.

5. Conclusion

In this study, we investigated the possibility of detecting the mental fatigue state with a convenient and inexpensive wearable ECG device. This device uses ADS1292R (developed by Texas Instruments) as the core chip to accurately acquire the ECG and multiple respiration states. Bluetooth is used to transmit data from the wearable ECG device to the smartphone. In this study, 35 healthy participants without heart disease were recruited from a public university in East China. Eight HRV indicators (i.e., the NN.mean, rMSSD, PNN50, TP, HF, LF, VLF and LF/HF) were collected at intervals of 5 min throughout the entire experiment. After the feature selection was performed, six indicators remained for further analysis, including the NN.mean, rMSSD, PNN50, TP, LF, and VLF. Four algorithms, SVM, KNN, NB, and LR, were used to build classifiers that automatically detect the fatigue state. The best performance achieved in this study was a CV accuracy of 75.5%. The NN.mean, PNN50, TP and LF were the key HRV indicators identified for the mental fatigue detection. KNN performed the best among the four algorithms and had an average CV accuracy of 65.37% on all of the possible feature combinations.

This study has several limitations. First, only 35 samples were collected, and only 29 samples were used in the analysis in this study. Due to the high cost of hiring subjects, a small sample size is a common limitation in similar studies [52]. However, future studies should include more subjects to increase the research reliability.

Second, the fatigue state was manipulated by a quiz with 30 logical referential and computing problems and 25 memory tests. The fatigue state that is manipulated by a quiz may be different from the state from the everyday work environment (e.g., overwork due to long hours spent at the workplace). Therefore, the validity of this research should be strengthened by an analysis in a real workplace in the future.

Third, our analysis results show that heart rate is an important indicator to the final mental fatigue diagnosis. However, the hear rate may be influenced by many factors. For example, various diseases (e.g., hyperthyroidism [59], myocarditis [60], pericarditis [61]) and conditions (e.g., high-intensity exercises [62]) can cause sinus tachycardia. Therefore, it would be important to test our proposed approach in populations of patients with chronic diseases (e.g. diabetes, hypertension, coronary heart disease), and in individuals with different level of physical fitness in the future.

This study demonstrates that the user's mental fatigue state can be detected with a reasonable accuracy using wearable ECG devices. Since wearable smart devices are inexpensive, convenient, popular and widely available today, real-time mental fatigue monitoring and overwork altering is a promising application in m-health. In addition, this study found that the fatigue state can be accurately predicted by using only 2 or 3 HRV indicators. This finding means that the ECG data can be collected and processed very quickly even when using low-cost, lightweight wearable chips.

Disclosure statement

The authors report no competing financial interests.

Authors' contributions

Shitong Huang: Data Collection.
Weiqiang Zhang: Data Processing.
Jia Li: Research Design, Writing.
Pengzhu Zhang: Research Idea, Research Design.

Acknowledgement

This research was supported by the National Natural Science Foundation of China with grants (71371005, 71471064, and 91646205).

Appendix A. Supplementary data

Supplementary material related to this article can be found, in the online version, at doi:https://doi.org/10.1016/j.ijmedinf.2018.08.010.

References

[1] T. Yamauchi, et al., Overwork-related disorders in Japan: recent trends and development of a national policy to promote preventive measures, Ind. Health 55 (3) (2017) 293–302.
[2] C. Weller, Japan is Facing a' Death by Overwork' Problem — Here's What It's All About, [cited 2018-4-26; Available from: (2017) http://www.businessinsider.com/what-is-karoshi-japanese-word-for-death-by-overwork-2017-10.
[3] Y. Nakao, 'Death by Overwork': Workaholic Japanese to Be Forced to Take Vacation Time, 2018-4-26]; Available from: (2015) https://www.rt.com/news/229943-japan-work-holiday-compulsory/.
[4] W. Sim, Death by Overwork: Will Japan Finally Face up to 'Karoshi'? [cited 2018-4-26; Available from: (2017) https://www.straitstimes.com/asia/east-asia/will-japan-finally-face-up-to-karoshi.
[5] S. Oster, Is Work Killing You? In China, Workers Die of Overwork at Desks, 2018-4-26; Available from: (2014) https://www.ohio.com/akron/business/is-work-killing-you-in-china-workers-die-of-overwork-at-desks.
[6] E. Grandjean, Fitting the Task to the Man: An Ergonomic Approach, Taylor & Francis, 1969, p. 210.
[7] D.S. Ke, Overwork, stroke, and karoshi-death from overwork, Acta Neurol. Taiwan 21 (2) (2012) 54–59.
[8] E.M. Backé, et al., The role of psychosocial stress at work for the development of cardiovascular diseases: a systematic review, Int. Arch. Occup. Environ. Health 85 (1) (2012) 67–79.
[9] M.M. Goedendorp, et al., Chronic fatigue in type 1 diabetes: highly prevalent but not explained by hyperglycemia or glucose variability, Diabetes Care 37 (1) (2014) 73–80.
[10] J.E. Bower, et al., Fatigue, cancer, J. Clin. Oncol. Off. J. Am. Soc. Clin. Oncol. 32 (17) (2014) 1840–1850.
[11] M.A. Staal, Stress, Cognition, and Human Performance: A Literature Review and Conceptual Framework, National Aeronautics & Space Administration, 2004.
[12] N.L. Galambos, B.J. Walters, Work hours, schedule inflexibility, and stress in dual-earner spouses, Canad. J. Behav. Revue Canad. Des. Du Comport. 24 (3) (1992) 290–302.
[13] D.C. Ganster, J. Schaubroeck, Work stress and employee health, J. Manag. 17 (2) (1991) 235–271.
[14] T. Chalder, et al., Development of a fatigue scale, J. Psychosom. Res. 37 (2) (1993) 147–153.
[15] E. Hoddes, et al., Quantification of sleepiness: a new approach, Psychophysiology 10 (4) (1973) 431–436.
[16] L.B. Krupp, The fatigue severity scale, JAMA Neurol. 46 (10) (1989) 1121.
[17] M.W. Johns, A new method for measuring daytime sleepiness: the Epworth

sleepiness scale, Sleep 14 (6) (1991) 540.
[18] K. Al-Shair, et al., Shortening, dimensions, reliability and validity of a novel COPD-specific fatigue scale (COPD-SFS), Eur. Respir. J. 34 (Suppl. 53) (2009) 1387.
[19] A.S. Urrila, et al., Psychomotor vigilance task performance during total sleep deprivation in young and postmenopausal women, Behav. Brain Res. 180 (1) (2007) 42–47.
[20] R.B. Sangal, L. Thomas, M.M. Mitler, Maintenance of wakefulness test and multiple sleep latency test. Measurement of different abilities in patients with sleep disorders, Chest 101 (4) (1992) 898–902.
[21] P. Artaud, et al., An on-board system for detecting lapses of alertness in car driving, The Fourteenth International Technical Conference on the Enhanced Safety of Vehicles, (1994).
[22] S.K.L. Lal, A. Craig, Electroencephalography activity associated with driver fatigue: implications for a fatigue countermeasure device, J. Psychophysiol. 15 (3) (2001) 183–189.
[23] Ashley Craig, Driver fatigue: electroencephalography and psychological assessment, Psychophysiology 39 (3) (2002) 313–321.
[24] S.K.L. Lal, et al., Development of an algorithm for an EEG-based driver fatigue countermeasure, J. Saf. Res. 34 (3) (2003) 321–328.
[25] J.E. Hong, K.C. Min, S.H. Kim, Electroencephalographic study of drowsiness in simulated driving with sleep deprivation, Int. J. Ind. Ergon. 35 (4) (2005) 307–320.
[26] S.K. Lal, A. Craig, A critical review of the psychophysiology of driver fatigue, Biol. Psychol. 55 (3) (2001) 173–194.
[27] P.S. Pandian, et al., Smart Vest: wearable multi-parameter remote physiological monitoring system, Med. Eng. Phys. 30 (4) (2008) 466–477.
[28] A.M.N.I.B. Blau, Forecast: Wearable Electronic Devices, Worldwide, 2017, Available from: (2017) https://www.gartner.com/doc/3778064?ref=SiteSearch&sthkw=global%20wearables&fnl=search&srcId=1-3478922254.
[29] K.-L. Tan, F.E.H. Tay, Memswear - incorporating mems technology into smart shirt for geriatric healthcare, Int. J. Comput. Eng. Sci. 4 (2) (2003) 289–292.
[30] R. Paradiso, Wearable healthcare system: new frontier on e-textile, J. Telecommun. Inf. Technol. 4 (2005) 105–113.
[31] M. Pacelli, et al., Sensing fabrics for monitoring physiological and biomechanical variables: E-textile solutions, IEEE/EMBS International Summer School on Medical Devices and BIOSENSORS, (2006).
[32] C. Otto, et al., System architecture of a wireless body area sensor network for ubiquitous health monitoring, J. Mob. Multimed. 1 (4) (2006) 307–326.
[33] L. Sherwood, Human Physiology: From Cells to Systems, Cengage Learning, Brooks/Cole, 2012.
[34] R. Ren, et al., A bluetooth-based portable design device with wireless power module for electrocardiogram and respiration measurement, Appl. Mech. Mater. 441 (2013) 129–132.
[35] T. Chalder, et al., Development of a fatigue scale, J. Psychosom. Res. 37 (2) (1992) 147–153.
[36] L.J. Mulder, Measurement and analysis methods of heart rate and respiration for use in applied environments, Biol. Psychol. 34 (2–3) (1992) 205–236.
[37] P.A. Hancock, W.B. Verwey, Fatigue, workload and adaptive driver systems, Accid. Anal. Prev. 29 (4) (1997) 495–506.
[38] S. Boonnithi, S. Phongsuphap, Comparison of heart rate variability measures for mental stress detection, Comput. Cardiol. 19 (1) (2011) 85–88.
[39] R. Costin, C. Rotariu, A. Pasarica, Mental stress detection using heart rate variability and morphologic variability of EeG signals, International Conference and Exposition on Electrical and Power Engineering, (2013).
[40] R.J. Winchell, D.B. Hoyt, Spectral analysis of heart rate variability in the ICU: a measure of autonomic function, J. Surg. Res. 63 (1) (1996) 11–16.
[41] N. Rentero, et al., Activity patterns of cardiac vagal motoneurons in rat nucleus ambiguus, Ajp Regul. Integr. Comp. Physiol. 283 (6) (2002) R1327–1334.
[42] P.A. Lanfranchi, V.K. Somers, Arterial baroreflex function and cardiovascular variability: interactions and implications, Am. J. Physiol. Regul. Integr. Comp. Physiol. 283 (4) (2002) R815.
[43] S.C. Malpas, Neural influences on cardiovascular variability: possibilities and pitfalls, Am. J. Physiol. Heart Circ. Physiol. 282 (1) (2002) H6.
[44] K. Aoki, D.P. Stephens, J.M. Johnson, Diurnal variation in cutaneous vasodilator and vasoconstrictor systems during heat stress, Am. J. Physiol. Regul. Integr. Comp. Physiol. 281 (2) (2001) R591.
[45] M.S. Blumberg, T.G. Knoot, R.F. Kirby, Neural and hormonal control of arterial pressure during cold exposure in unanesthetized week-old rats, Am. J. Physiol. Regul. Integr. Comp. Physiol. 281 (5) (2001) R1514–21.
[46] J.M. Overton, et al., Cardiovascular and metabolic responses to fasting and thermoneutrality are conserved in obese Zucker rats, Am. J. Physiol. Regul. Integr. Comp. Physiol. 280 (4) (2001) R1007.
[47] A.N. Braga, et al., Effects of angiotensins on day-night fluctuations and stress-induced changes in blood pressure, AJP Regul. Integr. Comp. Physiol. 282 (6) (2002) R1663–1671.
[48] G. Potter Jr, S.A. Rivkees, Ontogeny of humoral heart rate regulation in the embryonic mouse, AJP Regul. Integr. Comp. Physiol. 281 (281) (2001) R401–407.
[49] H. Han, X. Guo, H. Yu, Variable selection using mean decrease accuracy and mean decrease Gini based on random Forest, 2016 7th IEEE International Conference on Software Engineering and Service Science (ICSESS) (2016).
[50] S. Nizami, et al., Heart disease classification through HRV analysis using parallel cascade identification and fast orthogonal search, 2010 IEEE International Workshop on Medical Measurements and Applications Proceedings (MeMeA) (2010).
[51] P. Pradhapan, et al., Effect of heart rate correction on pre-and post-exercise heart rate variability to predict risk of mortality—an experimental study on the FINCAVAS cohort, Front. Physiol. 5 (2014) 208.
[52] A. Zhang, Z. Zhao, Y. Wang, Estimating VDT visual fatigue based on the features of ECG waveform, Proceedings of the International Workshop on Information Security, (2009).
[53] T. Song, et al., Research on the application of HRV in the evaluation of exercise fatigue for experimental teaching, Lab. Sci. (2017).
[54] M.K. Byeon, et al., A Study of HRV Analysis to Detect Drowsiness States of Drivers, (2006), pp. 153–155.
[55] M. Patel, et al., Applying neural network analysis on heart rate variability data to assess driver fatigue, Expert Syst. Appl. 38 (6) (2011) 7235–7242.
[56] N. Sato, et al., Power spectral analysis of heart rate variability in type A females during a psychomotor task, J. Psychosom. Res. 45 (2) (1998) 159.
[57] D. Nagae, A. Mase, Measurement of vital signal by microwave reflectometry and application to stress evaluation, Microwave Conference, 2009 (2009).
[58] S. Nizami, et al., Heart disease classification through HRV analysis using parallel Cascade identification and fast orthogonal search, IEEE International Workshop on Medical Measurements and Applications Proceedings, (2010).
[59] J. Echeverría, et al., The autonomic condition of children with congenital hypothyroidism as indicated by the analysis of heart rate variability, Auton. Neurosci. 167 (1-2) (2012) 7–11.
[60] N. Ling, et al., Heart rate variability in children with myocarditis presenting with ventricular arrhythmias, Eur. Rev. Med. Pharmacol. Sci. 22 (4) (2018) 1102–1105.
[61] C.M. de Oliveira Filho, et al., Increased rate of postoperative pericarditis with contact force sensing catheters in atrial fibrillation ablation, J. Am. Coll. Cardiol. 67 (13 Supplement) (2016) 758.
[62] P. Gonzalez, et al., Heart rate and excess post-exercise oxygen consumption of whole-body high intensity interval training Vs. moderate-intensity exercise, Int. J. Exerc. Sci.: Conf. Proc. 8 (4) (2016) 37.