Research on Abnormal Heart Rate Detection Method of Athletes Based on Wireless Sensing Multi-source Information Fusion

Youchun Zhang[✉]

School of Application Engineering, Anhui Business and Technology College, Hefei 231131, China
zhangyouchun@ahbvc.edu.cn

Abstract. Most of the existing heart rate anomaly detection methods only use a single signal to determine the abnormal heart rate state of college athletes, which is easily disturbed, and when the actual detection environment is not so ideal, it is impossible to collect enough accurate data, which can lead to a large deviation in the test results. This paper proposes a heart rate anomaly detection method based on the fusion of random forest and multi-source information, and uses multiple miniature sensors to synchronously collect the breathing, heartbeat, pulse and grip signals of college athletes, when a signal acquisition is biased The other signal is not affected, and then all the collected signals are integrated to establish a multi-source fusion heart rate abnormal state dataset as the basis for detection, avoiding the defect that a single signal sometimes has a larger error, and the detection accuracy rate is higher. Compared with classification algorithms such as SVM and GBDT, random forests can balance the error of the data set, are insensitive to missing values, and the classification speed is fast, and the optimal detection accuracy of the model is 89.16%, which can effectively avoid abnormal heart rate of college athletes.

All along, the world has attached great importance to how to avoid abnormal heart rate, and more and more companies and scientific research institutions have designed various devices to avoid abnormal heart rate. The detection method of abnormal heart rate is generally divided into three methods: based on physiological signals, based on exercise information, and based on the faces of college athletes [3]. Physiological information detection method through the direct contact with the human body sensor to collect the physiological signals of college athletes to judge the abnormal state of heart rate in college athletes in sports, this method of obtaining reliable data, and physiological signals directly from the human body, accurate and reliable when analyzing abnormal heart rate states, but usually to install a device to detect physiological signals on college athletes, the interference of college athletes is larger. The Motion Information Detection Method detects abnormal heart rate states of college athletes based on signal characteristics such as detecting the driver's control of the steering wheel [4], the change of rotation speed

and slowness, whether the pressure on the accelerator pedal is smooth, and the driving trajectory of the movement. This detection method is simple to implement, but it is susceptible to environmental factors, such as the good environment of the driving road surface, whether the climate is harsh, etc., and the anti-interference ability of the detection system is poor [5]. In addition, some normal driving operations such as hyper-movement, parallel lanes, etc., may also lead to misjudgment.

The face information detection method of college athletes installs high-definition cameras in sports, and when college athletes start sports, the cameras begin to work. Through computer vision, the changes in facial expression characteristics of college athletes are detected, such as blinking frequency, eyelid closure PERCLOS, eye tracking, pupillary response [6], head rotation amplitude and height lifting, whether to yawn and other behaviors Assess whether college athletes have entered abnormal heart rate states, which is less costly and simple to achieve.

However, due to the fact that the clarity of the shooting image is determined by the ambient lighting conditions, this method has higher requirements for light, and the camera has been shooting at the university athletes, which will have a certain impact on the privacy of the college athletes, and the college athletes are prone to resistance and irritability psychologically Affect driving safety. Most of the existing heart rate anomaly detection methods only use a single signal to judge the abnormal state of heart rate of college athletes, and the signal collection during the actual exercise process is easily disturbed by various factors, and sometimes the collected data is not accurate enough, due to the single basis for discrimination, the trained model is poorly robust, prone to false alarms, etc., but affects the safety of college athletes. Therefore, this paper combines random forests with multi-source information fusion for detection. The multi-source information fusion detection method uses multiple micro-sensors with high accuracy [7] to synchronously collect signals such as breathing, heartbeat, pulse, grip and other signals of college athletes, and when the collection of one signal is deviated, the acquisition of another signal is not interfered with This information is then filtered and processed by Fourier transform, and a multi-source heart rate anomaly state dataset is established, which avoids the defect of poor anti-interference of a single signal [8], and greatly improves the accuracy of heart rate anomaly detection. Compared with classification algorithms such as SVM (Support Vector Ma-chines) and GBDT (Gradient Boosting Decision Tree), random forests can balance the error of the data set and are not sensitive to missing values. It can operate in parallel and has a high classification speed, which meets the requirements of signal acquisition with certain errors and high real-time requirements in heart rate anomaly detection.

1   Multi-source Signal Acquisition Platform for Abnormal Heart Rate

The experimental environment set up is shown in Fig. 1, and the actual line movement environment is simulated by simulating the driver to ensure the safety of the tester. The signal acquisition platform mainly includes three parts: Doppler radar module, flexible grip sensor module [9] and photoplethysmography pulse sensor module. In addition, the facial images of cameras recording the driving process of college athletes were used as

the basis for expert criticism of the data set signal classification [10]. The hardware design of the three modules all use smaller sensors to minimize contact with college athletes and prevent influence on the operation of college athletes. After the initial signal is collected, the physiological signals of college athletes collected by Doppler radar are filtered by a zero-phase ellipse filter to separate the breathing and heartbeat signals. The low-frequency noise in the pulse signal and the fixed power frequency interference at 50 Hz are filtered out by the zero-phase Butterworth band-pass filter and the Chebyshev II band-stop filter to obtain a clear and complete pulse signal and facilitate feature extraction.

In order to increase the collection efficiency, the experiment was selected in the afternoon, and the human body's activity level was generally low during this period, and it was easy to be sleepy, so that the test college athletes could better enter the abnormal state of heart rate. After the system is built and debugged, the experimental personnel are arranged to collect data.

2 Detection of Heart Rate Anomalies in the Fusion of Random Forests and Multi-source Information

The random forest is one of the classical ensemble learning algorithms, and its basic unit is the same as the gradient boost tree algorithm (GBDT), which is also a decision tree, but the type of the underlying decision tree can vary according to the actual data set. It is not necessary to use the same base learner for all. In addition, unlike GBDT, it adopts the idea of Bagging, which improves the stability of decision tree classification.

2.1 Random Forest Algorithm Model

At the heart of the random forest algorithm are two words, one is "forest" and the other is "random". "Forest" is because it integrates many decision trees, and the random forest algorithm uses the underlying decision tree model as a weak learner. In this paper, the Stochastic Forest mainly uses the Gini coefficient to calculate the impurity of the model nodes, and the smaller the Gini coefficient calculated from the node sample, the lower the impurity of the data The better the selection of representative features. Suppose a classification problem has a total of K different categories, and the probability of the k-th category is pk, then its probability distribution Gini coefficient expression is: the node of each decision tree in a random forest, select the values to split the dataset among the features of the input sample to minimize the division of Gini impurities.

Because the sample and feature selection composition of each base learner are different, the generalization ability is greatly enhanced. As shown in Fig. 2, the random forest is trained multiple times by generating different decision trees, and after the results are obtained separately, the categories of the test signal data are selected by voting method, and the result with the highest vote is used as the output. Assuming that a training set has k samples and samples it randomly, the probability that the same sample will be picked each time it is sampled is $1/k$. The probability that k random samplings did not choose it is: Bagging sampled the samples of the training dataset, and then selected them in turn, and about 36.8% of the samples in the training set will not be selected, which is called

" Out of bag data". These "out-of-bag data" do not have learning data as models, and can be used for "outsourced estimation" to analyze whether the learned classification model has good generalization performance. For random forest generation, trees are independent of each other, there are no dependencies, and can be trained in parallel, so the operation speed is fast. Its node division feature attribute set is randomly selected, without manual selection, compared to SVM and GBDT algorithms, it is not afraid of the loss of certain data features It is insensitive to outliers and has a good performance for high-dimensional data with many features.

2.2   Experimental Results and Analysis

Comparing the performance of the stochastic forest algorithm under the single signal dataset and the multi-source fusion dataset, it is verified that the multi-source information fusion has a good effect on improving the accuracy of the heart rate anomaly detection model, and then the established multi-source information fusion dataset is further input to the classification model based on SVM, GBDT and random forest for analysis and comparison. The superiority of the random forest algorithm as a classification algorithm model for heart rate anomaly detection is verified, and the classification model of heart rate anomaly detection under the optimal parameters is obtained after the detection method is adjusted.

3   Advantages of Multi-source Information Fusion Datasets

The experiment used the single signals of heartbeat, respiration, pulse and grip force to train the random forest classification algorithm model, and compared and analyzed the performance of the classification model trained by the multi-source information fusion dataset in the detection of heart rate abnormality. The optimal detection classification effect obtained in the test is shown in Fig. 1 and Fig. 2.

The detection accuracy of each heart rate anomaly status level can be seen that the heart rate anomaly detection model based on a single signal may have better classification accuracy in a certain category, but the overall detection accuracy is far less than that of the heart rate anomaly detection model based on multi-source information fusion. The best detection accuracy of the stochastic forest algorithm model after multi-source information fusion is about 89%, while the detection accuracy of a single dataset is only about 75%. Compared with a single data set, the random forest model based on multi-source information fusion increased the detection rate of abnormal heart rate states by 14%, which verified the superiority of multi-source information fusion.

The multi-source information fusion dataset is established to train SVM, GB-DT, and random forest algorithm models, respectively. The sample dataset is randomly divided into training sets and test sets, and the GBDT classification model and random forest classification model are trained with the training set, and the test set tests the classification model effect. The main adjustment parameters in the training process of random forest classification model are: the maximum number of features established by the decision tree (max-feature) and the maximum depth of the decision tree (max-depth), The number of decision trees generated (n_estimators). After multiple trainings, the classification
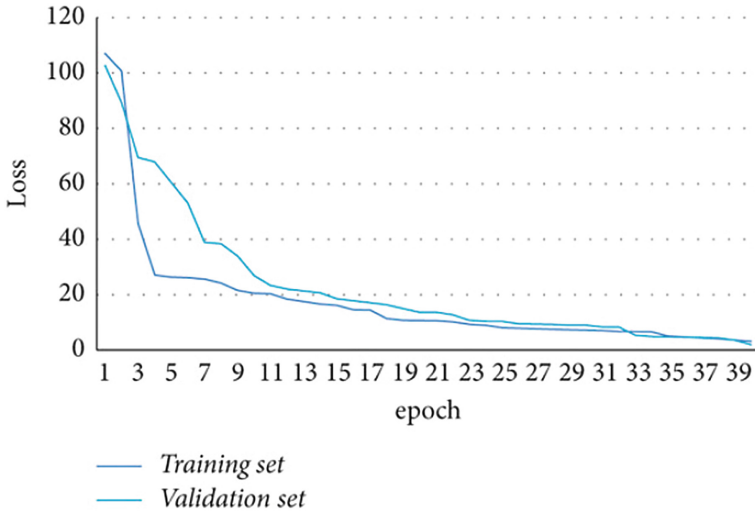
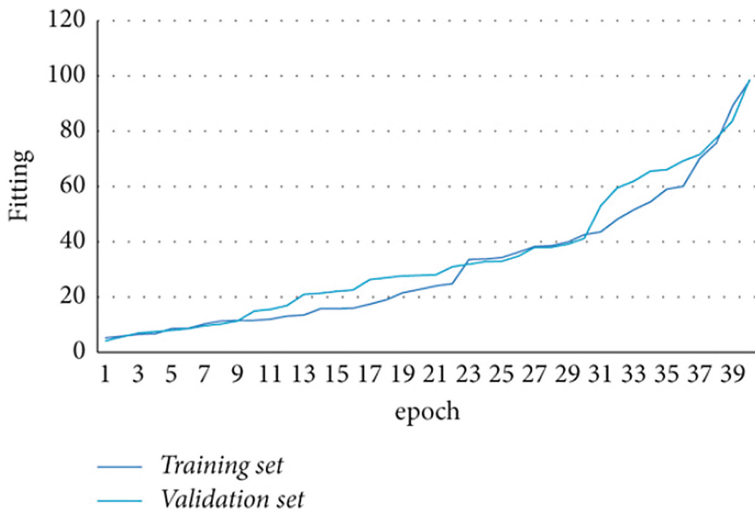**Fig. 1.** Training process loss convergence curve.

Fig. 2. Training process performance improvement diagram.

model is trained to obtain SVM, GBDT and random forest algorithm models under the multi-source information fusion dataset learning, and the entire data set is input into three abnormal heart rate driving detection models for testing. The output renderings of heart rate anomaly detection under different classification algorithms (see Fig. 3) and related accuracy parameters are compared.

Combined with the detection effect shown in Fig. 2, it can be seen that the random forest algorithm model has a good performance for the detection of abnormal heart rate
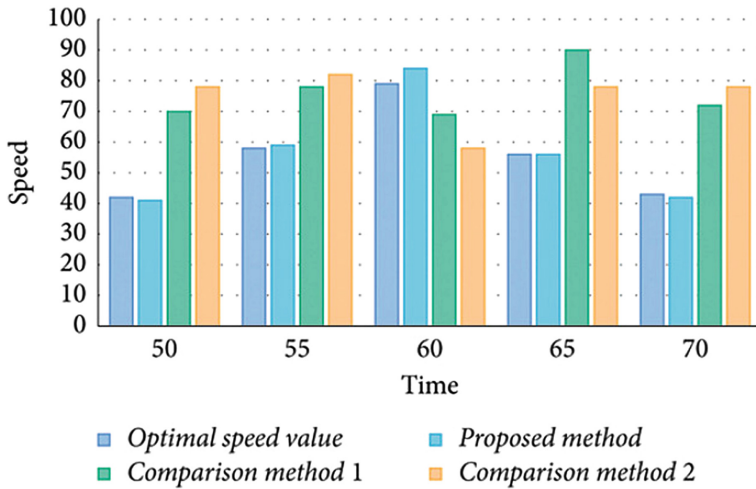
Fig. 3. Fatigue driving status detection effects of Different algorithm models

states at all levels, and the recognition accuracy is not much improved compared with SVM and GBDT. In addition, random forests can produce decision trees in parallel, and the detection model is also faster than SVM and GBDT for abnormal heart rate states. In order to avoid the deviation of the experimental effect caused by the accidental division of the test data set, the model automatically establishes the test set by random sampling multiple times, analyzes the detection accuracy of different classification algorithms, and takes the average of the detection accuracy of multiple detections in the experiment. The above experiments verify that the algorithm model based on random forest and multi-source information fusion can accurately detect abnormal heart rate states, and the best detection accuracy of the random forest algorithm model after multi-source information fusion is 89.18%. Compared with GBDT and classic SVM, its training speed is faster and more accurate, and it basically meets the requirements of real-time and high-precision detection of abnormal heart rate conditions.

4   Conclusion

Based on the heart rate anomaly detection method of random forest and multi-source information fusion, the signal acquisition module adopts a miniature sensor to reduce the deep contact with college athletes, and will not cause physical and mental interference to college athletes during exercise. After the collected signals are filtered wave processing, a fusion dataset with high reliability is established, and a random forest model is designed to learn and train the heart rate abnormal state data set, which realizes the high-precision and high-speed detection of heart rate abnormalities, and avoids the problem of large detection errors based on only a single signal. In addition, for the individual differences of athletes in different colleges and universities, the heart rate anomaly detection model is prone to deviations, and in the future, it is necessary to analyze more signal characteristics that can accurately reflect the abnormal heart rate state of college athletes, try to select

those feature values that are less affected by the driving environment and individual differences, and optimize the classification algorithm to make the classification model have self-learning ability, so as to achieve the optimal heart rate abnormal classification model that adapts to the individual through self-learning training.

References

1. ISI. Network Simulator. Development was supported by DARPA through the VINT project at LBL, Xerox PARC, UCB, and USC/ISI in 2020. http://www.isi.edu/nsnam/ns/
2. Meng, M., Wu, X., Jeong, B.-S., Lee, S., Lee, Y.-K.: Energy efficient routing in multiple sink sensor networks. In: Proceedings of the 5th International Conference on Computational Science and Its Applications (ICCSA 2007), pp. 561–566, IEEE, Kuala Lumpur, Malaysia, August 2021
3. Wang, C., Wu, W.: A load-balance routing algorithm for multi-sink wireless sensor networks. In: Proceedings of the International Conference on Communication Software and Networks (ICCSN 2009), pp. 380–384, Macau, China, February 2019
4. Robert, S., Wayne, K.: Ghaps. In: Algorithms, chapter 4. Princenton Editorial, 4th edn. (2019)
5. Chang, X., Wu, J., He, Z., Li, D., Sun, H., Wang. Understanding user's travel behavior and city region functions from station-free shared bike usage data. Transp. Res. Part F: Psychol. Behav. (2020)
6. Xing, Y., Wang, K., Lu, J.J.: Exploring travel patterns and trip purposes of dockless bike-sharing by analyzing massive bike-sharing data in Shanghai, China. J Transp. Geography (2020)
7. Lee, D., Derrible, S., Pereira, F.C.: Comparison of Four Types of Artificial Neural Network and a Multinomial Logit Model for Travel Mode Choice Modeling. Transp. Res. Record (49) (2020)
8. Ma, X., Liu, C., Wen, H., Wang, Y., Wu, Y.-J.: Understanding commuting patterns using transit smart card data. J. Transp. Geography (2019)
9. Paulssen, M., Temme, D., Vij, A., Walker, J.L.: Values, attitudes and travel behavior: a hierarchical latent variable mixed logit model of travel mode choice. Transp. (4) (2019)
10. Srinivasan, K.K., Mahmassani, H.S.: Analyzing heterogeneity and unobserved structural effects in route-switching behavior under ATIS: a dynamic kernel logit formulation. Transp. Res. Part B (9) (2021)