Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Application of Random Forest Algorithm Based on Fiber Optic Sensors in Fatigue Detection in Sports Competitions

Zhaoyu Chen

zhaoyuchen202212@126.com

Xinjiang University Of Political Science And Law

Research Article

Keywords: Fiber optic sensors, Random Forest Algorithm, Sports competitions, Fatigue detection

Posted Date: January 16th, 2024

DOI: https://doi.org/10.21203/rs.3.rs-3859351/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Application of Random Forest Algorithm Based on Fiber Optic Sensors in Fatigue Detection in Sports Competitions

Zhaoyu Chen
Department of Basic Subjects Teaching, Xinjiang University Of Political Science And Law, Tumshuk City Xinjiang, 843900 China
Corresponding Author: Zhaoyu Chen, e-mail: zhaoyuchen202212@126.com

Abstract:
In the field of competitive sports, increasing training intensity is mostly used to improve sports and competition levels. However, after high-intensity training, the function of the body muscles will decrease, which is known as muscle fatigue. If there is a lack of reasonable control over exercise intensity, athletes may experience muscle fatigue and sports injuries. Based on this, this article takes the random forest algorithm of fiber optic sensors as the design basis and develops a fatigue detection system for sports competitions. This article first analyzes the relevant principles of fiber optic sensors, and based on their corresponding mode coupling theory, derives formulas for the output content of sensors. Through experiments, the advantages of fiber optic sensors are demonstrated. Then, the random forest algorithm was analyzed and improved, which belongs to the classic ensemble learning algorithm. The model used is universal, easy to understand, and not prone to overfitting. However, in terms of dynamic data classification, the performance is poor. The improved random forest algorithm not only has excellent detection performance, but also higher prediction accuracy than before. Finally, this article successfully developed a fatigue detection application system for sports competitions, which can detect the fatigue level of athletes in real time, adjust the intensity of exercise, and effectively avoid potential safety hazards during the exercise process.

Key words: Fiber optic sensors; Random Forest Algorithm; Sports competitions; Fatigue detection

1 Introduction

With the increasing richness of research on sports fatigue, corresponding analysis and detection methods have become more diverse. They have evolved from subjective detection methods (questionnaire surveys) and biological detection methods to sensor signal detection today (McKee et al. 2018). However, early sensor detection required the tester to be in a stationary state, and the degree of human motion fatigue determined was not accurate, and fatigue detection during movement was also impossible. And when collecting various types of data, it heavily relies on professional equipment, most of which are very large in size and weight, and there are difficulties in connecting and complex operations. Fiber optic sensors not only have a simple internal structure and high sensitivity, but their response speed is unmatched by most other types of sensors, which effectively promotes the development of new sensing technologies (Du et al. 2020). At the same time, the development of new technologies has placed higher demands on this sensor,

and people's demand for lightweight and portable devices is constantly increasing. Therefore, the random forest algorithm based on fiber optic sensors has become a hot research topic (Sarica et al. 2017). This algorithm can quickly and in real-time analyze the user's physical fatigue status through monitoring data. The emergence of fiber optics has led to the development of fiber optic sensing technology, and with the development of new fiber optic materials, various new fiber optic sensors have been widely used in people's lives, promoting the development and growth of the fiber optic sensing field, and also providing assistance for exploring unknown areas of technological research and development (Bao et al. 2019). Driven by this demand, the random forest algorithm has emerged, which performs well in data ensemble learning and classification and gradually develops into a hot research topic in the field of information technology (Zhu et al. 2019). And the models involved in this algorithm need to be combined in order to run, which also reduces the occurrence of overfitting and has a much better performance than decision tree algorithms. In addition, the random forest algorithm can effectively handle problem data, such as data with missing values, noise, and outliers (Everingham et al. 2016). It is precisely because of its excellent performance that this algorithm is widely used in many practical fields. In summary, this article uses a combination of fiber optic sensors and random forest algorithm to design and develop a sports competition fatigue detection system, which is used to analyze the degree of sports fatigue, thereby assisting athletes in real-time grasping the physical fatigue status during the exercise process, achieving scientific and reasonable exercise intensity control, and reducing unnecessary sports injuries (Thorpe et al. 2015).

2 Related work

The literature emphasizes that in order to help athletes achieve scientific training, the primary condition is to accurately grasp the specific sports characteristics and corresponding laws of sports projects (Clarke et al. 2021). The level of understanding of sports projects is a major determining factor in improving the level of sports training. This is because only by comprehensively understanding and mastering the characteristics of sports can we choose reasonable training content, training methods, and programs to alleviate sports fatigue that meet the needs of specialized competitions. The literature indicates that in addition to its inherent advantages, the sensitivity index corresponding to fiber optic sensors is also a key parameter to consider in the manufacturing process, because sensitivity has a very direct impact on the resolution and measurement accuracy of the sensor (Zheng et al. 2020). In addition, the method of improving the sensitivity index of fiber optic sensors no longer relies solely on adding sensitizing substances to the fiber optic, but on improving and optimizing the internal structure of the fiber optic on this basis. The literature provides a brief overview of the internal structure of optical fibers, which mainly includes three major components: internal fiber core, external cladding layer, and coating layer (Addanki et al. 2018). The refractive index of the internal fiber core must exceed that of the external wrapping layer in order to achieve continuous reflection and propagation of light within the internal fiber core. The literature mentions that when calculating the importance of features, the random forest algorithm mainly analyzes the error before and after obtaining features through random permutation of out of bag data (Simsekler et al. 2020). The larger the value obtained, the more important the corresponding features become, and thus determines the combination of feature sequences. The relevant models of the random forest algorithm are far superior to other

classical types of sorting combinations in reflecting the interaction relationship between features. However, it should also be noted that as the total number of features included in the dataset increases, the computational efficiency of the random forest algorithm will decrease. Therefore, how to efficiently extract valuable feature information using this algorithm and quickly construct a reasonable feature sequence is also a problem that many scholars in this field have been paying attention to and researching. The literature indicates that using the generalization error in the random forest algorithm can effectively calculate the specific generalization ability of the model (Advani et al. 2020). The smaller the generalization calculation error value of this algorithm, the stronger its generalization ability, and vice versa.

3 Theoretical basis of random forest algorithm based on fiber optic sensors

3.1 Principle analysis of fiber optic sensors

The corresponding expression for step type optical fibers is as follows:

$$n(r) = \begin{cases} n_1, r \le a \\ n_2, r > a \end{cases} (1)$$

Essentially, fiber optic mode is a type of physical electromagnetic field field distribution that can be expressed mathematically in the form of fundamental solutions. The propagation of light in vacuum and transparent dielectric waveguides is fundamentally different, so multiple vectors need to be used when describing. The required Maxwell square program is as follows, where D and B represent the electromagnetic induction intensity, and the charge density in the field is represented by ρ To represent, the Hamiltonian calculation symbol is π.

$$\nabla \cdot D = \rho, \nabla \cdot B = 0 (2)$$

Simplifying the Maxwell equation above yields the following expression, also known as the Helmholtz computational equation.

$$\nabla^2 E + k^2 E = 0, \nabla^2 E + k^2 E = 0 (3)$$

In this equation, the refractive index of the optical fiber is n, and the wavelength of the light wave is $\lambda_{\circ}$

The formula for calculating the normalized frequency in optical fibers is as follows (4), which is also a major related parameter in the mode field distribution.

$$V = \frac{2\pi a}{\lambda}\sqrt{n_1^2 - n_2^2} (4)$$

By using the calculation results of the above equation, the mode propagation data inside the fiber can be obtained. The larger the calculation result, the greater the total number of guided modes. When defining the effective refractive index of guided modes, the propagation constant can be used β, Obtain:

$$n_{eff} = \frac{\beta}{k_0} (5)$$

However, to ensure the effectiveness of the guided mode, the constant β Must meet k0n2 $\leqslant \beta \leqslant$ K0n1.

Figure 1 clearly shows the energy coupling changes of the fiber optic sensor. It can be seen that when the optical power enters through the fiber core A, the energy in A will be completely coupled to B, because the distance between the two fiber cores A and B is relatively close. And the energy in core B will also be coupled to core A, ultimately outputting energy from both cores.

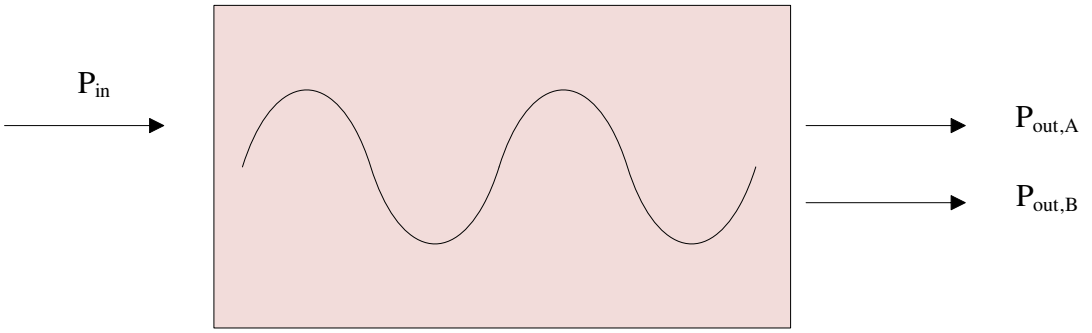


Figure 1 Schematic diagram of energy coupling for fiber optic sensors

When exploring the energy changes of the fiber core in fiber optic sensors, coupling theory will be used, and the corresponding equation system is as follows:

$$\begin{cases} \frac{da_A(z)}{dz} = j\beta_A a_A(z) + jK_{BA} a_B z \\ \frac{da_B(z)}{dz} = j\beta_B a_B(z) + jK_{AB} a_{QA} \end{cases} (6)$$

As shown in equation (6), a (z) represents the mode field amplitude of the fiber core at position Z, β Then it corresponds to the propagation constant, and KAB and KBA correspond to the coupling coefficient between the two fiber cores.

By utilizing the theory of light related interference and phase calculation conditions, the wavelength calculation formula for the resonant peaks and valleys in the interferometer can be obtained, which is:

$$\lambda_m = \frac{2\Delta n_{eff}^0 L}{2m+1} (7)$$

Δ Neff0 represents the effective refractive index difference between the core and cladding, unaffected by curvature. The formula for calculating the strain difference between the core and cladding of optical fibers is as follows:

$$\Delta\varepsilon = D/R = D * C (8)$$

At this point, the effective refractive index difference between the two is expressed as:

$$\Delta n_{eff} = \Delta n_{eff}^0 + k\Delta\varepsilon = \Delta n_{eff}^0 + kD * C (9)$$

In the above formula, the refractive index coefficient of strain is expressed as k. By combining equations (7) and (9) above, the wavelength of the resonant peak and valley in the interferometer after variation can be obtained:

$$\lambda_m = \frac{2(\Delta n_{eff}^0 + r*C)L}{2m+1} (10)$$

According to the above formula, it can be concluded that λ M will change with the value of curvature C.

By analyzing the specific changes in light intensity in the sensor, the degree and direction of bending can be calculated. The light intensity sensing parameters are as follows:

$$M = 10\lg\frac{I_{max}}{I_{min}} (11)$$

Then we can obtain:

$$I_{max} = I_{CO} + I_{cl} = 2\sqrt{I_{CO} I_{cl}}, I_{min} (12)$$

By analyzing the above expressions (11) and (12), it is found that when the bending direction of the sensor is different, the measurement of the curvature of the sensor can be completed by

observing the changes in the resonant peak wavelength and light intensity sensing parameters of the transmission spectrum.

3.2 Construction of Random Forest Algorithm

The decision tree is the most basic unit in the random forest algorithm. Combining and arranging multiple decision trees can effectively improve the classification accuracy of the algorithm. The result of the combination arrangement is usually a binary or multi tree, and it forms a random forest classifier. The detailed structure is shown in Figure 2. If we need to solve the problem of regression classes, we need to obtain the average value of all decision trees in the classifier.

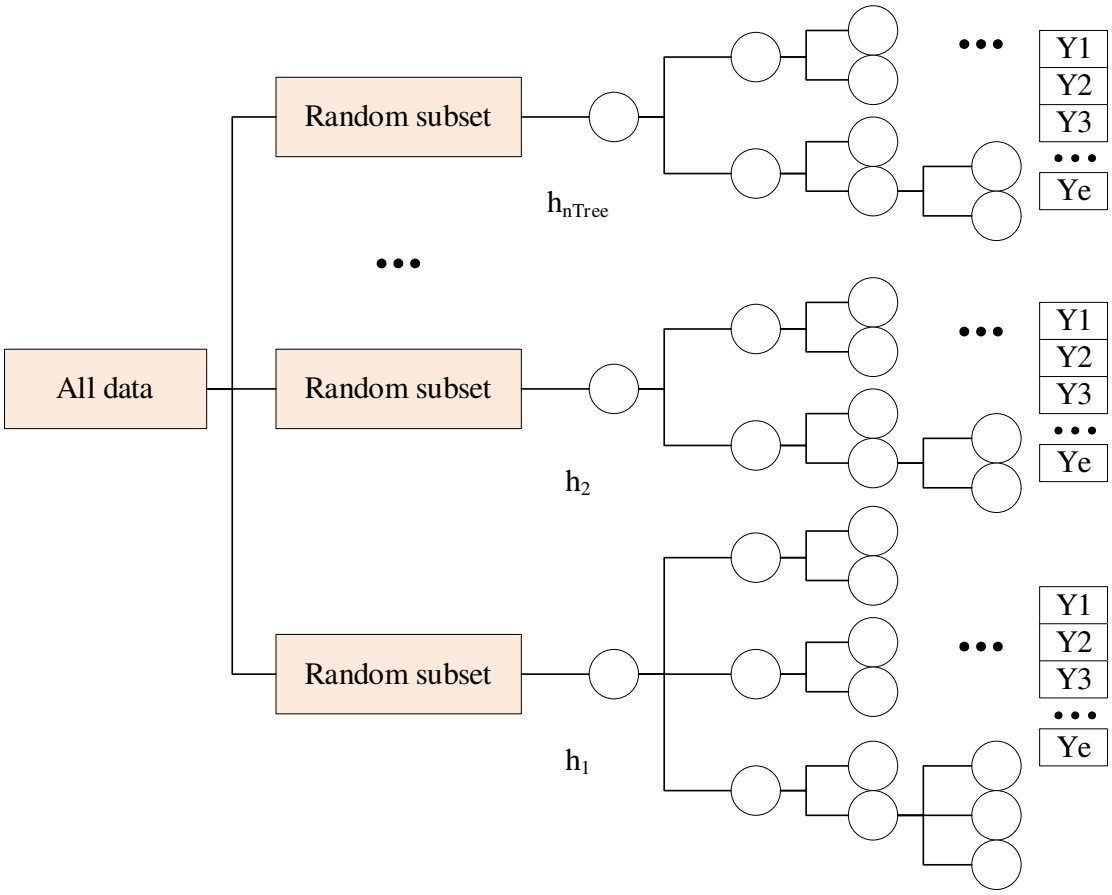


Figure 2 Schematic diagram of random forest

The most widely used algorithm in parallelization processing is the random forest algorithm, which has excellent predictive performance. Known weak learning algorithm H:

$H = \{h_1, h_2, \ldots, h_m\}(13)$

Train the learning model using weak learning algorithms and combine the final prediction results to obtain:

$H(x) = \text{majority}\{h_i(x)|i = 1,2,\ldots,m\}(14)$

If it involves a simple regression problem, simply simplify the average calculation.

$H(x) = \frac{1}{m}\sum_{i=1}^{m} h_i(x)(15)$

Train weak learning models using training datasets and adjust the distribution of training samples appropriately based on specific performance. Substitute the adjusted training sample distribution into the new learning model, perform repeated iterations, and weight the prediction results of the

model. The resulting expression is as follows:

$$H(x) = \frac{1}{m}\sum_{i=1}^{m} w_i h_i(x) (16)$$

By inputting sample x, the final output can be obtained as follows:

$$f(x) = \text{majority}\{h(x, \theta_k) | k = 1,2,\ldots,k\} (17)$$

For a certain event, if the total number of possible values is K, the probability of the k-th possible value occurring can be represented by pk, and the total entropy value of the possible values corresponding to a certain event can be expressed in the form of information entropy. In the actual calculation process, empirical probability can be used to replace the value of pk, and empirical information entropy can be expressed as:

$$H(D) = -\sum_{k=1}^{k} p_k \log_2 p_k (18)$$

The conditional entropy of an event refers to the entropy of a certain event that needs to be calculated based on other similar events, and the corresponding expected expression is as follows:

$$H(D|A) = \sum_{i,k} P(A_i) H(D_k|A_i) (19)$$

For a certain dataset D, entropy H (D) can be used to characterize its corresponding uncertainty. When the samples in the dataset belong to the same category:

$$H(D) = 0 (20)$$

By using information gain, the degree to which feature A reduces the classification confusion of dataset D can be characterized. The greater the impact of event A on D, the smaller the value of conditional entropy. In the variable selection of root or internal nodes, it is necessary to select the feature with the highest information gain.

The gain rate refers to the normalization process carried out on the basis of information gain to overcome the defect of biased selection values in information gain. The corresponding formula for calculating the information gain rate gR (D, A) is:

$$g_R(D, A) = H(D) (21)$$

In the above equation, the more values contained in event A, the greater the information gain rate may be.

The Gini index, as the optimal partitioning attribute, can be seen as the only criterion for reducing the probability of misclassification. The purity of dataset D can be measured using Gini values, and the relevant calculation formula is as follows:

$$\text{Gini}(D) = 1 - \sum_{k=1}^{K} p_k^2 (22)$$

3.3 Performance testing

To further investigate the specific performance of the fiber optic sensor used in this article, we conducted strain experiments on a single fiber optic sensor. During the experiment, the adjustment of the spiral micrometer was limited each time. The fiber optic sensor fixed on the three-dimensional moving platform moved a distance of 0.5mm for each rotation, with a total of four movements and corresponding strains of 3.5m ε、 7m ε、 10.5m ε、 14m ε。 After each adjustment, it is necessary to wait for about 3 minutes before proceeding to the next operation to ensure effective storage of spectral data. The contrast of the transmission spectrum is relatively high, reaching 18dB. Using the valley within the wavelength range of 1530-1545nm as a reference point, it can be observed that the longer wavelength of the valley shifts with increasing strain. By linearly fitting the corresponding wavelengths under different strain conditions in the data, the fitting result graph shown in Figure 3 can be obtained. The fitting equation is y=0.243x+1537.58, R2=0.9992, which further demonstrates the good linearity of a single fiber optic sensor. By using

the fitting equation for calculation, the sensitivity value of a single fiber optic sensor is approximately 0.24, corresponding to a resolution value of up to 83.3.

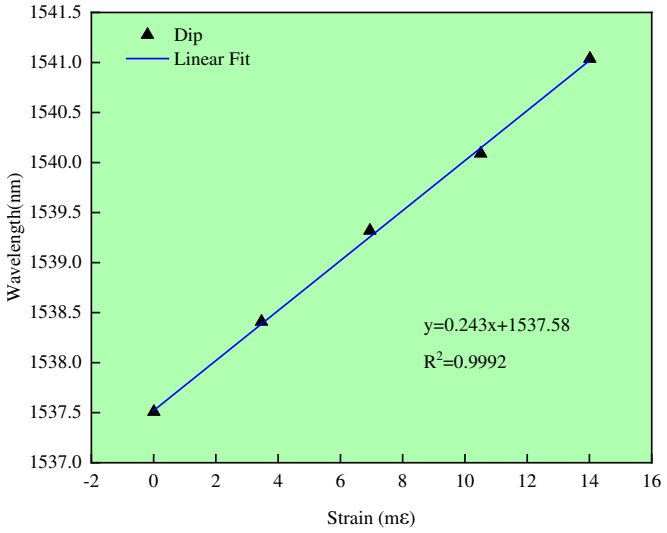


Figure 3 Linear fitting of strain and trough

In order to comprehensively test the performance and reliability of the random forest algorithm, support vector machine algorithm was selected for experimental comparison with this algorithm. Firstly, ensure that the parameter configurations of both are consistent, and that the selected data comes from the Blood dataset, with a random feature variable value of 3. In the end, a total of 5 combinations were obtained, each combination underwent 1000 experiments, and the OOB (Out of Bag) estimated mean and OOB standard deviation obtained from the experiments were compared. The detailed comparison results are shown in Table 1.

Table 1 Performance Comparison and Analysis of Two Algorithms

Combination		1	2	3	4	5
Mean of OOB estimation	random forest algorithm	0.211827	0.211830	0.211826	0.211816	0.211787
	Support Vector Machine Algorithm	0.221977	0.219743	0.218475	0.217407	0.21665
Standard value of OOB estimation	random forest algorithm	0.0048395	0.0044753	0.2215127	0.0038900	0.0037642
	Support Vector Machine Algorithm	0.002771	0.0018877	0.0016005	0.0013416	0.001259

The number of trees in the random forest algorithm directly affects the performance of the algorithm, but the degree of impact is slightly smaller compared to the support vector machine algorithm. The comparison of the changes in the OOB estimation mean of these two algorithms is shown in Figure 4.

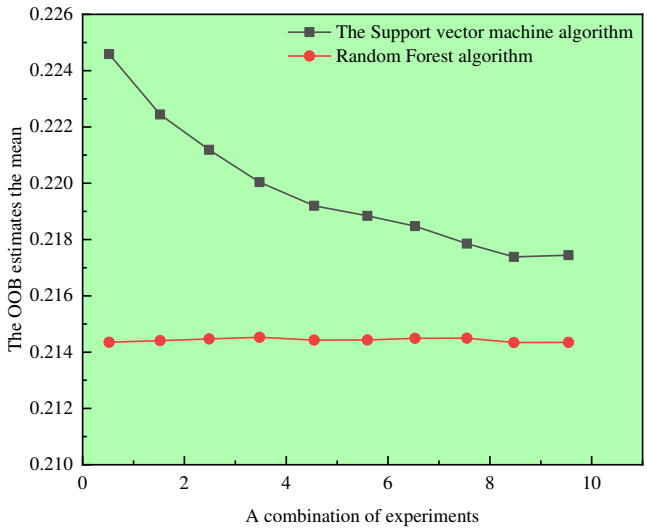


Figure 4 Comparison of mean values for OOB estimation using two algorithms

Analyzing the graphical results, it is found that the mean OOB estimation of the random forest algorithm is always lower than that of the support vector machine algorithm. Therefore, under the same conditions, the classification accuracy of the random forest algorithm is higher, which successfully proves the reliability of the random forest algorithm proposed in this paper.

4 Overall design of fatigue detection system for sports competitions

4.1 Overall System Design

Nowadays, many sports competitions incorporate technological means in their training, with the aim of assisting athletes in improving their training effectiveness and achieving better competition results. However, there are also numerous cases of physical injuries caused by excessive training. When athletes show signs of fatigue, they should immediately adjust the intensity of training to avoid unnecessary sports injuries. It can be seen that real-time monitoring of fatigue status during exercise is crucial. Based on monitoring data, reasonable adjustment of exercise intensity can be made to prevent potential safety hazards. In the field of sports science research, fatigue detection of athletes is also a major research hotspot. In competition training, it is necessary to continuously increase training intensity to improve training effectiveness and results, but if the speed of increase is too fast, it can lead to physical fatigue. If the body is already in a state of fatigue and has not received timely rest or reduced training intensity, sports injuries will occur, reducing training efficiency. So, timely fatigue testing not only protects the physical health of athletes, but also is the key to improving athletic performance.

Each sports event is composed of different intensity forms of exercise, which to some extent determine the characteristics of sports fatigue that athletes need to endure during the competition. Effectively understanding and mastering the characteristics of sports fatigue can help us understand the requirements of competitions for the physical fatigue level of athletes, thereby better guiding the arrangement and design of training plans and content during the training process,

ensuring that training is carried out on the basis of meeting scientific requirements, reducing or avoiding the randomness and blindness of training, and making training more specialized.

The upper computer interface of the system should have the function of displaying the fiber optic sensor signals, characteristic parameters, and athlete fatigue level of the three major channels. In summary, the overall architecture of the sports competition fatigue detection system designed in this article is proposed, as shown in Figure 5. Among them, it includes the acquisition modules of three fiber optic sensors and the communication, display, and signal analysis modules in the upper computer.

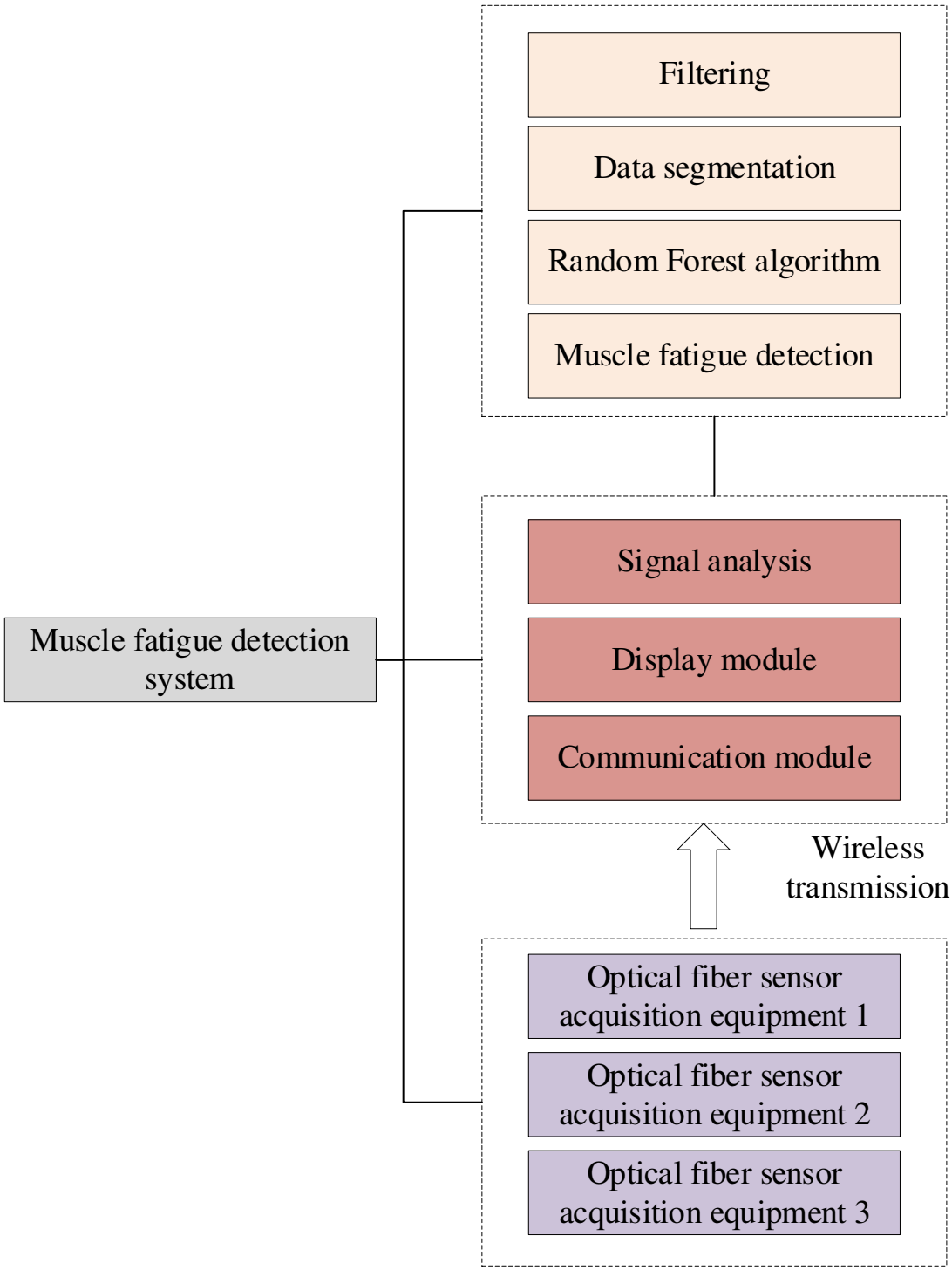


Figure 5 Overall System Block Diagram

The signal amplitude values corresponding to fiber optic sensors vary from a few microvolts to a few millivolts. Therefore, before using fiber optic sensors for signal acquisition, it is necessary to amplify the signal. Due to the influence of coupling interference signals on the human body, there will be 50Hz power frequency interference in the collected signals, which requires notch analysis. The signal acquisition circuit in ordinary fiber optic sensors mainly includes three major circuits, namely the front-end amplification circuit, low-pass filtering circuit, and high pass filtering circuit. However, due to the excessive number of components, the portability of the module may not be good. Therefore, in the design process, this article converts hardware into software and utilizes a microprocessor with a 12 bit built-in circuit. This operation can not only reduce the total number of components in the module, reduce the battery loss of the sensor, but also extend the continuous working time of the module.

4.2 Actual System Testing

In order to further verify the computational accuracy of the random forest algorithm used in this article, a comparative experiment was conducted on the accuracy of the test set of the algorithm and the support vector machine algorithm, and the results obtained were presented in the form of a table in Table 2. Through experiments, it was found that using the signals from the three directional axes of temperature and fiber optic sensors to fit linear calculations performed well in determining the fatigue level of the body. The test set accuracy of the fatigue detection model constructed using the random forest algorithm in this article is over 95%, and the effect is the best.

Table 2 Algorithm Accuracy

Supervised learning	Accuracy of Random Forest Testing	Support Vector Machine Testing Accuracy
temperature	73.64%	67.80%
X-axis	43.83%	28.63%
Y-axis	41.49%	31.55%
Z-axis	49.68%	39.74%
X-axis+Y-axis	50.85%	91.33%
X-axis+Z-axis	52.59%	45.58%
Y-axis+Z-axis	54.94%	50.85%
X-axis+Y-axis+Z-axis	57.28%	53.77%

5 Conclusion

Since the 21st century, people's attention to sports competitions has been increasing day by day. However, there are also many cases of physical injuries caused by excessive exercise during sports, especially among athletes. So it is necessary to adjust the exercise intensity in real time based on the body's exercise fatigue status, which can effectively reduce or avoid potential sports safety hazards. Physical fatigue is an extremely common phenomenon among athletes during sports training. In order to timely and effectively detect the physical fatigue of athletes, and to pause or reduce the overall intensity of training in a state of continuous fatigue, in order to avoid unnecessary sports injuries, this paper uses fiber optic sensors and random forest algorithm as the basis to conduct research on fatigue detection methods related to sports competitions. Meanwhile,

how to efficiently process the signals collected by fiber optic sensors is also crucial. This article uses the random forest algorithm to process data signals. This algorithm not only enhances the data's ability to resist overfitting, but also ensures the accuracy of detection on the basis of parallel processing, making the evaluation results more stable and effectively preventing noise interference. In summary, the sports competition fatigue detection system designed in this article using the random forest algorithm based on fiber optic sensors can effectively detect the physical fatigue level of athletes during the competition training process, thereby reducing or avoiding sports injuries.

Author contributions

Zhaoyu Chen has contributed to the paper’s analysis, discussion, writing, and revision.

Fund

The authors have not disclosed any funding.

Declarations

Data availability

The data will be available upon request.

Conflict of interest

The authors declare that they have no competing interests.

Ethical approval

Not applicable.

References

[1] Khalid, M. K., Agha, M. H., Shah, S. T. H., Akhtar, M. N.: Conceptualizing audit fatigue in the context of sustainable supply chains. Sustainability 12(21), 9135 (2020)

[2] McKee, K., Fee, D., Haney, M., Matoza, R. S., Lyons, J.: Infrasound signal detection and back azimuth estimation using ground-coupled airwaves on a seismo-acoustic sensor pair. Journal of Geophysical Research: Solid Earth 123(8), 6826-6844 (2018)

[3] Du, C., Dutta, S., Kurup, P., Yu, T., Wang, X.: A review of railway infrastructure monitoring using fiber optic sensors. Sensors and Actuators A: Physical 303, 111728 (2020)

[4] Sarica, A., Cerasa, A., Quattrone, A.: Random forest algorithm for the classification of neuroimaging data in Alzheimer's disease: a systematic review. Frontiers in aging neuroscience 9, 329 (2017)

[5] Bao, Y., Huang, Y., Hoehler, M. S., Chen, G.: Review of fiber optic sensors for structural fire engineering. Sensors 19(4), 877 (2019)

[6] Liu, X., Jin, B., Bai, Q., Wang, Y., Wang, D., Wang, Y.: Distributed fiber-optic sensors for vibration detection. Sensors 16(8), 1164 (2016)

[7] Zhu, L., Qiu, D., Ergu, D., Ying, C., Liu, K.: A study on predicting loan default based on the random forest algorithm. Procedia Computer Science 162, 503-513 (2019)

[8] Everingham, Y., Sexton, J., Skocaj, D., Inman-Bamber, G.: Accurate prediction of sugarcane yield using a random forest algorithm. Agronomy for sustainable development 36, 1-9 (2016)

[9] Thorpe, R. T., Strudwick, A. J., Buchheit, M., Atkinson, G., Drust, B., Gregson, W.: Monitoring fatigue during the in-season competitive phase in elite soccer players. International journal of sports physiology and performance 10(8), 958-964 (2015)

[10] Clarke, F., Jones, A., Smith, L.: Building peace through sports projects: a scoping review. Sustainability 13(4), 2129 (2021)

[11] Shan, M., Zhao, L., Liu, F., Qi, D., Zhang, J.: Revealing the competitive fatigue failure behaviour of CFRP-aluminum two-bolt, double-lap joints. Composite Structures 244, 112166 (2020)

[12] Zheng, Y., Zhu, Z. W., Xiao, W., Deng, Q. X.: Review of fiber optic sensors in geotechnical health monitoring. Optical Fiber Technology 54, 102127 (2020)

[13] Addanki, S., Amiri, I. S., Yupapin, P.: Review of optical fibers-introduction and applications in fiber lasers. Results in Physics 10, 743-750 (2018)

[14] Bulatov, M. I., Shatsov, A. A.: Strength and fracture resistance of quartz fibers with polyimide coatings. Russian Journal of Non-Ferrous Metals 62, 756-762 (2021)

[15] Simsekler, M. C. E., Qazi, A., Alalami, M. A., Ellahham, S., Ozonoff, A.: Evaluation of patient safety culture using a random forest algorithm. Reliability Engineering & System Safety 204, 107186 (2020)

[16] Advani, M. S., Saxe, A. M., Sompolinsky, H.: High-dimensional dynamics of generalization error in neural networks. Neural Networks 132, 428-446 (2020)