Research Article

Random Forest and LightGBM-Based Human Health Check for Medical Device Fault Detection

Weiwei Wang

Department of Equipment Pingdu People’s Hospital, Pingdu, Qingdao, Shandong 266700, China

Correspondence should be addressed to Weiwei Wang; qdpdwww@sina.com

Received 13 January 2022; Accepted 11 February 2022; Published 17 March 2022

Copyright © 2022 Weiwei Wang. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Medical devices are items used directly or indirectly in the human body and are a prerequisite for hospital treatment of patients, and their quality can have a direct impact on the health of patients, so strengthening the quality control of medical device use is a hot spot of concern in the clinic. Current medical device testing can reduce the occurrence of adverse events, but it cannot be completely avoided, and its work still needs to be further strengthened. In this paper, we design a two-way feature selection algorithm based on PSO_RF. We use random forest to calculate the importance of the feature attributes of the sample data and sort the results in descending order, where a particle swarm algorithm is introduced to optimize the parameters of the random forest algorithm. The 245 medical device adverse event reports received by the testing center were selected, the occurrence and types of adverse events were analyzed retrospectively, and quality control countermeasures for medical device use were formulated.

1. Introduction

With the rapid development of science and technology in the medical field, the medical equipment used is becoming more and more sophisticated and complex and the normal and stable operation of the equipment becomes more and more important in the process of using the instruments and equipment. The problem of accidental failure of the equipment will cause a great loss of scientific research results and reduce the efficiency of using the equipment [1].

Now, most universities, in the actual experiments, mostly use the way of regular testing and maintenance. Such testing cannot be based on the state of the equipment itself for maintenance and repair but according to people’s own experience, which means that the equipment’s potential failure points cannot be accurately maintained and, at the same time, will lead to a serious waste of funds and time [2–4]. With the development of complex and sophisticated instruments and equipment, the traditional manual detection method has become infeasible and the manual fault detection method has the problems of difficult fault detection, time consumption, inaccuracy, and high cost. In addition, especially for some precision instruments, it is more difficult for the management personnel of the equipment to accurately understand the operation of the equipment, so that the equipment is not timely maintained and repaired, which may have certain impact on the resulting experimental outcomes, affecting the final results of the experiment [5, 6]. At the same time, this also caused the low utilization rate of instruments and equipment, maintenance and repair costs, and other problems. Therefore, traditional testing methods have become difficult to adapt to the development of current instrumentation.

For the aforementioned problems, equipment failure detection technology was proposed. As early as the 1960s, the National Aeronautics and Space Administration of the United States set up a failure prediction team, specializing in data collection and calculation, etc., so as to make up for the shortcomings of the traditional manual mode to a certain extent [7]. Then, the United Kingdom and other countries also followed the research of equipment failure detection technology. With the continuous development of technologies, the stability of the equipment becomes more and more
high and the fault detection technology is constantly moving forward. Nowadays, the computing power of computers has been improved as never before, and the data from the operation of the equipment are constantly collected, so how to use the collected information to achieve effective management of the operation of equipment has become a popular issue [8–11].

Most of those who use medical devices are clinical healthcare workers, so those in control of their quality should actively communicate with clinicians to understand the potential problems that exist in their use and put forward certain adjustment suggestions [12]. At the same time, it is necessary to establish a perfect quality of medical equipment use rules and regulations, requiring that each contact with the use of medical equipment can follow the rules and regulations to reduce the adverse events of medical devices caused by human factors [13]. Routine and regular maintenance of equipment can prolong the service life of hospital equipment. For example, after clinicians use it, maintenance personnel needs to fill in maintenance records [14]. The contents of the records can dynamically grasp the use of medical equipment, analyze the adverse events that may occur, and exclude failures in the dangerous period, which in turn can ensure the normal operation of the instruments and equipment [15]. At the same time, the equipment should be regularly dusted and cleaned, its performance should be tested, the vulnerable parts should be replaced in a timely manner, and all records should be made. And a direct responsible person for medical devices should be established, so that adverse events that occur during maintenance or use can be traced directly by the individual, which in turn improves the requirements of each person in contact with medical equipment for themselves [16].

In summary, analysis of medical device adverse events and strengthening control over the quality of their use can reduce the occurrence of adverse events.

2. Related Work

Fault detection methods can be broadly classified into three categories: model-based, knowledge-based, and data-driven. Among them, model-based fault detection methods can provide a deeper understanding of the system nature and more real-time fault detection. In [17], a generalized graceless Kalman filter algorithm was used to detect and separate faults in the phase current and rotor position sensors of a three-phase permanent magnet synchronous motor, which handled the nonlinear data well [18]. By using the fault identification model of a knowledge vector machine, a hybrid reasoning model of knowledge reasoning and information fusion can be obtained. The authors of [19] used least squares support vector machine for fault detection and classification of regulating valves. Firstly, cleaning the experimental data and then using LS-SVM multiplicative for classification experiments on regulating valve samples achieved better results.

The machine learning method is a method that gives a computer a human way of thinking so that it can have the ability to process complex data, by training the algorithm model, using test data to verify the accuracy of the algorithm, and finally using the trained model to make effective decisions. In [20], the original data were subjected to feature extraction by PCA algorithm to obtain the feature vector of the original data, and finally, a multiclassification algorithm combining binary tree and SVM was used to achieve fault detection during vibration sensor operation, whose experimental results showed that the improved method not only increased the accuracy of fault detection but also accelerated the classification speed. In [21], a multimodal SVM learning method was proposed and applied to the fault detection problem of gearboxes, and the effectiveness of the method was verified in experiments on gearboxes with two structures: straight gearboxes and helical gearboxes.

3. Fault Detection Based on Random Forest and LightGBM

3.1. Analysis of Equipment Failure Detection Problems.

LightGBM is chosen for its higher efficiency and accuracy, lower memory usage, and support for parallelized learning, so it is chosen to build the equipment fault detection model and to test the effectiveness of the PSO_RF-based bidirectional feature selection method based on the accuracy and precision of the confusion matrix [22]. The data after feature selection are then input to the LightGBM algorithm for learning, and a grid search method is used to optimize the parameter search process to produce the final classification results. In this paper, an equipment fault detection model is developed through the following steps:

Step 1: data preprocessing, including the deletion of sample data with problems such as missing and duplicate data, data transformation, and other operations

Step 2: feature selection using PSO_RF’s bidirectional feature selection method

Step 3: Initializing the parameters of LightGBM, inputting the processed data into the model for training, and performing parameter optimization using the grid method

Step 4: inputting the test data into the model and then evaluating and analyzing the final output results

The flow chart of the model is shown in Figure 1.

3.2. Data Acquisition and Normalized Preprocessing.

To detect equipment failure problems, the first thing to do is to conduct a comprehensive study of the equipment’s failure information and the factors that affect it. The first step is to conduct a comprehensive study of the equipment fault information and its influencing factors. However, there are various problems in the data that do not meet the requirements as input variables for the model [23]. The data are preprocessed according to the following four steps.

Data collection: the real data used in the model come from the laboratory’s independent project “Large-Scale instrument sharing platform,” which mainly solves the problem of low
utilization rate of instruments in universities and is currently serving many universities. The data set contains 1200 pieces of X-ray camera usage information, including 955 pieces of faulty data and 245 pieces of normal data. There is more information in the design, and although there are many factors to consider [24], there are some uncontrollable and other reasons that cannot be considered, such as improper human operation and other factors that are not included in the study. In summary, this data set can be used for the training of the equipment fault detection model.

Based on the analysis of the intuitive factors affecting the equipment, a total of 16 attribute values, E, F, G, H, I, J, K, L, M, N, O, P, Q, R, S, and T, are extracted as influencing factors [25].

The data in the sample have problems such as missing and duplicate data. Due to the relatively small amount of data, this paper fills in the values of the missing attributes by calculating the mean values of the attributes, and for the duplicate data, they are deleted because there are few duplicates. The same method of calculating the mean value of attributes is used to fill in the case of handling abnormal values.

Although the features are reduced using intuitive factors such as knowledge and experience in the previous section, there is still redundancy among the features. Redundant features can have a great impact on the model results and training efficiency during the model building process, so it is necessary to analyze the data and remove the useless information in this step. If an attribute value has a large range of values, then it will have a certain impact on the training of the model, so the transformation of the data is needed. In the experiments for E, F, G, H, I, J, K, L, M, N, O, P, Q, R, S, and T, data were normalized using the Max-Min (maximum-minimum) approach to normalize their attribute values. The attribute values are normalized to the interval from 0 to 1. For the fault types, Err1, Err2, Err3, Err4, and Err5 are used for power failure, dome failure, heat dissipation failure, imaging failure, and cable failure, respectively. Normal is N.

After the above four steps, the problems of data duplication, missing values, and inconsistency in the original data are solved. This improves the quality of the data to a large extent, which is important for the subsequent performance improvement of the model [26].

3.3. Bidirectional Feature Selection Algorithm for Random Forests. The main idea is to add noise to a relevant feature and then judge the importance of a feature based on the change of the result before and after adding noise to the feature. The main idea is to add noise to a relevant feature and then judge the importance of a feature based on the change of the result before and after adding noise [27]. The procedure for calculating the importance of the random forest attribute X is as follows:

Step 1: for each tree, use its corresponding out-of-bag (OOB) data.
Step 2: among all samples of out-of-bag (OOB) data, randomly select the feature attribute X, add some noise to it, then calculate the out-of-bag error once again, and record it as errOOB.
Step 3: if the random forest is composed of N trees, then the importance of the feature X is

$$\text{importance} = \frac{\sum (\text{errOOB}_2 - \text{errOOB}_1)}{N}$$  \hspace{1cm} (1)

If the accuracy of attribute X is significantly lower after adding noise than before, then it can be said that feature X has a great influence on the learning effect of the model.

The random forest algorithm has many parameters, but there is no fixed method of parameter selection for different sample data. In order to solve this problem, this paper uses particle swarm optimization to optimize the parameter search process of the random forest algorithm, so that the random forest can find the optimal combination of parameters more quickly and efficiently and the algorithm can further improve the performance of the model; the process of the algorithm is as follows:

Step 1: initialize the parameters of the random forest and particle swarm based on experience
Step 2: Generate a decision tree by randomly selecting k samples from the sample data according to the bootstrap algorithm
Step 3: Compute the output of the model
Step 4: The above classification results are used as the fitness values, and the particle swarm algorithm is used to continuously iterate, perform parameter optimization, and compare with historical results to finally output the optimal model parameters
Step 5: Based on the obtained model parameters, the random forest is trained and the importance score of feature attributes are finally derived
In this paper, a two-way feature selection algorithm based on PSO_RF is used to calculate the importance of the feature attributes of the sample data using random forest and the results are sorted in descending order. Then, the search starts from the full set of sample features, and each time, the features with the lowest degree of importance are removed from the current subset of features to form a new subset; finally, the part of backward selection is performed, and the accuracy of the current subset of features is calculated using LightGBM based on the confusion matrix; if the accuracy of the final result decreases after removing the features, the features just removed are recycled and so on until the end of the cycle. In this way, we can reduce the volatility of feature attributes and ensure that the selected subset has less redundancy and does not lose classification accuracy by adding the prediction results of the current subset as a factor to evaluate the feature subset on top of the feature importance [28].

4. Case Study

4.1. General Information. 245 cases of medical device adverse event reports received by the testing center were selected for retrospective analysis of the occurrence and types of adverse events and implementation of quality control countermeasures for medical device use. These included five major categories of medical devices: nonwoven surgical gowns, mercury thermometers, monitors, single-use sterile syringes, and OCU intrauterine devices, and all personnel involved in testing, maintenance, and quality control were in the same group.

245 cases of medical device adverse events were analyzed, including nonwoven surgical gowns, mercury thermometers, monitors, single-use sterile syringes, and OCU IUDs, and the total number of medical device adverse events from September 2018 to August 2019 and from September 2019 to August 2020 was counted [29].

4.2. Results. Among the 245 medical device adverse events, nonwoven surgical gowns accounted for 22.1%, mercury thermometers accounted for 35.5%, monitors accounted for 14.7%, single-use sterile syringes accounted for 16.7%, and OCU IUDs accounted for 10.4%, as shown in Table 1. The number of cases of medical device adverse events after the implementation of quality control in September 2019 ~ August 2020 was significantly less than the number of cases that occurred in September 2018 ~ August 2020. The number of cases occurring without implementation from September 2019 ~ August 2019 (P < 0.05) is shown in Table 2.

5. Fault Detection Model Training and Evaluation

In this paper, a particle swarm algorithm is used to optimize the bidirectional feature selection based on random forest and LightGBM for device fault detection. In this paper, we use the bidirectional feature selection algorithm based on the particle swarm optimization random forest described in this paper for feature selection. After collecting and preprocessing the original data in the above section, the processed data are input to the model as the input variables, 70% of the data set is used as training data for model training, and the remaining data are used as test data for model testing. In the experiments, CFS (correlation-based feature selection) is used as a comparison experiment, the search strategy of CFS for feature subsets is best-first search, the respective feature subsets are selected by both the feature selection algorithm and CFS, and then the results are predicted using LightGBM. Table 3 shows the results of the feature selection process using the two algorithms to select subsets separately and then using LightGBM for training, where X_num indicates the number of features.

As shown in Figure 2, the curves in the figure specifically represent the line graph of the variation of the accuracy of the fault detection model constructed from the data processed by the two feature selection algorithms with the number of features.

As can be seen from Table 3, the classification accuracy of the fault detection model built by the particle swarm optimized random forest bidirectional feature selection method selected in this paper is 89.73% and the F1 value is 90.26%, which is higher than that of the CFS feature selection method in terms of accuracy and F1 value, and the time cost of the feature selection algorithm used in this paper is smaller in terms of model time overhead. Finally, 12 features that have an impact on the device are selected as the input of the final model, including 12 features such as temperature, tube voltage, current, half-valence layer, and output repeatability.

The optimal feature subset was selected through the above experiments, then the fault detection model was constructed using the LightGBM algorithm and trained and tested in groups using the ten-fold cross-validation method, LightGBM, GBDT, and random forest were used for training and testing, and the average value was taken as the final result.

Figure 3 shows the accuracy of random forest, GBDT, and LightGBM algorithms using cross-validation each time, where the horizontal coordinates indicate the serial number of cross-validations and the vertical coordinates indicate the classification accuracy.

As shown in Table 4 and Figure 3, LightGBM achieves a good level of accuracy and F1 value in prediction results, which are better than the other two models, and the training efficiency of the model is better than the other two models. In summary, the equipment fault detection method used in this paper is more accurate and reliable in terms of results through feature selection and fault model building and it effectively improves the computational efficiency of the fault detection model and improves the performance of the model.
Table 1: 245 medical device adverse event occurrence comparison.

| Group                        | Number of cases | %   |
|------------------------------|-----------------|-----|
| Nonwoven surgical garment    | 55              | 22.1|
| Mercury thermometer          | 87              | 35.5|
| Monitor                      | 36              | 14.7|
| Disposable sterile syringe   | 41              | 16.7|
| OCU IUD                      | 26              | 10.4|

Table 2: Comparison of the number of cases of adverse events after the implementation of quality control (n).

| Group            | Number of cases |
|------------------|-----------------|
| 2018.9–2019.8    | 245             |
| 2019.9–2020.8    | 184             |
| $X^2$            | 4.894           |
| $P$              | $<0.05$         |

Table 3: Analysis of the operation results of different feature selection algorithms.

|               | $X_{num}$ | Accuracy rate | F1 value |
|---------------|-----------|---------------|----------|
| CFS           | 13        | 0.8625        | 0.8589   |
| Bidirectional feature selection based on PSO RF | 12 | 0.8972 | 0.9025 |

Figure 2: Effect of the number of features of the two algorithms on the results.
6. Conclusions

This paper presents the complete process of building a fault detection model for equipment, firstly preprocessing the data, then selecting the optimal feature subset by a two-way feature selection method based on PSO_RF, finally using the selected feature subset as input to build a fault detection model, and verifying that the fault detection model used in this paper is realistic and effective by comparing and analyzing with other algorithms.

Data Availability

The data underlying the results presented in the study are included within the manuscript.

Conflicts of Interest

The author declares no conflicts of interest.

Authors’ Contributions

The author has seen the manuscript and approved for submission.

References

[1] D. D. Rufo, T. G. Debelee, A. Ibenthal, and W. G. Negera, “Diagnosis of diabetes mellitus using gradient boosting machine (LightGBM),” *Diagnostics*, vol. 11, no. 9, p. 1714, 2021.

[2] L. Kopitar, P. Kocbek, L. Cilar, A. Sheikh, and G. Stiglic, “Early detection of type 2 diabetes mellitus using machine learning-based prediction models,” *Scientific reports*, vol. 10, no. 1, pp. 11981–12012, 2020.

[3] S. H. Kassania, P. H. Kassanib, M. J. Wesolowskic, K. A. Schneidera, and R. Detersa, "Automatic detection of coronavirus disease (COVID-19) in X-ray and CT images: a machine learning based approach," *Biocybernetics and Biomedical Engineering*, vol. 41, no. 3, pp. 867–879, 2021.

[4] C. H. Cao, Y. N. Tang, D. Y. Huang, and W. Gan, "IBE: An improved identity-based encryption Algorithm for WSN security," *Security and Communication Networks*, pp. 2021–8, Article ID 8527068, 2021.

[5] D. Wu, C. Zhang, L. Ji, R. Ran, H. Wu, and Y. Xu, "Forest fire recognition based on feature extraction from multi-view images," *Traitement du Signal*, vol. 38, no. 3, pp. 775–783, 2021.

[6] L. Wang, C. Zhang, Q. Chen, S. Liu, Z. Yang, and H. Li, "A communication strategy of proactive nodes based on loop theorem in wireless sensor networks," in *Proceedings of the 2018 9th International Conference on Intelligent Control and
7

Information Processing (ICICIP), pp. 160–167, IEEE, Wanzhou, China, November 2018.

[7] H. Li, D. Zeng, L. Chen, Q. Chen, M. Wang, and C. Zhang, “Immune multipath reliable transmission with fault tolerance in wireless sensor networks,” in Proceedings of the International Conference on Bio-Inspired Computing: Theories and Applications, pp. 513–517, Springer, Singapore, June 2016.

[8] C. H. Cao, Y. N. Tang, D. Y. Huang, G. WeiMin, and Z. ChunJiong, “IIBE: an improved identity-based encryption algorithm for wsn security,” Security and Communication Networks, vol. 2021, Article ID 8527068, 8 pages, 2021.

[9] X. I. E. Tao, C. Zhang, and Y. Xu, “Collaborative parameter update based on average variance reduction of historical gradients,” Journal of Electronics and Information Technology, vol. 43, no. 4, pp. 956–964, 2021.

[10] D. R. Wallace and D. R. Kuhn, “Lessons from 342 medical device failures,” in Proceedings of the 4th IEEE International Symposium on High-Assurance Systems Engineering, pp. 123–131, IEEE, Washington, DC, USA, November 1999.

[11] J. Zhou, A. Guo, B. Celler, and S. Su, “Fault detection and identification spanning multiple processes by integrating PCA with neural network,” Applied Soft Computing, vol. 14, pp. 4–11, 2014.

[12] J. Lin, M. J. Zuo, and K. R. Fyfe, “Mechanical fault detection based on the wavelet de-noising technique,” Journal of Vibration and Acoustics, vol. 126, no. 1, pp. 9–16, 2004.

[13] I. Oliver, “An approach to combining medical device fault analysis with trusted computing forensics,” in Proceedings of the 2020 IEEE International Conference on Big Data (Big Data), pp. 1831–1837, IEEE, Atlanta, GA, USA, December 2020.

[14] B. W. Bequette, “Fault detection and safety in closed-loop artificial pancreas systems,” Journal of diabetes science and technology, vol. 8, no. 6, pp. 1204–1214, 2014.

[15] Y. Yang, Q. Liu, Z. Gao, X. Qiu, and L. Meng, “Data fault detection in medical sensor networks,” Sensors, vol. 15, no. 3, pp. 6066–6090, 2015.

[16] L. Meneghetti, M. Terzi, S. Del Favero, G. A. Susto, and C. Cobelli, “Data-driven anomaly recognition for unsupervised model-free fault detection in artificial pancreas,” IEEE Transactions on Control Systems Technology, vol. 28, no. 1, pp. 33–47, 2018.

[17] P. Ahir, M. Mozaffari-Kermani, and R. Azarderakhsh, “Lightweight architectures for reliable and fault detection Simon and Speck cryptographic algorithms on FPGA,” ACM Transactions on Embedded Computing Systems, vol. 16, no. 4, pp. 1–17, 2017.

[18] H. Zhang, J. Liu, and R. Li, “Fault detection for medical body sensor networks under bayesian network model,” in Proceedings of the 2015 11th International Conference on Mobile Ad-hoc and Sensor Networks (MSN), pp. 37–42, IEEE, Shenzhen, China, December 2015.

[19] Y. Jia and T. A. Reddy, “Characteristic physical parameter approach to modeling chillers suitable for fault detection, diagnosis, and evaluation,” Journal of Solar Energy Engineering, vol. 125, no. 3, pp. 258–265, 2003.

[20] H. Zhang, J. Liu, and N. Kato, “Threshold tuning-based wearable sensor fault detection for reliable medical monitoring using Bayesian network model,” IEEE Systems Journal, vol. 12, no. 2, pp. 1886–1896, 2018.

[21] J. E. Braun, “Automated fault detection and diagnostics for vapor compression cooling equipment,” Journal of Solar Energy Engineering, vol. 125, no. 3, pp. 266–274, 2003.