Frontiers in Computing and Intelligent Systems
ISSN: 2832-6024 | Vol. 11, No. 3, 2025

Research on Medical Device Measurement and Inspection System Based on Big Data Analysis

Yanqing Yu, Qiuyan Wang

Jixiang Fire (Hangzhou) Calibration and Testing Co., Ltd. Hangzhou, Zhejiang, 310012, China

Abstract: The accuracy and stability of medical equipment have a crucial impact on the diagnostic results and therapeutic effects of patients, so the measurement and testing technology of medical equipment has become particularly important. This thesis is based on big data analysis of medical equipment measurement and testing technology, through the comprehensive use of data mining, machine learning and other big data analysis technology, in-depth analysis of medical equipment measurement and testing data, to achieve accurate assessment of the performance of medical equipment, fault prediction and timely maintenance, so as to improve the efficiency and safety of the use of medical equipment, and to provide a strong support for the development of the medical industry. This thesis provides new ideas and methods for the development of medical equipment measurement and testing technology, which helps to promote the scientific and intelligent management of medical equipment.

Keywords: Big Data Analytics; Medical Devices; Metrology and Testing; Quality Control.

1. Introduction

In today's digital era, the amount of data generated in the medical field is exploding. Medical equipment, as an important support for medical services, is increasing in variety and number, and its functions are becoming increasingly complex. Traditional medical equipment measurement and testing methods have been difficult to meet the needs of modern healthcare for equipment accuracy, reliability and efficient management. The emergence of big data analysis technology has brought new opportunities for medical equipment measurement and testing [1]. Through the collection, storage, processing and analysis of a large amount of medical equipment data, it can achieve more accurate detection, predictive maintenance, improve the use of medical equipment safety and reliability, reduce the impact of equipment failure on patient care, maintain the continuity and high quality of medical services, provide a scientific basis for decision-making for healthcare institutions, and drive the process of digitalization and intelligent transformation of the medical industry [2].

This thesis conducts in-depth research on medical equipment measurement and testing technology based on big data analysis, and builds equipment performance assessment model and fault prediction mechanism by integrating medical equipment measurement and testing data with big data analysis technology to realize the fine management and quality control of medical equipment. Using big data analysis technology to mine the potential information in the medical equipment measurement and testing data, revealing the pattern of change in equipment performance and failure modes, developing a failure prediction model for medical equipment based on big data, improving the accuracy and timeliness of the failure early warning, reducing the failure rate of the equipment, and combining the results of the analysis of big data to provide medical institutions with scientific and reasonable recommendations on equipment maintenance strategies and procurement decisions, and improving the medical equipment management level.

2. Research on Measurement and Detection Systems

2.1. System Architecture Design

The architecture of this thesis metrology testing system is mainly divided into data acquisition layer, data transmission layer, data processing layer and application layer. The data acquisition layer collects real-time operating data, metrology data, etc. of medical devices through sensors, interfaces, and other means. The data transmission layer uses wired or wireless communication technologies, such as 5G, Wi-Fi, etc., to transmit the collected data to the data center safely and quickly. For the data processing layer, this study adopts the Hadoop ecosystem and Spark computing engine to build a data processing platform, which realizes multi-source heterogeneous data purification, distributed persistent storage, multi-dimensional feature analysis and knowledge graph construction [3], so as to effectively explore the potential data value. The application layer provides users with a visualization interface, which is convenient for managers to view the status of equipment, generate inspection reports, predict equipment failures, etc., in order to realize the intelligent management of medical equipment metrology and inspection, improve the accuracy and safety of medical equipment, and safeguard the quality of medical services.

2.2. Data Acquisition and Preprocessing

2.2.1. Sources and Modes of Data Collection

The source of data acquisition covers all kinds of medical equipment, including CT, MRI, ultrasound and other diagnostic equipment, radiotherapy, surgical navigation systems and other therapeutic equipment, electrocardiographic monitors, blood pressure monitors and other monitoring equipment, and biochemical analyzers, blood analyzers and other testing equipment. Acquisition methods include data interface, sensor monitoring, network communication protocols, etc. The data interface can be used to directly acquire the operating parameters, working status, fault codes and other information of the equipment. Sensor monitoring can be used to collect the temperature, pressure, vibration and other physical quantities of the equipment, and the network communication protocol ensures the stable transmission of data, realizing remote data collection and centralized management.

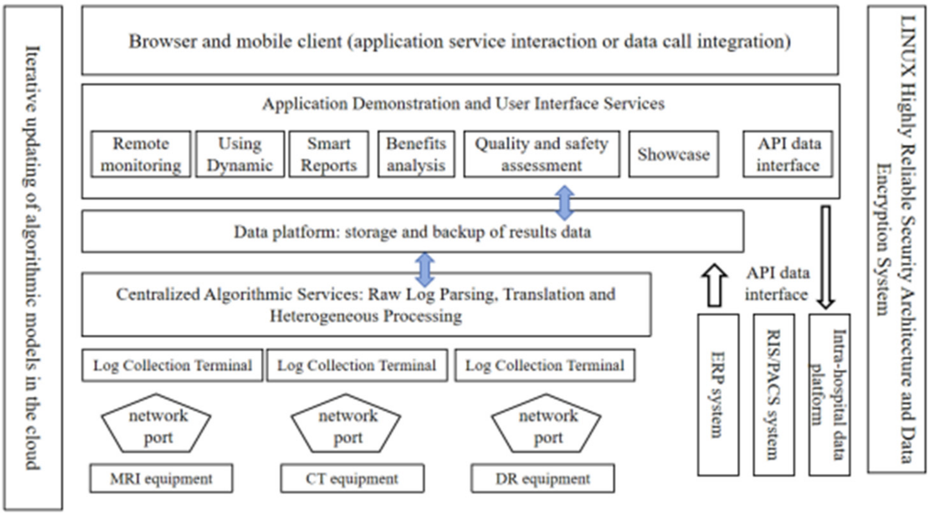

Figure 1. Logical relationship diagram of real-time monitoring data acquisition, transmission and interface for medical devices

2.2.2. Data Preprocessing Methods and Processes

The collected data may have problems such as noise, missing values and outliers, which require preprocessing. Preprocessing methods include data cleaning, data integration, data transformation and data normalization. Data cleaning mainly removes noise and outliers, using filtering algorithms, statistical methods and threshold setting techniques [4]; data integration through the standardization and integration of heterogeneous medical device data, to achieve format conversion and coding unification of multi-source information, to ensure the consistency and completeness of the dataset, to provide a reliable basis for the subsequent analysis; in this paper, the data are standardized and normalized to achieve data transformation to make them adapt to the requirements of the data analysis model; data generalization is achieved through the standardization of data, and data transformations are achieved through the data analysis model; data generalization is achieved through the data analysis model. This paper standardizes and normalizes the data to achieve data transformation, so that it can adapt to the requirements of the data analysis model; data normalization reduces the amount of data and improves the efficiency of data analysis through methods such as feature selection and data sampling, while retaining the key information of the data.

2.3. Big Data Analytics Modeling

2.3.1. Machine Learning Algorithms

According to the needs and data characteristics of medical equipment measurement and testing, decision trees, support vector machines, neural networks and other machine learning algorithms for equipment fault classification and prediction are selected [5]. The clustering algorithm is used for equipment performance evaluation and classification management, and the regression analysis algorithm is used for the prediction and calibration of equipment parameters. The accuracy, interpretability, computational efficiency and adaptability of the algorithms are fully considered in the selection process, and the optimal algorithms are determined through experimental comparison and performance evaluation.

2.3.2. Model Training and Optimization

The selected machine learning model is trained using historical metrological inspection data and equipment operation data, and the dataset is divided into training, validation and test sets, and methods such as cross-validation are used to optimize the model parameters and improve the model's generalization ability. During the training process, regularization techniques, hyperparameter adjustment and integrated learning are applied to prevent overfitting and enhance the stability and accuracy of the model [6]. Through continuous iterative training and optimization, the model is able to accurately predict equipment failures, evaluate performance and provide reasonable maintenance recommendations.

3. Hadoop-based Big Data Storage Optimization

3.1. HDFS-based Data Storage Distribution Strategy

As the distributed storage core component of the Hadoop ecosystem, HDFS is designed with a distributed architecture, which can effectively support the storage needs of medical data of petabyte scale and above, and provide efficient data access services while ensuring data security and reliability.

HDFS adopts a master-slave cluster architecture, which consists of two types of nodes, NameNode (master node) and DataNode (slave node). Among them, the master node is responsible for maintaining the metadata information of the file system, including the location mapping of data blocks; the deputy node performs the actual data storage tasks. To ensure the high availability and data reliability of the system, HDFS adopts a multi-copy storage strategy with a default configuration of three copies, which can be dynamically distributed to any sub-node in the cluster [7]. Assuming a cluster size of N nodes, when executing a computation job containing M Mapij tasks, the network transmission overhead incurred by the system can be quantitatively analyzed by the following notation, considering the random distribution characteristics of the data blocks.

$$F=\sum_{i=1}^{M}\sum_{k=1}^{d_i}\left[\frac{C_{d_i}^{k}B^{k}}{N^{d^i}(N-B+1)^{d_i-k}}(d_i-k)\ s\right]$$

B represents the number of data copies, s represents the data block capacity, and di represents the number of data blocks that need to be processed by each branch task at the node. When the required data is not stored in the local computing node, additional network transmission overhead will be incurred. According to the characteristics of cluster architecture, data access can be divided into two typical scenarios: one is that the target data block and the execution task are located in the same rack, and the probability of this event is p1, and the corresponding data transmission delay is q1; the other is that the target data block and the execution task belong to different racks, and the probability of this scenario is p2, and the overhead of its network communication is q2. Based on the above analysis, the total time consumed for the whole task is:

$$T=\frac{F}{p_1q_1+p_2q_2}=\frac{[\frac{C_{d_i}^{k}B^{k}}{N^{d^i}(N-B+1)^{d_i-k}}(d_i-k)\ s]}{p_1q_1+p_2q_2}$$

In HDFS-based distributed storage systems, data access efficiency is mainly affected by the storage topology. By aggregating data sets with strong correlation and storing them in the same compute node, the cross-node data transfer overhead can be significantly reduced, thus enhancing the execution efficiency of data-intensive applications. This data locality optimization strategy can realize the co-location of computing tasks and storage resources, so that both query processing and data computation operations can be completed locally, effectively avoiding network transmission bottlenecks. In this paper, an optimization algorithm based on association-aware hash-bucket storage is proposed, which provides a new idea for optimizing the data layout of distributed storage systems.

3.2. Data Storage Optimization Based on Hash Bucket Algorithm

The distributed data storage system based on HDFS architecture adopts a master-slave structure to achieve centralized management of data resources. In this system, metadata is uniformly maintained and managed by the master node, while the actual data is distributed and stored on the slave nodes, and this physical separation mechanism of metadata and entity data effectively improves the data management efficiency. To ensure high availability and fault tolerance, the system adopts a three-copy redundancy strategy to prevent data loss or damage from affecting cluster operation. For the internal data of medical organizations, the device ID is used as the key field for data association, and for external medical data, the area code is used as the primary key for data association. This differentiated association field design takes into full consideration the diversity and complexity of medical data, which is conducive to the accurate management and efficient use of data, and meets the needs of the medical industry in terms of data storage, querying and analysis and other diversified applications.

Hash operations are performed on medical datasets to generate unique identifiers, and copies of data subjects with the same hash value are divided into different storage units. In the HDFS cluster architecture, physical storage aggregation of associated data is achieved by establishing a hierarchical directory structure that maps data sets with associated relationships under specific namespaces.

4. System Testing and Verification

4.1. Experimental Design

4.1.1. Test Data Set Selection

The actual operation data of several different types of medical equipment in a hospital in Zhejiang Province are selected as the test data set, including CT machines, MRI equipment, cardiac monitors, ventilators, etc. The data covers the normal operation status of the equipment, the different degree of failure status, and the data before and after the maintenance of the equipment. The data covers the normal operation status of the equipment, different degrees of fault status, and the data before and after equipment maintenance, to ensure the diversity and representativeness of the test data set, and to be able to comprehensively validate the performance and functionality of the system.

4.1.2. Setting of Test Indicators

Accuracy, recall, F1 value, etc. are set as the performance evaluation indexes of the fault diagnosis and prediction model to measure the accuracy of the system in detecting and predicting equipment faults. Set average absolute error, root mean square error, etc. as the evaluation indexes of equipment performance evaluation and prediction model to assess the accuracy of the system in predicting equipment parameters. Setting the system response time, warning timeliness, etc. as the system operation performance indicators to assess the real-time and effectiveness of the system in practical applications.

4.2. Experimental Results and Analysis

4.2.1. Model Performance Evaluation

The experimental results show that the fault diagnosis and prediction model based on big data analytics achieves high accuracy and recall on the test dataset. The diagnostic accuracy for common CT machine faults reaches more than 90%, and the recall rate for MRI equipment fault prediction can reach more than 85%, indicating that the model can effectively identify equipment faults and issue timely warnings. The prediction errors of the equipment performance evaluation model for key parameters of the equipment are within the acceptable range, and the average absolute error of the heart rate parameter prediction of the ECG monitor is less than 0.5 beats/minute, which meets the requirements for clinical use.

4.2.2. System Functional Verification

The system's equipment status monitoring and early warning function can accurately display the equipment operating status in real time, and the warning time for abnormal situations is more than 30 minutes before the failure on average, providing sufficient preparation time for maintenance personnel. The maintenance management module realizes efficient processing of maintenance work orders, and the average processing time of maintenance work orders has been shortened by about 20%, which improves the efficiency of equipment maintenance. The quality control and assessment function effectively identifies abnormal values in the testing data, ensuring the accuracy and reliability of the metrological testing results.

5. Conclusion

This thesis applies big data analysis technology to the field of medical equipment metrology and inspection, and constructs a complete inspection system to realize real-time monitoring, fault early warning and predictive maintenance of equipment operation status. By optimizing the process of data acquisition, storage and analysis, it improves the detection efficiency and accuracy, and provides a strong guarantee for the safe operation of medical equipment. The introduction of real-time data processing technology realizes immediate monitoring and feedback of equipment status and enhances the system's responsiveness. It helps medical institutions to improve the level of equipment management, reduce operating costs and improve the quality of medical services.

References

[1] Yi Yi-ye. Application of metrological testing in quality control of medical devices[J]. Branding and Standardization, 2024 (3):192-194.

[2] Wang Jun. Discussion on the problems and optimization countermeasures of medical equipment measurement and testing[J]. Popular standardization,2023(13):183-185. DOI:10. 3969/j. issn.1007-1350.2023.13.064.

[3] Li Ligong. Design and realization of hospital medical equipment information management system [D]. Xi'an University of Electronic Science and Technology.

[4] HUO Hongbo, HE Boshi,WU Bin,et al. Statistical analysis of medical device performance based on data warehouse[J]. China Medical Device,2013,000(009):27-30.DOI:10. 3969/j. issn. 1674-1633.2013.09.009.

[5] CHEN Xing, XIA Ke, ZHANG Yang. Full-cycle management of medical equipment in the context of information integration[J]. Information and Computer,2019(9): 2.DOI: CNKI: SUN: XXDL.0.2019-09-078.

[6] Guo Mingquan. Analysis of medical equipment informatization management in the era of Internet+[J]. China Health Industry, 2015,12(13):2. DOI:CNKI: SUN:WSCY. 0. 2015-13-050.

[7] Xing, Hongbo. Research on medical data storage based on Hadoop [D]. Shenyang University of Technology,2022.