

## *Retraction*

# **Retracted: The Role of Medical Big Data and Artificial Intelligence in the Diagnosis of Knee Osteoarthritis with Cardiovascular and Cerebrovascular Diseases**

### **Wireless Communications and Mobile Computing**

Received 26 September 2023; Accepted 26 September 2023; Published 27 September 2023

Copyright © 2023 Wireless Communications and Mobile Computing. 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.

This article has been retracted by Hindawi following an investigation undertaken by the publisher [1]. This investigation has uncovered evidence of one or more of the following indicators of systematic manipulation of the publication process:

- (1) Discrepancies in scope
- (2) Discrepancies in the description of the research reported
- (3) Discrepancies between the availability of data and the research described
- (4) Inappropriate citations
- (5) Incoherent, meaningless and/or irrelevant content included in the article
- (6) Peer-review manipulation

The presence of these indicators undermines our confidence in the integrity of the article's content and we cannot, therefore, vouch for its reliability. Please note that this notice is intended solely to alert readers that the content of this article is unreliable. We have not investigated whether authors were aware of or involved in the systematic manipulation of the publication process.

In addition, our investigation has also shown that one or more of the following human-subject reporting requirements has not been met in this article: ethical approval by an Institutional Review Board (IRB) committee or equivalent, patient/participant consent to participate, and/or agreement to publish patient/participant details (where relevant).

Wiley and Hindawi regrets that the usual quality checks did not identify these issues before publication and have since put additional measures in place to safeguard research integrity.

We wish to credit our own Research Integrity and Research Publishing teams and anonymous and named external

researchers and research integrity experts for contributing to this investigation.

The corresponding author, as the representative of all authors, has been given the opportunity to register their agreement or disagreement to this retraction. We have kept a record of any response received.

### **References**

- [1] X. Huan, P. Wan, and Y. Zhou, "The Role of Medical Big Data and Artificial Intelligence in the Diagnosis of Knee Osteoarthritis with Cardiovascular and Cerebrovascular Diseases," *Wireless Communications and Mobile Computing*, vol. 2022, Article ID 1839280, 12 pages, 2022.

## Research Article

# The Role of Medical Big Data and Artificial Intelligence in the Diagnosis of Knee Osteoarthritis with Cardiovascular and Cerebrovascular Diseases

Xin Huan,<sup>1</sup> Pan Wan,<sup>1</sup> and Yonghuan Zhou <sup>2</sup>

<sup>1</sup>Wuhan No. 1 Hospital, Wuhan, 430000 Hubei, China

<sup>2</sup>People's Hospital of Dongxihu District, Wuhan, 430040 Hubei, China

Correspondence should be addressed to Yonghuan Zhou; [161847112@masu.edu.cn](mailto:161847112@masu.edu.cn)

Received 15 February 2022; Revised 9 May 2022; Accepted 20 May 2022; Published 9 June 2022

Academic Editor: Rashid A Saeed

Copyright © 2022 Xin Huan et al. 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.

In this Internet era, people use more and more network data, and hospital diagnosis is even more important. The improvement of medical standards has brought about the rapid growth of medical data. How to analyze related symptoms from a large amount of data performing an analysis and accurately diagnosing its symptoms has become one of the difficult problems. This article is aimed at studying the role of medical huge amount of data and artificial intelligence in the diagnosis of knee osteoarthritis combined with cardiovascular and cerebrovascular diseases. To this end, this article proposes to improve data mining methods and proposes a comprehensive analysis of patient data in the field of artificial intelligence for intelligent diagnosis and treatment so that doctors can make accurate judgments about the symptoms of patients and designed related experiments to explore its specific effects. The experimental results in this article show that the improved data mining speed has increased by 17%, the data integrity has increased by 31%, and the proportion of valid data has also increased by 23%, which is very effective for clinical diagnosis.

## 1. Introduction

With the rapid development of cloud computing, huge amount of data, and other technologies, the process of informatization construction of the medical industry in various countries in the world is accelerating. Therefore, the hospital information system based on the electronic medical record subsystem and the image archive communication subsystem has been widely used as never before. However, the traditional hospital information system with a single-node data warehouse as the main body is limited by hardware capabilities, and it is easy to encounter performance bottlenecks when processing large amounts of unstructured data, and it is difficult to achieve two-way expansion of storage and computing capabilities.

Due to the continuous advancement of modern medical auxiliary detection methods, a large number of two-dimensional data such as medical images, images, and sounds have emerged. The nature of audio-visual data is unstructured data, which is difficult to express in the form of traditional database. Therefore, when storing records, a new value needs to be created to describe its structure, which increases the difficulty of data storage and processing. Although most hospitals currently use PACS to store various types of digital medical unstructured data, which partially solves the storage problem of multidimensional data, they still face cross-analysis of multiple data structures and efficient extraction of practical information. Significantly weak and huge amount of data processing technology breaks the concept of “complete records” in traditional databases and

divides all kinds of data in the form of files and stores and manages them in a distributed system. At the same time, a data association is created for each data block, and then, a specific mapping/merging algorithm is used to read and write operations and integrated analysis of the distributed file set, which greatly improves the real-time processing capability of the huge heterogeneous data set.

With the rapid development of cloud computing and distributed processing technology, research projects and application engineering based on huge amount of data have received extensive attention from scholars from all over the world. Godinho et al. proposed that the amount of medical imaging data produced has increased substantially in the past few decades. So, they proposed a method to create a simulated medical imaging database, which is based on the index of model data sets, the extraction of patterns, and the modeling of research and production [1]. Hu et al. localized the median of Down syndrome (DS) screening markers and compared the impact of localization and built-in median data on the efficiency of DS screening. He examined data from early and midpregnancy screening (FTS and STS) from DS in a retrospective study, which consisted of selected pregnancies that were considered normal. The localized regression analysis is calculated by using five models to perform statistical fitting on the original data [2]. Inomata et al. proposed that through medical huge amount of data analysis, mining information from multiomics research, and mobile health applications, it may provide solutions for the management of chronic diseases [3]. WN Price and Cohen proposed that the rapid development of machine learning technology and artificial intelligence is expected to revolutionize medical practice, from resource allocation to diagnosis of complex diseases. Among other topics, they also discussed how to best consider health privacy [4]. Chen uses a hybrid methodology to investigate Taiwan's healthcare industry, aiming to evaluate, predict, and summarize the main applications of medical huge amount of data and to establish a strategic path followed by medical institutions for applications in different dimensions [5]. In order to improve the intelligence of the medical system, Zhang and Wang designed and implemented a secure medical huge amount of data ecosystem on the Hadoop huge amount of data platform. It is designed in the context of the increasingly severe trend of the current safe medical huge amount of data ecosystem [6]. Ilyasova et al. introduced the main research results of data mining in the medical field. The application of huge amount of data technology in the proposed medical diagnosis system can improve the quality of the learning set and reduce classification errors [7]. Xiao-Qing et al. proposed that artificial intelligence research based on computer technology is one of the effective ways to solve this problem. In the research of intelligent diagnosis and treatment path, the characteristics of "four diagnosis," "comprehensive examination," "combination of disease and syndrome," "treatment according to person and cause," and other TCM holistic views and syndrome differentiation and treatment are embodied by artificial intelligence in TCM diagnosis or the key to successful research in clinical practice [8]. The above-mentioned literatures are very

detailed on the technical points of the relevant content, and the depth of understanding of some technologies is also in place, which has a good reference for the research of this article. But basically all the literature does not have a good test for the performance of the algorithm involved, lack of experimental demonstration.

The innovation of this article lies in the use of data mining in the medical huge amount of data technology, combined with the intelligent diagnosis and treatment system in artificial intelligence, to conduct a comprehensive data analysis of the patient's illness. Through a large amount of data analysis, it can be concluded that the patient is more comprehensive; the data is applied to the intelligent diagnosis and treatment system to diagnose the disease.

## 2. Intelligent Medical Diagnosis Method

*2.1. Multidimensional Mining of Medical Huge Amount of Data.* Its sources mainly include four categories: one is pharmaceutical companies/life sciences; the second is clinical decision support and other clinical applications [9].

For example, cerebral ischemia and cerebral infarction brain tissue is very sensitive to ischemia and hypoxia. It is particularly important to obtain cerebral hemodynamic parameters during the early diagnosis and treatment of ischemic stroke. The development of rapid magnetic resonance technology makes it possible. The display of the ischemic penumbra and the measurement of its perfusion parameters can provide a direct basis for arterial thrombolysis and detect the recovery of damaged tissue [10].

Aiming at the source of these data, huge amount of data analysts can create application value in different scenarios. The first is individualized treatment. The same medicine is not the same disease for everyone. Therefore, individual differences must be considered and the most appropriate treatment method should be adopted. The other is the monitoring of medical insurance fraud. For the research methods of social medical insurance fraud, a variety of methods can be tried to combine, such as based on game theory, incorporating measurement models and statistical models, and so on.

The main point of this article is medical huge amount of data. There are five major application systems for modern hospital informatization construction. Hospital information system (HIS) refers to the use of electronic computers and communication equipment to provide various departments of the hospital with the ability to collect, store, process, extract, and exchange data for patient diagnosis and treatment information and administrative management information and to meet all authorizations and user functional requirements. With the rapid development of medical technology and information technology, the hospital information system is also constantly developing and improving. It is foreseeable that the scope and level of application of the hospital information system in the hospital will have an increasingly important impact on the reform and modernization of the hospital [11]. The structure of the hospital's information system is shown in Figure 1. There were a total

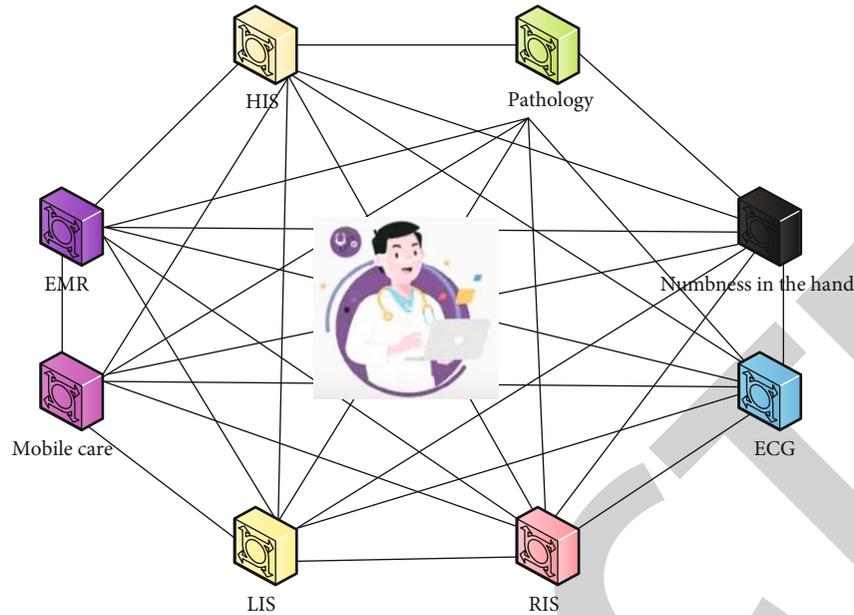


FIGURE 1: The current situation of the hospital information system.

of 563 patients with varying physical conditions, all in between qualifying conditions.

Specifically, the huge amount of data generated by these medical systems has the following four characteristics, namely,

- (1) Large data volume (volume): for example, a hospital in Shanghai receives tens of thousands of patients every day, and the application system will generate text data of tens of thousands of medical reports and hundreds of thousands of inspection reports. If it is an image report such as filming, the generated image data will be larger in scale [12]
- (2) Data structure diversity (variety): a hospital in Shanghai has 44 large and small systems with different patient information and medical equipment information that generate 60 TB of data each year
- (3) Rapid data growth (velocity): data from related research shows that by 2020, medical data will increase sharply to 35 zettabytes, including patient information, medical equipment information, and system information, which is equivalent to 44 times the amount of data in 2009
- (4) The value of data is huge (value): undoubtedly, data is oil, resources, and assets; new drug research and development and the ability to overcome chronic diseases [13]

In the traditional data model, databases are often used as data storage, and use SQL queries to read and write data. General databases (Oracle, SQLSERVER, DB2, etc.) do not have data mining functions; huge amount of data engineers need to do data mining, and the client obtains data from the database. There may be various problems such as informa-

tion omission, duplication, and inappropriate processing. Secondly, open-source data mining algorithms are mixed. The more well-known mining algorithms are weak, which has insufficient scalability and weak processing capabilities for huge amount of data. Secondly, there is Spark's ML algorithm, but Spark has higher requirements for machines, and large-scale clusters need to be built to achieve good computing results. The traditional data mining architecture is shown in Figure 2.

As the name implies, data mining [14] is the process of exploring and analyzing large amounts of data to discover meaningful patterns and rules. Data mining is a technique to find its regularity from a large amount of data by analyzing each data. It mainly includes three steps: data preparation, regularity search, and regularity representation. Data preparation is to select the required data from relevant data sources and integrate them into a data set for data mining; regular search is to find out the regularity contained in the data set by a certain method; regular expression is to be as user-readable as possible. The way of comprehension expresses the found law. The tasks of data mining include association analysis, cluster analysis, classification analysis, anomaly analysis, specific group analysis, and evolution analysis [15]. It has been widely used in many organizations. Especially in the medical field, data mining is becoming more and more important. It is a method of analyzing a large amount of data stored in an enterprise through mathematical models to find out different customers or market segments and analyze consumer preferences and behaviors.

For data mining, data analysis is an unnecessary prerequisite. But the complete data analysis process is helpful for the later data mining. The system composition framework [16] is mainly composed of four parts, as shown in Figure 3.

In the figure, data source 1 and data source 2, respectively, represent the thyroid information table and the diabetes score table, and then, these two data sources are imported

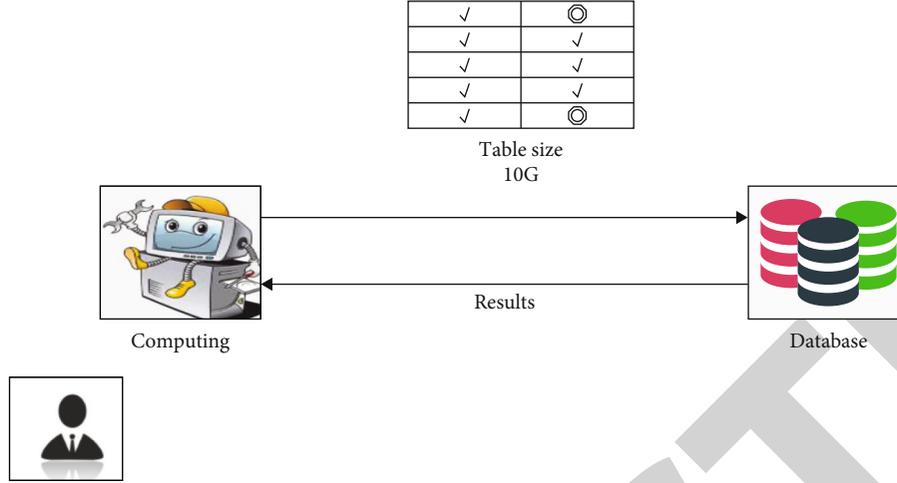


FIGURE 2: Traditional data mining.

into the database through the ETL tool, and then, the Hana modeling tool generates a specific view and finally provided by the BO visualization tools for corresponding queries and calculations. The functional structure of the system mainly includes four modules, and the four modules are, respectively, introduced below. The system function structure diagram is shown in Figure 4.

The system is mainly composed of four modules, which are data preprocessing module, data integration module [17, 18], correlation algorithm implementation module, and external query module.

**2.1.1. Data Preprocessing Module.** This module prepares data for the entire data mining, including the deletion of wrong data, the completion of repairable data, and the correction of inconsistent data.

**2.1.2. Data Modeling Module.** This module integrates the effective and reasonable data after cleaning and generates a new data table file with a specific organization.

**2.1.3. Association Algorithm Implementation Module.** This algorithm is the core of the entire analysis system. It traverses and queries the integrated data table, accumulates the number of occurrences of candidate item sets, and stores them in the intermediate data table.

**2.1.4. Provide Query Module Externally.** This module is designed for the display of mining results.

**2.2. Data Mining Algorithm.** Data mining is a cross-discipline; it needs to synthesize different research in multiple fields. This optimization is mainly aimed at its association analysis, discovering frequent itemsets from the data set and extracting the association rules between each data. Data mining is the nontrivial process of obtaining valid, novel, potentially useful, and ultimately understandable patterns from large amounts of data. Others view data mining as a fundamental step in the process of knowledge discovery in databases. The knowledge discovery process consists of the following steps: data cleaning, data integration, data

selection, data transformation, data mining, pattern evaluation, and knowledge representation. Data mining can interact with users or knowledge bases.

**2.2.1. Association Rule Mining.** For a given transaction set  $D \in I$ ,  $X$  is a subset of  $I$ . If  $X \in D$  is satisfied, then  $D$  is said to contain  $X$ . If the number of items in set  $X$  is  $k$ , then set  $X$  is called a set of  $k$  items. Definition:

$$I = \{i_1, i_2, \dots, i_n\}. \quad (1)$$

Definition. If the item set  $X \in I$ ,  $Y \in I$ , and  $X \cap Y = \Phi$ , then the expression of the form  $(X \rightarrow Y)$  is called an association rule. If the transaction set containing itemset  $X$  has a probability of  $c\%$  and also contains itemset  $Y$ , then the confidence level of  $(X \rightarrow Y)$  is called  $c\%$ ; if there are  $s\%$  transactions in the association rule set  $D$  that contain itemset  $X \cap Y$ , the support degree of the association rule  $(X \rightarrow Y)$  is called  $s\%$ , and the specific expressions of the confidence degree and the support degree are as follows:

$$\text{Supp}(X \Rightarrow Y) = P(X \cup Y), \quad (2)$$

$$\text{Conf}(X \Rightarrow Y) = P(X|Y). \quad (3)$$

When this association rule is a strong association rule, the specific conditions are as follows:

$$\text{Supp}(X \Rightarrow Y) \geq \text{min\_sup}, \quad (4)$$

$$\text{Conf}(X \Rightarrow Y) \geq \text{min\_conf}. \quad (5)$$

Among them, the minimum support threshold [19] and the minimum confidence threshold [20] can be defined according to specific applications. When the mined association rules meet these two conditions at the same time, the association rule is valid; otherwise, the association rule is invalid.

Since the first step needs to scan the entire transaction set, its time complexity is very large, so the first step is to find that frequent itemsets are usually the bottleneck of the

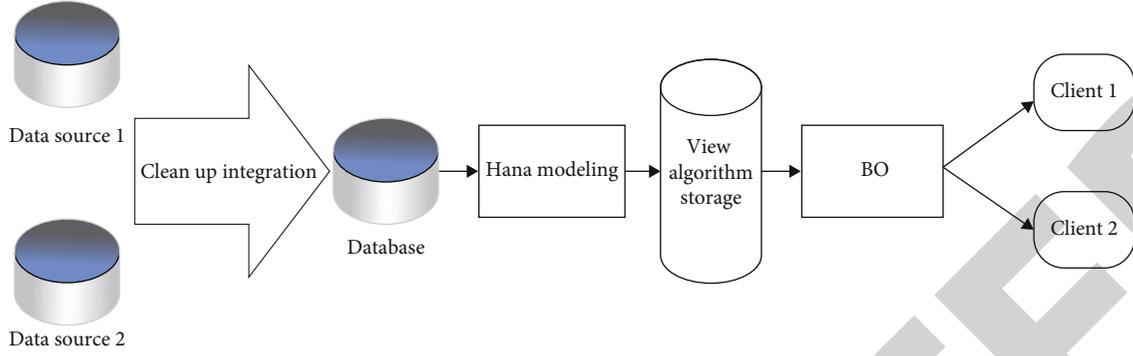


FIGURE 3: System composition framework.

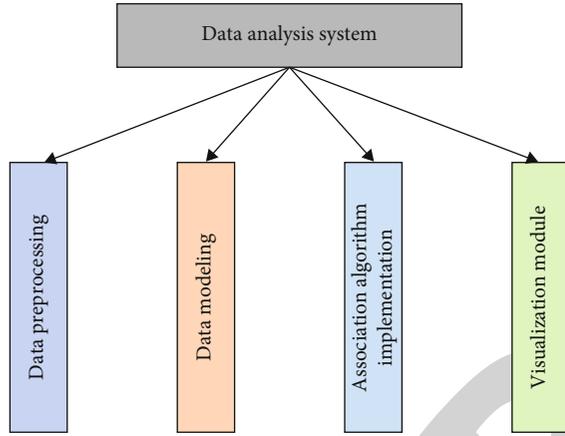


FIGURE 4: The functional structure diagram of the data analysis system.

performance of the entire algorithm. The Apriori algorithm finds frequent items in two steps, namely, the connection step and the pruning step. This section refers to the description of the connection steps.

**2.2.2. Connection Step.** If the items in transaction set  $D$  and all itemsets are sorted in alphabetical order, then any item in the  $(k-1)$  itemset meets the following conditions:

$$I_i[1] < I_i[2] < \dots < I_i[k]. \quad (6)$$

If it meets:

$$(I_m[1] = I_n[1]) \&\& (I_m[2] = I_n[2]), \quad (7)$$

$$(I_m[k-1] = I_n[k-1]) \&\& (I_m[k] < I_n[k]). \quad (8)$$

It is considered that  $I_m$  and  $I_n$  are connectable. The result of connecting  $I_m$  and  $I_n$  is:

$$\{I_m[1], I_m[2], \dots, I_m[k-1], I_n[k-1]\}. \quad (9)$$

**2.2.3. Improved Apriori Algorithm.** According to the many shortcomings of the Apriori algorithm [21] proposed in the previous section, combined with the characteristics of the medical huge amount of data mining environment, this paper optimizes the Apriori algorithm by introducing inter-

est to improve its performance. The following is an analysis and comparison of the existing interest degree models.

The probability-based interest model is as follows:

$$\text{Int}(X \Rightarrow Y) = \frac{1 - P(Y)}{(1 - P(X))(1 - P(Y|X))}. \quad (10)$$

The difference-based interest model is as follows:

$$\text{Int}(X \Rightarrow Y) = \frac{\text{Conf}(X \Rightarrow Y) - \text{Supp}(Y)}{\max(\text{Conf}(X \Rightarrow Y), \text{Supp}(Y))}, \quad (11)$$

in:

$$\max(\text{Conf}(X \Rightarrow Y), \text{Supp}(Y)). \quad (12)$$

as an evaluation standard.

The interest degree model based on correlation is as follows:

$$\text{Int}(X \Rightarrow Y) = \frac{\text{Supp}(X \cup Y)}{\text{Supp}(X)\text{Supp}(Y)}, \quad (13)$$

in:

$$\text{Supp}(X \cup Y) = \frac{\text{Count}(X \cup Y)}{N}, \quad (14)$$

$$\text{Supp}(X) = \frac{\text{Count}(X)}{N}, \quad (15)$$

$$\text{Supp}(Y) = \frac{\text{Count}(Y)}{N}. \quad (16)$$

The interest degree model based on the amount of information defines the interest degree of the association rule  $(X \rightarrow Y)$  as:

$$\text{Int}(X \Rightarrow Y) = P(X) \left[ P(Y|X) \log \left( \frac{P(Y|X)}{P(Y)} \right) + (1 - P(Y|X)) \log \left( \frac{1 - P(Y)}{1 - P(Y)} \right) \right], \quad (17)$$

in:

$$P(X) = \frac{\text{Count}(X)}{N}, \quad (18)$$

$$P(Y) = \frac{\text{Count}(Y)}{N}, \quad (19)$$

$$P(Y|X) = \frac{\text{Count}(X \cup Y)}{N}. \quad (20)$$

The influence-based interest model is as follows:

$$\begin{aligned} \text{Int}(X \Rightarrow Y) &= \log \frac{\text{Conf}(X \Rightarrow Y)/\text{Conf}(X \Rightarrow \bar{Y})}{\text{Supp}(Y)/\text{Supp}(Y)} \\ &= \log \left( \frac{N - \text{Count}(X)}{\text{Count}(X) - \text{Count}(X \cup Y)} \times \frac{\text{Count}(X \cup Y)}{\text{Count}(Y)} \right). \end{aligned} \quad (21)$$

The influence-based interest degree model determines the interest degree of association rules through the influence of the previous item  $X$  on the association rules ( $X \rightarrow Y$ ), which can effectively measure the strong and weak association rules close to the threshold. Therefore, this association rule is a rule that users are not interested in and needs to be deleted. This effectively removes those wrong strong association rules, so the optimization is meaningful.

**2.3. Intelligent Auxiliary Diagnosis and Treatment Model.** This research proposes an intelligent diagnosis and treatment model based on multigraph neural network [22] to realize computer intelligent recommendation from symptom data to clinical medication.

In the process of diagnosis and treatment, there are often some summarized relationships among symptoms, syndrome types, and status elements. Reasonably mining this relationship information will help improve the performance of the auxiliary diagnosis and treatment model. It is difficult for traditional algorithms to mine these multidimensional relationships, but this study builds a data feature aggregation module based on graph neural network, which can solve this problem well.

The prediction module is the module with the “learning” ability in the model, which is constructed based on the MLP algorithm. In the feature aggregation module, the output of each sample is a new feature representation that fuses the information of symptoms, syndrome types, and state elements. In the prediction module, it can be used as the input of MLP, and the final output is a set of probability values, corresponding to the probability that the drug is recommended. Through the embedding of the above two modules, the simulation process of intelligent diagnosis and treatment can be realized [23].

In this study, the performance of the MGNN intelligent diagnosis and treatment model was evaluated by three evaluation indicators commonly used in recommendation algorithms. The three evaluation indicators are accuracy, precision, and recall. All indicators are positively correlated, the larger the value, the better the experimental result.

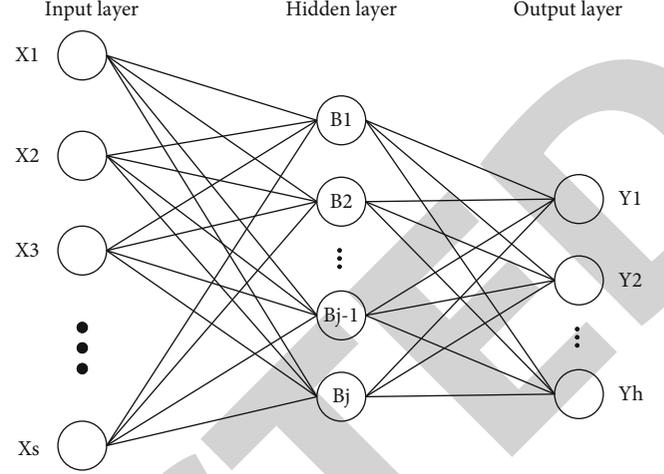


FIGURE 5: Multilayer perceptron model diagram.

Among them, the accuracy of the model represents the proportion of correct predictions among the drugs predicted by the model, which can intuitively indicate the performance of the model. The recall rate of the model represents the proportion of drugs in the actual label that are correctly predicted by the model and can indicate the coverage of all correct labels by the correct predicted label. The  $F1$  score is a weighted and harmonic average of the precision rate and the recall rate. As the third evaluation criterion in the model test, it can alleviate the objective contradictions in mathematical calculations to a certain extent.

**2.3.1. Multilayer Perceptron Algorithm.** Multilayer perceptron is a feed-forward artificial neural network model, which can learn from the structure and function of neurons in the human brain, and abstract an information processing model that transmits and responds to information characteristics. In practical applications, MLP can fit the mapping relationship from data characteristics to a certain category from data samples by cooperating with learning algorithms, so it is widely used in classification tasks. The model diagram of the multilayer perceptron is shown in Figure 5.

In order to enable MLP to update the network parameters and have the learning ability and to train the model, the BP algorithm calculates the error between the model output, and the expected output after the network is propagated forward and then back-propagates to update the network parameters by calculating the gradient.

**2.3.2. MGNN Intelligent Diagnosis and Treatment Model Construction.** The MGNN model proposed in this research is a computer multilabel classification model, which is mainly constructed based on the computer graph convolutional neural network algorithm and the principle of multilayer perceptron technology. It consists of two modules: medical data feature aggregation module and medical prediction module. The composition and model diagram are shown in Figure 6.

In recent years, the auxiliary diagnosis and treatment model constructed by knowledge representation and more

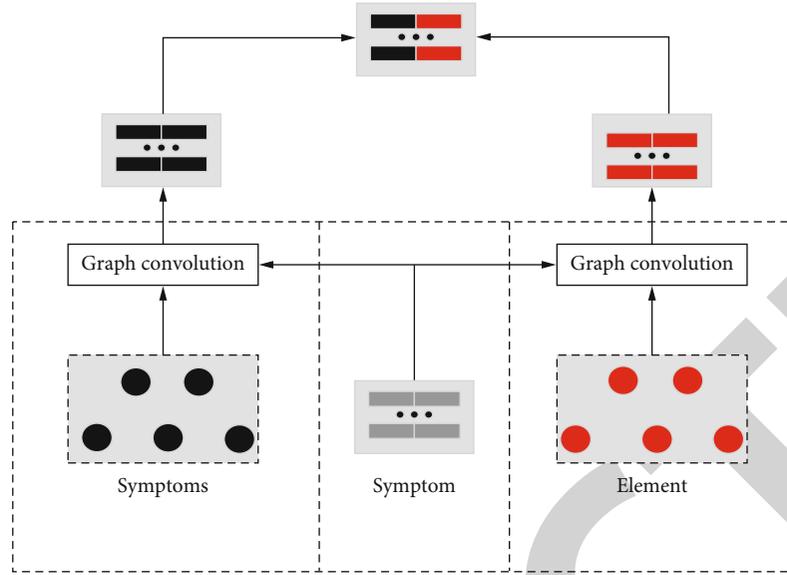


FIGURE 6: MGNN model diagram.

advanced artificial intelligence technology has made great progress, but it has not yet fully met the needs of clinical applications. The reason is that clinical applications require certain accuracy and timeliness to ensure data integrity. The research and development of the new generation of auxiliary diagnosis and treatment system is to solve the defects in the previous two generations of diagnosis and treatment models as much as possible and make full use of the four factors to comprehensively empower, so as to better solve the unsafe information, the uncertain environment, and the dynamic change environment and inferior problems and then realize the real sense of intelligent auxiliary diagnosis and treatment. Information insecurity refers to issues such as privacy leakage.

Reasonable mining of these relationship information will help improve the performance of the auxiliary diagnosis and treatment model. Traditional algorithms are difficult to mine such multidimensional relationships, but this research builds a data feature aggregation module based on graph neural network, which can better solve this problem.

### 3. Huge Amount of Data Application Experiment in General Hospitals

#### 3.1. Investigation Experiment

**3.1.1. Survey Object.** Using a combination of cluster sampling and simple random sampling, the representatives of 64 member units of the Medical Branch of the Chinese Research Hospital Association were selected for investigation.

**3.1.2. Investigation Method.** On the basis of literature study, I designed a survey scale of factors affecting the application of huge amount of data in hospitals. After two rounds of expert consultation, the survey scale was further revised and improved.

TABLE 1: Reliability analysis of factors affecting the application of huge amount of data in hospitals.

Scale	Dimension	Number of items	Cronbach alpha value
Influencing factors	Data base	3	0.805
	Management method	4	0.686
	Application promotion	4	0.809

**3.1.3. Reliability Analysis.** This study uses internal consistency reliability for analysis. The final measurement results are shown in Table 1.

**3.1.4. Validity Analysis.** In addition, this study also combines the variance interpretation rate, factor loading coefficient, and other indicators for analysis. The analysis results are shown in Table 2.

**3.1.5. Analysis of Influencing Factors.** Regarding the three primary variables involved in the influencing factors that limit the popularization of huge amount of data applications in hospitals, the survey subjects all show a high degree of acceptance. The specific situation is shown in Table 3.

**3.1.6. Correlation Analysis.** Analyze the correlation between the popularity of huge amount of data applications in hospitals and the three variables of data foundation, management methods, and application promotion. However, there is no significant correlation between the popularity of huge amount of data applications in hospitals, data foundation, and application promotion. The specific situation is shown in Table 4.

**3.1.7. Regression Analysis.** On the basis of correlation analysis, the three primary variables of data foundation, management

TABLE 2: Validity analysis of factors affecting the application of huge amount of data in hospitals.

Factor	Project	Factor loading factor		
		1	2	3
Dimension	Large volume	0.028	0.837	0.154
	Many types	0.152	0.917	-0.016
	Low standard	0.114	0.731	0.161
Number of items	Low degree of openness	0.351	0.224	0.432
	Insufficient information	0.333	-0.031	0.749
	Lack of professionals	0.527	-0.007	0.489
	There is a security risk	0.028	0.271	0.781
	Lack of management	0.816	0.002	0.182
Cronbach alpha value	Lack of platform	0.689	0.036	0.316
	Insufficient analytical techniques	0.798	0.223	0.123
	Weak foundation	0.755	0.146	0.068

TABLE 3: Overall analysis of factors affecting the application of huge amount of data in hospitals.

Influencing factors	Sample	Minimum	Maximum value	Average value	Standard deviation
Dimension	103	1.01	5.03	3.87	0.89
Number of items	104	1.74	5.03	4.08	0.61
Cronbach alpha value	105	2.51	5.06	4015	0.69

TABLE 4: Correlation analysis between influencing factors and popularity of huge amount of data application in hospitals.

	Popularity	Data base	Management method	Application promotion
Popularity	1			
Dimension	.040	1		
Number of items	-.227*	.294**	1	
Cronbach alpha value	-.075	.248*	.574**	1

\* $p < 0.05$ ; \*\* $p < 0.01$ .

TABLE 5: Regression analysis of influencing factors and popularity of huge amount of data application in hospitals.

Dependent variable	Independent variable	Nonstandard coefficient		Standard coefficient Beta	$t$	$p$	$R^2$	Adjust $R^2$	$F$
		$B$	Standard error						
Popularity	Constant	3.297	0.601	—	5.497	$\leq 0.001$			
	Data base	0.115	0.097	0.125	1.523	0.251	0.06	0.038	2.23*
	Management method	-0.382*	0.165	-0.304*	-2.368	0.018			
	Application promotion	0.075	0.145	0.067	0.501	0.626			

\* $p < 0.05$ ; \*\* $p < 0.01$ .

method, and application promotion are used as independent variables, and the popularity of hospital huge amount of data application is used as the dependent variable for multiple linear regression analysis. The specific situation is shown in Table 5.

Although there have been many successful application cases of huge amount of data in retail, finance, and other fields, it is also being gradually researched and explored in

the medical industry. Before the situation of huge amount of data leading the reform of the medical industry has not yet formed, some reported medical huge amount of data applications. The impression formed by medical staff is still in the laboratory stage, and there are uncertainties in accessibility and reliability. Therefore, it did not bring actual changes to the daily management and medical business of the hospital. Therefore, the importance of medical huge

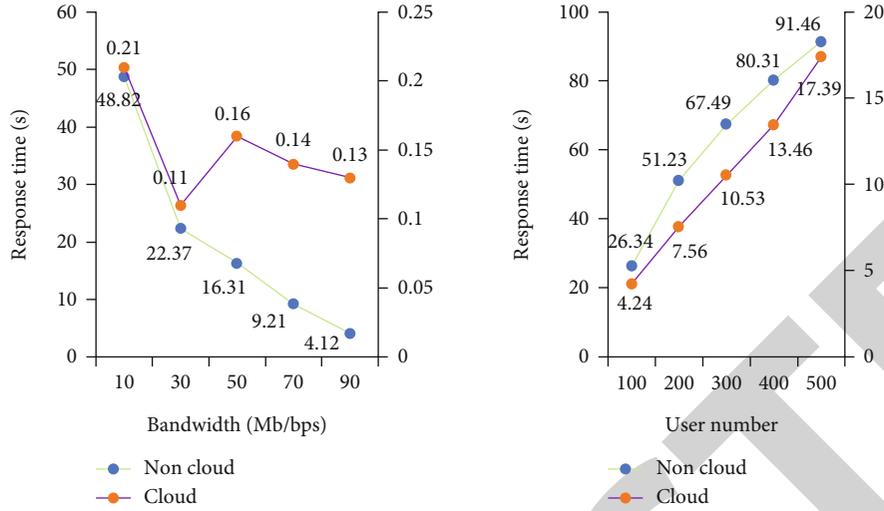


FIGURE 7: Comparison of average response time between cloud and noncloud environments.

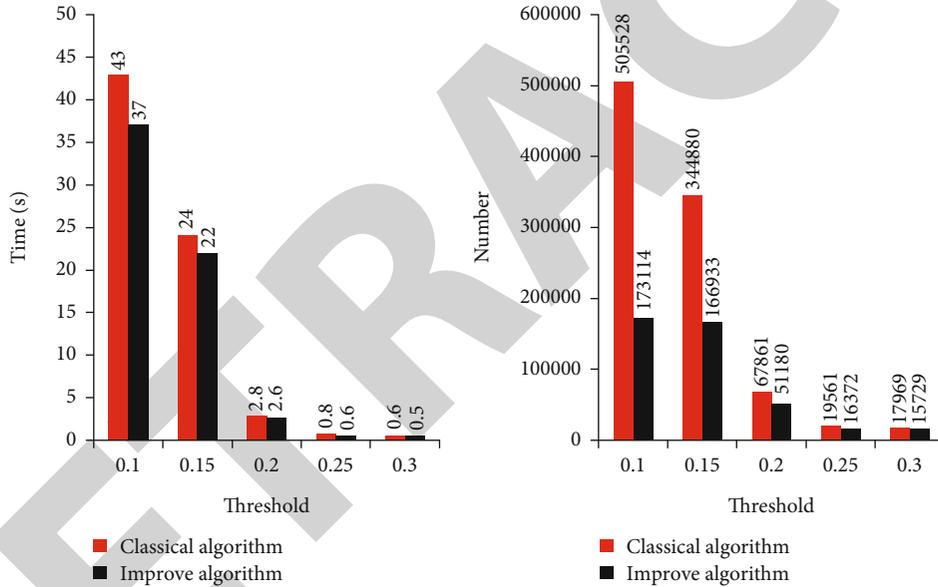


FIGURE 8: Comparison of data mining time and mining number between the improved algorithm and the classic algorithm.

amount of data researched in this article is self-evident, and the barrier between traditional hospitals, and the use of huge amount of data needs to be broken.

#### 4. The Effect of Medical Huge Amount of Data

4.1. *Performance Evaluation of Medical Huge Amount of Data.* For this reason, this article simulates the interaction process of EMRS and PACS to realize the transmission of DICOM image files and designs two sets of control experiments in cloud and noncloud environments: (1) average response time under different network bandwidths; (2) different concurrency. The average response time under the number of users is used to compare the performance difference between the two, as shown in Figure 7.

We can see from the above figure that compared with the traditional IT architecture, the hospital private cloud

shows its strong client load capacity and concurrency. First of all, under the same number of concurrent users, the data transmission efficiency between systems under the hospital private cloud architecture is higher than under the traditional architecture. Secondly, as the number of concurrent users increases, the average response time increase in the figure is greater under the traditional architecture than under the hospital private cloud architecture, indicating that the latter has better concurrency performance than the former.

4.2. *Improved Algorithm Simulation.* In order to verify the performance of the improved algorithm, the improved algorithm is now simulated with the classic Apriori algorithm.

First, test the time comparison between the improved algorithm and the classic Apriori algorithm under the same degree of support. The left side of the figure below shows the comparison between the improved Apriori algorithm and

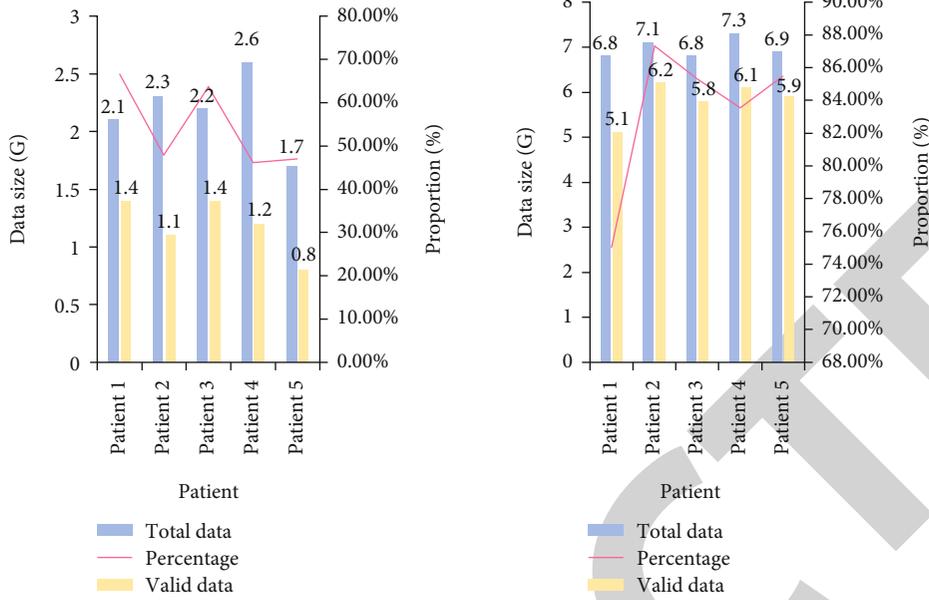


FIGURE 9: Comparison of comprehensiveness of knee osteoarthritis diagnosis data before and after improvement.

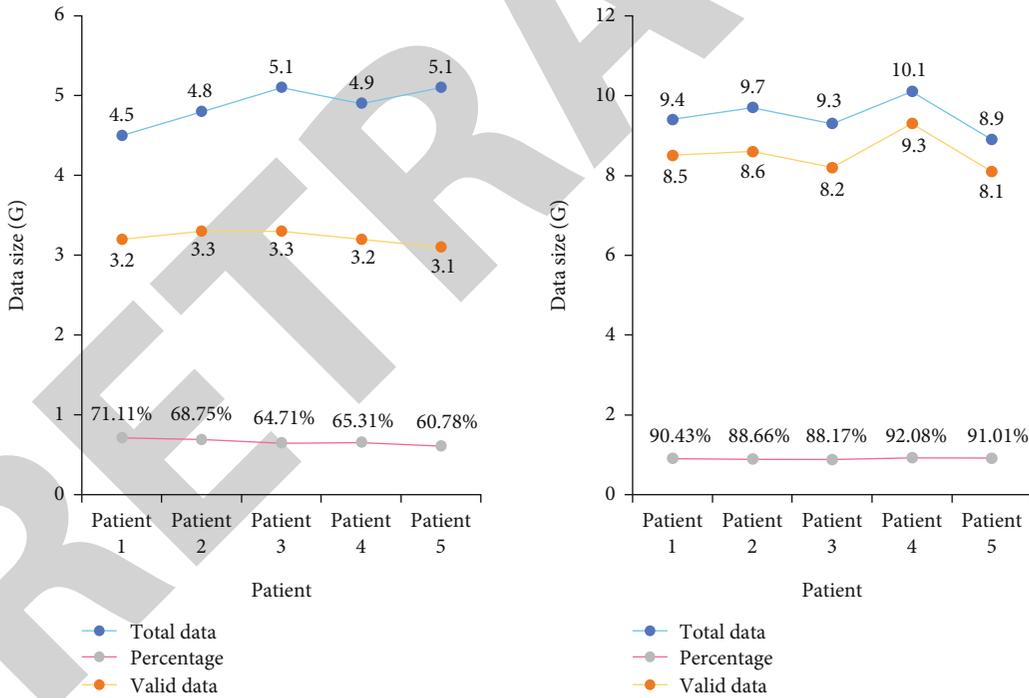


FIGURE 10: Comparison of the comprehensiveness of the diagnosis data of cardiovascular and cerebrovascular diseases before and after the improvement.

the classic algorithm in mining time. Because the introduction of interest is improved in the calculation part of mining association rules, and the main time complexity of the Apriori algorithm is to find frequent items, so in terms of time complexity, the improved algorithm does not improve the performance.

Secondly, carry out simulation test on the mining association rules. By setting different degrees of support, the classic Apriori algorithm, the improved algorithm with an interest threshold of 0.3, and an improved algorithm with an interest threshold of 0.45 are compared. The specific data comparison is shown in Figure 8.

From the above two algorithm simulation comparison figures, we can see that in practical applications, relevant experts can set an appropriate minimum interest threshold through experience, so as to filter out association rules that users are not interested in. The algorithm overcomes the classic Apriori. The algorithm may dig out the drawbacks of invalid strong association rules and optimize the algorithm.

#### 4.3. Comparison of the Effect of Medical Huge Amount of Data on Disease Diagnosis

**4.3.1. Comparison of Comprehensiveness of Knee Osteoarthritis Diagnosis Data before and after Improvement.** Through the analysis of the above system value, the improved algorithm is reliable in terms of stability and performance. In order to explore the practical application effect of the system, the data analyzed by the algorithm before and after the improvement in this article are in five cases of knee bone. Perform a statistical analysis on the comprehensiveness of the clinical diagnosis data of different patients with arthritis and compare the comprehensiveness of the data based on the statistically calculated clinical data size of the patient. The results are shown in Figure 9.

From the figure above, we can see that the diagnostic data of the knee osteoarthritis of the improved patient is 2.1 G, the effective data is 1.4 G, the diagnostic data of the second patient is 2.3 G, the effective data is 1.1 G, the third diagnostic data of the patient is 2.2 G, the valid data is 1.4 G, the diagnostic data of the fourth patient is 2.6 G, the valid data is 1.2 G, the diagnostic data of the fifth patient is 1.7 G, and the valid data is 0.8 G. The improved diagnosis data of knee osteoarthritis of patient no. 1 is 6.8 G, the effective data is 5.1 G, the diagnosis data of the no. 2 patient is 7.1 G, the effective data is 6.2 G, and the diagnosis data of the no. 3 patient is 6.8 G. The effective data is 5.8 G, the diagnostic data of the fourth patient is 7.3 G, the effective data is 6.1 G, the diagnostic data of the fifth patient is 6.9 G, and the effective data is 5.9 G.

Through the comparative analysis of the data, it can be seen that the data efficiency before the improvement is generally distributed between 50% and 70%, and the data efficiency after the improvement is between 74% and 88%, which is significantly higher than that before the improvement. And the amount of effective data after the improvement is obviously superfluous to the data before the improvement.

**4.3.2. Comparison of the Comprehensiveness of the Diagnosis Data of Cardiovascular and Cerebrovascular Diseases before and after the Improvement.** After comparison and comparison of knee osteoarthritis, it is found that the efficiency after improvement has been significantly improved. However, the comprehensiveness of the data in cardiovascular and cerebrovascular diseases is not limited for the time being; this section will explore it. The specific data is shown in Figure 10.

From the above figure, we can see that the data acquisition and data analysis capabilities of the improved medical

huge amount of data have been significantly improved. Before the improvement, the data obtained by the patient 1 was 4.5 G, the effective data was 3.3 G, the diagnostic data of the second patient is 4.8 G, the valid data is 3.3 G, the diagnostic data of patient 3 is 5.1 G, the valid data is 3.3 G, the diagnostic data of patient 4 is 4.9 G, the valid data is 3.2 G, and the diagnostic data of patient 5 is 5.1 G. The effective data is 3.1 G. The improved diagnostic data for cardiovascular and cerebrovascular diseases of patient no. 1 is 9.4 G, the effective data is 8.5 G, the diagnosis data of patient no. 2 is 9.7 G, the effective data is 8.6 G, and the diagnosis data of patient no. 3 is 9.3 G, which is valid. The data is 8.2 G, the diagnostic data of patient no. 4 is 10.1 G, the valid data is 9.3 G, the diagnostic data of patient no. 5 is 8.9 G, and the valid data is 8.1 G.

After a comprehensive analysis, we have concluded that the completeness of data acquisition for related diseases has increased by 31%; after the algorithm is improved, the proportion of effective data has increased by 23%, and the speed of data analysis has also increased by 17%. It is very effective and faster to obtain patient data and apply it to clinical symptom diagnosis.

## 5. Conclusions

The main research content of this paper is to diagnose knee osteoarthritis combined with cardiovascular and cerebrovascular diseases through medical huge amount of data and artificial intelligence. First, this paper improves the key algorithms of medical huge amount of data and proposes intelligent diagnosis and treatment for artificial intelligence diagnosis. The plan is to conduct a comprehensive analysis of relevant medical huge amount of data and make an accurate judgment on its symptoms. Finally, the experiment and analysis part are designed. The performance of the algorithm is monitored through the experiment, and the reliability of the monitoring system is also in the analysis part. A comprehensive analysis and judgment were made for the relevant effects.

### Data Availability

The data that support the findings of this study are available from the corresponding author upon reasonable request.

### Conflicts of Interest

The authors declare that they have no conflicts of interest.

### Authors' Contributions

Pan Wan contributed to this work as co-first author.

### References

- [1] T. M. Godinho, C. Costa, and J. L. Oliveira, "Intelligent generator of big data medical imaging repositories," *IET Software*, vol. 11, no. 3, pp. 100–104, 2017.
- [2] Z. M. Hu, L. L. Luo, L. Li, S. D. Dai, H. G. Zhang, and R. Z. Liu, "Indigenization of the median of markers for Down syndrome

- screening based on statistical analysis of medical big data,” *Taiwanese Journal of Obstetrics and Gynecology*, vol. 59, no. 4, pp. 556–564, 2020.
- [3] T. Inomata, J. Sung, M. Nakamura et al., “Using medical big data to develop personalized medicine for dry eye disease,” *Cornea*, vol. 39, no. 1, pp. S39–S46, 2020.
- [4] W. N. Price and I. G. Cohen, “Privacy in the age of medical big data,” *Nature Medicine*, vol. 25, no. 1, pp. 37–43, 2019.
- [5] P. T. Chen, “Medical big data applications: intertwined effects and effective resource allocation strategies identified through IRA-NRM analysis,” *Technological Forecasting and Social Change*, vol. 130, pp. 150–164, 2018.
- [6] X. Zhang and Y. Wang, “Research on intelligent medical big data system based on Hadoop and blockchain,” *EURASIP Journal on Wireless Communications and Networking*, vol. 2021, no. 1, 21 pages, 2021.
- [7] N. Ilyasova, A. Kupriyanov, R. Paringer, and D. Kirsh, “Particular use of big data in medical diagnostic tasks,” *Pattern Recognition and Image Analysis*, vol. 28, no. 1, pp. 114–121, 2018.
- [8] Z. Xiao-Qing, T. Tian-Hao, Z. Yi-Di, and Z. Lu, “Clinical intelligent diagnosis path based on the chief complaint,” *Digital Chinese Medicine*, vol. 3, no. 1, pp. 44–49, 2020.
- [9] H. Hu, B. Tang, X. Gong, W. Wei, and H. Wang, “Intelligent fault diagnosis of the high-speed train with big data based on deep neural networks,” *IEEE Transactions on Industrial Informatics*, vol. 13, no. 4, pp. 2106–2116, 2017.
- [10] Q. Qin, X. Yang, R. Zhang, M. Liu, and Y. Ma, “An application of deep belief networks in early warning for cerebrovascular disease risk,” *Journal of Organizational and End User Computing*, vol. 34, no. 4, pp. 1–14, 2022, (forthcoming).
- [11] Y. Zhao, H. Li, S. Wan et al., “Knowledge-aided convolutional neural network for small organ segmentation,” *IEEE Journal of Biomedical and Health Informatics*, vol. 23, no. 4, pp. 1363–1373, 2019.
- [12] K. Chen, “Intelligent pre-diagnosis method and experimental research based on patient’s personal information,” *Journal of Clinical and Experimental Medicine*, vol. 2, no. 3, pp. 14–18, 2018.
- [13] V. P. Singh and R. Srivastava, “Content-based mammogram retrieval using wavelet based complete-LBP and K-means clustering for the diagnosis of breast cancer,” *International Journal of Hybrid Intelligent Systems*, vol. 14, no. 1-2, pp. 31–39, 2017.
- [14] T. L. Shen and X. L. Fu, “Application and prospect of artificial intelligence in cancer diagnosis and treatment,” *Zhonghua zhong liu za zhi [Chinese journal of oncology]*, vol. 40, no. 12, pp. 881–884, 2018.
- [15] V. Roy, P. K. Shukla, A. K. Gupta, V. Goel, P. K. Shukla, and S. Shukla, “Taxonomy on EEG artifacts removal methods, issues, and healthcare applications,” *Journal of Organizational and End User Computing (JOEUC)*, vol. 33, no. 1, pp. 19–46, 2021.
- [16] F. Y. Wang, C. Li, Y. Guo et al., “Parallel gout: an ACP-based system framework for gout diagnosis and treatment,” *Moshi Shibie yu Rengong Zhineng/Pattern Recognition and Artificial Intelligence*, vol. 30, no. 12, pp. 1057–1068, 2017.
- [17] M. U. Ozturk, A. R. Arman, G. C. Bulut et al., “Statistical analysis and multimodal classification on noisy eye tracker and application log data of children with autism and ADHD,” *Intelligent Automation & Soft Computing*, vol. 24, no. 4, 2018.
- [18] P. D. Kaur and P. Sharma, “IC-SMART: IoTCloud enabled seamless monitoring for Alzheimer diagnosis and rehabilitation system,” *Journal of Ambient Intelligence and Humanized Computing*, vol. 11, no. 8, pp. 3387–3403, 2020.
- [19] L. Peng, Y. Na, D. Changsong, L. I. Sheng, and M. Hui, “Research on classification diagnosis model of psoriasis based on deep residual network,” *Digital Chinese Medicine*, vol. 4, no. 2, pp. 92–101, 2021.
- [20] F. Qian, J. Guo, Z. Jiang, and B. Shen, “Translational bioinformatics for cholangiocarcinoma: opportunities and challenges,” *International Journal of Biological Sciences*, vol. 14, no. 8, pp. 920–929, 2018.
- [21] T. Katsuda and H. Isshiki, “Contribution of medical big data analysis technology to medical innovations,” *Fujitsu Scientific & Technical Journal*, vol. 54, no. 2, pp. 59–65, 2018.
- [22] L. I. Dan, W. Hongdong, and L. Xiaofeng, “Bayesian network based approach for diagnosis of modified sequencing batch reactor,” *Journal of Shanghai Jiaotong University (Science)*, vol. 24, no. 4, pp. 417–429, 2019.
- [23] S. Wang, F. Li, F. Xia, and D. Ling, “Design and synthesis of dynamic nano assemblies for biomedical applications,” *Scientia Sinica Chimica*, vol. 49, no. 9, pp. 1168–1178, 2019.