

# Research on the application of data mining in the field of healthcare

Wanwan Ding, Juntao Fang\*

School of Management, Tianjin University of Traditional Chinese Medicine, Health Industry Park, Jinghai District 301617, Tianjin, China

**Abstract:** The healthcare big data industry is rapidly developing globally, and data mining and knowledge services in the healthcare field have become one of the core demands for its development. Data mining in healthcare is beneficial to improve the efficiency of diagnosis and treatment of patients, which is helpful to formulate more effective treatment plans and reduce medical costs. In this paper, we searched the core journals on China Knowledge Network and web of science by subject terms, and eliminated the irrelevant articles for literature counting. In this paper, the commonly used models and algorithms of data mining in healthcare are firstly elaborated; then the progress of the application of this technology in assisting medical tasks, optimizing resource allocation and improving health information services are respectively reviewed, summarizing the segmentation, classic algorithms and representative studies implied by each application. However, the application of data mining technology in healthcare also faces some problems, from data collection, to data cleaning, preprocessing, visualization, to the selection of algorithms and evaluation of results, each link is full of difficulties and challenges. Finally, this paper proposes future research directions such as diversifying data sources, strengthening security and privacy protection, developing visualization and analysis tools, accurately using big data to improve the service level of healthcare institutions, semanticizing electronic medical records mining, and improving cancer prevention. At the same time, data mining is deeply integrated with cloud computing, artificial intelligence and other fields to jointly promote scientific and technological progress in the field of health care.

**Keywords:** Data mining; healthcare; medical tasks; resource allocation; health information services.

## 1. Introduction

Big data in healthcare refers to the large amount of data generated by medical activities, medical research and medical treatment and other activities. This type of data is characterized by large volume, complex type, low value rate and fast speed, which makes conventional data analysis and processing difficult. Data mining techniques have been widely researched in health care and other areas because of their ability to process large-scale, heterogeneous data. Deep mining and analyzing big data can enhance the process of diagnosis and treatment, increase its application value in health medicine, develop targeted diagnosis and treatment plans for doctors, and lay the foundation for cooperation among patients, healthcare workers, researchers and other parties. In this paper, by organizing the research progress at home and abroad, we describe some important applications of data mining technology in the field of health medicine, point out the challenges and problems, and discuss the future research direction, which lays the foundation for future in-depth research on big data in health medicine.

The source databases and search methods of the review papers are shown in Table 1, which yielded 61 Chinese literature and 352,679 English literature. Since there were too many English literature, the top 100 most relevant ones were firstly sorted according to relevance; and then 161 literature were initially screened based on the titles and abstracts of the papers, and literature that was not in the field of healthcare and those that did not use data mining algorithms were excluded for the Analyze. Through full-text reading, the literature was categorized according to specific application areas and purposes, and the more recent and representative literature among similar papers was selected for review.

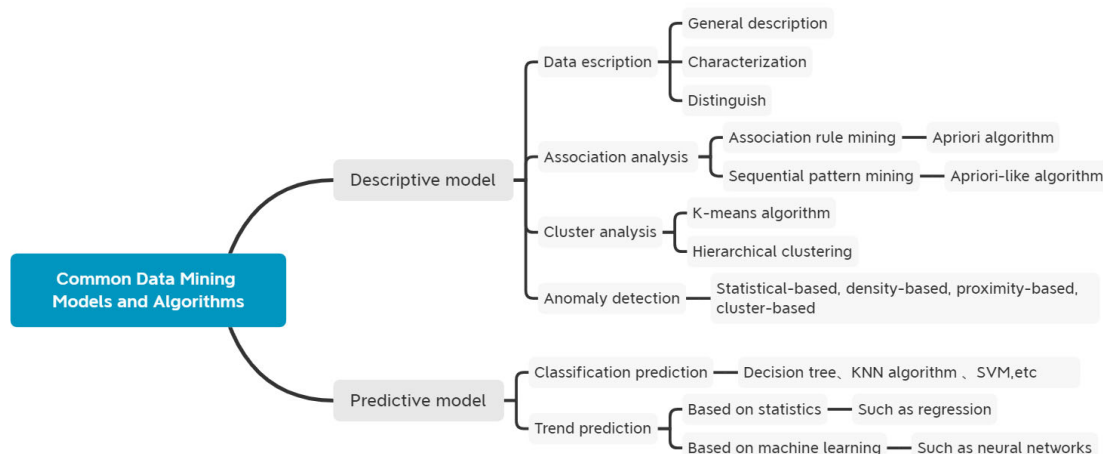


**Table 1 Source databases and search methods for review papers**

Comprehensive database	Web of science core collection	CNKI
Search method	TOPICS=data mining AND medicine OR medical	Topic=Data Mining AND Medical
publication date	2019-2023	2019-2023
Source category	SCIExpanded, CCR, IC	SCI, EI, PKU, CSSCI
Search results	352679 articles	61 articles

## 2. Data Mining Models and Algorithms

Data mining emerged in 1989, also known as knowledge discovery in databases. It is a product of the integration of knowledge from multiple disciplines, including machine learning, database application techniques, statistics, artificial intelligence, and other disciplinary areas of research results[1]. Data mining is the process of mining interesting patterns and knowledge from a large amount of data, and common data mining models and algorithms are shown in Figure 1.



**Fig.1 Summary of common data mining models and algorithms**

Descriptive modeling focuses on describing and summarizing past data by analyzing historical data to reveal patterns and relationships in the data. Descriptive analysis is the foundation of predictive analytics, enabling predictive models to more effectively capture correlations in data and make accurate predictions. Predictive modeling, on the other hand, focuses on using historical data to predict the future. Specifically, models and algorithms commonly used in healthcare include:

(1) Data description is the basic description, feature summary and comparison of all datasets. Zheng et al.[2] screened the top 20 most frequently used herbs in databases such as PubMed and Embase, described the validation information including location, site of use, family, and genus, and analyzed the classification, medicinal properties, and flavors of all the common medications used to treat endometriosis.

(2) Association analysis is used to explore the correlation between research objects, which is generally realized by mining the frequently occurring itemsets in the data, such as Min Shen et al.[3] used the association rule Apriori algorithm to analyze the data mining on the medication pattern and the effect of traditional Chinese medicine treatment of allergic rhinitis in children.

(3) Cluster analysis divides similar objects into different groups or subsets by static categorization, where objects within a group are similar to each other and objects in different groups are unrelated. Hong M et al.[4] studied the correspondence between etiology and cancer symptom clusters

in Chinese medicine based on cluster analysis, and found that the core etiology of clinical symptoms in patients with middle- to late-stage lung adenocarcinoma included qi stagnation and phlegm obstruction, phlegm and stasis coexistence, qi and yin deficiency, heat damage to the lung collaterals, and phlegm and stasis, with heat and deficiency as the main pathologic factors; the seven basic types of symptoms that were common in the classification of the clinical syndromes of patients with late-stage cancer, including the liver-kidney double-deficiency, spleen-kidney double-deficiency syndromes, lung deficiency syndromes, and liver-depression syndromes, and so on.

(4) Anomaly detection, in data mining, refers to the identification of items, events, or observations that do not match the expected pattern or other items in the data set.

(5) Classification prediction reflects how to identify features of common nature of similar things and features of differences between different things. Classification involves building a classification model through guided learning training and using the model to classify instances of unknown classification. Xing Wang et al.[5] obtained a simpler decision tree model for distinguishing between cold and cough diseases through the description of the algorithm, and used the model to predict other historical cases, which can be used to achieve diagnostic prediction and assist in diagnosis, and provide support for medical business as well as scientific research.

(6) Trend forecasting, also known as time series forecasting, is a method of applying mathematical and statistical methods to arrange historical information in chronological order, and then applying a certain mathematical model to predict the future. Tapak et al.[6] used support vector machines, artificial neural networks and random forest time series models to predict the timing of influenza outbreaks, and the time series models used had good performance.

### **3. Current status of data mining application in healthcare field**

#### **3.1 Paramedical tasks**

##### **(1) Prevention**

Their prevention can be achieved through effective intervention and early screening for associated risk factors. The general causative factors are demographic characteristics, family history, patient signs and lifestyle habits. Chen et al.[7] The Apriori algorithm of association rules for cervical cancer sample data set showed that age, sexual behavior, smoking status, hormonal contraceptive use, and birth control placement can lead to the occurrence of STDs or cancers with different levels of support and confidence, and there is a strong link between STDs and cancer diagnosis, which provides a new way of thinking about screening and diagnosis of cervical cancer.

##### **(2) Diagnosis**

The use of text information and images in electronic medical records can assist in disease diagnosis. Zheng et al.[8] Epidemiological characteristics of real cases in Sina Weibo were obtained by analyzing 690 valid cases in Sina Weibo's "Pneumonia patients' help hypertext" from February 4 to 22, 2020 based on Selenium data mining technology.

##### **(3) Treatment**

Based on the analysis of big data, hospitals can provide personalized treatment plans for patients, which is also conducive to the improvement of hospital healthcare. Zhao et al.[9] used data mining techniques to analyze drug treatment plans for hypertension, identify the key factors affecting the success of each treatment plan, and consider them as criteria to guide clinicians in prescribing drugs for patients in order to alleviate patients' pain, improve the quality of healthcare, and save limited health resources. Chen et al.[10] proposed a data-driven approach to mine typical treatment duration patterns for rational medication use from electronic medical records, which can recommend appropriate treatment plans for patients based on their admission information. Wang et al.[11] by constructing a TCM medical database utilizing multidisciplinary synergistic research to mine the database for information on the patterns between prescriptions, evidence and diseases, to clarify the

targeting mechanisms and effect pathways in the TCM diagnostic and treatment process, and to precision TCM.

### **3.2 Optimizing resource allocation**

Medical resources are the sum of various elements required for medical and health care. In the face of increasing medical needs, health management is facing serious challenges, and how to effectively allocate health services is an urgent problem in the current health field.

#### **(1) Hospitalization management**

Data mining can be used to analyze the factors influencing hospitalization costs, which in turn leads to cost prediction and helps in patient decision making. Mo Chen et al.[12] used K-means clustering and support vector machine to analyze the hospitalization information of urban residents' basic medical insurance patients with major diseases in a city in Hubei province from January 2016 to May 2017, and the result of the clustering superiority test suggests that the hospitalization cost is divided into 3 categories of the best, and the main factors of hospitalization cost are the main diagnosis disease, hospitalization day, hospital level, medical insurance business category and hospital type. The main factors of hospitalization cost are main diagnosis disease, hospital day, hospital level, medical insurance business category and hospital type.

#### **(2) Medical cost management**

Medical cost reflects the consumption of medical service resources, and is also the most concerned issue for patients. Chen Rui et al.[13] Based on the LIME-BP neural network, a medical cost prediction model was established to predict medical costs and explain their sources of costs, which is an important reference value for reducing medical costs and improving the doctor-patient relationship. Lixia Hu et al.[14] Using data mining methods to analyze the changes of various types of costs and the impact of drug costs on other costs using the cost consumption percentage, monthly sub-average cost changes, and Pearson's correlation coefficient, respectively, the results of the sub-average total cost, sub-average cost of western patent medicine, and sub-average cost of traditional Chinese herbs show an overall decreasing trend, and the costs of various other tests, surgeries, and treatments show an increasing trend. Jiaying Zhang et al.[15] established a multiple linear stepwise regression model based on the information of hypertensive hospitalized patients to screen the influencing factors of hospitalization cost, and used Apriori algorithm to mine the potential strong correlation rules between each factor and high hospitalization cost. The results showed that the number of hospitalization days had the greatest impact on the hospitalization costs of hypertensive patients, and 10 of the 12 strong rules associated with high hospitalization costs mined by the association rule model were related to the number of hospitalization days, which can provide a reference for reducing the burden of patients' medical care.

#### **(3) Medical insurance management**

Hengliang Wu et al.[16] Combined with the practice of Hubei and other regions to health insurance audit as an example from the three dimensions of technology, management and modeling, put forward the "top-down" big data audit implementation path, including big data culture construction, overall planning, team formation, platform construction, process design and breakthroughs and other five aspects, aims to provide a reference for the regions to carry out the health insurance big data audit work. The purpose is to provide reference for regions to carry out health insurance big data audit work. Yingliang Zhou et al.[17] Based on the audit practice in the field of medical insurance, analyze the current status of big data environment faced by the audit of medical insurance fund, explore the change of auditing method and method innovation, summarize the experience model of auditing experts, and explore the auditing application of relational network analysis method and data mining method.

#### **(4) Medical quality and equipment safety**

Jie Chen et al.[18] Combining the characteristics of medical quality and safety adverse events, applying ontology knowledge modeling method to design medical quality and safety adverse events

knowledge graph, which can be an integrated and fusion tool of medical quality and safety adverse events related knowledge, and provide support for big data analysis and inference prediction of medical quality and safety adverse events in the future. Yutong Wu et al.[19] Adopted Apriori model in association rule mining to mine the device category, department of use, hospital category, whether it is overdue for use, and the region to which the reporting unit belongs, to explore the risk associated with medical device adverse events.

### 3.3 Improving health information services

#### (1) Provision of health information needs

Health information needs contribute to more targeted health information services. Lin et al.[20] introduced Natural Language Processing (NLP) to quantify the importance of different mental disorder descriptions relative to the five reservoirs and two palaces, the stomach and the gallbladder through the classic medical text Huangdi Neijing, which showed that the most relevant organ to mental disorders is the heart, while the two most important emotional factors associated with mental disorders are anger and worry. Therefore, based on this it can be a reference for people who are concerned about their mental health to pay attention to controlling their emotions in order to avoid developing mental disorder diseases.

#### (2) Research on Health Decision Making

Conducting health decision-making research can help physicians decide on treatment options more efficiently. Xinyu Yang et al.[21] Using random forest and other methods to construct an electronic medical record data-driven lung cancer metastasis prediction model was compared with the empirical level of constructing a lung cancer metastasis prediction model using real electronic medical record data, and the prediction model performs better, which can provide support for clinical treatment. Zhang et al.[22] Capturing the movement trajectories of sportsmen to determine whether they visit healthcare centers frequently, healthcare organizations can send personalized and relevant medical information to people who visit certain healthcare centers regularly, and it will help to detect and diagnose the patient's disease. Bao Juan et al.[23] Bao Juan et al. proposed a new model of “multimodal fusion correlation mining analysis and predictive intelligent decision making” for intelligent healthcare analysis and decision making, and explored a new way of intelligent drug decision making by researching “intelligent drug decision making method”.

#### (3) Improved health information mining algorithm

Yu et al.[24] proposed a new medical big data clustering algorithm based on an improved immuno-evolutionary approach in a cloud computing environment, and this new approach can improve the accuracy of data classification, reduce the error rate, and improve the performance of data mining and feature extraction for medical data clustering. Yi et al.[25] proposed an improved version of the traditional Apriori algorithm, which is mainly based on the computer's fast response to bit-string logic operations, and is especially superior to the traditional algorithm in terms of running time, frequent itemset mining, and strong association rules. The improved Apriori algorithm outperforms the existing algorithms when it is applied to asthma medication and comorbid symptomatic medication data for association analysis. Yang[26] Taking chronic atrophic gastritis (CAG), a typical gastrointestinal disease in the plateau region, as a research object, a new BP network model for classification and prediction of Tibetan medicine symptoms was proposed. In order to overcome the typical shortcomings of BP networks, including slow convergence and easy overfitting, they used a Gaussian distribution-based method to improve the initialization of weights, and dynamically adjusted the learning rate by using the learning rate exponential decay method. The prediction accuracy of the model is significantly improved.

#### (4) Securing health information

Data mining, data pre-rule, and all-round digital monitoring of healthcare big data have made the dilemma of controlling personal privacy prominent, and the protection of health information is crucial

in the context of the big data era. Hai Gu et al.[27] Addressing Individual Privacy by Enhancing Value Transparency of Big Data Technologies, Returning and Reinventing Humanism, and Mining Shared Values to Reduce Conflicts of Interest in Healthcare Big Data Applications.

#### **4. Problems of Data Mining on Healthcare Domain**

There are still more problems and limitations in the application of data mining techniques in the healthcare field, as follows:

##### **(1) Complexity of data types and lack of high-quality data**

Currently, healthcare data is often incomplete, unstandardized, and complex in terms of data type, so an initial assessment of the quality of the data is needed before analysis. Data silos are another major problem constraining data collection. On the one hand, there are still some healthcare organizations that are unable to interoperate with each other; on the other hand, most people are reluctant to disclose their data on mHealth or health monitoring platforms because their willingness to share information drops sharply when they believe that the extent of their benefit from mHealth is not as great as the risk of privacy leakage[28,29] .

##### **(2) Dimensional disaster and data imbalance**

Healthcare data tends to be of high dimensionality and the data distribution is sparse, so the preprocessing stage requires the removal of less relevant attributes to improve mining speed or accuracy. In addition, in the classification prediction task, healthcare data often faces the problem of extremely unbalanced sample category distribution. Utilizing unbalanced samples for data mining can lead to higher prediction accuracy for categories with large sample sizes and lower prediction accuracy for categories with small sample sizes.

##### **(3) Inadequate analytical capabilities for big data in the medical field**

Since there are not enough relevant professionals in the healthcare field at present, it will be impossible to summarize and analyze the big data in a targeted way, so it is impossible to apply the big data accurately. Mining algorithms are the core of data mining, and when applied to the healthcare field, researchers should consider the characteristics of the dataset and the mining task to choose appropriate algorithms. For example, support vector machines are suitable for high-dimensional datasets, and their performance on low-dimensional datasets is relatively weak; decision trees and logistic regression have more interpretable results.

##### **(4) Selection of indicators for cross validation and assessment**

Model evaluation is an important step in the data mining process. For unsupervised models, one evaluation method is intuitive analysis, such as randomly drawing samples from each category for manual evaluation; the other is to analyze its internal information, such as measuring the distance between sample points in a cluster and measuring whether the distance from a sample to other clusters is far enough. For supervised models, balanced samples can be obtained by reducing the number of samples in large sample categories, and k-fold cross-validation methods can also be used to fully utilize each sample.

##### **(5) Data security issues**

At present, the protection of big data in the healthcare field is insufficient and confidentiality is inadequate, and the development of the network era has made it easier for some hackers to obtain data related to the healthcare field, so it is increasingly important to strengthen the privacy protection of big data.

##### **(6) Visualization issues**

Analytical tools based on advanced technologies are indispensable for the use of big data in the healthcare field, and visual analytical tools are now widely used in western developed countries. The

traditional relational database model handles data in multiple tables according to the usual procedures, and it is not possible to generate complex reports by processing big data due to the time-consuming queries.

## 5. Conclusion and prospect

This paper summarizes the progress of the application of data mining in the field of healthcare by reviewing the relevant literature at home and abroad. Data mining can not only assist medical tasks such as prevention, diagnosis, and treatment, but also assist the management of hospitalization, medical insurance, and medical expenses, so as to optimize the allocation of medical resources, but also assist doctors in decision-making, and thus improve health information services.

In the future, further research could be conducted in the following areas.

(1) Diverse data sources. There are many types of healthcare data, and most of the medical data come from the hospital information system (HIS). medical science research data come from specific medical research and monitoring programs, and their data are of high quality. Measurement data of human body is mainly based on human physiological characteristics, which is usually acquired in real time by devices such as wearable terminals. Textual data is usually obtained from health communities, social networks, etc., and in-depth research on it will be conducive to improving the level of “people-centered” health services. Most of the existing studies only analyze a single data source, but there are also studies that combine public health data and search engine data to predict influenza[30]. There are also studies that combine public health data and search engine data to predict influenza, and the results show that search data can reflect the latest influenza mutation trends that cannot be predicted by traditional data, indicating that the information contained in different data is complementary to a certain extent. How to integrate and mine multi-source data to improve the accuracy and real-time performance is one of the future research directions.

(2) Development of visual analysis tools. Visual analysis tools store data in a multidimensional database model that enables online multidimensional processing of big data using an analysis system, simplifying query procedures and reducing execution time.

(3) Precise use of big data to improve the service level of medical institutions. The proper application of big data in the field of healthcare cannot be separated from precision medicine. Precision medicine is an emerging method of disease prevention and treatment that designs personalized treatments for patients, taking into account the differences in their genes, living environments, and lifestyles. To make precision medicine highly implementable it is necessary to mine and analyze the big data of healthcare organizations, using text mining or natural language processing to mine patients' electronic health information.

(4) Semantic mining of electronic medical records. Currently there are fewer studies on semantic mining based on electronic medical records, on the one hand, because most of the information in electronic medical records is stored in the form of unstructured text, which cannot be understood and processed by computers; on the other hand, because it is of little help to medical activities if only superficial data mining is performed on electronic medical records. It is possible to construct a thesaurus of medical terms, establish semantic relationships between medical entities, and perform data mining at the conceptual level to provide clinical decision-making and support for medical personnel.

(5) Improving cancer prevention. Healthcare organizations need to pay more attention to the prevention and treatment of cancer when carrying out precision medicine, fully taking into account the specificity of the cancer group, and providing targeted treatment for cancer patients through data mining. At the same time, the lesions of patients are fully grasped, so as to develop a set of personalized treatment plans for patients.

(6) From a macro perspective, in the future, the application foundation of healthcare data mining should be strengthened, a comprehensive basic database centered on residents' electronic health

records, electronic medical records, etc. should be developed, and key technologies such as healthcare data storage and cleaning, analysis and mining, and security and privacy protection should be strengthened. At the same time, data mining and cloud computing, artificial intelligence and other areas of deep integration, together to promote scientific and technological progress in the field of health care.

## Acknowledgments

This research was supported by Natural Science Foundation of China (NSFC) with Grant No. 71902139, 72342017 and the Youth Foundation of the Ministry of Education under Grant No. 18YJC630029.

## Conflict of interest

The authors declare that they have no competing interest.

## References

- [1] Yang Xiaojuan. A review of domestic research on data mining[J]. Computer Programming Skills and Maintenance,2020,No.422(08):115-117.DOI:10.16184/j.cnki.comprg.2020.08.041.
- [2] Zheng Weilin, Wu Jiayi, Gu Jiangyong, Weng Heng, Wang Jie, Wang Tao, Liang Xuefang, Cao Lixing. Modular Characteristics and Mechanism of Action of Herbs for Endometriosis Treatment in Chinese Medicine: A Data Mining and Network Pharmacology-Based Identification.[J]. Frontiers in pharmacology,2020,11.
- [3] Shen Min,Liu Jianghai,Wang Kuanhong. Effect of Traditional Chinese Medicine on Allergic Rhinitis in Children under Data Mining[J]. Computational and Mathematical Methods in Medicine,2022,2022.
- [4] Hong M,Sun WH,Zhu Y.A cluster analysis-based study on the traditional Chinese medicine whose pathogenesises corresponding to lung cancer symptom groups [J]. Asian Journal of Surgery,2023,5.
- [5] WANG Xing,LIU Xiaoyan. Research on disease prediction modeling in medical big data environment[J]. Manufacturing Automation,2022,44(07):24-27.
- [6] Tapak L ,Hamidi O ,Fathian M , et al.Comparative evaluation of time series models for predicting influenza outbreaks: application of influenza-like illness data from sentinel sites of healthcare centers in Iran[J].BMC Research Notes,2019,12(1):1-6.
- [7] CHEN Zhi,WU Juanjuan. Research on medical data mining based on association rules[J]. Statistics and Decision Making,2020,36(06):174-177.DOI:10.13546/j.cnki.tjyj.2020.06.040.
- [8] ZHENG W, ZHAO Y, LI Zekun et al. Epidemiological characterization of COVID-19 based on Web data mining[J]. Journal of University of Electronic Science and Technology,2020,49(03):408-414.
- [9] Yan Zhao,Jiaying Le,Lifeng Zhu,Ming Zuo. Study on the effect of hypertensive treatment based on drug factor analysis model under the background of big data.[J]. Journal of Intelligent and Fuzzy Systems,2019,37(3).
- [10] Jingfeng Chen,Chonghui Guo,Leilei Sun,Menglin Lu. Mining Typical Treatment Duration Patterns for Rational Drug Use from Electronic Medical Records [J]. Journal of Systems Science and Systems Engineering,2019,28(5).
- [11] WANG Jianchao,YU Ye,CAO Xian et al. Application of big data in precision research of traditional Chinese medicine[J]. Shizhen Guomao Guomao,2019,30(02):437-439.
- [12] CHEN Mo,CAI Miao,HUANG Ahong et al. Research on influencing factors and control strategies of hospitalization cost for patients with major diseases based on K-means clustering and support vector machine[J]. China Hospital Management,2019,39(05):45-47+53.
- [13] Rui Chen, Li J, Guo DW et al. Research on medical cost prediction based on LIME-BP neural network[J]. China Health Statistics,2020,37(05):698-700.
- [14] HU Lixia,CHEN Chen,ZHANG Hong et al. Analysis of chronic disease outpatient costs in a hospital after zero markup using data mining[J]. Chinese Journal of Hospital Pharmacy,2020,40(09):1054-1057.DOI:10.13286/j.1001-5213.2020.09.20.
- [15] ZHANG Jiaying,LI Xiangrong,GAO Hua et al. Influencing factors of high hospitalization costs in hypertensive patients[J]. Chinese Journal of Hypertension,2020,28(03):271-275.DOI:10.16439/j.cnki.1673-7245.2020.03.021.
- [16] WU Hengliang, YU Benhai, ZHAI Yuxiang et al. Research on the implementation path of health insurance fraud audit in big data environment[J]. Finance and Accounting Monthly,2019,No.868(24):105-110.DOI:10.19641/j.cnki.42-1290/f.2019.24.013.

- [17] Research Group of Theoretical Research Society of Kunming Office of the Audit Commission, Zhou Yingliang, Chen Bo, etc. Research on Big Data Audit Practice in the New Era--Taking the Audit of Medical Insurance Fund as an Example[J]. *Audit Research*,2020,No.214(02):7-13.
- [18] CHEN Jie,FENG Dan,LIU Yuehui. Design and preliminary exploration of knowledge mapping for medical quality and safety adverse events[J]. *China Hospital Management*,2023,43(02):53-56+67.
- [19] WU Yutong,WU Sijia,YANG Jianwei et al. Analyzing the association of medical device adverse events in Shandong Province in 2021 based on Apriori algorithm[J]. *Journal of Shandong University (Medical Edition)*,2022,60(12):111-118.
- [20] Lin WanLing,Liang YuChi,Chung KuoHsuan,Chen PingHo,Chang YungChun. Using Text Mining and Data Visualization Approaches for Investigating Mental Illness from the Perspective of Traditional Chinese Medicine[J]. *Medicina*,2023,59(2).
- [21] YANG Xinyu,MOU Dongmei,PENG Hao et al. Research on data-driven clinical decision-making model of electronic medical record based on data characteristics[J]. *Intelligence Theory and Practice*,2022,45(05):181-188.DOI:10.16353/j.cnki.1000-7490.2022.05.023.
- [22] Zhang Dongzhi,Lee Kyungmi,Lee Ickjai.Multi-level medical periodic patterns from human movement behaviors.[J].*Health information science and systems*,2019,7(1).
- [23] Juan Bao,Min Tuo,Ting Meng Hou,Yu Xi Li,Qun Wang. Research on Intelligent Medical Engineering Analysis and Decision Based on Deep Learning[J]. *International Journal of Web Services Research (IJWSR)*,2022,19(1).
- [24] Yu Jing,Li Hang,Liu Desheng. Modified Immune Evolutionary Algorithm for Medical Data Clustering and Feature Extraction under Cloud Computing Environment.[J]. *Journal of healthcare engineering*,2020,2020.
- [25] Yi Z,Peipei C,Biyu C, et al. Application of Apriori Improvement Algorithm in Asthma Case Data Mining.[J]. *Journal of healthcare engineering*,2021,2021.
- [26] Yang Shenghao,Zhu Xiaolan,Zhang Lei.Classification and Prediction of Tibetan Medical Syndrome Based on the Improved BP Neural Network.[J]. *Institute of Electrical and Electronics Engineers*,2020,2020.
- [27] GU Hai,JIA Cangcang,WU Di et al. Exploration of personal privacy issues in the context of healthcare big data application[J]. *China Health Care Management*,2021,38(02):117-120.
- [28] Richardson J E,Ancker J S.Public Perspectives of Mobile Phone Effects on Healthcare Quality and Medical Data Security and Privacy A 2-Year Nationwide Survey [C]IIAMIA.AMIA Annual Symposium proceedings. San Francisco, 2015:1076-1082.
- [29] Gaylin D S, Moiduddin A, Mohamoud S, et al. Public Attitudes about Health Information Technology, and its Relationship to Health Care Quality, Costs, and Privacy[J].*Health Services Research*, 2011,46(4):92093 and Privacy[J].*Health Services Research*, 2011,46(3):920-938.
- [30] WANG RuoJia,LI Pei. Comparison and optimization of influenza surveillance models based on internet search data[J]. *Library and Intelligence Work*,2016(18):122-132.