

# Research on hospital information data statistics under the background of big data and artificial intelligence

Xufan Chen<sup>1\*</sup>, Tianyi Chai<sup>2</sup>

{Email: chen\_xufan@outlook.com, tianyi\_chai@outlook.com}

<sup>1</sup> Mathematics, Sichuan University, Chengdu, Sichuan 61000, China

<sup>2</sup> Financial Mathematics, Beijing Normal University Hong Kong Baptist University United International College, Zhuhai, Guangdong 519000, China

**Abstract.** With the continuous increase of online social media data, data has become a powerful assistant for the competition and development of various industries, and various data statistics and processing analysis tools in big data technology have also developed into necessary tools for enterprises. Artificial intelligence technology enables electronic devices to have intelligent thinking ability after being trained by data sets, and is widely used in various industries that require complex computing processes. The hospital is an inseparable place for every citizen, and the amount of data it generates is extremely huge. Therefore, it will be extremely difficult for the staff to manually count the data in the hospital system. Applying big data and artificial intelligence technology to hospital information data statistics will reduce the complicated manual statistical process and reduce labor costs. This paper firstly conducts a basic analysis of hospital information data statistics requirements, big data and artificial intelligence technology, and then in order to investigate the current hospital information data statistics, through the questionnaire star online platform, the information data statistics staff of 30 hospitals have been surveyed. After the interview, through the analysis of the questionnaire data, it was found that there are problems in the current hospital information data statistics that the quality of data statistics is poor, the data statistics take a long time and the process is complicated. Finally, based on big data and artificial intelligence technology, the pertinence of the hospital information data statistics problem is designed. Solution: After a second visit to the research object, it was found that the statistical efficiency, cost and accuracy of hospital information using related strategies were greatly improved.

**Keywords:** big data artificial intelligence hospital information data statistics.

## 1 Introduction

### 1.1 Hospital Information Data Statistics

Hospitals generate a large amount of valuable data every day, and statistics can help hospitals understand how employees are doing, and thus improve the way the hospital operates[1]. From another perspective, statistical data can also provide protection for the storage and analysis of patients' medical data [2]. The current hospital information data statistics are mainly completed by using the hospital's own office automation system and manual statistics [3]. Hospital information data is the basis for the hospital's human resources and financial management [4].

An effective information statistics method can ensure that the hospital's economic benefits are maximized, while reducing the hospital's labor costs and reducing the pressure on statisticians [5]. However, there are various problems in the current hospital data statistics, so that the statistical results can not meet the actual data needs [6]. Moreover, the use of new computer technology in the current hospital information data statistics is low, which cannot make the information data statistics work achieve the most convenient effect [7].

### 1.2 Big Data

With the development of information technology, big data has become an indispensable tool for processing and analyzing large amounts of data [8-9]. It becomes particularly important to find patterns or screen out valuable information in redundant and mixed massive information [10-11].

### 1.3 Artificial Intelligence

Artificial intelligence first came to the public eye when a chess-playing artificial intelligence called alpha defeated the world champion Go[12]. Later, physicist Hawking repeatedly predicted that the future development of artificial intelligence may lead to the extinction of human beings, which has triggered a long-standing debate in the academic and business circles[13]. In fact, artificial intelligence technology has not yet reached the ability to fully simulate human behavior, but the application of artificial intelligence in various industries has become the current trend[14]. Institutions predict the development trend of artificial intelligence technology as shown in Figure 1.

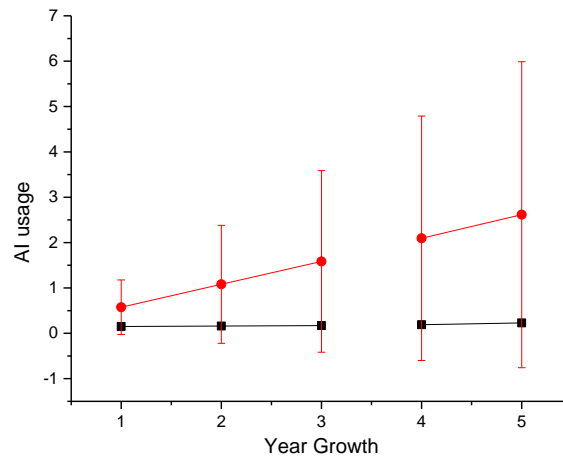


Fig. 1 The development trend of artificial intelligence technology with age

## 2 Questionnaire and Analysis

In order to investigate the current hospital information and data statistics, the information and data statistics staff of 30 hospitals were interviewed through the Star Online Questionnaire platform. A total of 300 questionnaires were distributed and 299 were recovered. The recovery rate met the statistical standards of the questionnaire. Among them, the quality of data statistics has a greater impact on the accuracy of data archiving and information management.

**Table 1** Statistical quality of data

hospital number	Statistical quality [0,1]	Statistics speed (relative)
A	0.21	0.12
B	0.19	0.17
C	0.23	0.13

From the data display in Table 1, it can be seen that the statistical quality and statistical speed of the displayed hospitals A, B, and C belong to a relatively low level, in which the relative value of statistical quality does not exceed 0.25, and the statistical speed does not exceed 0.20. The hospital serves a large number of people every day. Trying to save working time for each staff member is an important way to improve the operation efficiency of the hospital and ensure the healthy operation of the hospital. The time spent in data statistics is related to the personal status and work enthusiasm of the employees in the hospital data statistics. It is highly correlated, which indirectly affects the operation of the hospital. The time of hospital information data statistics will also be related to the hospital human resource management and financial data, which will further affect the operation of other contents. Table 2 shows the time-consuming content of data statistics in this questionnaire.

**Table 2** Time-consuming data statistics

hospital number	Labor costs	Statistics Time
A	5 people on average	6 days
B	4 people on average	7 days
C	Average of 6 people	5 days

From the data in Table 2, it can be seen that the labor cost of each hospital's information statistics is more than 4 people, and the average cost of 5 people belongs to the higher labor cost in the industry. In addition, the minimum time for statistics is more than 5 days. In many hospitals, after patients go to check their bodies, they can get the results after a week. In fact, most of the time is wasted in the process of data statistics.

In most hospital hot clinics, it is not easy to get a quota from queuing in the morning to the end of get off work. To obtain the qualification to see a doctor, you need to make an appointment online in advance. This shows that the hospital receives a large number of patients every day, and the amount of data generated is also extremely large. Among them, it is the norm that there are many types of data and the process of data statistics is complex. In most cases, manual statistics are performed step by step.

**Table 3** Data Statistics Process

hospital number	registration stage	Before hospitalization
A	Online data export	Check data export
B	All-in-one data export	Outpatient data export
C	Counter data export	Manual statistical calculation

The construction of hospital informatization is accelerating with the changing needs of the times, and the technologies and tools used in the process of hospital information statistics are also constantly changing, and the types of tools and technologies used will also determine the final statistical efficiency and speed. This questionnaire The contents of the techniques and tools involved in the survey are shown in Table 4.

**Table 4** Usage of statistical tools for hospital data

hospital number	Use of statistical tools
A	Excel、SQL
B	OA、Excel
C	Echart、Hadoop

From the data in Table 4, it can be seen that currently, hospital A mainly uses manual Excel and database operation SQL language to conduct hospital information statistics. Hospital B mainly uses the data export function of the office automation system to complete the statistics with Excel. Hospital C has tried data visualization. The way of processing and data warehouse processing to carry out data statistics.

In order to prove the credibility of the hospital data statistical questionnaire survey, through the in-depth analysis of the questionnaire reliability theory, it is found that it is mainly affected by specific parameters. The details are shown in Table 5.

**Table 5** Reliability and validity parameters of hospital data statistical questionnaire

parameter	reliability	validity
Value	Alpha=0.81	KMO=0.90

Through the analysis in Table 5, it can be found that the alpha parameter and KMO parameter of this questionnaire are high. After combining with the reliability and validity theory analysis again, it is found that the parameters of this questionnaire conform to the standard values stipulated in the theory, so the hospital information of this time is determined. The statistical questionnaire is valid and credible.

### 3 Problems existing in the current hospital information data statistics

#### 3.1 Poor statistical quality of hospital information data

At present, some of the work of hospital information data statistics adopts the office automation system purchased by the hospital, and some work is completed by manual statistics[15]. Hospital information data statistics mainly include four types of data statistics, involving employees, outpatients, inpatients, and medical record management[16]. Among them, the outpatient data statistics use the computer used for registration as the collection tool, which is mainly used to count the number of registered patients and the data of patient visits[17]. However, in real life, there is a situation where a patient corresponds to multiple registration and diagnosis and treatment data[18]. This situation is mainly caused by the patient not providing his own medical record or failing to seek medical treatment after queuing up for registration. In addition, after making an appointment with a doctor online, The situation of not seeing a doctor in the follow-up will cause errors in the statistics of the information and data in the outpatient department. In the hospitalization process, there is also a situation where the information provided by the patient does not match the information stored in the hospital, mainly because the number of patients in major hospitals is large, and the amount of data that the statistician needs to review every day is correspondingly large, and the relevant nurses and on-duty review Doctors also failed to accurately review their information, resulting in errors in information statistics, and subsequent patients will not be able to query their own diagnosis and treatment data after they are discharged from the hospital. In the process of archiving medical records, there are multiple roles such as doctors, nurses, and archivists who operate on a patient's data at the same time[19]. In this process, errors in archived data often occur due to personal

calculation errors, which is not conducive to subsequent patient medical records. Reference, and the current hospital has not adopted a more intelligent way to solve the problem of poor statistical quality of information data[20].

### **3.2 Data statistics take a long time**

Through the questionnaire data, it can be found that the current hospital information data statistics generally take a long time, and patients are often eager to obtain the diagnosis information of the disease, and the short-term assessment of the hospital staff also needs to update the data in a more timely manner. The reason that data statistics takes a long time is that there is still no intelligent statistics strategy established for data statistics[21]. The current process is mainly to export the daily data of patients and staff from the office automation system, and then perform manual excel calculation. It takes a lot of time in the process of screening and exporting, and it will take some time in the process of manual statistics and verification[22]. The entire process above is completed according to specific rules, but requires more manual operations[23]. At present, statisticians have also considered using the method of setting functions in excel for quick statistics, but most of the time is still wasted in the process of repetitive calculations[24]. At present, the hospital has not combined some intelligent algorithms to further complete the intelligentization of the whole process[25]. The above reasons lead to the problem of long time-consuming hospital information data statistics[26].

### **3.3 The process is complicated**

From the statistical results of the questionnaire data, it can be found that the current hospital information data statistical process not only involves exporting data from OA, but also requires manual operations to complete excel calculations[27]. The complexity of which is mainly due to the manual identification and calculation of various data. The higher the error rate in the input stage, the more errors occur in manual statistics, resulting in the problem that the data cannot be matched during manual calculation. This process of manual statistical verification not only results in a waste of time, but also has other costs. In addition, the process of exporting data from machines used in various hospitals is more complicated, because hospital information data statistics involve various types of data such as registration, inspection, medical treatment, prescription of medicine, hospitalization, and the administration of hospital employees themselves. The type of information data involves different statistical equipment. For example, the registration can be made by online reservation, offline all-in-one registration, and counter registration. The three methods are stored in three different systems. In addition, the patient examination may involve multiple departments, and the medical record recording system involved in each department is also different. This multi-process and multi-device situation makes the operation of the background information and data statistics personnel extremely difficult.

## **4 Statistical countermeasure design of hospital information data based on big data and artificial intelligence**

### **4.1 Strategies for improving the statistical quality of hospital information data based on big data and artificial intelligence**

The intelligent analysis and processing capabilities of big data and artificial intelligence can solve the problem of poor statistical quality of current hospital information data. For different hospital information data statistics, different AI algorithms and big data technology can be used for intelligent design. For example, in the stage of outpatient data entry and statistics, the data processing method of big data technology can be combined to register patients in pharmacies or other hospitals. If the gap is too large, use the machine learning algorithm in the AI algorithm to conduct further review through the comparison and screening function in the machine learning algorithm, and use the intelligent algorithm to ensure that the accuracy of the data in the input stage is Guarantee the premise of follow-up hospital data statistics. In addition, the AI algorithm code can be developed according to the current rules of the hospital's manual information and data statistics work, and each variable can be defined step by step according to the steps of daily medical information statistics, and the data that needs to be extracted manually can be extracted, saving labor costs and reducing labor statistics. Errors in the process, while effectively improving the quality of data statistics. After a second survey of the research subjects, it was found that the speed and labor cost of hospital information data statistics combined with big data and artificial intelligence have improved. The specific situation is shown in Figures 2 and 3.

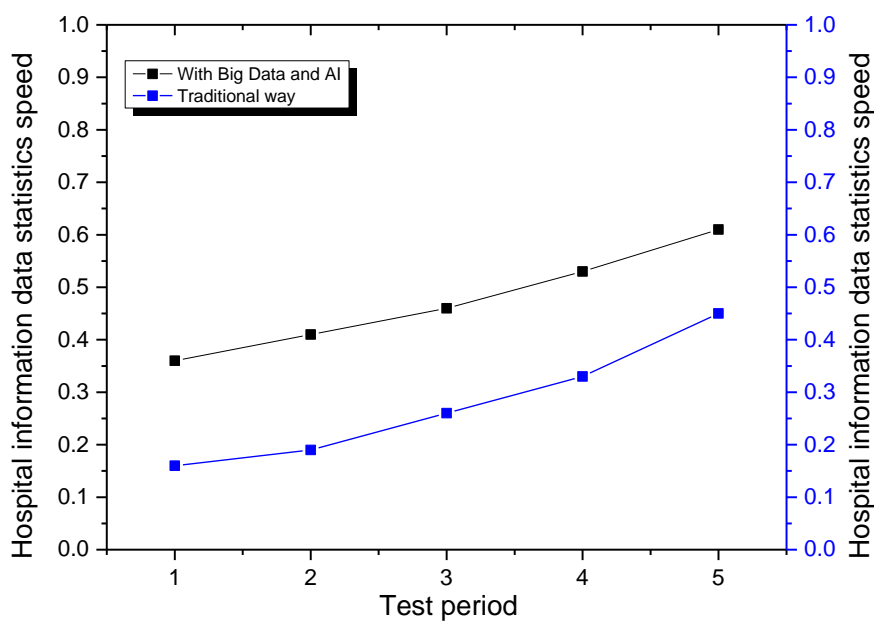
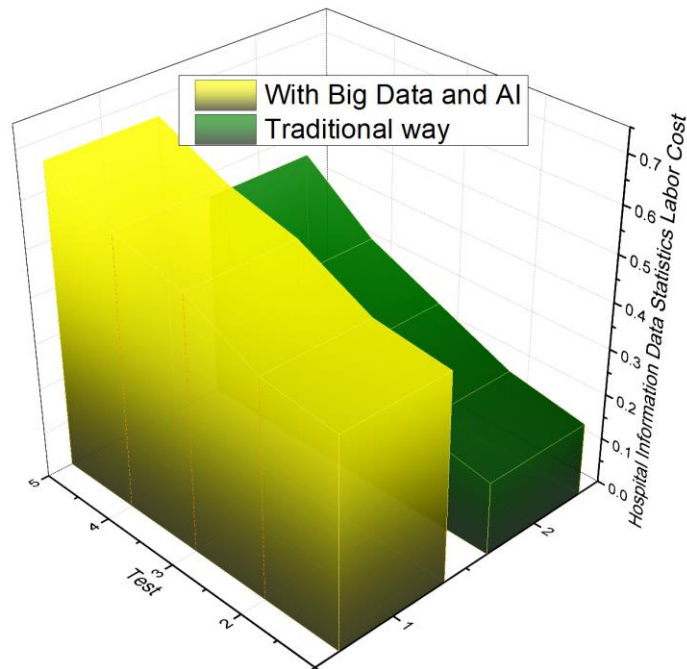


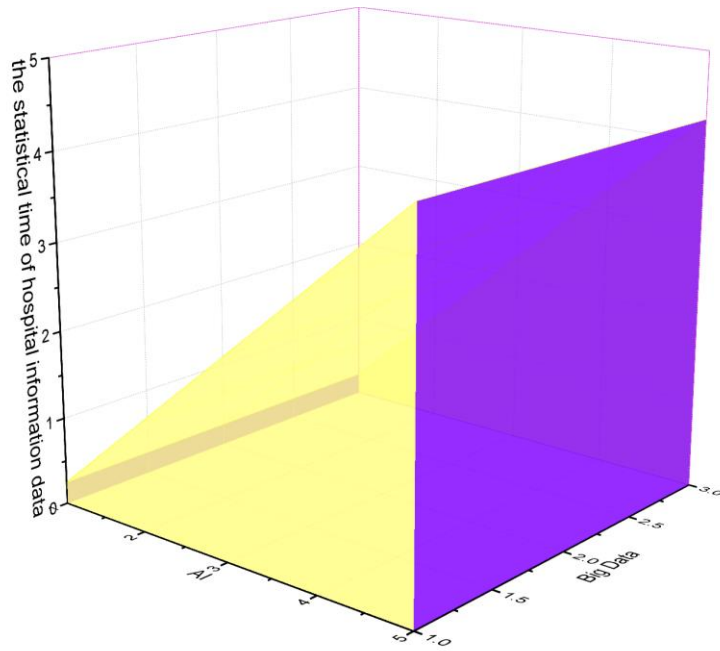
Fig. 2 Statistics speed of hospital information data



**Fig. 3** Labor cost of hospital information data statistics

#### 4.2 Data statistics time optimization combined with AI algorithm

Aiming at the problem that the current hospital information data statistics are time-consuming due to the complex process but more repetitive labor, an intelligent solution strategy can be designed in combination with AI algorithms. The algorithm is mainly used to deal with events that have a long time span but need to be memorized. It can remember the content of the previous and subsequent steps in the hospital information data statistics, so that the information data statistics process has continuity. In view of the process complexity of hospital information statistics, it can be designed in combination with the three-gate structure of LSTM. The input gate of LSTM is the unrecorded and recorded data of a certain process of hospital information data statistics, and the outputs of LSTM are mainly used for the following The data and unrecorded data to be recorded in a calculation process, and the structure of the forget gate of LSTM is S-shaped, which forgets the state between the input and the output to facilitate the jump between the two states. LSTM is mainly used for jumping between processes, such as hospital OA system jumping to the calculation of manual statistical process. In addition to LSTM, other AI algorithms can also be combined to realize the intelligence of the entire process. For example, in the subsequent data calculation, if intelligent analysis of patient data is required to obtain the preliminary results directly, big data and AI algorithms can be used to achieve and In this way, the time in hospital information statistics can be rationalized. After combining with AI algorithms such as LSTM, the length of hospital information data statistics will gradually decrease. The specific situation is shown in Figure 4.



**Fig. 4** Changes in the statistical time of hospital information data

#### 4.3 Using AI algorithms to improve statistical processes

The convolutional neural network in the AI algorithm can be designed according to the needs, and is mainly used to simplify and intelligentize the complex process. Aiming at the complex problem of the current hospital information statistics process, the RNN and CNN convolution methods in the AI algorithm can be used to develop codes for each step of manual operation, so that the computer can learn the behavior of human operation and achieve the goal of reducing the workload of manual operation. Among them, RNN can realize the effect of historical information recording and learning. Hospital statisticians can convert the steps in historical statistics into text, and combine RNN to learn the sequence of text content. Its cyclic structure has a great special effect on text processing. . After improving the process of hospital statistical data through algorithms such as RNN, it is found that the complexity of the hospital's information data statistical process is greatly reduced. The specific situation is shown in Figure 5.



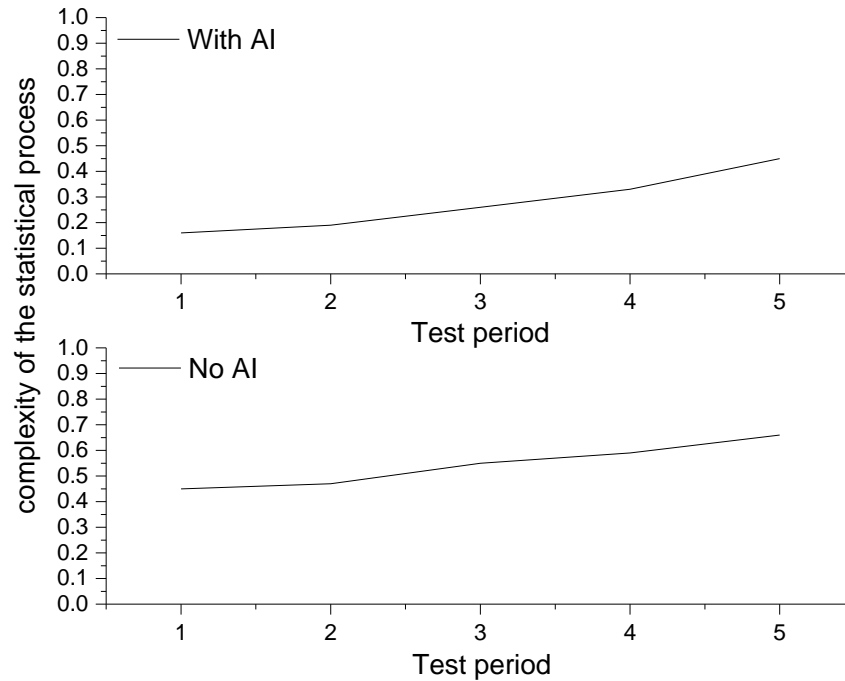


Fig. 5 Changes in the complexity of the hospital information data statistical process

## 5 Conclusion

The statistical work of hospital information data is not only related to the work intensity and cost of the hospital's own staff, but also to the timely processing and effective storage of hospital-related patient data. Big data technology can comprehensively analyze various data to be processed through some software similar to hadoop and spark, while artificial intelligence mainly optimizes various processes through algorithms, and its main goal is to let computers perform human-like operations. Although the current hospital information data statistics work has partially used computer technology, it has not been combined with big data and artificial intelligence technology, and its statistical intelligence is still insufficient, resulting in increased labor costs.

Based on the analysis of the demand and necessity of hospital information data statistics, this paper conducts a questionnaire survey of hospital information statisticians, and finds that there are problems of poor quality, time-consuming and complicated processes at present. The above-mentioned problems can be solved by adopting the strategy for improving the quality of hospital information data statistics, optimizing the data statistics duration combined with AI algorithms, and using AI algorithms to improve the statistical process. After a post-investigation on the subjects who participated in the questionnaire, it was found that the statistical process, time-consuming and quality of hospital information data combined with AI technology and big data have been improved.

## References

- [1] Xu Lixiang. Research on innovation of hospital statistical work mode in the era of big data[J]. *The Frontiers of Society, Science and Technology*,2021,3.0(5.0).
- [2] Xie Hua,Cui Xin,Ying Xiaohua,Hu Xiaohan,Xuan Jianwei,Xu Su. Development of a Novel Hospital Payment System – Big Data Diagnosis & Intervention Packet[J]. *Health Policy OPEN*,2022(prepublish).
- [3] Mao Zhongliang,Feng Li,Lou Jingsheng,Cao Jiangbei,Mi Weidong. [Design and Application of Perioperative Multi-center Data Center].[J]. *Zhongguo yi liao qi xie za zhi = Chinese journal of medical instrumentation*,2021 , 45(3).
- [4] Xiaohong Li, Yanling Zhang, Yujuan Li, Ke Yu, Yihua Du. Study of E-business applications based on big data analysis in modern hospital health management [J]. *Information Systems and e-Business Management*, 2021(prepublish) .
- [5] Lopes João, Braga João, Santos Manuel Filipe. Adaptive Business Intelligence platform and its contribution as a support in the evolution of Hospital 4.0 [J]. *Procedia Computer Science*, 2021, 184.
- [6] Chen, Talen, Madanian, Samaneh, Airehrour, David, Cherrington, Marianne. Machine learning methods for hospital readmission prediction: systematic analysis of literature[J]. *Journal of Reliable Intelligent Environments*, 2022(prepublish).
- [7] Jain Ravi. Introducing Artificial Intelligence Applications at Our Community Hospital: A Contrarian Approach.[J]. *Journal of the American College of Radiology: JACR*,2021,18(11).
- [8] Guo Mingyue. Mathematical Multivariate Statistical Analysis and Big Data Information System Establishment on Urban Pollutant through Intelligent Computational Method[J]. *Journal of Physics: Conference Series*,2021,2083(4).
- [9] Ashofteh Afshin, Bravo Jorge M. Data science training for official statistics: A new scientific paradigm of information and knowledge development in national statistical systems [J]. *Statistical Journal of the IAOS*, 2021, 37(3).
- [10] Rao Zhiwei, Yuan Jie. Data mining and statistics issues of precision and intelligent agriculture based on big data analysis[J]. *Acta Agriculturae Scandinavica, Section B — Soil & Plant Science*, 2021,71(9).
- [11] Yan Li. Research on the Statistical Model System of Each Balanced Game Based on Big Data [J]. *Journal of Physics: Conference Series*, 2021, 1952(4).
- [12] Hao Yuhong, Huang Ying, Feng Yu. Research on the Application of Big Data Technology in Information Statistics Research System [J]. *Journal of Physics: Conference Series*, 2021, 1865(4).
- [13] Wang Yanru. Research on Big Data Integration Method for Investment Statistics Based on Artificial Intelligence Technology [J]. *Journal of Physics: Conference Series*, 2021, 1757(1).
- [14] Zhang Bichun. Analysis on Application Issues of Economic Statistics in the Era of Big Data [J]. *The Frontiers of Society, Science and Technology*, 2020, 2.0(15.0).
- [15] Xia Wu. The Development Trend of Accounting Audit in the Era of Big Data Statistics [J]. *International Journal of Intelligent Information and Management Science*, 2020, 9(6).
- [16] Lifeng Jiang. Big Data Statistical Analysis Model based on Hadoop System [J]. *International Journal of Intelligent Information*, 2020, 9(5).
- [17] S B Yu, Yang X C, Wang H, Wang Z X. Three-dimensional classification and zoning method of surrounding rock and its application based on big data geomechanical information[J]. *IOP Conference Series: Earth and Environmental Science* , 2020, 570(5).
- [18] Aihua Chen. Research on Application of Big Data Analysis in Economic Statistics [J]. *International Journal of Education and Technology*, 2020, 1(3).

- [19] Ran Xu,Zhou Feng,Zhong Ming,Liu Yang,Zhang Jun. Innovative Applications of Patient Experience Big Data in Modern Hospital Management Improve Healthcare Quality[J]. Chinese Medical Sciences Journal,2020,35(4).
- [20] Zhuan Hongjuan. Innovation Research on Hospital Human Resource Performance Management in the Era of Big Data [J]. Heilongjiang Human Resources and Social Security, 2022(11):55-57.
- [21] Shan Tao. Innovation and development of hospital archives management under big data [J]. Heilongjiang Human Resources and Social Security, 2022(12):88-90.
- [22] Shang Wenhui. Analysis of hospital financial management in the era of big data [J]. Business News, 2022(13):29-32.
- [23] Wei Chuanyu. Analysis of key points of hospital informatization construction under the background of big data [J]. Electronic Components and Information Technology, 2022,6(04):92-95+99.
- [24] Deng Jihong. Innovative research on information construction of hospital archives management in the era of big data [J]. Lantai Neiwai, 2022(11):1-3.
- [25] Wu Yongsheng. Visualization analysis of hospital information management in the past 10 years based on CiteSpace [J]. Fujian Medical Journal, 2022, 44(02): 135-137.
- [26] Qu Linlin. An Analysis of the Strategy of Hospital Personnel File Management in the New Era [J]. Science and Technology Vision, 2022(11):53-55.
- [27] Wang Gang. How to further improve the management of hospital publicity files under the background of big data [J]. Economist, 2022(04):250-252.