Research on the application of statistical methods based on big data in the medical and health field

Wenyan Yang

Shandong University of Science and Technology, Qingdao, China

15610552907@163.com

Abstract. In the current era of information, the rise of big data technology has become a key force driving the development of the medical and health field. Statistical methods, as core tools for data analysis, have demonstrated unique advantages in processing vast medical datasets. This paper comprehensively analyzes the various applications of big data statistical methods in the medical and health field, focusing particularly on their practical effects and potential value in disease prediction, epidemiological research, medical resource optimization, and personalized medical services. Through a review of relevant literature and in-depth discussions of multiple cases, this paper reveals how statistical methods play a crucial role in improving diagnostic accuracy and medical service efficiency. It also addresses the challenges faced in actual applications, such as data privacy protection and technical standardization.

Keywords: Big data, statistical methods, medical health, disease prediction, personalized medicine.

1. Introduction

With the rapid development of big data technology, the medical and health field is experiencing an unprecedented transformation. Big data not only provides the ability to handle complex health information but also brings profound challenges and innovative opportunities to traditional medical models. Statistics, as a science of dealing with data relationships and analyzing data structures, has now become a bridge for interpreting these complex medical data. Through precise data processing techniques, disease prediction has become more accurate, epidemic control more effective, resource allocation more reasonable, and immense potential has been shown in personalized medicine. Against the backdrop of global health governance, effectively utilizing statistical methods can not only enhance the quality and efficiency of medical care but also play a crucial role during public health crises. However, as technology rapidly advances, how to use these data responsibly while ensuring individual privacy, and how to enhance the accuracy and reliability of data analysis, remain significant challenges in the medical and health field.

2. Basic Concepts and Developments of Big Data and Statistical Methods

2.1. Definition and Characteristics of Big Data Technology

Big data technology refers to the techniques for processing large-scale, multi-type data collections at high speeds, allowing valuable information to be extracted from vast amounts of data. In the medical

and health field, these data often come from electronic health records, medical imaging, genomic data, patient-reported information, and the internet. The core characteristics of big data include large volume, a wide variety of types, rapid processing speed, and low value density, which make big data technology a key tool for transforming medical and health information [1].

2.2. Application of Statistical Methods in Big Data Analysis
The application of statistical methods in big data analysis in the medical field mainly focuses on data organization, analysis, and interpretation. This includes using descriptive statistics to summarize the basic characteristics of data, employing inferential statistics to generalize from sample data to larger populations, and utilizing predictive models and machine learning algorithms to predict future events or determine correlations between variables. For example, by applying regression analysis, researchers can understand and predict the effects of certain drugs on different patient groups, or identify patient groups with similar disease presentations through cluster analysis.

2.3. Development Trends and Technological Innovations
As technology continuously advances, the development of big data and statistical methods is also accelerating. The integration of cloud computing, the Internet of Things (IoT), and artificial intelligence technologies makes data collection and analysis more efficient and precise. In the medical field, this means faster disease monitoring and alerting, and more accurate design of personalized treatment plans. Furthermore, with enhanced computing capabilities and innovative algorithms, future statistical methods will be able to handle more complex datasets, providing deeper insights, thus promoting more efficient and personalized medical services.

3. Applications of Big Data Statistical Methods in Disease Prediction and Epidemiology

3.1. Disease Risk Assessment Using Big Data
Big data statistical methods play a crucial role in disease risk assessment. By analyzing vast amounts of data on patients' historical health records, lifestyle habits, genetic information, etc., statistical models can identify high-risk groups and predict the likelihood of specific diseases. For example, using logistic regression analysis, researchers can determine which factors are closely associated with chronic diseases such as heart disease and diabetes, thereby providing a basis for early intervention and preventive treatment.

3.2. Data Mining Techniques in Epidemiological Research
Data mining techniques are increasingly used in epidemiological research, especially in the analysis of outbreak and spread patterns of diseases. By mining and analyzing data from social media, medical devices, and online health portals, researchers can monitor disease transmission trends in real-time, identify hotspots of outbreaks, and evaluate the effectiveness of control measures. Additionally, machine learning techniques like random forests and support vector machines are used to predict disease outbreaks and trends, enhancing the precision and speed of responses [2].

3.3. Case Study: Using Statistical Methods to Track the Spread of Epidemics
During the recent COVID-19 pandemic, statistical methods played a crucial role in analyzing and tracking the virus's transmission pathways. Through integrated models and network analysis, researchers were able to map out the virus's transmission networks, identify super-spreaders, and assess the effectiveness of various public health interventions. For instance, using time-series analysis to predict the peaks and declines of the pandemic, thereby providing a scientific basis for the allocation of medical resources and the formulation of public health policies.

4. Applications of Big Data in Medical Resource Optimization and Management

4.1. Data-Driven Models for Medical Resource Allocation

In the area of medical resource optimization, big data statistical methods provide an efficient data-driven decision-making model. These models help medical institutions and policymakers optimize resource distribution by analyzing data on medical service usage, patient geographical distribution, and disease prevalence trends. For example, predictive models forecast future patient numbers and medical needs in a specific area, allowing for the proactive deployment of medical personnel and equipment to manage potential medical pressures.

4.2. Optimization of Medical Costs and Service Efficiency

Big data technology also plays a key role in controlling medical costs and enhancing service efficiency. Through statistical analysis, medical institutions can identify cost-driving factors, assess the cost-effectiveness of different treatment options, and optimize service processes. For instance, by analyzing data from patient treatment processes, it is possible to determine which medical procedures are necessary and which may lead to resource wastage, thereby improving the overall efficiency and quality of medical services.

4.3. Case Study: The Effectiveness of Statistical Models in Hospital Management

In a case study at a comprehensive hospital, the introduction of data analysis and statistical models enabled the hospital's management to more accurately predict daily patient flows and optimize manpower resources in emergency and outpatient departments. Moreover, detailed analysis of medical service processes helped the hospital identify efficient and inefficient service elements, significantly improving service quality and reducing operational costs [3].

5. Personalized Medicine and the Integration with Big Data

5.1. From Data Analysis to Personalized Treatment Plans

Personalized medicine is one of the significant applications of big data technology. By analyzing a patient's genetic information, lifestyle, medical history, and other health-related data, doctors can tailor specific treatment plans for each patient. Statistical methods play a key role in this process by providing precise data analysis, predicting the effectiveness of drugs for specific patients, or identifying potential side effects of treatment methods. This approach not only enhances treatment effectiveness but also significantly reduces medical costs and patient risks.

5.2. Predictive Models Based on Patient Data for Treatment Outcomes

Predictive models developed using statistical methods can accurately forecast treatment outcomes, which is vital for disease management and treatment decisions. For example, by analyzing clinical data and genotypes of cancer patients, predictive models can forecast chemotherapy responsiveness, assisting doctors in choosing the most suitable treatment plan. This method not only enhances the level of treatment personalization but also improves survival rates and quality of life for patients [4].

5.3. Technological Innovations and Future Trends: Intelligent Health Monitoring Systems

With the development of the Internet of Things and artificial intelligence technologies, intelligent health monitoring systems are becoming a significant trend in personalized medicine. These systems continuously monitor a patient's health status, collecting and analyzing various physiological parameters, such as heart rate and blood pressure. By statistically analyzing this data, the systems can promptly alert potential health issues and automatically adjust treatment plans. This not only enhances the efficiency of medical services but also greatly facilitates patients' daily lives.[5]

6. Conclusion

This study, through in-depth exploration of the application of big data and statistical methods in the medical and health field, reveals how these technologies revolutionize modern medical practices, especially in areas like disease prediction, epidemiological research, medical resource optimization, and personalized medical services. Big data not only optimizes the medical decision-making process, enhances the precision and efficiency of treatment plans but also provides robust support for public health management. However, as technology rapidly evolves and its applications expand, ensuring data security and privacy, addressing ethical issues in data analysis, and further enhancing the accessibility and acceptability of analysis techniques remain critical challenges for future research and practice. These challenges also offer new research directions, indicating that the field of medical data science will continue to evolve and bring more innovations.

References

[1]   Vahid S , Gholamreza M , Leila O , et al. An MCDM approach to assessing influential factors on healthcare providers' safe performance during the COVID-19 pandemic: Probing into demographic variables [J]. Journal of Safety Science and Resilience, 2023, 4 (3): 274-283.

[2]   Faye C , David P , Dorothea N . A systematic review of statistical methodology used to evaluate progression of chronic kidney disease using electronic healthcare records. [J]. PloS one, 2022, 17 (7): e0264167-e0264167.

[3]   Health and Medicine - Medical Statistics; Findings from National Cancer Institute (NCI) in the Area of Medical Statistics Described (An Imputation Approach for Fitting Two-part Mixed Effects Models for Longitudinal Semi-continuous Data) [J]. Computer Weekly News, 2020, 364-.

[4]   Alex B , Paul A . Statistical Methods for Healthcare Performance Monitoring[M]. Taylor and Francis;CRC Press: 2016-08-05. DOI:10.1201/9781315372778.

[5]   John D, Robert S, Emily W, et al. Big Data Analytics in Healthcare: Statistical Methods and Applications [J]. International Journal of Medical Informatics, 2022, 145: 104280. DOI:10. 1016/j.ijmedinf.2021.104280.