Scalable Computing: Practice and Experience, ISSN 1895-1767, http://www.scpe.org
© 2025 SCPE. Volume 26, Issues 2, pp. 899–906, DOI 10.12694/scpe.v26i2.4066

A PRECISE HEALTH FOLLOW-UP MANAGEMENT INFORMATION SYSTEM FOR COMMUNITY CHRONIC DISEASES BASED ON BIG DATA ANALYSIS

QINGTIAN MIAO[*]

Abstract. To enhance the efficiency of medical big data utilization, the author introduces a study on a precise health follow-up management information system tailored for community chronic diseases, leveraging advanced big data analytics. The system includes a meticulously designed chronic disease health record management framework, optimized for efficient ETL (Extract, Transform, Load) processes of medical data. This approach aims to minimize indexing overhead, enhance query execution speed and search capabilities, and maximize the use of aggregated computing resources. The system can be decomposed into four modules: ELT, data block creation, index creation, and querying. As an intermediate layer between users and distributed data management systems, this system can provide data upload, query execution mechanisms, and provide indexes to facilitate data search operations. The experimental results show that when the simulation step size is 50t, the maximum amount of data can reach 10 × 104. This method has a much higher retrieval efficiency than heuristic algorithms when processing massive data, further verifying the effectiveness and practicality of the chronic disease health record management system.

Key words: Big data, Chronic diseases, Medical health, Archive management, data retrieval

1. Introduction.

Chronic diseases, due to their long incubation period and slow onset, are often overlooked by people. However, with the increasing desire and pursuit for a better life, people's emphasis on health is becoming more and more important. In the medical field, it's widely recognized that the majority of chronic disease patients are low-risk individuals. Traditional approaches to chronic disease management typically emphasize patient self-care to prevent disease progression. However, the development of a chronic disease health management system utilizing big data offers promising prospects for enhancing the overall management and care standards for these patients [1,2,3].

Big data has revolutionized the current situation of disease and health management, gradually shifting from a traditional doctor centered management model to a patient-centered model. This transformation truly realizes patient-centered health management. In the context of big data, patients in the chronic health management system upload their health sign information data to the cloud platform through smart devices they wear. Health care organizations and community health service institutions can link to the cloud platform server of health management data to obtain various physical health data of designated patients. The disease and health management network built with cloud servers as the core reduces the difficulty of managing diseases with large amounts of data. Through the constructed chronic disease health management system, medical institutions can carry out risk assessment on patients' health information at any time without the regional restrictions of time and space, and for patients who need medical guidance, they can achieve point-to-point health guidance for patients who need medical guidance by means of Internet social platform, telephone, SMS and other communication methods [4]. We provide comprehensive intelligent health management and medical services for patients with long incubation periods and high risk randomization rates through long-term tracking, testing, and risk warning. Big data has promoted the integration of medical institutions at all levels, reducing management costs and communication costs for patients. The implementation of a big data-driven chronic disease health management system has facilitated seamless communication across all levels of medical institutions. It has enabled automated and intelligent connectivity among healthcare providers, significantly enhancing the management and care coordination for chronic disease patients. When patients with chronic diseases change diagnosis and treatment institutions, the current reception medical structure can directly obtain a series of

[*]Institute of Population Health, Faculty of Health & Life Sciences, University of Liverpool, Liverpool, UK (miaoqingtian0528@163.com)

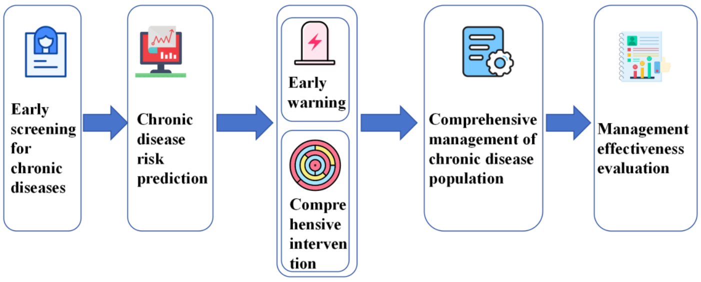
Early screening for chronic diseases
Chronic disease risk prediction
Early warning
Compre hensive interven tion
Comprehensive management of chronic disease population
Management effectiveness evaluation

Fig. 3.1: Schematic diagram of chronic disease management

health signs and treatment data from the cloud server database, greatly reducing the medical cost of patients, eliminating repetitive inspections and laboratory processes, and greatly improving treatment efficiency [5,6].

2. Literature Review.

As medical information technology advances, the growth rate and scope of medical data are rapidly increasing. The advent of the big data era is unlocking fresh opportunities in health and healthcare sectors, paving the way for transformative advancements. The concept of big data itself is not new in the field of healthcare. Healthcare providers not only improve the quality and details of handling large amounts of medical records, especially chronic diseases, but also continuously increase their scale due to technological advancements. Medical big data primarily originates from sources such as medical records, MRI scans, CT scans, health monitoring data, and genomic data. However, these datasets can often be incomplete, biased, and contaminated with noise. Factors like fuzzy information, redundancy, noise, and high dimensionality significantly hinder the effective utilization of medical data [7]. Li, Q. and colleagues aim to gather extensive data from the sports industry and employ data mining techniques along with neural network methods. Their objective is to comprehensively analyze and predict correlations within sports economic data, offering valuable management insights for companies in the sports industry [8]. Son, J. et al. have devised an analytical model for crafting an optimal alert strategy in asthma management. Their research findings offer actionable insights that can benefit patients, healthcare providers, and healthcare companies specializing in technical support [9]. Kiryu, Y. et al. introduced a domain driven design and development example of an artificial intelligence analysis system for clinical practitioners of traditional Chinese medicine to effectively interpret computational data and minimize noise, with the aim of further utilizing medical big data and artificial intelligence analysis [10].

However, traditional indexing techniques often underperform when applied to medical big data. Hence, the author's focus is on studying ETL management of medical data, aiming to reduce index overhead, enhance query execution, improve search performance, and meet the processing demands of medical and health record management systems while optimizing the utilization of computing resources. This research considers both the creation time and size of indexes to minimize overhead. Additionally, a simulation case study evaluates index traversal time and data retrieval time to assess query performance and search efficiency.

3. Method.

3.1. Definition of Chronic Disease Management.

Chronic disease management involves a systematic approach of regular testing, continuous monitoring, evaluation, and comprehensive intervention aimed at managing chronic non-communicable diseases and their risk factors. It encompasses early screening and risk prediction, proactive early warning systems, comprehensive interventions, population-wide management, and effectiveness evaluations. Figure 3.1 illustrates this process. The goal of chronic disease management is to provide holistic, continuous, and proactive care for patients, promoting healthy lifestyles, delaying disease progression, lowering disability rates, improving quality of life, and reducing healthcare costs.

3.2. Characteristics of chronic disease management.

(1) Long term. The development of chronic diseases is mainly caused by exposure to occupational and environmental factors, lifestyle and behavioral patterns. The course of chronic diseases is long, and as the

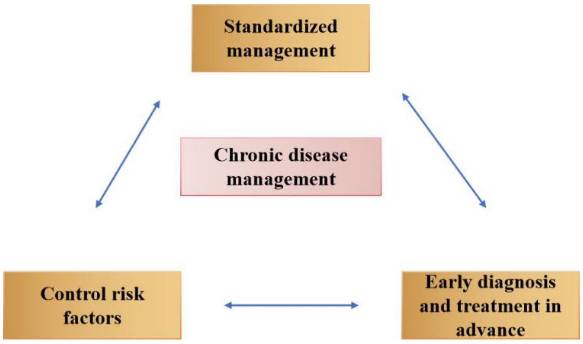
Standardized management
Chronic disease management
Control risk factors
Early diagnosis and treatment in advance

Fig. 3.2: Key Points of Chronic Disease Management

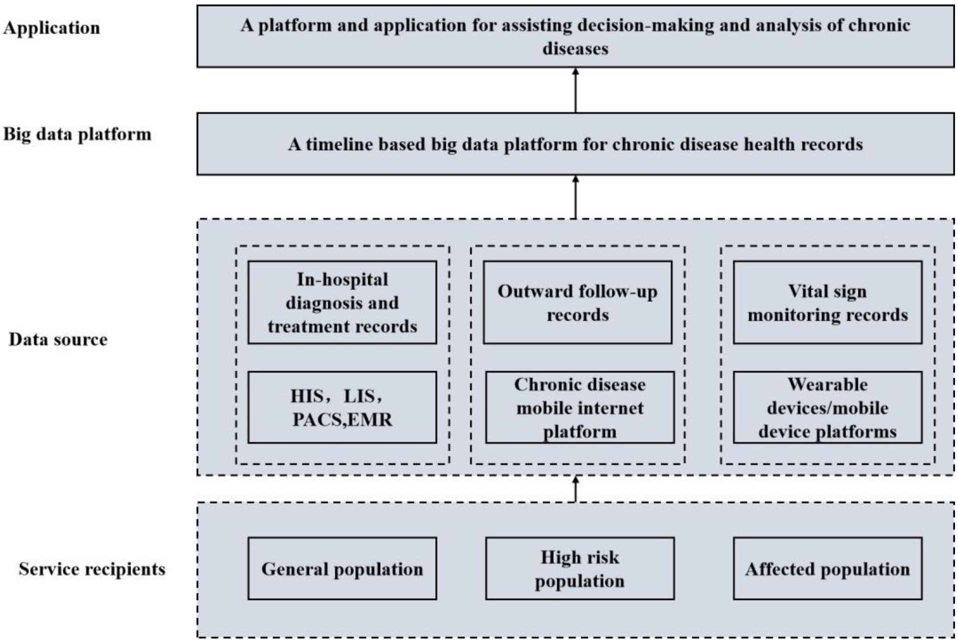
Application
A platform and application for assisting decision-making and analysis of chronic diseases
Big data platform
A timeline based big data platform for chronic disease health records
Data source
In-hospital diagnosis and treatment records
HIS，LIS，PACS,EMR
Outward follow-up records
Chronic disease mobile internet platform
Vital sign monitoring records
Wearable devices/mobile device platforms
Service recipients
General population
High risk population
Affected population

Fig. 3.3: Conceptual model of chronic disease big data

disease progresses, it manifests as progressive functional impairment or disability, causing serious damage to health.

(2) Normativeness. The prevention and treatment of chronic diseases is a continuous process, and chronic disease intervention needs to focus on controlling risk factors, early diagnosis and treatment, and standardized management. See Figure 3.2. Among them, standardized management is the key to the management and treatment of chronic diseases.

3.3. Conceptual model of chronic disease big data . From the perspective of chronic disease management, chronic diseases mainly target three types of population: General population, high-risk population, and diseased population, achieving management of pre hospital, in hospital, and post hospital processes. The conceptual model of chronic disease big data is shown in Figure 3.3.

The management and service targets for chronic diseases include three categories of population: general population, high-risk population, and diseased population. Different intervention measures can be taken to achieve the benefits of chronic disease intervention [11,12].

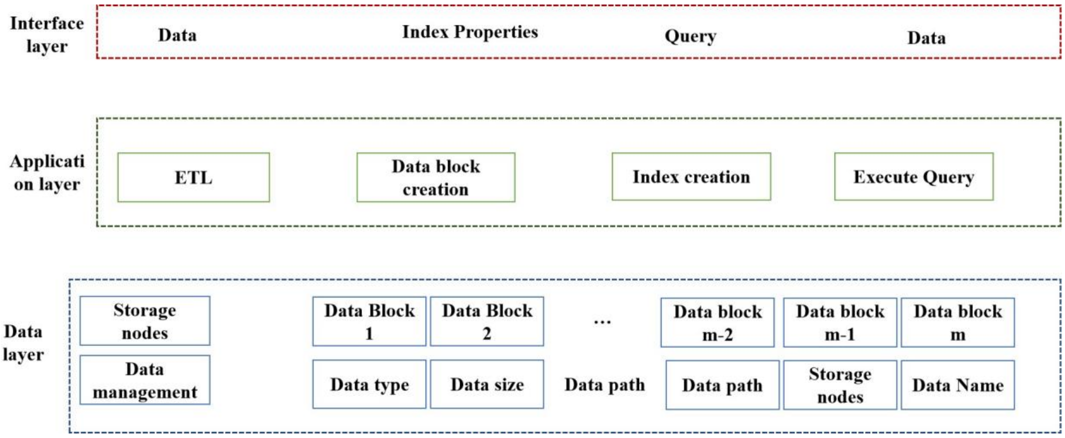

Fig. 3.4: Architecture of Health Record Management System

The main sources of big data on chronic diseases come from hospital information systems (HIS, LIS, PACS, EMR, etc.) within the hospital, as well as monitoring data collected by mobile internet platforms for chronic diseases outside the hospital and various wearable devices. It mainly includes structured, semi-structured, and unstructured multimodal data [13].

The chronic disease health big data platform mainly consists of in-hospital diagnosis and treatment health records, out of hospital disease follow-up records, and long-term monitoring data platforms for vital signs.

- Internal chronic disease diagnosis and treatment health records: mainly include patient diagnosis and treatment information, medication information, medical record information, etc., forming patient diagnosis and treatment health records. The data is sourced from information systems such as HIS, LIS, PACS, and EMR within the hospital.
- Outward follow-up health records: Mainly for the population under chronic disease management, including medication use, chronic disease development, and recovery. The data is mainly collected digitally and automatically through the Internet plus chronic disease platform [14].
- Vital sign monitoring files: mainly for the daily vital sign monitoring data of chronic disease management population, including commonly used indicators such as blood pressure, blood oxygen, respiration, pulse, etc., to achieve continuous and regular collection of vital sign data.

The application of chronic disease assisted decision-making analysis mainly integrates in hospital, out of hospital, and daily monitoring data of chronic disease management population to form patient health record data based on timeline, intelligently realizing the auxiliary diagnosis and treatment decision-making of chronic disease management.

3.4. Architecture. The Health Record Management System is a universal framework for patient health record management and big data indexing, which can be implemented on any distributed system. As an intermediate layer between users and distributed data management systems, this system can provide data upload, query execution mechanisms, and provide indexes to facilitate data search operations. The system architecture is shown in Figure 3.4.

It consists of three layers: (1) User Interface (UI) layer; (2) Application layer; (3) Data layer. Users initiate data upload and index creation operations through the interface layer, while the results of queries and index searches are returned from the application layer to the interface layer for users to browse. The application layer receives data upload and indexing instructions from the interface layer, and calls data block creation to store data in the data layer, while calling index creation to create indexes in the data warehouse. At the application layer, queries are received through the interface and processed by initiating index searches on data stored within the file system. The retrieved data is then delivered back to the user via the interface. Meanwhile, at the data layer, the system manages data blocks and ensures redundancy by storing designated replicas across available storage resources, thereby safeguarding data integrity and security [15].

3.5. System composition and functions. According to the system architecture diagram, the health record management system can be decomposed into four modules: ETL, data block creation, index creation, and querying. The ETL module will extract data from multiple sources, such as chronic disease and health monitoring data sources, and clean, customize, and insert it into the data warehouse; The second module focuses on efficiently organizing data into blocks to enhance storage and retrieval efficiency. The index creation module employs a B-tree structure to store key-value pairs derived from these data blocks. Lastly, the query execution module retrieves the desired data, showcasing enhanced search performance especially beneficial for handling larger datasets. Below is a detailed introduction to the system composition and functions.

ETL (Extract, Transform, Load) is a fundamental database operation process within medical data warehouses. It plays a crucial role in ensuring data accuracy and reliability, as inaccuracies can potentially impact medical decision-making. Data for these processes is sourced from diverse origins, including various operational databases for chronic disease management and health monitoring across different organizational departments, as well as external suppliers. These sources often present data of varying quality, utilizing inconsistent representations, codes, and formats.

Contemporary big data processing systems offer distributed storage solutions that enhance data reliability through replication. Furthermore, each system incorporates its own data segmentation approach, defining the size and placement of data blocks within the data warehouse. When data is split into fixed size data blocks, the last record will face the threat of interruption, resulting in data corruption. Therefore, accessing multiple sites to retrieve corrupt records will increase the overall data loading time. Thus, to minimize access time for result records, it's essential to retrieve each record in its entirety from a single location. That is to say, the pattern of introducing data blocks ensures that the last record in each block is never segmented [16].

During the data block creation process, records are sequentially read and stored until the block reaches its storage capacity. The author suggests adjusting the block size based on the typical record size in the dataset or the default block size of the ETL system. Let's denote the dataset as D containing x records, as described by equation 3.1.

$$D = \sum_{c=1}^{x} record_c \tag{3.1}$$

Among them, $record_c$ represents the c-th record. Furthermore, Algorithm 1 provides the process of creating data blocks. In a distributed data system, the creation of data blocks occurs before data upload and divides the data into smaller blocks. Then load each block into the data warehouse with an adjustable replication factor.

Algorithm 1 Creating Data Blocks

```
Input: block_l=D_, which means the data block capacity limit is $D\_s$; flag_c=true, The identifier for whether the data block capacity is full; block_n=0, indicates the number of initialized data blocks;
While reading data do
If flagtrue and block
Add data to a data block
Else
Load data block, return block,block
blockblock
flagtrue
End if
End while
Load data block, return block,block_n
```

The index creation process occurs after the data is loaded into the data warehouse, and creating an index can shorten the data retrieval time. Furthermore, it is crucial to optimize the index creation process to minimize both the delay between data upload and query execution and the additional space required by the index. Utilizing the B-Tree structure for indexing helps in managing space and time overhead efficiently. During index creation, each attribute record is assigned as a key, with its position serving as the corresponding value in the index structure.

The process of creating an index is shown in Algorithm 2.

Algorithm 2 Create Index

```
Input: Index attribute index_attr, data block block_id; Data block content
block_con;
For index_attr do
    Create an empty B-Tree structure
End for
Valueblock_id
While reading data do
    for index_attr do block_con
        Add<key, Value>to B-Tree
    End for
End while
Store all established B-Tree structures
```

As mentioned earlier, indexes play an important role in big data processing and can lead to increased system overhead. Hence, the search efficiency gained from the index must outweigh the overhead incurred during index creation. This section will showcase the functionality of the system's query execution module. This module conducts index searches to efficiently retrieve both indexed and non-indexed attributes. The reliability and availability of data blocks are ensured by the underlying data layer. As long as the data blocks and indexes stored in the underlying data warehouse remain accessible, the system's query execution module will operate seamlessly. In addition, using this module to complete the query execution and data retrieval process depends on the choice of predicates in the query. The query execution module does not provide services for queries that use non indexed attributes as selection predicates [17-18].

The process of using indexes to execute queries is shown in Algorithm 3.

Algorithm 3 Execution Query

```
Input: Error message err msg; Query; Index; Target name; Target attribute attr; Block position blockreloc;
B-Tree <key,Value>; Data block
If query error then
    Return err_msg
End if
Get name, attr, block_loc from query
If attr has index do
    Key
    Valueblock loc
    For block do
        Get the index of all blocks
        Get the key in the middle index
    End for
End if
```

Initially, the system analyzes incoming queries to validate syntax and parameters, promptly notifying users of any errors like typos or syntax issues. Queries that don't match existing files in the file system are automatically discarded. By accurately parsing the query string, the system ensures the availability of query indexes, resorting to full-scan operations only when necessary. It loads and navigates through indexes to pinpoint record locations, facilitating direct access to data from the file based on the specified query. The subsequent section will present findings highlighting the efficiency of search operations [19].

Table 3.1: System Dataset

Data	Size	Number of items	Number of attributes	Number of data blocks
Medical records	77.2	13362	10	3
MRI	405	121850	15	7
CT	1500	33133	15	25
Health monitoring	6450	2298707	15	103
genome	16110	19291856	36	250

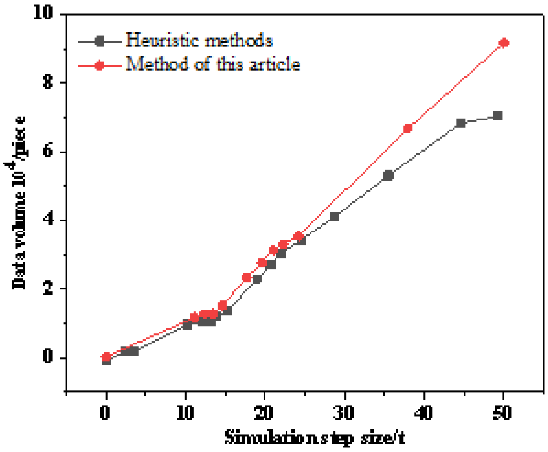

Fig. 4.1: Comparison of Data Classification and Retrieval Efficiency of Different Algorithms

3.6. Simulation analysis. In order to verify the performance of the architecture proposed by the author, a simulation analysis will be conducted using a case study. All modules and related algorithm development environments proposed by the author were developed using Python under the Ubuntu system. The system dataset for evaluating the monitoring and management system is shown in Table 3.1.

The dataset used for system validation includes medical record data, MRI data, CT data, health monitoring data, and genomic data. These datasets are of varying sizes, formats, and contain varying amounts of information. These characteristics of the dataset can affect data loading overhead, indexing overhead, and ultimately search performance.

4. Results and Discussion. Furthermore, compare the retrieval algorithm proposed by the author with the results of heuristic search. The efficiency comparison results of different algorithms for data classification and retrieval are shown in Figure 4.1.

As shown in Figure 4.1, when the amount of data is small, the performance of the two algorithms is almost the same. But as the amount of data continues to increase, the retrieval efficiency of this method continues to improve, and it is far higher than heuristic algorithms. Especially when there are a large number of indexed files, the retrieval rate of heuristic algorithms becomes increasingly low and overwhelming. Therefore, this method is more suitable for the retrieval of massive data. When the simulation step size is 50t, the maximum amount of data can reach 10 × 104 [20].

5. Conclusion. The author proposes a research on a precise health follow-up management information system for community chronic diseases based on big data analysis. The author has conducted research on medical data preprocessing, data modeling, security, data retrieval, etc., and proposed a health record management system with data management and efficient retrieval. The system provides the ability to create multiple indexes on a dataset with minimal index overhead, fast creation and traversal, and less space occupation, thereby improving data management and search performance.

REFERENCES

[1] Qianqian, C., & Lijuan, T. (2023). Study on the management of chronic diseases in american and british community pharmacy, 18(2), 157-164.

[2] Lin, X., Lei, Y., Chen, J., Xing, Z., Yang, T., & Wang, Q., et al. (2023). A case-finding clinical decision support system to identify subjects with chronic obstructive pulmonary disease based on public health data. Tsinghua Science and Technology, 28(3), 525-540.

[3] Sisodia, A., & Jindal, R. (2022). An effective model for healthcare to process chronic kidney disease using big data processing. Journal of Ambient Intelligence and Humanized Computing, 14(10), 1-17.

[4] Zhou, X., Lee, E. W. J., Wang, X., Lin, L., Xuan, Z., & Wu, D., et al. (2022). Infectious diseases prevention and control using an integrated health big data system in china. BMC Infectious Diseases, 22(1), 1-9.

[5] Ed-Daoudy, A., Maalmi, K., & Ouaazizi, A. E. (2023). A scalable and real-time system for disease prediction using big data processing. Multimedia Tools and Applications, 82(20), 30405-30434.

[6] Pakhale, S., Visentin, C., Tariq, S., Kaur, T., Florence, K., & Bignell, T., et al. (2022). Lung disease burden assessment by oscillometry in a systematically disadvantaged urban population experiencing homelessness or at-risk for homelessness in ottawa, canada from a prospective observational study. BMC Pulmonary Medicine, 22(1), 1-9.

[7] Debal, D. A., & Sitote, T. M. (2022). Chronic kidney disease prediction using machine learning techniques. Journal of Big Data, 9(1), 1-19.

[8] Barbanti, P., Egeo, G., Aurilia, C., Fiorentini, G., Proietti, S., & Tomino, C., et al. (2022). The first report of the italian migraine registry (i-graine). Neurological sciences : official journal of the Italian Neurological Society and of the Italian Society of Clinical Neurophysiology, 43(9), 5725-5728.

[9] Li, Q., & Pan, W. T. (2022). Application of multisource big data mining technology in sports economic management analysis. Mathematical Problems in Engineering: Theory, Methods and Applications(Pt.16), 2022.

[10] Son, J., Kim, Y., & Zhou, S. (2022). Alerting patients via health information system considering trust-dependent patient adherence. Information technology & management, 18(7), 719-730.

[11] Kiryu, Y. (2023). Development of a medical big data analysis system utilizing artificial intelligence analytics in clinical pharmacy. YAKUGAKU ZASSHI, 143(6), 501-505.

[12] Shafqat, S., Majeed, H., Javaid, Q., & Ahmad, H. F. (2022). Standard ner tagging scheme for big data healthcare analytics built on unified medical corpora, 2(4), 152-157.

[13] Hulsen, T., Friedecky, D., Renz, H., Melis, E., Vermeersch, P., & Fernandez-Calle, P. (2023). From big data to better patient outcomes. Clinical Chemistry and Laboratory Medicine CCLM , 61(4), 580-586.

[14] Zhou, Y., & Varzaneh, M. G. (2022). Efficient and scalable patients clustering based on medical big data in cloud platform. Journal of Cloud Computing, 11(1), 1-10.

[15] Senhao, C., Yingnan, C., Jingran, G., & Yuwen, C. (2023). Analysis and enlightenment of big data platform for adverse drug reaction supervision in china and the united states, 18(3), 213-220.

[16] Fazel-Najafabadi, A., Abbasi, M., Attar, H. H., Amer, A., Taherkordi, A., & Shokrollahi, A., et al. (2024). High-performance flow classification of big data using hybrid cpu-gpu clusters of cloud environments. Tsinghua Science and Technology, 29(4), 1118-1137.

[17] Kunnumakkara, A. B., Hegde, M., Parama, D., Girisa, S., Kumar, A., & Daimary, U. D., et al. (2023). Role of turmeric and curcumin in prevention and treatment of chronic diseases: lessons learned from clinical trials. ACS Pharmacology And Translational Science, 6(4), 447-518.

[18] Chimezie, R. O. (2023). Health awareness: a significant factor in chronic diseases prevention and access to care. Journal of Biosciences and Medicines, 11(2), 16.

[19] Filippo, A. D., Perna, S., Pierantozzi, A., Milozzi, F., Fortinguerra, F., & Caranci, N., et al. (2022). Socio-economic inequalities in the use of drugs for the treatment of chronic diseases in italy. International journal for equity in health, 21(1), 157.

[20] Sisodia, A., & Jindal, R. (2022). An effective model for healthcare to process chronic kidney disease using big data processing. Journal of Ambient Intelligence and Humanized Computing, 14(10), 1-17.

[21] Lin, X., Lei, Y., Chen, J., Xing, Z., Yang, T., & Wang, Q., et al. (2023). A case-finding clinical decision support system to identify subjects with chronic obstructive pulmonary disease based on public health data. Tsinghua Science and Technology, 28(3), 525-540.

Edited by: Hailong Li
Special issue on: Deep Learning in Healthcare
Received: May 30, 2024
Accepted: Jul 27, 2024