Journal of Multidisciplinary Healthcare

Dovepress
open access to scientific and medical research

Open Access Full Text Article

REVIEW

Bibliometric Analysis of Development Trends and Research Hotspots in the Study of Data Mining in Nursing Based on CiteSpace

Rui Zhang[1,2,*], Yingying Ge[3,*], Lu Xia[4], Yun Cheng[5]

[1]Department of Nursing, Huadong Hospital Affiliated to Fudan University, Shanghai, 200040, People's Republic of China; [2]Department of Nursing, Fudan University, Shanghai, 200433, People's Republic of China; [3]Yijiangmen Community Health Service Center, Nanjing, 210009, People's Republic of China; [4]Day Surgery Unit, Huadong Hospital Affiliated to Fudan University, Shanghai, 200040, People's Republic of China; [5]School of Medicine, The Chinese University of Hong Kong, Shenzhen, 518172, People's Republic of China

*These authors contributed equally to this work

Correspondence: Lu Xia, Day Surgery Unit, Huadong Hospital Affiliated to Fudan University, 221 of Yanan West Road, Jingan District, Shanghai, 200040, People's Republic of China, Tel +86 21-62483180-530401, Email yihaibeilu@126.com; Yun Cheng, School of Medicine, The Chinese University of Hong Kong, Shenzhen, 518172, People's Republic of China, Tel +86 755-23516157, Email cloud2011930@qq.com

Backgrounds: With the advent of the big data era, hospital information systems and mobile care systems, among others, generate massive amounts of medical data. Data mining, as a powerful information processing technology, can discover non-obvious information by processing large-scale data and analyzing them in multiple dimensions. How to find the effective information hidden in the database and apply it to nursing clinical practice has received more and more attention from nursing researchers.
Aim: To look over the articles on data mining in nursing, compiled research status, identified hotspots, highlighted research trends, and offer recommendations for how data mining technology might be used in the nursing area going forward.
Methods: Data mining in nursing publications published between 2002 and 2023 were taken from the Web of Science Core Collection. CiteSpace was utilized for reviewing the number of articles, countries/regions, institutions, journals, authors, and keywords.
Results: According to the findings, the pace of data mining in nursing progress is not encouraging. Nursing data mining research is dominated by the United States and China. However, no consistent core group of writers or organizations has emerged in the field of nursing data mining. Studies on data mining in nursing have been increasingly gradually conducted in the 21st century, but the overall number is not large. Institution of Columbia University, journal of *Cin-computers Informatics Nursing*, author Diana J Wilkie, Muhammad Kamran Lodhi, Yingwei Yao are most influential in nursing data mining research. Nursing data mining researchers are currently focusing on electronic health records, text mining, machine learning, and natural language processing. Future research themes in data mining in nursing most include nursing informatics and clinical care quality enhancement.
Conclusion: Research data shows that data mining gives more perspectives for the growth of the nursing discipline and encourages the discipline's development, but it also introduces a slew of new issues that need researchers to address.
Keywords: data mining, nursing, bibliometric analysis, global trends, hotspots

Introduction

The medical and health fields have steadily moved into the big data era due to the acceleration of information.[1,2] Big data is the term for information assets that are high volume, speed, and diversity that require specialized technologies and analytical techniques to be valuable.[3] Traditional database retrieval and statistical techniques are insufficient to meet our current needs for data extraction and analysis.[4] In light of this, data mining technology is introduced as a fresh approach to data analysis, helping us explore more of the information hidden behind the data.[5]

The process of extracting potentially useful information and knowledge from a large amount of imprecise, complex, hazy, and random practical application data is known as data mining.[6] Data mining is an interdisciplinary field combining

Received: 11 January 2024
Accepted: 4 April 2024
Published: 10 April 2024

the application of mathematical science, statistics, artificial intelligence, and machine learning to find connections between variables from massive data sets.[7] Trends, associations, meaningful patterns, anomalies, and features of interest are all discovered through these linkages.[8] Data mining offers distinct benefits in clinical big data research, such as assisting medical personnel in forecasting, diagnosing, and treating diseases, hence improving service quality and minimizing costs.[9]

As a practical discipline, nursing is vital to the medical field and holds a distinct place in it. Nursing big data refers to huge amounts of data connected to nursing and health.[10] These nursing-related big data were recorded in large electronic repositories. Mined it can control the dynamic changes of nursing as a whole, discover patterns hidden in data, help make nursing decisions, and provide great opportunities for nursing management, thus significantly improving patient care and management practices.[11,12] Chiang et al surveyed and cluster analyzed the symptoms of patients with systemic lupus erythematosus to help healthcare professionals clarify the focus and direction that should be addressed in clinical care.[13] Utilizing a big data management platform based on data mining technology to timely and accurately assess the vital signs as well as physiological functions of tumor patients, so as to formulate personalized diagnosis and treatment, consulting services, follow-up plans, or clinical care can maximize the improvement of the quality of life of end-stage patients.[14,15] However, as research into data mining in nursing grows, it is becoming increasingly challenging for academics to determine the most recent research hotspots as well as future trends in the area.

Currently, the literature study on the use of data mining in nursing frequently concentrates on a single type of data mining technique,[16] or a limited part of the area.[17] More relevant research is an integrated literature review on the application of data mining in nursing quality management.[18–20] These related studies can't clearly help nursing researchers to understand the current research hotspots and future research directions of data mining in the nursing field. Bibliometric analysis is an interdisciplinary discipline that statistically analyzes all knowledge carriers using statistical and mathematical techniques.[21] Through graphical depiction, it may assist us in rapidly comprehending the development context and frontier hotspots of connected academic subjects, even though it is not able to fully display all the nuances of the topic.[22,23] As a result, this study uses a bibliometric approach based on CiteSpace to conduct an in-depth analysis of data mining research applied to the nursing field over the last two decades, with the goal of revealing the current research status, hotspots, and trends in the field from a visual perspective, which will provide new ideas and clues for future related research work.

CiteSpace is a Java-based information visualization program, created by Dr. Chaomei Chen and his team at Drexel University in Philadelphia, Pennsylvania, USA's School of Information Science and Technology, in the beginning of 2004.[24] It is an interactive analytical tool that makes visualization tasks in science mapping possible through the use of data mining algorithms, visual analytic techniques, and bibliometrics.[25] It has recently emerged as a feature and significant information visualization tool for information analysis.

The purpose of this study was to conduct bibliometrics and visualization analysis of nursing data mining research in the past two decades. The following five questions are discussed in this paper: (i) what are the overall publication trends of nursing data mining research around the world? (ii) Which countries or regions are dominant in nursing data mining? (iii) Which institutions, journals and authors are most influential in nursing data mining research? (iv) What are research hotspots in nursing data mining? (v) What is the future development trend of nursing data mining, and what suggestions can be put forward to scholars and decision-makers?

Methods

Data Acquisition

The data for this investigation were gathered from the Web of Science Core Collection (WOSCC). This database is a typical citation database and a widely used database for bibliometric analysis, containing literature of sufficient size to reflect the current status of research in a certain topic.[26] To eliminate bias caused by database upgrades, on August 31, 2023, all literature was searched and downloaded. Following the completion of the search, two researchers worked independently to exclude irrelevant material and extract data, and if there was dispute, a third researcher was invited to discuss it until a consensus was established.

Search Strategy

All data were searched on August, 31, 2023, the data retrieval strategy was as follows: (i) Topic = data mining AND Topic = nurse OR nursing. (ii) Document type = Article OR Review Article. (iii) Publication date (custom year range) = January, 1, 2002 - August, 31, 2023. (iv) Language = English. A total of 279 were obtained. After screening these articles for eligibility using the title, abstract, and full text, 194 papers were found to be eligible.

Bibliometric Visualization and Analysis

In this study, version6.2.R6 of CiteSpace was used for all visual analyses. After data acquisition, the data set was exported to CiteSpace for further analysis. We set the overall time span from January 2002 to August 2023, Slice length: 2 years, g-index k=25, choice Pathfinder and Pruning the merged network, then ran CiteSpace to the generate networks. All of the other necessary parameters were set to the default values provided by CiteSpace.

The number of articles published each year was calculated using Microsoft Excel 2020 program based on the number of articles screened after retrieval, and a bar chart was created. We have created seven visualization maps: a co-occurrence network map of countries, institutions, authors, and keywords and a burst of keywords, a clustering of keywords, a timeline of keywords.

In the co-occurrence and clustering map, different colors represent different years. The frequency of keyword is represented by the circle size. The higher the frequency, the larger the circle. The thicker the line between the nodes, the closer the two keywords work together. Centrality is the degree of nodes that is part of the path connects any pair of nodes in the network, with greater than 0.1 being the key node. The purple rings in the outside indicate that these indicators have greater centrality.[27] In the burst map, the blue line depicts the time interval, while the red line depicts the time when a keyword burst.

Results

Time Characteristics Analysis

We analyzed the final obtained 194 literatures related to nursing data mining. As shown in Figure 1, the number of relevant articles published each year has been steadily increasing, from 0 in 2002 to 32 in 2022, with tone significant jump in 2021. The number of published articles available by August 31, 2023 is 19.

Three research phases can be used to classify the total number of articles on nursing data mining. The first phase, known as the nascent development stage (2003–2009), is characterized by fewer noteworthy research outputs—fewer than two publications annually, on average. The second stage, known as the sluggish development stage (2010–2020), sees an average of 8.6 articles produced annually and a total of 103 papers released. During this time, the amount of literature fluctuates and increases. The third stage, known as the rapid growth period (2021–2023), is characterized by an

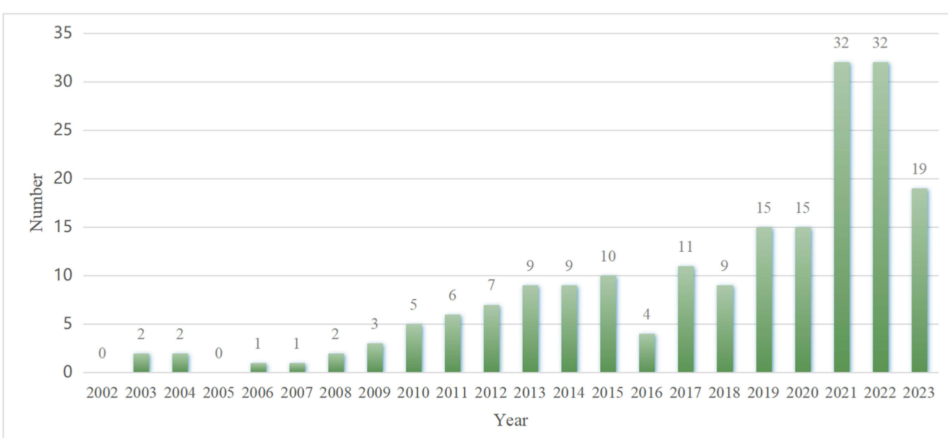

Figure 1 Trend chart of the number of articles published on data mining in nursing.

Table 1 Top 5 Journal Published Analysis (2002–2024)

No.	Journal Title	IF	Amount	Country	Research Area (JCR Partition)
1	Cin-computers Informatics Nursing	1.300	14	USA	Computer Science, Interdisciplinary Applications (Q4) Medical Informatics (Q4) Nursing (Q4)
2	Journal of Healthcare Engineering	3.822	7	Ireland	Health Care Sciences & Services (Q2)
3	International Journal of Medical Informatics	4.9	6	USA	Computer science, information systems (Q2) Health care science & services (Q1) Medical Informatics (Q2)
4	Nursing Research	2.5	5	USA	Nursing (Q2)
5	PLOS One	3.7	5	USA	Multidisciplinary science (Q2)

average of over 30 publications annually. This finding indicates that nursing data mining research is still in its early phases, but it is growing and becoming one of the current hotspots.

Publication of Active Journals

Table 1 lists the top 5 journals that published the largest number of papers regarding data mining in nursing from 2002 to 2023. *Cin-computers Informatics Nursing* published about 14 papers, ranking the first. Overall, the specific subject scope comprises Computer Science, Interdisciplinary Applications, Medical Informatics, Nursing, Health Care Sciences & Services, Computer science, information systems, Multidisciplinary science, and so on. In the listed top 5 journals, four journals from the United States and one from Ireland. The journal of highest impact factor is *International Journal of Medical Informatics*, nearly 4.9.

Distribution of Countries/Regions

As is shown in Figure 2, the USA (75) ranks first in the publication quantity, which is followed by China (52) and Japan (14). The top 10 prolific countries/regions in this research field are shown in Table 2. The centralities of USA, China,

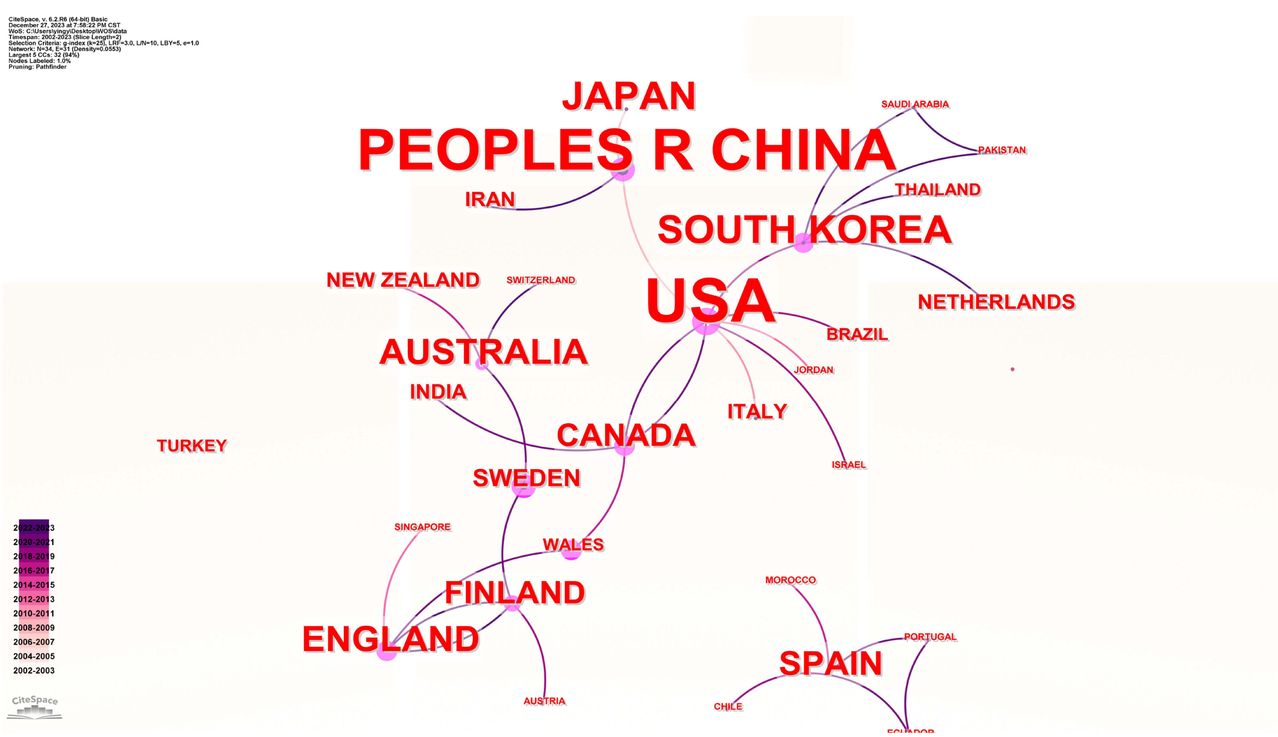

Figure 2 Co-countries' network (2002–2023).

Table 2 Top 10 Countries/Regions on Data Mining in Nursing

NO.	Country	Publications	Centrality
1	USA	75	0.70
2	China	52	0.16
3	Japan	14	0.00
4	South Korea	13	0.31
5	Australia	10	0.45
6	England	10	0.16
7	Spain	9	0.02
8	Canada	7	0.53
9	Finland	7	0.36
10	Sweden	4	0.23

South Korea, Australia, England, Canada, Finland and Sweden are greater than 0.1, indicating that these eight countries/regions are the most influential countries in nursing data mining research.

Distribution of Institutions

As Figure 3 illustrates, Columbia University (8) and University of Illinois System (7) ranks first in the publication quantity, which is followed by Harvard University (7). The top 10 Institutions prolific in this research field are shown in Table 3. All of the institutions' centralities are less than 0.1, indicating that a stable circle of collaboration among global research institutions has not yet been formed.

Influential Authors

As seen in Figure 4, Diana J Wilkie (4), Muhammad Kamran Lodhi (4), Yingwei Yao (4) all have the most publications. The top 10 authors prolific in this research field are shown in Table 4. The centralities of all authors are less than 0.1, indicating that there was a lack of influential authors.

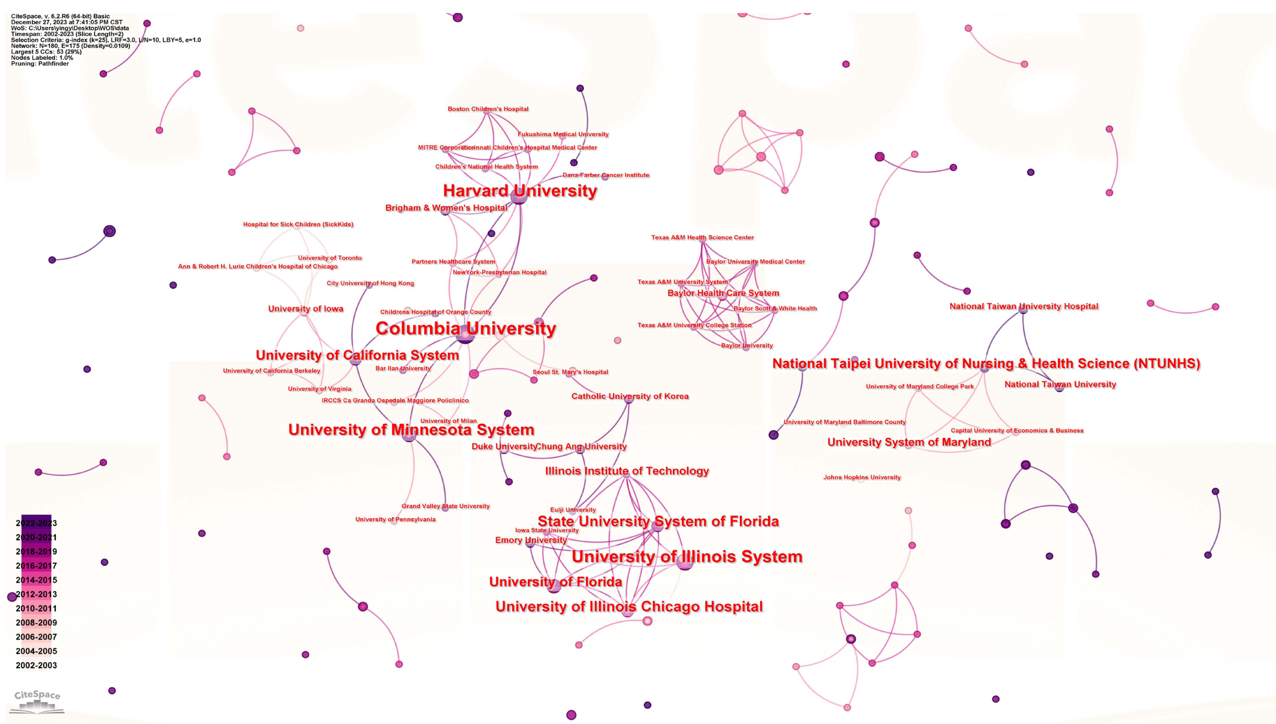

Figure 3 Co- Institutions' network (2002–2023).

Table 3 Top 10 Institutions on Data Mining in Nursing

NO.	Institution	Country	Publications	Centrality
1	Columbia University	USA	8	0.02
2	University of Illinois System	USA	7	0.00
3	Harvard University	USA	7	0.02
4	University of Minnesota System	USA	6	0.02
5	Gachon University	Korea	5	0.00
6	State University System of Florida	USA	5	0.00
7	University of Illinois Chicago Hospital	USA	5	0.00
8	University of Florida	USA	4	0.00
9	National Yang Ming Chiao Tung University	China	4	0.00
10	National Taipei University of Nursing & Health Science	China	4	0.00
11	University of California System	USA	4	0.02

Analysis of Hotspots and Trends

High-Frequency Keywords Analysis

As Figure 5 reveals, high-frequency keywords are crucial markers of the research hotspots. Apart from the most basics phrases in this field of research, Data mining (n=53) and care (n=20), the top three high-frequency keywords were electronic health records (n=17), text mining (n=17) and management (n=12). According listed in Table 5, adverse events (centrality = 0.43), data mining (centrality = 0.38), risk (centrality = 0.27), classification (centrality = 0.22), and electronic health records (centrality = 0.14) were the top five terms with high centrality.

Keywords with Citation Bursts

Burst detection technology was utilized to investigate how research trends changed over time. The term "burst" refers to a certain period of time during which there is an abrupt shift in frequency, and "burst" became the focus of attention during that time. The top 10 keywords with the most significant citation bursts are shown in Figure 6. The blue line

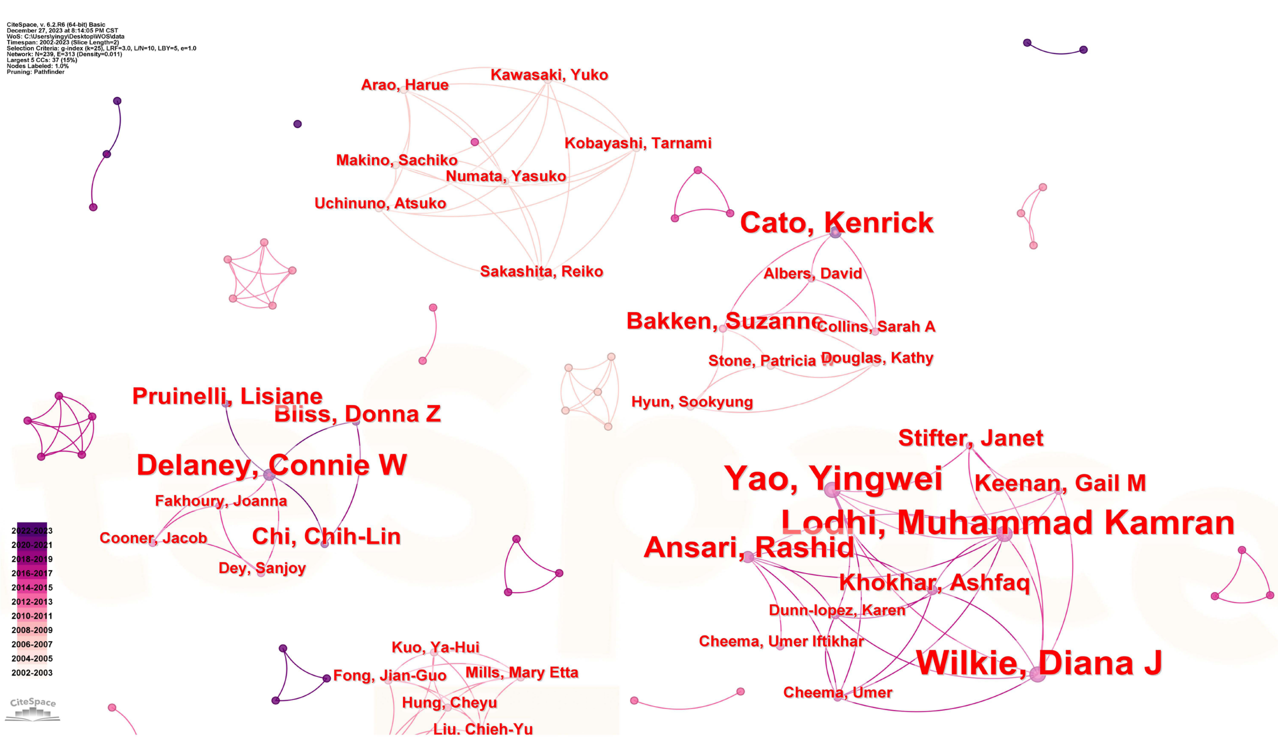

Figure 4 Co-Authors' network (2002–2023).

Table 4 Top 10 Authors on Data Mining in Nursing

NO.	Author	Publications	Centrality
1	Wilkie, Diana J	4	0.00
2	Lodhi, Muhammad Kamran	4	0.00
3	Yao, Yingwei	4	0.00
4	Delaney, Connie W	3	0.00
5	Ansari, Rashid	3	0.00
6	Cato, Kenrick	3	0.00
7	Chu, Woei-Chyn	2	0.00
8	Pasupathy, Kalyan S	2	0.00
9	Stifter, Janet	2	0.00
10	Keenan, Gail M	2	0.00

depicts the time interval, while the red line depicts the time when a keyword burst. In the study of data mining in nursing, the keywords with large mutation values that lasted until 2023 are highlighted: machine learning.

It was vital to note that "machine learning" "natural language processing" "text mining" and "nursing informatics" were recently developing hot issues. Potential trends and research areas for upcoming data mining in the nursing field are suggested by these themes.

Keywords with Cluster Analysis

Two indicators based on network structure and clustering clarity: Modularity Q and Weighted Mean Silhouette S. It can be used as a benchmark against which we can assess the mapping effect. In general, the Modularity Q falls within the range of [0, 1]. Q > 0.3 indicates that the community structure is significantly split. Clustering is efficient and persuasive when the S value is 0.7; if it is greater than 0.5, clustering is typically deemed acceptable. The visualization map obtained N = 254, E = 527 (density = 0.0164), the Modularity Q score was 0.8231, the Mean Silhouette score was 0.9509, as presented in Figure 7.

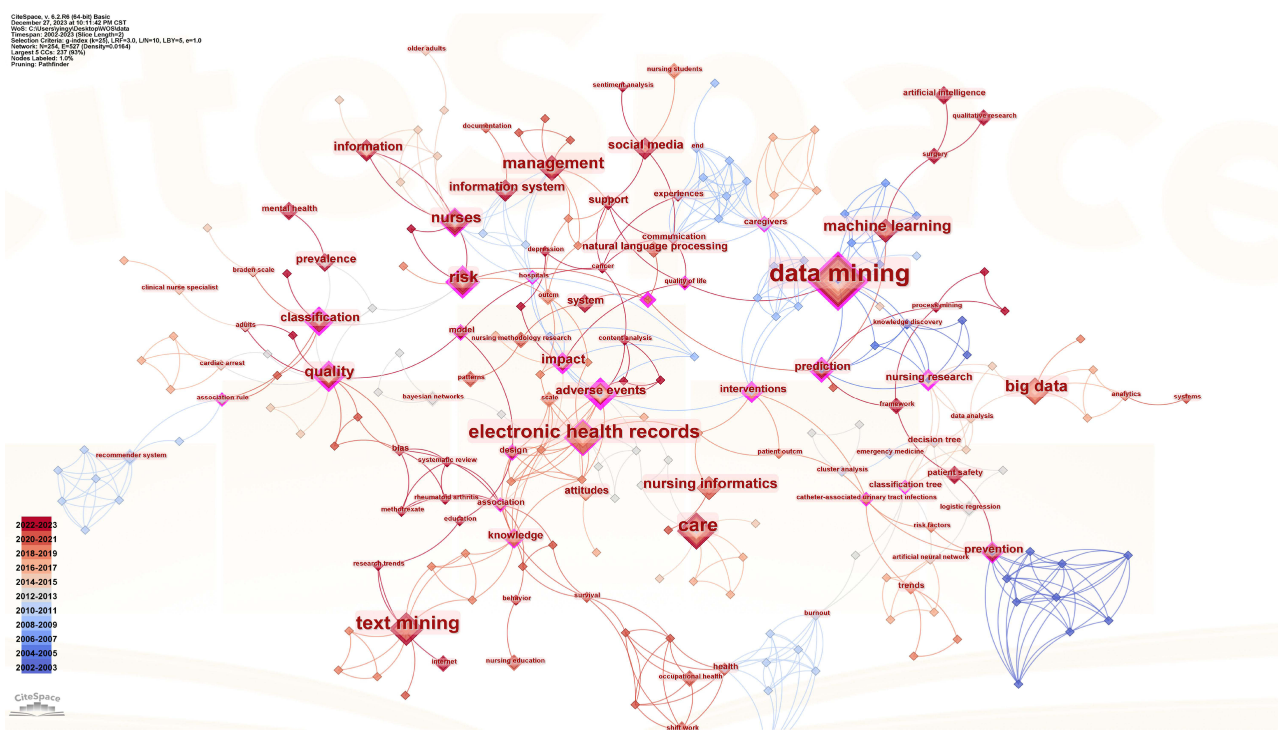

Figure 5 Keyword co-occurrence map (2000–2022).

Table 5 High-Frequency Keywords on Data Mining in Nursing (Frequency ≥ 5)

NO.	Keyword	Frequency	Centrality	NO.	Keyword	Frequency	Centrality
1	Data mining	53	0.38	11	Nursing informatics	8	0.01
2	Care	20	0.07	12	Impact	7	0.22
3	Electronic health records	17	0.14	13	Adverse events	6	0.43
4	Text mining	17	0.06	14	Classification	6	0.22
5	Management	12	0.07	15	Information	6	0.01
6	Big data	11	0.07	16	Information system	6	0.01
7	Risk	11	0.27	17	Social media	6	0.03
8	Machine learning	10	0.04	18	Prediction	5	0.3
9	Nurses	10	0.16	19	Prevalence	5	0.01
10	Quality	10	0.25	20	Prevention	5	0.13

The summary of the largest eleven clusters is listed in Table 6. There are two major research themes related to data mining in nursing. The first theme is nursing informatics (for example, #1 informatics, #6 recommender system, and #7 knowledge base). The other one is about quality improvement of clinical care (for example, #4 palliative care, #8 prevention, #9 disorders of consciousness, and #10 nurse-patient assignments).

As Figure 8 clearly shows, a timeline view of keywords shows how high-frequency keywords have changed over time. As search phrases for this study, "#0 nursing research" and "#5 data mining" first surfaced at the start of the temporal evolution of clustering. During the nascent development stage, clusters #1 informatics, #2 text mining, and #4 palliative care initially surfaced. The majority of the residual clusters initially emerged in the sluggish development stage.

Discussion

Research Status of Data Mining in Nursing

According to published literature, the number of papers on data mining in nursing has gradually increased since 2003, which might be connected to the fact that research experts are investing more in data mining research.[28] During the nascent development stage, which had been expanding slowly until then. The publishing of a paper on Dynamo systems by Amazon in 2007 and the publication of Nature in 2008 on big data which showed that how big data is gradually gaining research experts traction.[29,30] When the number of data mining related studies enter the sluggish development

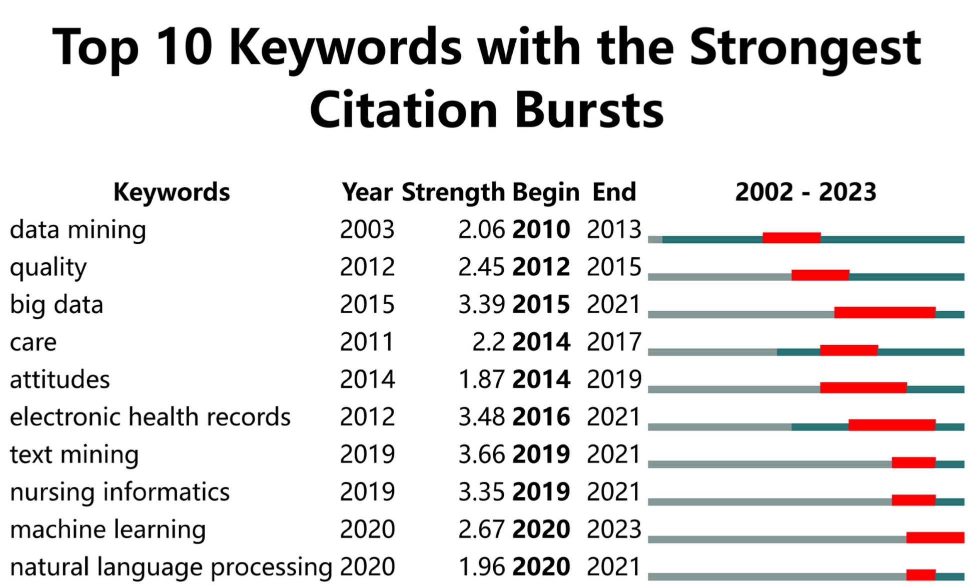

Figure 6 Burst map of keywords (2002–2023).

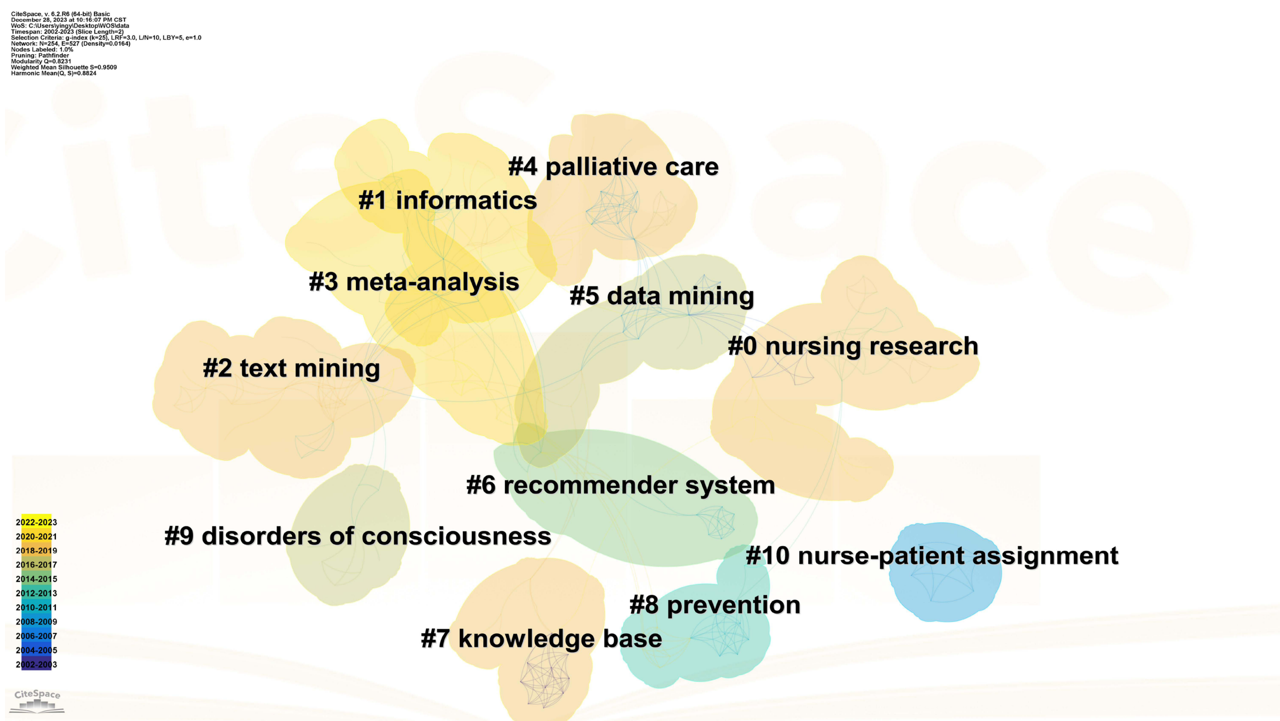

Figure 7 Clustering map of keywords (2002–2023).

stage, the goal of the Nursing Knowledge Big Data Science Initiative has been put forward, which is to create a plan for getting "sharable and comparable" nursing data and to make sure big data techniques are quickly adopted throughout the nursing disciplines.[31]

With the onset of the era of digital intelligence,[32] a growing amount of nursing research has focused on big data in recent years.[20,33] These have immediately resulted in a notable upsurge in data mining research based on big data concepts. Put it another way, research related to the application of data mining in nursing has entered a phase of rapid

Table 6 Summary of the Largest Eleven Clusters

Cluster ID	Size	Silhouette	Mean (Year)	Log-Likelihood Ratio
#0 nursing research	32	0.897	2014	Nursing research; knowledge discovery; data analysis; classification tree; big data
#1 informatics	31	0.878	2015	Informatics; quality improvements; nursing notes; ovarian cancer; impact
#2 text mining	30	0.981	2018	Text mining; research trends; medication error; data mining; critical care
#3 meta-analysis	23	0.996	2017	Meta-analysis; systematic review; complementary and alternative medicine; east Asian herbal medicine; rheumatoid arthritis
#4 palliative care	20	0.956	2014	Palliative care; social media; civility; grief; big data mining
#5 data mining	19	0.955	2010	Data mining; nursing informatics; hospital management; care providers; big data science
#6 recommender system	17	0.975	2013	Recommender system; association rule; nursing care; healthcare integrated information systems; correlation
#7 knowledge base	16	0.963	2008	Knowledge base; expert nurses; KDD; Scientometrics; data mining issues
#8 prevention	14	0.953	2014	Prevention; work schedule; work-family conflict; random forests; shiftwork
#9 disorders of consciousness	8	0.967	2014	Disorders of consciousness; nursing intervention; undergraduate; assessment; anatomy and physiology
#10 nurse-patient assignment	5	1	2009	Nurse-patient assignment; simulation; nurse workload; data mining; text mining

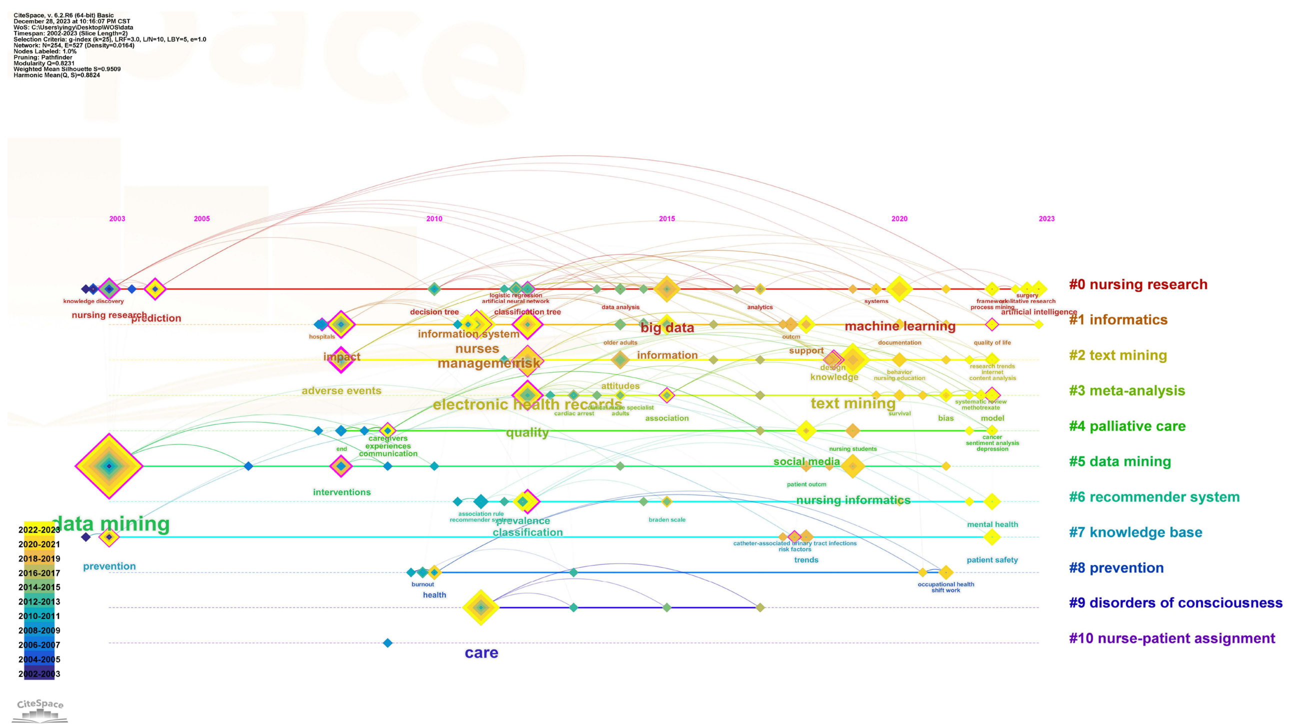

Figure 8 Timeline map of keywords (2002–2023).

development. However, just 194 publications were published in total, demonstrating that nursing data mining research is still in its early stages. It might be because the global information nurse has not been widely publicized.

The Journal Citation Report presents the first Journal Impact Factor list for journal management. The JCR can be of reference value within the overall framework, taking into account other subjective and objective factors.[34] Most of the research articles related to data mining in nursing were published in the Q1 and Q2 region related to health, indicating the overall importance of this research direction. However, based on the top five journals' impact factors, it appears that there is still plenty of opportunity for related research to expand in the future.

Nursing data mining research is conducted practically everywhere, although it has mostly concentrated on nations with effective health systems, such as the United States, China and Korea. However, due to low economic and information technology levels, the growth of nursing data mining may be limited in some developing countries/nations. More transnational assistance and collaboration will be required in the future to enhance the growth of global nursing data mining research.

The direction of a subject area is determined by the institutions that conduct important research in that area. Our findings indicate that the universities in the United States have the greatest influence and research capacity in data mining in nursing and are the core research institutions in the field. It might be due to the fact that America was the first country to train information nurses. Collaboration among researchers may improve not just the productivity of scientific knowledge, but also profits, wealth, and economic growth.[35] However, the finding of this study indicates that a core group of writers for nursing data mining research has yet to emerge.

Research Hotspots of Data Mining in Nursing

The majority of the data mined at the moment comes from electronic health record, social media, and several important systems or databases, such as the Federal Adverse Event Reporting System. An electronic health record is a data repository for a subject of care's health and healthcare, in which all information is maintained on electronic medium.[36] Electronic health records are mandatory and legally meaningful in many countries.[37] The volume of healthcare data managed and stored electronically will continue to expand as the digital transformation progresses.[38] These data are reliable data mining sources since they include a large amount of potentially useful information. Social media including Twitter, Blogs, Instagram, Communities and so on. Twitter is a prominent social media platform that allows anyone, including people and government officials, to communicate brief messages (tweets), with about 500 million tweets each day.[39] It may represent a large amount of data that is communicated in real time.[40]

Text mining is the process of extracting information from unstructured text.[41] Unstructured text refers to narrative text, such as nursing records, which include a wealth of valuable nursing information but suffer from a lack of formal expression, resulting in a wide range of expression with the same mean, semantic error and so on. This circumstance makes the raw data more complicated, making data mining more challenging.

Machine learning methodology innovation has become a related research spot in contemporary research because nursing researchers want to be able to employ more and more appropriate research approaches to handle the difficult research challenges of today. The most often used data mining algorithms in nursing are association analysis, cluster analysis, artificial neural network, decision tree. Association analysis and cluster analysis are descriptive algorithms that identify unknown patterns or relationships in data by assessing the similarity of objects.[42] Association analysis analyzes all variables by setting the minimum support and confidence threshold, which can clearly describe the interrelationship between one thing and other things, and obtain potential and valuable rules.[43] Qin Li' team use association analysis found that the most important stroke risk factor is atrial fibrillation.[44] Cluster analysis is the process of dividing a collection into several similar objects and discovering new relationships by creating classifications of homogeneous groups.[45] It aims to reveal relationships and classifications of homogeneous groups that are not otherwise obvious in the data set. Through cluster analysis, the Oh WO Team discovered the features of each of the symptom clusters linked with moyamoya illness, helping in the development of therapies for the symptom characteristics of adolescent moyamoya disease.[46] The prediction algorithms artificial neural network and decision tree can derive prediction rules (classification/prediction models) from (training) data and apply the rules to unpredictable/unclassified data.[42] Artificial neural network simulates the information processing process of the brain with a widely interconnected structure and effective learning mechanism. Each node in the network can be regarded as a neuron and can store and process information. Each node on the neural network can process information and output it to other nodes, which receive it and output it again until all the work of the neural network is completed, and finally output the result.[47] Tingting Lee's team through artificial neural network analysis of the available data in hospital information system, a prediction model was established, and it was found that the use of fall-related nursing assessment, anti-psychotics and diuretics might be the related factors of patients' falls.[48] Decision tree is an analysis method to judge the feasibility on the basis of the known probability of the occurrence of various situations.[49] It uses a dendritic to explain the influence of each variable on the prediction model.[50] The team of Philip Zachariah used decision tree model and neural network to predict urinary tract infections in hospitalized patients.[51]

Furthermore, a review of papers revealed that there are various other data mining algorithms, such as support vector machine, Bayesian classification, logistic regression and so on.[52,53] However, none of them are listed in high-frequency words, for the following reasons: Data mining in nursing began late, and the algorithms utilized are mostly basic and traditional, with limited usage of newer algorithms. Newer algorithms will be necessary in the future to analyze larger, more complicated data in order to improve data processing speed and accuracy. In conclusion, the data mining algorithm is a vital tool for sorting through large amounts of information. Because it is not limited by the assumptions of traditional statistical methods, they may solve complex issues and handle vast amounts of data quickly.[54] When utilized correctly, it can help clinical nursing personnel discover nursing patterns and make reasonable predictions quickly and accurately. Data mining algorithms are tools for extracting massive data hidden rules with potential value, and play an important role in nursing risk prediction, clinical decision support, disease development prediction, accurate nursing intervention implementation, and improving the quality of nursing management and nursing education level.

The arrival of the big data era has opened up a wealth of data sources for nursing research. How to carry out natural language processing is now a research spot. The proper use of this data can aid in the discovery of rules and the advancement of nursing. At the same time, data can be also mined from numerous systems at once. Multiple data sources increase the diversity of word expression, which complicates the following specification step of data mining. As a result, figuring out how to provide unified and consistent international general data is critical. At present, some researchers have processed natural language using language modeling, word embedding, and two phrase mining algorithms (Text-Rank and NC-Value) to discover target terms in nursing notes. This approach, in comparison to human judgment, can swiftly extract high-quality phrases from a huge number of nursing notes.[55] Furthermore, this was the first study to assess automatic phrase identification systems on nursing notes, and it could serve as a model for future text mining. In the future, it will be required to support the creation of international nursing professional words in order to ease the standardization issues created by inconsistent big data expression.

Research Trends of Data Mining in Nursing

The main research objectives of data mining include revealing hidden knowledge in databases, investigating disease-causing factors, developing clinical nursing intervention programs, developing early prediction and early warning models, and improving clinical care quality. Quality improvement of clinical care is the most significant aim, and related research on the application of data mining in nursing will be a research trend. Hsiu-lan Li's team used the decision tree approach to evaluate 2062 end-of-life medical records, found that history of pressure injuries, non-cancer diagnosis, excretion, activity/mobility, and skin condition/ circulation are the predictive factors in pressure injuries for patients at the end of life.[54] This information can assist nursing staff in predicting the presence of pressure injuries, communicating with caregivers and patients towards the end of life, and developing care goals. To improve disease care and nursing human resource management, the Park JI team used different data sources and data mining approaches to discover parameters related with hospital acquired catheter-associated urinary tract infections.[56] Xia Li's team examined the data interaction and integration of intelligent nursing clinical decision support systems using nursing big data and data mining technology, and built a big data platform based on a data warehouse on this foundation, promoting the application and development of data-driven intelligent nursing decision support systems.[57] Currently, medical data mining is primarily used in the prediction of disease early warning models, the exploration of prognostic factors in cancer patients, the derivation of a dietary pattern and so on.[38] We can see data mining in medical research scope is wide. However, data mining research in nursing is restricted in scope, with few studies on nursing education, psychological nursing, or dietary nursing. To maximize the use of nursing big data, it is advised that future nursing data mining research be conducted from many perspectives and at multiple levels.

Nursing informatics, the creation and improvement of nursing-related systems based on data mining algorithms are the research hotspots, which are also data mining trends in nursing. The rise of nursing informatics is an unavoidable consequence of the Big Data era. Nursing informatics refers to the integration of nursing science, computer science, and information science to manage and communicate data, information, knowledge and wisdom in nursing practice through the use of information structures, information processes and information technology to support decision-making by consumers, patients and providers in all roles and environments.[58] Data mining is closely related to nursing informatics, and the process of data mining demands nurses gather the necessary computer and information knowledge reserves. Therefore, only when an information nurse has such knowledge can she better grasp and use massive data to promote the development of nursing. The construction and improvement of a data-mining -based nursing system is beneficial to supporting nurses in making good judgments, as well as standardizing nursing behavior, continuous monitoring, and real-time control.[59] For example, in a review of dietary management for individuals with stress damage, Sandra W. Citty's team analysis revealed that optimization of electronic health record systems can enhance management, monitoring and evaluation of nutritional therapies.[3]

Conclusion

This study used bibliometrics and visualization approaches to demonstrate the features of the knowledge distribution of nursing data mining research based on the gathered literature data by data mining during the previous two decades. Meanwhile, the status, hotspots and trends of nursing data mining research were investigated.

The following issues in data mining technology for nursing research need to be investigated further. (i) There are currently few studies on nursing data mining throughout the world, and no collaborative network of authors or institutions has been established. (ii) The ability to deeply mine nursing data and extract effective information is not enough. (iii) The current state of nursing data mining research is insufficient, with few studies on nursing education, psychological nursing, and nutritional nursing. (iv) Many nursing data lack the representation of worldwide standards words, and the data is too fragmented and unstructured, reducing direct data usage and making integration difficult. In light of these issues, it is advised that future studies expand transnational collaboration and exchange across institutions, forming a multi-way collaborative research. Improve the information technology level of nursing personnel by strengthening professional training for information nurses. An international general information nursing platform and international common nursing terms should be built to extend the supply of nursing data, minimize the difficulty of data standardization, and enhance the data use rate.[60,61]

Association analysis, cluster analysis, artificial neural network and decision tree are common data mining algorithms in the field of nursing, which are widely used in nursing clinical research. The application of association analysis in nursing clinic can find the hidden relationship, reduce the risk of disease and shorten the time of disease prognosis by analyzing its influencing factors and improving it.[62] Cluster analysis is more used in the analysis of clinical symptom clusters of diseases in nursing research to identify the characteristics of different clusters, so as to formulate different nursing interventions and achieve precision nursing.[63,64] Artificial neural network has more advantages in the application of clinical nursing management and nursing education. By constructing an accurate risk assessment model, it can optimize the loopholes in talent management and talent training, so as to improve the quality of management and education.[65,66] Decision trees can be used to build predictive evaluation models to support clinical decision-making, thereby reducing medical costs and saving medical resources.[67,68]

The information technology level of nursing staff can be improved further by incorporating a health informatics major into nursing school, enhancing professional training for information nurses, and creating dedicated positions for information nurses in clinics. To create an international common information nursing platform and terminology, enhance the supply of nursing data, reduce the difficulty of data standardization, and increase the utilization rate of nursing clinical data. Improve the decision support and health recommendation systems based on the health information system, as well as clinical nursing decision-making abilities and health education effectiveness. In general, data mining opens up new avenues for the ongoing development and expansion of nursing disciplines, as well as encouraging the development of disciplines to adapt to the information age.

Funding

This research was funded by Shanghai Hospital Development Center Foundation (No. SHDC12023612), Shanghai Municipal Science and Technology Commission Program (No. 16411951200).

Disclosure

The authors declare no conflicts of interest in this work.

References

1. Hulsen T, Friedecký D, Renz H, et al. From big data to better patient outcomes. *Clin Chem Lab Med*. 2023;61(4):580–586. doi:10.1515/cclm-2022-1096
2. Flanagan J. Editorial: big data: nursing's moment. *Int J Nurs Knowl*. 2023;34(3):169. doi:10.1111/2047-3095.12438
3. Citty SW, Cowan LJ, Wingfield Z, et al. Optimizing nutrition care for pressure injuries in hospitalized patients. *Adv Wound Care*. 2019;8(7):309–322. doi:10.1089/wound.2018.0925
4. Drawz PE, Archdeacon P, McDonald CJ, et al. CKD as a model for improving chronic disease care through electronic health records. *Clin J Am Soc Nephrol*. 2015;10(8):1488–1499. doi:10.2215/cjn.00940115
5. Zeng QT, Fodeh S. Clinical data mining. *Comput Biol Med*. 2015;62:293. doi:10.1016/j.compbiomed.2015.05.014
6. Wu WT, Li YJ, Feng AZ, et al. Data mining in clinical big data: the frequently used databases, steps, and methodological models. *Mil Med Res*. 2021;8(1):44. doi:10.1186/s40779-021-00338-z
7. Lan K, Wang DT, Fong S, et al. A survey of data mining and deep learning in bioinformatics. *J Med Syst*. 2018;42(8):139. doi:10.1007/s10916-018-1003-9
8. Saberi-Karimian M, Khorasanchi Z, Ghazizadeh H, et al. Potential value and impact of data mining and machine learning in clinical diagnostics. *Crit Rev Clin Lab Sci*. 2021;58(4):275–296. doi:10.1080/10408363.2020.1857681
9. Islam MS, Hasan MM, Wang X, et al. A systematic review on healthcare analytics: application and theoretical perspective of data mining. *Healthcare*. 2018;6(2):54. doi:10.3390/healthcare6020054
10. Zhu R, Han S, Su Y, et al. The application of big data and the development of nursing science: a discussion paper. *Int J Nurs Sci*. 2019;6(2):229–234. doi:10.1016/j.ijnss.2019.03.001
11. Clancy TR, Gelinas L. Knowledge discovery and data mining: implications for nurse leaders. *J Nurs Adm*. 2016;46(9):422–424. doi:10.1097/nna.0000000000000369
12. Dunbar P, Keyes LM, Browne JP. Determinants of regulatory compliance in health and social care services: a systematic review using the Consolidated Framework for Implementation Research. *PLoS One*. 2023;18(4):e0278007. doi:10.1371/journal.pone.0278007
13. Westwell-Roper C, Williams KA, Samuels J, et al. Immune-related comorbidities in childhood-onset obsessive compulsive disorder: lifetime prevalence in the obsessive compulsive disorder collaborative genetics association study. *J Child Adolesc Psychopharmacol*. 2019;29(8):615–624. doi:10.1089/cap.2018.0140
14. Li D, Huang Q, Zhang W, et al. Effects of routine collection of patient-reported outcomes on patient health outcomes in oncology settings: a systematic review. *Asia Pac J Oncol Nurs*. 2023;10(11):100297. doi:10.1016/j.apjon.2023.100297

15. Cai T, Yuan C. Insight into the dyadic driving force of theory and big data to advance cancer care. *Cancer Nurs*. 2023;46(3):248–249. doi:10.1097/ncc.0000000000001218
16. Atay E, Bahadır Yılmaz E, Atay M. Analysis of dementia research trends in nursing using text mining approach. *Psychogeriatrics*. 2024. doi:10.1111/psyg.13097
17. Jing X. The unified medical language system at 30 years and how it is used and published: systematic review and content analysis. *JMIR Med Inform*. 2021;9(8):e20675. doi:10.2196/20675
18. van Velzen M, de Graaf-Waar HI, Ubert T, et al. 21st century (clinical) decision support in nursing and allied healthcare. Developing a learning health system: a reasoned design of a theoretical framework. *BMC Med Inform Decis Mak*. 2023;23(1):279. doi:10.1186/s12911-023-02372-4
19. Fu J, Li C, Zhou C, et al. Methods for analyzing the contents of social media for health care: scoping review. *J Med Internet Res*. 2023;25:e43349. doi:10.2196/43349
20. Bernardi FA, Alves D, Crepaldi N, et al. Data quality in health research: integrative literature review. *J Med Internet Res*. 2023;25:e41446. doi:10.2196/41446
21. Hood WW, Wilson CS. The literature of bibliometrics, scientometrics, and informetrics. *Scientometrics*. 2001;52(2):291–314. doi:10.1023/A:1017919924342
22. Wu Z, Guo K, Luo E, et al. Medical long-tailed learning for imbalanced data: bibliometric analysis. *Comput Methods Programs Biomed*. 2024;247:108106. doi:10.1016/j.cmpb.2024.108106
23. Dol J, Campbell-Yeo M, Leahy-Warren P, et al. Bibliometric analysis of published articles on perinatal anxiety from 1920 to 2020. *J Affect Disord*. 2024;351:314–322. doi:10.1016/j.jad.2024.01.231
24. Chen C. Searching for intellectual turning points: progressive knowledge domain visualization. *Proc Natl Acad Sci USA*. 2004;101(Suppl 1):5303–5310. doi:10.1073/pnas.0307513100
25. Luo H, Cai Z, Huang Y, et al. Study on pain catastrophizing from 2010 to 2020: a bibliometric analysis via CiteSpace. *Front Psychol*. 2021;12:759347. doi:10.3389/fpsyg.2021.759347
26. Zhang XL, Zheng Y, Xia ML, et al.. Knowledge domain and emerging trends in vinegar research: a bibliometric review of the literature from WoSCC. *Foods*. 2020;9(2):2719.
27. Liu S, Sun YP, Gao XL, et al. Knowledge domain and emerging trends in Alzheimer's disease: a scientometric review based on CiteSpace analysis. *Neural Regen Res*. 2019;14(9):1643–1650. doi:10.4103/1673-5374.255995
28. de la Torre Díez I, Cosgaya HM, Garcia-Zapirain B, et al. Big data in health: a literature review from the year 2005. *J Med Syst*. 2016;40(9):209. doi:10.1007/s10916-016-0565-7
29. DeCandia G, Hastorun D, Jampani M, et al. Dynamo: Amazon's highly available key-value store. *ACM SIGOPS Operat Systems Rev*. 2007;41(06):205–220. doi:10.1145/1323293.1294281
30. Buxton B, Hayward V, Pearson I, et al. Big data: the next Google. Interview by Duncan Graham-Rowe. *Nature*. 2008;455(7209):8–9. doi:10.1038/455008a
31. Delaney CW, Weaver C. 2018 Nursing knowledge big data science initiative. *Comput Inform Nurs*. 2018;36(10):473–474. doi:10.1097/cin.0000000000000486
32. Yuan C. Data quotient: the future competence of oncology nurses. *Cancer Nurs*. 2021;44(4):261–262. doi:10.1097/ncc.0000000000000961
33. Seibert K, Domhoff D, Bruch D, et al. Application scenarios for artificial intelligence in nursing care: rapid review. *J Med Internet Res*. 2021;23(11):e26522. doi:10.2196/26522
34. Krampl A. Journal Citation Reports. *J Med Libr Assoc*. 2019;107(02):280–283. doi:10.5195/jmla.2019.646
35. Bornmann L, Leydesdorff L. Topical connections between the institutions within an organisation (institutional co-authorships, direct citation links and co-citations). *Scientometrics*. 2015;102(1):455–463. doi:10.1007/s11192-014-1425-1
36. Negro-Calduch E, Azzopardi-Muscat N, Krishnamurthy RS, et al. Technological progress in electronic health record system optimization: systematic review of systematic literature reviews. *Int J Med Inform*. 2021;152:104507. doi:10.1016/j.ijmedinf.2021.104507
37. Aguirre RR, Suarez O, Fuentes M, et al. Electronic health record implementation: a review of resources and tools. *Cureus*. 2019;11(9):e5649. doi:10.7759/cureus.5649
38. Subrahmanya SVG, Shetty DK, Patil V, et al. The role of data science in healthcare advancements: applications, benefits, and future prospects. *Irish J Med Sci*. 2022;191(4):1473–1483. doi:10.1007/s11845-021-02730-z
39. Karami A, Zhu M, Goldschmidt B, et al.. COVID-19 vaccine and social media in the US: exploring emotions and discussions on Twitter. *Vaccines*. 2021;9(10):1059.
40. Thorlton J, Catlin AC. Data mining for adverse drug events: impact on six learning styles. *Comput Inform Nurs*. 2019;37(5):250–259. doi:10.1097/cin.0000000000000513
41. Labrosse J, Lam T, Sebbag C, et al. Text mining in electronic medical records enables quick and efficient identification of pregnancy cases occurring after breast cancer. *JCO Clin Cancer Inform*. 2019;3:1–12. doi:10.1200/cci.19.00031
42. Alonso SG, de la Torre-Díez I, Hamrioui S, et al. Data mining algorithms and techniques in mental health: a systematic review. *J Med Syst*. 2018;42(9):161. doi:10.1007/s10916-018-1018-2
43. Dorling L, Carvalho S, Allen J, et al. Breast cancer risk genes - association analysis in more than 113,000 women. *N Engl J Med*. 2021;384(5):428–439. doi:10.1056/NEJMoa1913948
44. Li Q, Zhang Y, Kang H, et al. Mining association rules between stroke risk factors based on the Apriori algorithm. *Technol Health Care*. 2017;25(S1):197–205. doi:10.3233/thc-171322
45. Dalmaijer ES, Nord CL, Astle DE. Statistical power for cluster analysis. *BMC Bioinf*. 2022;23(1):205. doi:10.1186/s12859-022-04675-1
46. Oh WO, Shim KW, Yeom I, et al. Features and diversity of symptoms of moyamoya disease in adolescents: a cluster analysis. *J Adv Nurs*. 2021;77(5):2319–2327. doi:10.1111/jan.14723
47. Sabir Z, Ben Said S, Al-Mdallal Q. An artificial neural network approach for the language learning model. *Sci Rep*. 2023;13(1):22693. doi:10.1038/s41598-023-50219-9
48. Lee TT, Liu CY, Kuo YH, et al. Application of data mining to the identification of critical factors in patient falls using a web-based reporting system. *Int J Med Inform*. 2011;80(2):141–150. doi:10.1016/j.ijmedinf.2010.10.009

49. Schöning V, Hammann F. How far have decision tree models come for data mining in drug discovery?. *Expert Opin Drug Discov*. 2018;13 (12):1067–1069. doi:10.1080/17460441.2018.1538208
50. Luo X, Wen X, Zhou M, et al. Decision-tree-initialized dendritic neuron model for fast and accurate data classification. *IEEE Trans Neural Netw Learn Syst*. 2022;33(9):4173–4183. doi:10.1109/tnnls.2021.3055991
51. Zachariah P, Sanabria E, Liu J, et al. Novel strategies for predicting healthcare-associated infections at admission: implications for nursing care. *Nurs Res*. 2020;69(5):399–403. doi:10.1097/nnr.0000000000000449
52. Park JI, Bliss DZ, Chi CL, et al. Knowledge discovery with machine learning for hospital-acquired catheter-associated urinary tract infections. *Comput Inform Nurs*. 2020;38(1):28–35. doi:10.1097/cin.0000000000000562
53. Christopher JJ, Nehemiah HK, Kannan A. A Swarm Optimization approach for clinical knowledge mining. *Comput Methods Programs Biomed*. 2015;121(3):137–148. doi:10.1016/j.cmpb.2015.05.007
54. Li HL, Lin SW, Hwang YT. Using nursing information and data mining to explore the factors that predict pressure injuries for patients at the end of life. *Comput Inform Nurs*. 2019;37(3):133–141. doi:10.1097/cin.0000000000000489
55. Korach ZT, Yang J, Rossetti SC, et al. Mining clinical phrases from nursing notes to discover risk factors of patient deterioration. *Int J Med Inform*. 2020;135:104053. doi:10.1016/j.ijmedinf.2019.104053
56. Park JI, Bliss DZ, Chi CL, et al. Factors associated with healthcare-acquired catheter-associated urinary tract infections: analysis using multiple data sources and data mining techniques. *J Wound Ostomy Continence Nurs*. 2018;45(2):168–173. doi:10.1097/won.0000000000000409
57. Xia LX, Wang R, Lin Z. Research on the construction of the data platform of the intelligent nursing decision support system from the perspective of big data. *Chin Digital Med*. 2022;17(03):55–62. doi:10.3969/j.issn.1673-7571.2022.3.012
58. Wang J, Gephart SM, Mallow J, et al. Models of collaboration and dissemination for nursing informatics innovations in the 21st century. *Nurs Outlook*. 2019;67(4):419–432. doi:10.1016/j.outlook.2019.02.003
59. Liu QF, Zhao CS, Ying YP, et al. Development and practice of incontinence associated dermatitis information intelligent prevention decision support system. *Chin Nurs Manage*. 2021;21(10):1549–1553.
60. Gaudet-Blavignac C, Foufi V, Bjelogrlic M, et al. Use of the systematized nomenclature of medicine clinical terms (snomed ct) for processing free text in health care: systematic scoping review. *J Med Internet Res*. 2021;23(1):e24594. doi:10.2196/24594
61. Austin RR, Lu SC, Geiger-Simpson E, et al. Evaluating systemized nomenclature of medicine clinical terms coverage of complementary and integrative health therapy approaches used within integrative nursing, health, and medicine. *Comput Inform Nurs*. 2021;39(12):1000–1006. doi:10.1097/cin.0000000000000764
62. Kamal AN, Clarke JO, Oors JM, et al. The role of symptom association analysis in gastroesophageal reflux testing. *Am J Gastroenterol*. 2020;115 (12):1950–1959. doi:10.14309/ajg.0000000000000754
63. Mathew A, Tirkey AJ, Li H, et al. Symptom clusters in head and neck cancer: a systematic review and conceptual model. *Semin Oncol Nurs*. 2021;37(5):151215. doi:10.1016/j.soncn.2021.151215
64. Fanelli S, Bellù R, Zangrandi A, et al. Managerial features and outcome in neonatal intensive care units: results from a cluster analysis. *BMC Health Serv Res*. 2020;20(1):957. doi:10.1186/s12913-020-05796-0
65. Zhang X, Zhang L. Risk prediction of sleep disturbance in clinical nurses: a nomogram and artificial neural network model. *BMC Nurs*. 2023;22 (1):289. doi:10.1186/s12912-023-01462-y
66. Ladstätter F, Garrosa E, Moreno-Jiménez B, et al. Expanding the occupational health methodology: a concatenated artificial neural network approach to model the burnout process in Chinese nurses. *Ergonomics*. 2016;59(2):207–221. doi:10.1080/00140139.2015.1061141
67. Vera-Salmerón E, Domínguez-Nogueira C, Romero-Béjar JL, et al.. Decision-tree-based approach for pressure ulcer risk assessment in immobilized patients. *Int J Environ Res Public Health*. 2022;19(18):11161.
68. Flaks-Manov N, Shadmi E, Yahalom R, et al. Identification of elderly patients at risk for 30-day readmission: clinical insight beyond big data prediction. *J Nurs Manag*. 2022;30(8):3743–3753. doi:10.1111/jonm.13495