scientific reports

Check for updates

OPEN

Systematic analysis of healthcare big data analytics for efficient care and disease diagnosing

Sulaiman Khan[1], Habib Ullah Khan[1✉] & Shah Nazir[2]

Big data has revolutionized the world by providing tremendous opportunities for a variety of applications. It contains a gigantic amount of data, especially a plethora of data types that has been significantly useful in diverse research domains. In healthcare domain, the researchers use computational devices to extract enriched relevant information from this data and develop smart applications to solve real-life problems in a timely fashion. Electronic health (eHealth) and mobile health (mHealth) facilities alongwith the availability of new computational models have enabled the doctors and researchers to extract relevant information and visualize the healthcare big data in a new spectrum. Digital transformation of healthcare systems by using of information system, medical technology, handheld and smart wearable devices has posed many challenges to researchers and caretakers in the form of storage, minimizing treatment cost, and processing time (to extract enriched information, and minimize error rates to make optimum decisions). In this research work, the existing literature is analysed and assessed, to identify gaps that result in affecting the overall performance of the available healthcare applications. Also, it aims to suggest enhanced solutions to address these gaps. In this comprehensive systematic research work, the existing literature reported during 2011 to 2021, is thoroughly analysed for identifying the efforts made to facilitate the doctors and practitioners for diagnosing diseases using healthcare big data analytics. A set of rresearch questions are formulated to analyse the relevant articles for identifying the key features and optimum management solutions, and laterally use these analyses to achieve effective outcomes. The results of this systematic mapping conclude that despite of hard efforts made in the domains of healthcare big data analytics, the newer hybrid machine learning based systems and cloud computing-based models should be adapted to reduce treatment cost, simulation time and achieve improved quality of care. This systematic mapping will also result in enhancing the capabilities of doctors, practitioners, researchers, and policymakers to use this study as evidence for future research.

Healthcare around the world is under high pressure due to limiting financial resources, over-population, and disease burden. In this modern technological age the healthcare paradigm is shifting from traditional, one-size-fits-all approach to a focus on personalized individual care[1]. Additionally, the healthcare data is varying both in type and amount. The healthcare providers are not only dealing with patient's historical, physical and namely information, but they also deal with imaging information, labs, and other digital and analogue information consists of ECG, MRI etc. This data is voluminous, varying in type and formats, and of differing structure. These are the capabilities of Big Data to handle not only different types of and forms of data, but can handle 10 V structure including volume, variety, venue, varifocal, varmint, vocabulary, validity, volatility, veracity and velocity. Thus, the doctors facing an increasing burden of rising patient numbers coupled with progressively less time to spend with each patient. In other words, we are facing more patients, more data, and less time.

Big data has significantly attracted the researchers to explore different research fields including healthcare, banking, imaging, smart cities, internet of things (IoT) based smart applications, tracking and transportation system etc.[2]. Software engineers constantly develops new applications for patient's health and well-being. Both government and non-government organizations develop infrastructure using big data analytics for improved decision making capabilities of both doctors and managers[3]. It was recorded that 80% increase in big data is due to cloud sources, big data analytics, mobile technology and social media technologies[4]. A number of research articles proposed using big data analytics in varying domains especially in healthcare such as Kumar et al.[5] proposed a cognitive technology-based healthcare evaluations system using big data analytics. Chen et al.[6] presented an

[1]Department of Accounting and Information Systems, College of Business and Economics, Qatar University, Doha, Qatar. [2]Department of Computer Science, University of Swabi, Swabi, Pakistan. ✉email: habib.khan@qu.edu.qa

nature portfolio

1

intelligent healthcare application for brain hemorrhage detection using Big Data analytics and machine learning (ML) techniques. Smart health appointment system is developed by Liang and Zhao using big data analytics is[7].

Some researchers explored big data analytics in healthcare domain in different ways. They presented survey papers and review papers to understand the meanings of big data analytics in healthcare such as Galetsi and Kasaliasi performed a review of healthcare big data analytics[8] while Lindell defined big data analytics in terms of accounting and business perspectives[9]. Alharthi proposed a review article on healthcare challenges facing in Saudi Arabia by performing analysis of the available literature[10]. Lee et al.[11] presented a survey paper to explore the applications and challenges of healthcare big data analytics. From the literature it is concluded that multiple new applications are developed for big data analysis. Review and survey papers are presented to outline the published literature, but most of these papers are region specific or limited to a few numbers of papers. On the other side systematic review process formulate multiple research question and identifies keywords to explore the available literature from different angles. Systematic analysis of the available literature is presented in many fields like PMIPv6 domain[12], in smart homes[13], navigation assistants[14], and many others, but there is no significant work reported on systematic analysis for healthcare big data domain to find the gaps in the available literature and suggest future research directions.

The inspirational point that led us to pursue this systematic analysis was the pervasive and ubiquitous nature of big data. Efficient management and timely execution are the dire needs of big data, to extract enriched information regarding a certain problem of interest[15]. Many factors involved behind this systematic research work, but the most eminent reasons are:

i. The exiting research reported on big data does not provide significant information about the key features that should be considered to integrate both structured and unstructured big data in healthcare domain. The pervasiveness of big data features challenging the researchers in pursuing research in this specialized domain. The underlying research on finding the key features will not only help in integrating big data in healthcare domain, but it will also assist in findings new gateways for future research directions.

ii. Digital transformation of healthcare systems after the integration of information system, medical technology and other imaging systems have posed a big barrier for the research community in the form of a vast amount of information to deal with. While the over-population, limited data access, and disease burdens have restricted the doctors and practitioners to check more patients in a limited time. So, finding a suitable model that can efficiently process healthcare big data to extract information for a certain disease symptoms will not only helps the practitioners to suggest accurate medication and check more patients in timely manners, but it will open future research directions for the industrialists and policymakers to develop optimal healthcare big data processing models.

iii. Accurate disease diagnosing by processing of gigantic amount of data, especially a plethora of types of data, within an interested processing domain is a key concern for both researchers and practitioners. Developing an efficient model that can accurately diagnose a certain by classifying images or other historical details of patients will not only helps the doctors to diagnose disease in timely manner and suggest medicine accordingly, but it will encourage the researchers and developers to develop an accurate disease identification model.

The remaining research paper of the paper is organized as follows. Section 2 of the paper outlines the related work reported in the proposed field. Section 3 presents the research framework followed for this systematic research work. Quality assessment is detailed in Sect. 4. Section 5 outlines the discussion on findings of the proposed systematic research work. Section 6 provides the limitations of this systematic study traced by the conclusion and future work in Sect. 7 of the paper.

Literature review

From the last few decades, we experienced an unprecedented transformation of traditional healthcare systems to digital and portable healthcare applications with the help of information systems, medical technology and other imaging resources[16]. Big data are radically changing the healthcare system by encouraging the healthcare organizations to embrace extraction of relevant information from imaginary data and other clinical records. This information will produce high throughput in terms of accurate disease diagnosing, plummeting treatment cost increase availability. In data visualization context the term 'big data', is firstly introduced in 1997[17], posed an ambitious and exceptional challenge for both policy-makers and doctors with special emphasis on personalized medicine. Nonetheless, data gathering moves faster than both data analysis and data processing, emphasizing the widening gap between the rapid technological progress in data acquisition and the comparatively slow functional characterization of healthcare information. In this regard, the historical information (phonotypical and other genomic information) of an individual patient form electronic health records (EHR) are becoming of critical importance. Figure 1 represents the primary sources of big data.

Significant research work has been reported in the domains of healthcare big data analytics. To process this vast amount of information in timely manner and identify someone's health condition based on his her is more difficult. Researchers proposed numerous applications to address this problem such as; Syed et al.[18] proposed a machine learning-based healthcare system for providing remote healthcare services to both diseased and healthy population using big data analytics and IoT devices. Venkatesh et al.[19] developed heart disease prediction model using big data analytics and Naïve Bayes classification technique. Kaur et al.[20] suggested a machine learning (ML) based healthcare application for disease diagnosing and data privacy restrictions. This model works by considering different aspects like activity monitoring, granular access control and mask encryption. Some researchers presented review and survey papers to outline the recent published work in a specific directions such as Patel
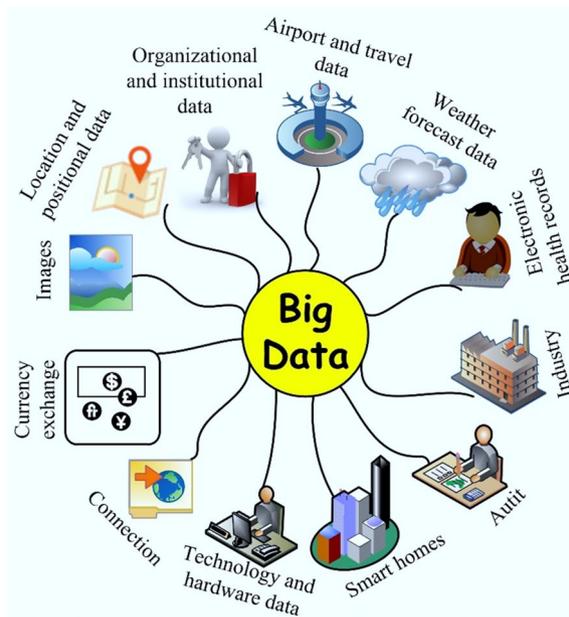
Figure 1. Main steps of the research protocol.

and Gandhi reviewed the literature for identifying the machine learning approaches proposed for healthcare big data analytics[21]. Rumbold et al.[22] reviewed the literature for find the research work reported for diabetic diagnosing using big data analytics.

From the above discussions, it is worth mentioning that most of the researchers and industrialists gave significant attention towards the development of new computational models or surveyed the literature in a specific research direction (heart disease detection, diabetes detection, storage and security analysis etc.), but no significant research work is reported to systematically analyze the literature with different perspectives. To address this problem, this research work presents a systematic literature review (SLR) work to analyze the literature reported in healthcare big data analytics domain. This systematic analysis will not only find the gaps in the available literature but it will also suggest new directions of future research to explore.

Research framework

Systematic literature reviews and meta-analysis has gained significant attention and became increasingly important in healthcare domain. Clinicians, developers and researchers follow SLR studies to get updated about new knowledge reported in their fields[23,24], and they are often followed as a starting point for preparing basic records. Granting agencies mostly requires SLR studies to ensure justification of further research[25], and even some healthcare journals follows this direction[26]. Keeping these SLR applications in mind the proposed systematic analysis is performed following the guidelines presented by Moher et al.[27] (PRISMA) and Kitchenham et al.[28]. This SLR work accumulates the most relevant research work from primary sources. These papers are then evaluated and analyzed to grab the best results for the selected research problem. Figure 2 represents the results after following the PRISMA guidelines. This systematic analysis are performed using the following preliminary steps:

- Identification of research questions to systematically analyze the proposed field from different perspectives.
- Selection of relevant keywords and queries to download the most relevant research articles.
- Selection of peer-reviewed online databases to download relevant research articles published in healthcare big data domain during the period ranging from 2011 – 2021.
- Perform inclusion and exclusion process based on title, abstract and the contents presented in the article to remove duplicate records.
- Assess the finalized relevant articles for identifying gaps in the available literature and suggest new research directions to explore.

Research questions. Selecting a well-constructed research question(s) is essential for a successful review process. We formulate a set of five research questions based on the Goal Questions Metrics approach proposed by Van Solingen et al.[29]. The formulated research questions are depicted in Table 1 below.

Search strategy. Search strategy is the key step in any systematic research work because this is the step that ensures the most relevant article for the analysis and the assessment process. To define a well-organized search strategy a search string is developed using the formulated relevant keywords. For the accumulation of most relevant articles for a certain research problem, only keywords are not sufficient. These keywords are concatenated in different strings for searching articles in multiple online repositories[30]. Inspired from the SLR work of Achimugu
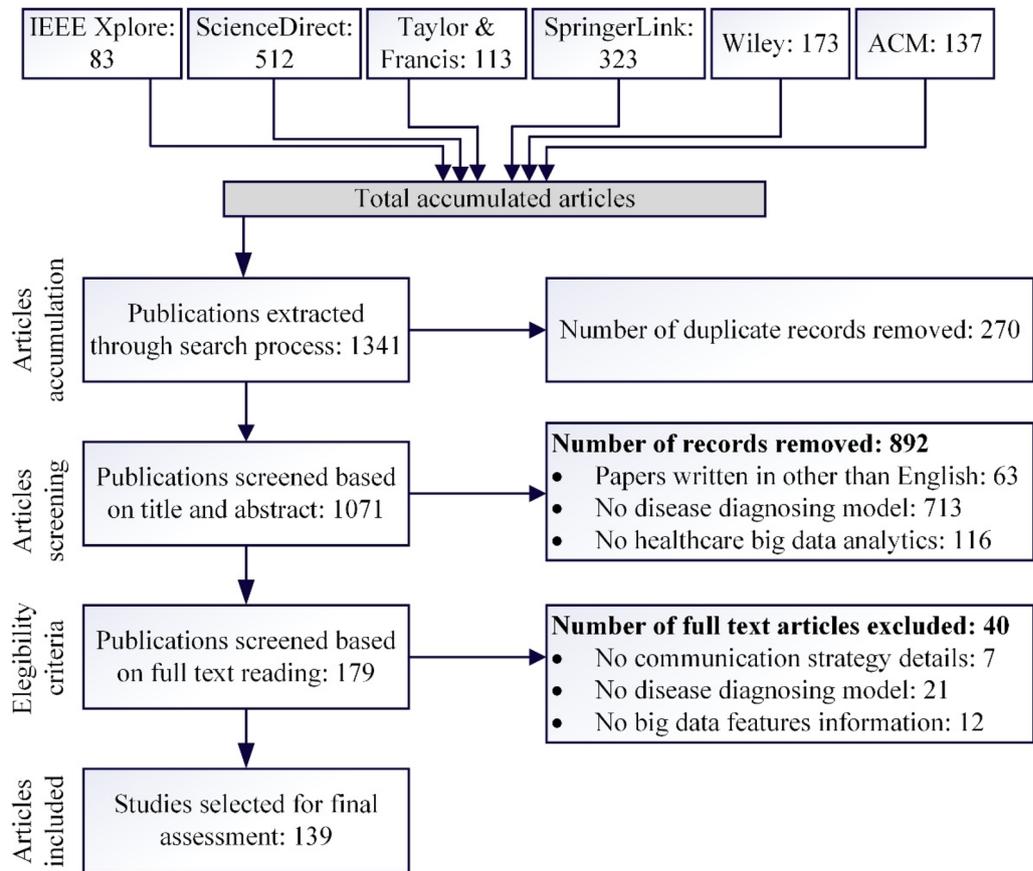
Figure 2. PRISMA process model for articles accumulation, screening, and final selection.

Research questions	Description and motivation
RQ1. What are the key features adapted to integrate structured and unstructured data in healthcare big data domain?	Big data can handle a plethora of types of data including images, temporal information, EHR data, business and audit relevant data, and many other structural and non-structural data. The prime concern of this research question is to outline the available techniques proposed in the literature to handle structured and unstructured data in healthcare domains
RQ2. What are different techniques proposed to provide an easy and timely data-access interface for doctors?	Big data comprises a gigantic amount of sensitive information. It is highly secure and only authentic person can access it. Normally it takes more time in verification, data providence and in simulations to extract relevant information. And in this over-populated real-life scenario, the caretakers face more patient in less timings. To overcome this problem, how many different techniques suggested and what are the gaps in this solution to be addressed in future are the key concern of this research question
RQ3. What are different ways to improve communication between the doctor and patient?	The prime objective of this research question is, to develop an optimal communication model for both doctor and patients by enlisting the communication problems (language barrier, lack of facilities, old age or handicap personnel etc.) in the available systems. This efficient model will change the overall healthcare system
RQ4. What are different types of classification models proposed for accurate disease diagnosing using patient historical information?	This question aims to outline multiple IoT-based and machine learning-based disease diagnosing applications suggested by researchers using patient's historical information. It also aims to discuss each model capabilities based on the information provided in the article
RQ5. What are different applications of big data analytics in health-care domain?	The prime focus of this research question is, to extend big data analytics to new healthcare applications by outlining the current available applications of healthcare big data analytics

Table 1. Set of Research questions.

(Features OR Characteristics OR Values) AND (Big Data OR Healthcare Big Data) AND (Management OR Process OR EFFICIENT MAN-AGEMNT) AND (Disease Detection OR "Disease Diagnosing") AND (Patient History OR Medical Records OR Treatment Record) AND (Applications OR Uses)

Table 2. Generalized query.

S. no	Online repository (s)	Permalink	Article downloading date
i	Science Direct	https://www.sciencedirect.com/	June 13, 2021
ii	Springer	https://link.springer.com/	June 15, 2021
iii	Taylor & Francis	https://www.tandfonline.com/	June 17, 2021
iv	ACM	https://dl.acm.org/	June 17, 2021
v	IEEE Xplore	https://ieeexplore.ieee.org/Xplore/home.jsp	June 18, 2021
vi	Wiley online	https://onlinelibrary.wiley.com/	June 18, 2021

Table 3. List of online datasets.

et al.[31], in software requirement domain, our search strategy consists of four main steps includes identification of keywords relevant to selected research problem, formulation of search string based on the keywords, and selection of online repositories to accumulate relevant articles to the problem selected.

Selection of keywords. List of keywords are defined for each research question to download all relevant articles. Some researchers defined a generic query[32] and starts downloading articles. Although it is simple for the accumulation of articles from online database but mostly it tends to skip some most relevant articles. So, the correct option is to define keywords for each research question. In fact, it is a hectic job, but it ensures the retrieval of each relevant article from online databases regarding a certain research problem.

Formulation of search string. Search strings (queries) are formulated using the keywords identified from the selected research questions. The search string is tested in online databases and was modified according to retrieve each relevant articles from these databases. Inspired from the guidelines proposed by Wohlin[33], following are the key steps undertaken to develop an optimal search string:

   i. Identification of key terms from the formulated topic and research questions
   ii. Selection of alternate words or synonyms for key terms
   iii. Use "OR" operator for alternating words or synonyms during query formation
   iv. Link all major terms with Boolean "AND" operator to validate every single keyword.

Following all these preliminary steps a generic query/search-string is developed that is depicted in Table 2. This generic query is further refined for each research question as depicted in Table 3 to retrieve each relevant article.

Selection of online repositories. After identifying keywords and formulating search strings the next step is to download relevant articles specific to the interested research problem. For the accumulation of relevant articles six well-known and peer-reviewed online repositories are selected, as depicted in Table 3.

Articles accumulation and final database development. For relevant articles accumulation and final database development we followed the guidelines suggested by Kable et al.[34]. After specifying the research questions, identifying keywords, and formulating search queries, and selecting online repositories, the next key step is to develop a relevant articles database for the analysis and assessment purposes that includes three prime steps: (1) identification of inclusion/exclusion criteria for a certain research article(s), and (2) Relevant articles database development. These steps are discussed in detail below.

Inclusion and exclusion criteria. After selecting online database and starts the articles downloading process, the most tedious task that the author (s) facing, is the decision about whether a certain paper should be included in the final database or not? To overcome this problem an inclusion and exclusion criteria is defined for the inclusion of a certain article in the final set of articles. Table 4 represents the inclusion and exclusion criteria followed for this systematic research work.

A manual process is followed by the authors for the inclusion and exclusion of a certain article. These articles are evaluated based on title, abstract and information provided in the overall paper. If more than half authors agree upon the inclusion of a certain article based on these parameters (title, abstract, and contents presented in the article), then that paper was counted in the final database otherwise rejected. A total of 134 relevant primary studies are selected for the final assessment process. To ensure no skip of relevant article snowballing is applied to retrieve each relevant article.

5

Inclusion criteria
1. Select papers that are only presented/written in English language
2. Select the papers that are published during the period 2011–2020
3. Include papers whose page limit size is greater than or equal to 3
4. Select only those book chapters, workshop papers or conference papers that has relevant information to our main topic
5. Select only primary research articles for the analysis purposes
Exclusion criteria
1. Gray papers should be excluded
2. Exclude those papers that has no relevant information regarding the selected topic
3. Elimination of duplicate research papers

**Table 4.** Inclusion of exclusion criteria.
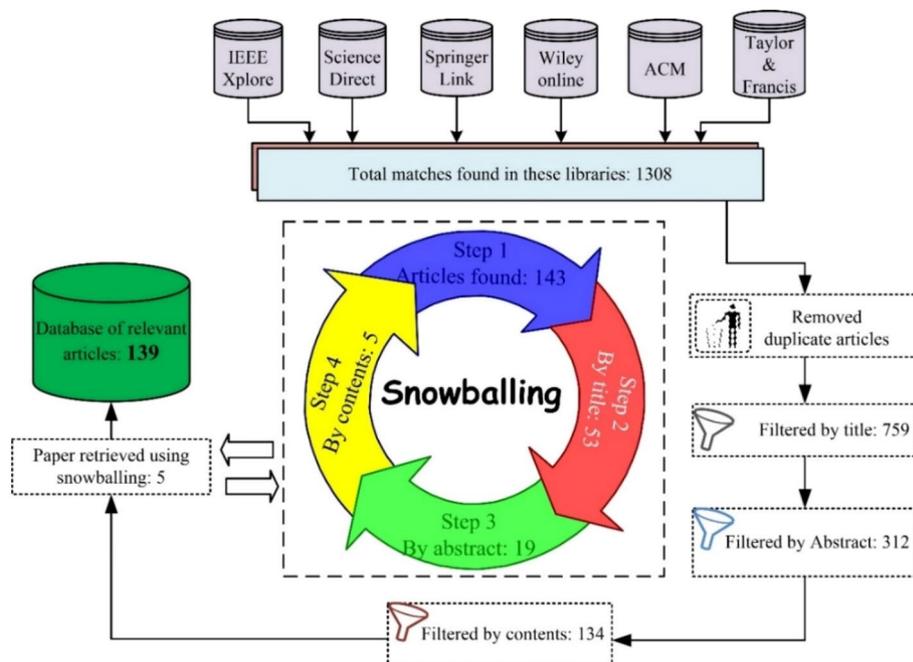
Figure 3. Extraction of each relevant article using snowballing.

- *Snowballing* To extract each relevant primary article snowballing is applied in the proposed research work[33]. In this systematic analysis both types of snowballing backward and forward snowballing is applied to ensure extraction of each relevant primary article. 145 relevant articles retrieved after applying snowballing process. These articles are then filtered by title and resulted for 53 relevant articles. After further processing by abstract resulted into 19 articles, and at last when filtered by contents presented in the paper resulted into only 5 relevant articles. This overall process is depicted in Fig. 3. After adding these articles to the accumulated relevant articles, a total of 139 articles added to the final database.

Relevant articles database development. After accumulating each primary article reported in the proposed field, a database of relevant articles is developed for the assessment and analysis work, to find the current available trends in healthcare big data analytical domain and investigate the gaps in these research articles to open new gates for future research work. A total of 139 relevant articles are added to the final database. The overall contribution of the selected online repositories in the relevant articles database development is depicted in Fig. 4.

From Fig. 4, it is concluded that IEEE Xplore and Science Direct contributing the more that reflects the interest of research community to present their work with.

Articles accumulation and final database development.
After developing a database of relevant articles, it is evaluated using different parameters like type of article (conference proceedings, journal article, book chapter etc.), publication year, and contribution of individual library. Figure 5 represents the information regarding the total contribution of articles by type in the final database.
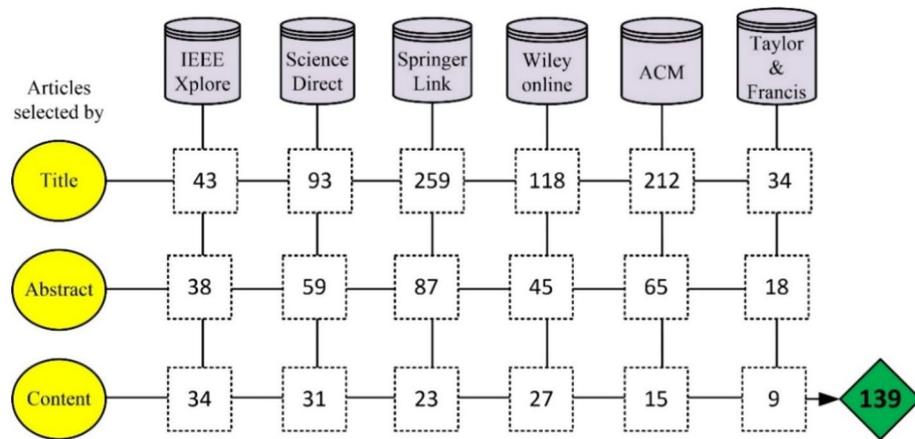
**Figure 4.** Distribution of primary studies.
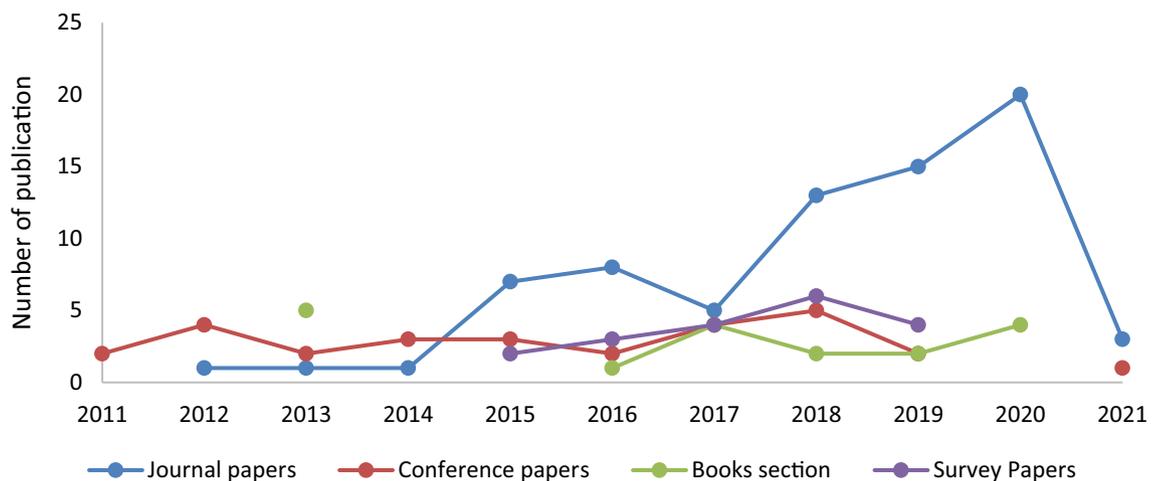
Figure 5. Evolution of final database by type of article and year.

Figure 5 concludes that the researchers paid significant attention towards the development of new healthcare systems instead of finding the gaps in the available systems and develop enhanced solutions accordingly. This enhanced solution can accurately identify and diagnose a certain disease based on patient's historical medical information. A small amount of work is reported using review articles, survey papers, but no systematic mechanism is followed to analyse the work in specific range of years followed by a set of research questions. The same problem can also be seen from Fig. 6 where highest percentage contribution is shown more comparative to book sections, conference papers etc.

Figure 7 depicts the percentage contribution of each library in the proposed assessment work.

Figure 8 represents the annual distribution of articles selected for the analysis and assessment purposes. Form Fig. 8 it is evident, that with passage of time number of articles increases, and that shows the maturity and interest of the researchers in this specific domain.

From Fig. 8, it is concluded that IEEE Xplore contributing the more in the final database of relevant articles that shows the trend of researchers to present healthcare relevant works in the IEEE journals. Figure 9 represents the total number of journal articles, survey papers, conference papers, and book sections in the selected relevant articles database.

From Fig. 9 it is concluded that significant attention is given towards the development of new healthcare models. This shows the maturity of the proposed field. Dealing with such a mature field and extracting useful information is hectic job for the researchers. A systematic analysis of this research field is needed to provide an overview of the work reported during a specific range of years. This analysis will not only save precious time of the researchers, but it will also open gates for the future research work in this field.

Table 5 represents the annual contribution of studies in the final relevant database.

Overall information regarding type of paper, publication year and number of records is depicted in Fig. 10 below.
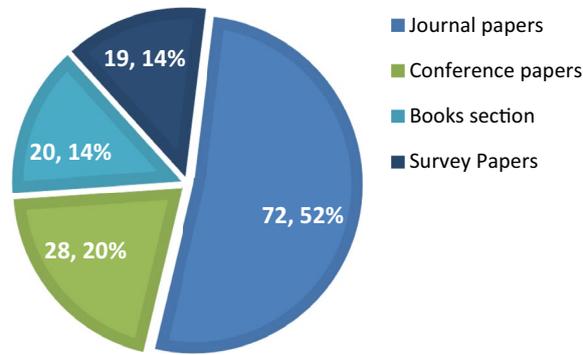
**Figure 6.** Percentage contribution by type of paper.
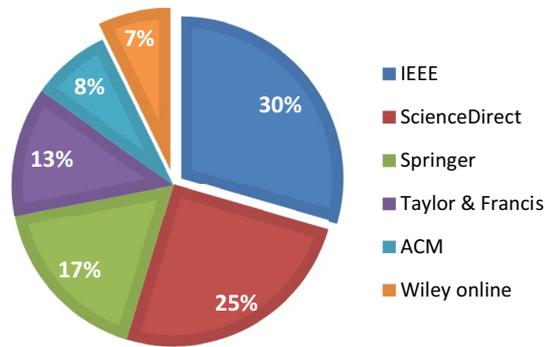
**Figure 7.** Percentage contribution of each library.
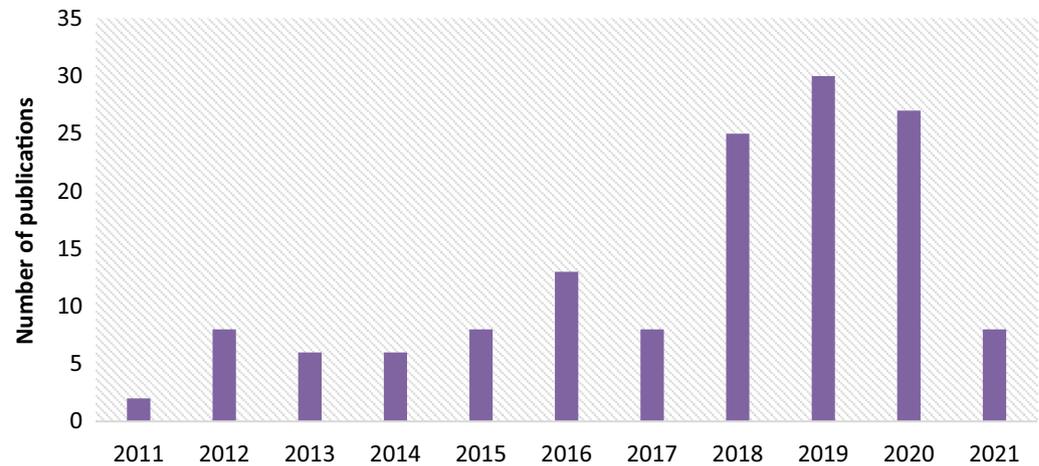
Figure 8. Annual distribution of articles.

Quality assesment

After executing exclusion and inclusion process, all the relevant articles in the database are manually assessed by authors to check the relevancy of each article with the selected research problem. A quality criterion is defined to check every research article against the formulated research questions. This quality criteria is defined in Table 6.

Weighted values are assigned against each quality criteria to check the relevancy of an article with a certain research question. These weighted values and description is depicted in Fig. 11.

After the assessment process, the relevancy of each article is decided based on its aggregated weighting score. If the score is greater than 3 it represents the most relevancy of an article to the selected research topic. Figure 12 represents the aggregate score values of each article based on the defined quality assessment criteria.
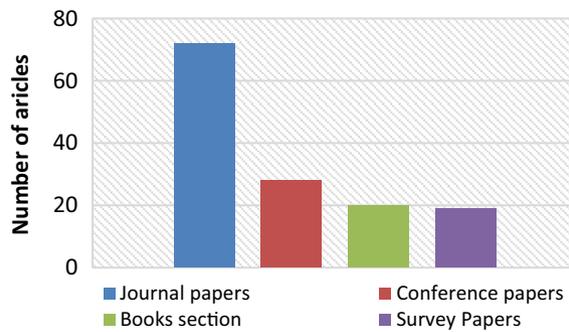
Figure 9. Evolution of database by number of articles by type.

Year	Paper ids
2011	35,36
2012	34,37–44
2013	3,45–49
2014	16,50–54
2015	4,55–65
2016	66–83
2017	11,84–94
2018	9,10,20,21,95–123
2019	7,8,18,19,22,124–152
2020	15,153–183
2021	167,184–190

Table 5. Inclusion or exclusion criteria.

Results and discussion

After executing the quality assessment work, the next key step of an SLR work is, to analyse all the relevant article to identify different techniques proposed for efficient communication between patient and practitioner, accurate feature extraction from healthcare big data and implement it in practical use.

This section of the paper performs a descriptive analysis of each article based on five research questions. In this systematic review process, a total of 139 research articles published during the period ranging from 2011 to 2021.

Healthcare big data. The researcher and data analysts suggested no contextual name for "big data" in healthcare, but for implementation and interpretation purposes they divided it into 5 V architecture. Figure 13 depicts a 5 V architecture of big data.

The exponential increase in IoT-based smart devices and information systems resulted a plethora of information in healthcare domain. This information increases exponentially on daily basis. These smart IoT based healthcare devices produces a huge of data. An alternated term "Big Data" is selected for this gigantic amount of data. This is the data for which scale, diversity, and complexities require innovative structure, variables, design, and analytics for efficient utilization and management, accurate data extraction and visualization, and to grab hidden stored information regarding a specific problem of interest. Main idea behind the implementation of healthcare big data analytics is to retrieve enriched information from huge amount of data using different machine leering and data mining techniques[191]. These techniques help in improving quality of care, reducing cost of care, and helps the practitioners to suggest medicines based on clinical historical information.

RQ1. What are the key features adapted to integrate the structured and unstructured data in healthcare big data domain? Big data comprises a huge amount of data to be processed, especially a plethora of types of data to process and extract enriched information regarding a problem of interest. Several features are assessed and analyzed especially in healthcare domain, to integrate both structural and non-structural data. Multiple researchers analyzed semantic based big data features for big data integration purposes while some researchers proposed behavior and structural based features for patient monitoring and activity management purposes[151,192]. While some performed real-time analysis using a group of people for data integrating and clustering purposes. Table 7 enlists the research work published for the structural and non-structural data integration purposes.
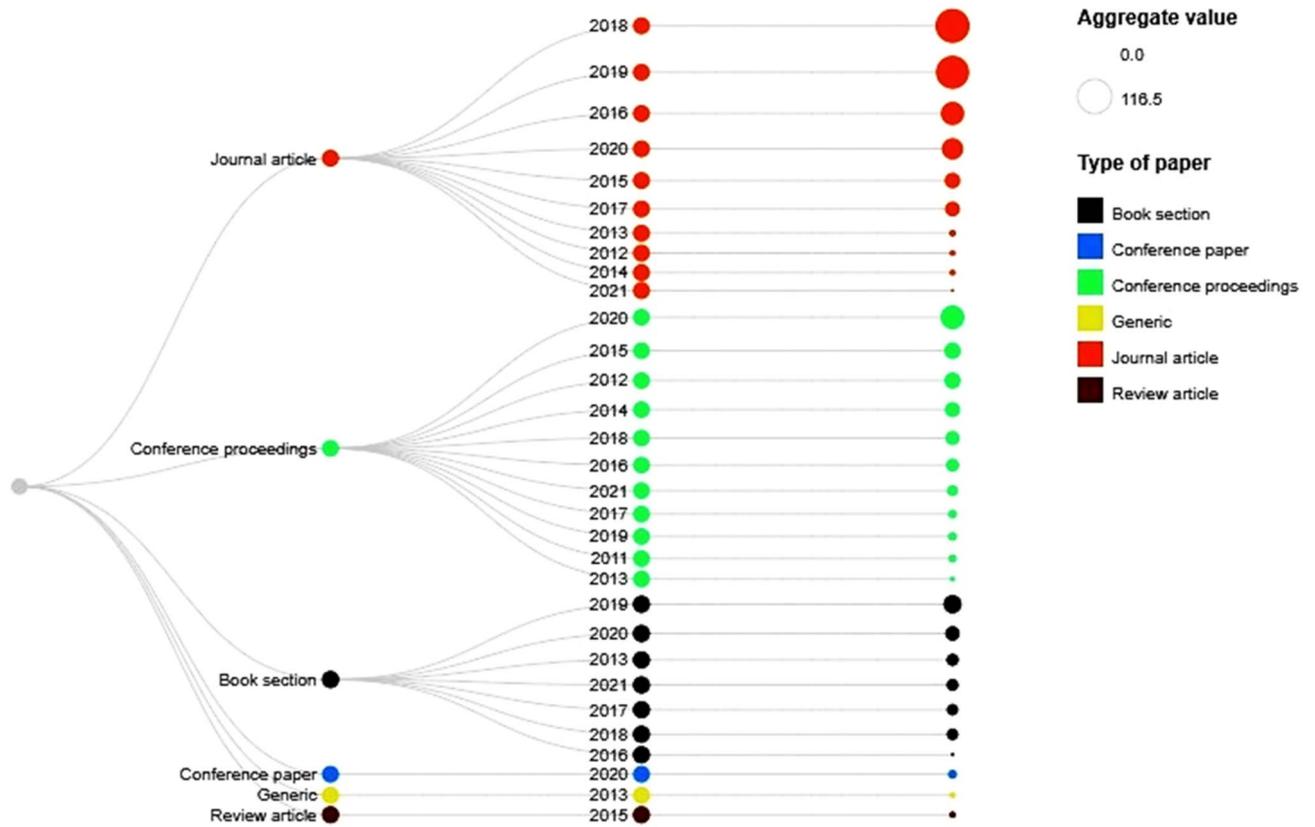
Figure 10. Evolution of final database.

S.no	Quality criteria
QC1	What are the key features selected for the optimum organization of both structural and non-structural data in healthcare big data domain?
QC2	Whether an algorithm proposed for the development of an optimal and timely data access interface for doctors to treat patients?
QC3	Whether any technique proposed to develop an accurate communication model for both patient and doctor?
QC4	Using patient historical information, whether a diagnosing model suggested for the accurate detection of a certain type of disease?
QC5	Whether healthcare big data analytics is proposed for certain type of application?

Table 6. Quality criteria.
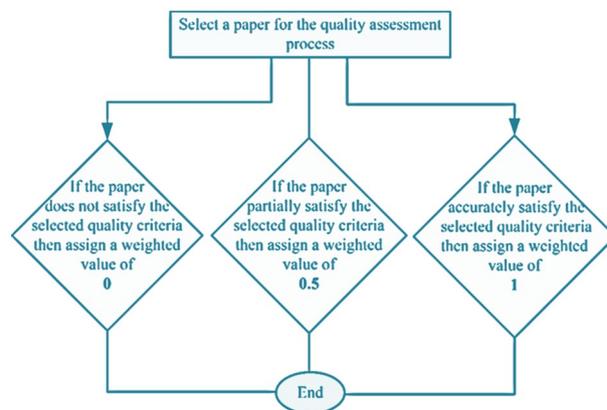
Figure 11. Quality criteria for the proposed SLR work.
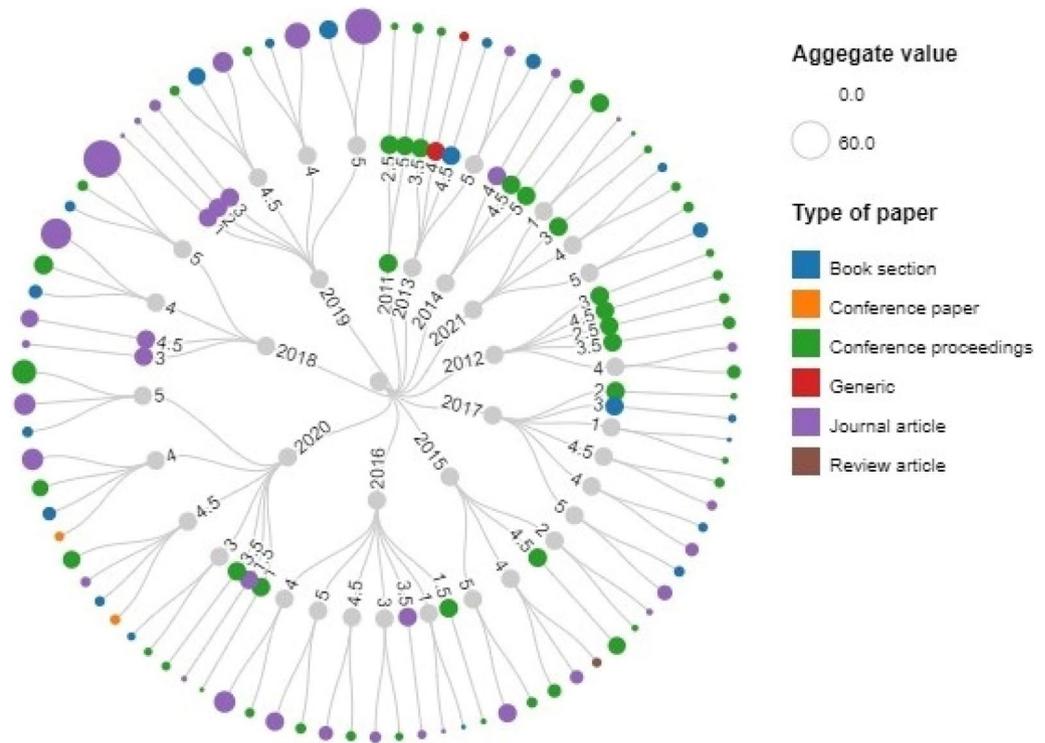
Figure 12. Quality assessment process.
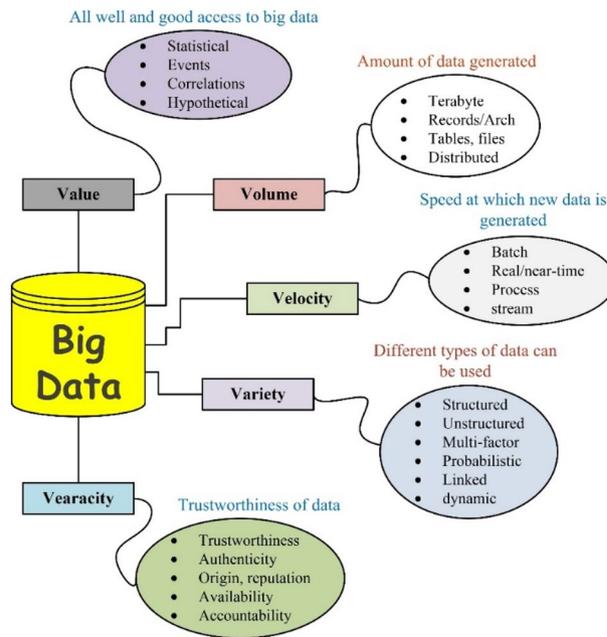
Figure 13. Big Data 5Vs[15].

After analysing the available literature in Table 8, it was concluded that mostly semantic based, structure-based, and real-time activity-based features are considered for the information extraction and organization purposes. If we consider geometric based feature and adapt clustering mechanism for data organization purposes, then this will not only integrate both structural and non-structural data efficiently, but it will improve the simulation capabilities of different applications.

S. no	Citation	Key features	Description
1	42	Structural features	Using the CT scan images and healthcare big data analytics, this paper identified the normal and abnormal brain structure condition with the help of structural based features
2	193	Behaviour-based features	This article summarizes the action-based (exemplar-based) features to improve the capabilities of nursing staff in some unwanted and emergency situations. It mainly focuses on the exploration of exemplars from nursing informatics research, to develop a smart mechanism that is feasible for nursing staff to participate in the big data revolution
3	60,119	Semantic based features	Around the world, a standard format of documents relevant to healthcare are suggested for the healthcare assessment and analysis process but the companies and the researchers face a big barrier during its comprehension and timely processing. To address this problem Hadoop based semantic transformation model is presented in this research work by using semantic based features with the help of CAD architecture
4	64	Heterogeneous data features	For monitoring quality of life of an individual patient several features are considered for the analysis and assessment processes. These features include non-textual information, heterogeneous data spaces, organ and organism scale, and specified analytics for identifying "physiological envelope" for a patient routine assessment purpose
5	67,83	Activity-based features	Social networking is an alternate source for retrieving information regarding daily healthcare activities. Social media facilitates the user by proving an easy access environment without any long time wait in queue or interruption compared to the traditional health pooling stations, where both younger and elder people wait for long time for their turn to poll. A Facebook application is proposed in this research work, where more than 1400 employees selected to accumulate patient routine-based activity information. This data can be collected in relatively small interval of time compared to the traditional arguing system without any promotional action
6	105	5 V features	5 V features are outlined in this book chapter to integrate both structural and non-structural data in healthcare big data domain. These features include volume, velocity, veracity, value, and variety-based features. It also consists other features such as: vulnerability and complexity
7	44	Real-Time analysis-based features	This research article presents a real time scenario for feature accumulation and disease diagnosing purposes. In this case study seven patient were contributed for the prediction of seizure disease using support vector machine
8	36	Integrated features	With rapid improvement and ever-increasing data in healthcare domain make it more complex to retrieve data regarding specific research oriented and disease diagnosing problem. To address this critical problem Tian[36] presented the concept of integrated features to integrate CT scan images in big data domain for efficient retrieval and disease diagnosing purposes
9	194	Functional features	This systematic review work has outlined the use of cognitive computing and functional features reported in healthcare, cybersecurity, and big data

Table 7. List of key features used for data integration purposes.

RQ2. What are different techniques proposed to provide an easy and timely data-access interface for doctors? Digital transformation of healthcare systems by using of information system, medical technology, handheld and smart wearable devices has posed many challenges for both the researchers and caretakers in the form of storage, dropping the cost of care and processing time (to extract relevant information for refining quality of care and reduce waste and error rates). Prime goal of healthcare big data analytics is, to process this vast amount of data using machine learning and other processing models to extract certain problem relevant information and use it for human well beings[195]. Several supervised and unsupervised classification techniques are followed for the said purposes. ML-based architectures and big data analytical techniques are integrated in healthcare domain for efficient information retrieval and exchange purposes, risk analysis, optimum decision-support system in clinics, and suggesting precise medicines using genomic information[196]. Table 8 represent the literature reported for the providence of an easy and timely data-access interface for the practitioners.

RQ3. What are different ways to improve communication between the doctor and patient? Healthcare around the world is under high pressure due to limiting financial resources, over-population, and disease burden. In this modern technological age, the healthcare paradigm is shifting from traditional, one-size-fits-all approach to a focus on personalized individual care [1]. Additionally, the healthcare data is varying both in type and amount. The healthcare providers are not only dealing with patient's historical, physical and namely information, but they also deal with imaging information, labs, and other digital and analogue information consists of ECG, MRI etc. This data is voluminous, varying in type and formats, and of differing structure. These are the capabilities of Big Data to handle not only different types of and forms of data, but can handle 5 V structure including volume, variety, value, veracity, and velocity. Thus, the doctors facing an increasing burden of rising patient numbers coupled with progressively less time to spend with each patient. In other words, we are dealing with more patients, more data, and less time.

Different techniques are proposed in the literature to provide an easy and timely communication interface for both doctors and patients. Table 9 depicts different information exchange tools/techniques reported in the literature.

RQ4. What are different types of classification models proposed for accurate disease diagnosing using patient historical information? This research question aims to outline different disease diagnosing models proposed in the literature using healthcare big data. Around the world diverse approaches are proposed by researchers for healthcare big data analysis to ensure accurate disease diagnosing capabilities, provide healthcare facilities at doorstep, development of eHealth and mHealth applications, and many others. Multiple statistical and ML-based approaches proposed for accurate diagnosing purposes. Figure 14 represents multiple techniques proposed for automatic disease diagnosing purposes using healthcare big data domain.

All these techniques perform the diagnosing process using semantic-based features or structural based features. But no attention is given towards geometric feature extraction techniques that are prominent in extracting enriched information from data and results in high identification rates. Also, no advanced hybrid neural network

12

No	Citation	Technique	Description
1	56	Network foresight approach	To provide an easy and timely data-access interface this research work used networked foresight approach for the prediction of accurate medicine using current and past clinical historical information. This approach provides an easier interface for the patients and doctors to communicate and exchange information regarding their health status
2	8–11,19,57,63,73,78,82,86,92,95,97,111–113,116,120,123–126,128,129,131–133,135,137,139–141,144–147,149–154,156,157	Big data analytics for timely data-access purposes	These survey papers presents applications of big data analytics in term of facilities like presenting an easier and timely-access mechanism for both patients and practitioners
3	70,197	Pig Latin Script programming tool	Pig Latin Script is applied to analyse the healthcare datasets against different queries, to enhance health facilities in India by assessing healthcare big data. It also provides an optimum and timely data-access mechanism for both patient and practitioner
4	83	Social media-based applications	This research article performed data analysis approach on data retrieved from Facebook walls. In this analysis process 153 public organization were contributed to accumulate information regarding patient activities and performance using unsupervised classification models
5	91	IoT-based systems	This paper outlined different IoT-based applications, cloud computing-based application, embedded systems, and other smart applications proposed for smart healthcare purposes at door step (without visiting the practitioners on daily basis)
6	94,109	eHealth applications	The overpopulation and complexities in medical and treatment makes the things complex for the patients to access their basic health providers on time. And this situation becomes more critical in 3rd world countries. To tackle this issue a smart phone-based eHealth application is developed to provide clinical facilities at homes. It also provides an easy interface for both doctors and patients to communicate easily and in timely fashion
7	100	Formal methods	Overpopulation, treatment errors, and overtreatment results in developing a disease burden situation and health disparities by combining with noncontagious disease population. Formal methods are applied to develop an optimum mechanism with zero defects for timely and accurate disease diagnosing and treatment purposes
8	104	Medicine suggestion interface	This paper presents an automatic medicine suggestion interface by extracting information from the patient's clinical records. This model not only enhanced the efficiency of doctors to check multiple patients in a timely fashion but also enhance his/her capabilities to advice accurate medicine for a certain type of disease
9	106	ML techniques	This paper surveyed the literature by finding the gaps in the available mHealth applications and developed a new mHealth2.0 application. This advanced version is developed using machine learning (ML) techniques to simulate big data in healthcare. This application provides enhance quality of care services to the patients
10	22,121	Statistical analysis-based models	This research work has developed a mathematical model for diabetic's analysis using big data. This analytical model helped the doctors by providing an easy simulation interface to process patient's clinical records in a short time to monitor patient's current diabetic situation remotely

Table 8. List of techniques proposed for easy and time data-access interface.

No	Citation	Technique	Description
1	[59]	Analytical framework	This research article proposed an analytical framework using big data analytics for the identification of Lupus disease. This model presents an easy interface for the doctors to diagnose Lupus disease on patient's clinical records. In this smart framework, the patients upload their EHR and the doctor performs the Lupus disease diagnosing
2	[69]	Bio Inspired ICT interface	In this modern technological age, the human health and life-loss will be purely dependant on data and its optimum organization. But organizing a huge amount of data generated on daily basis from industries, smart and portable IoT healthcare devices, and many others is a hectic job. This paper presents a Bio inspired ICT interface to present an intelligent healthcare application to both doctors and the patient. This application makes the doctor and patient convenient to exchange the information easily and in timely fashion
3	[76]	Cloud-based architecture	An optimal data exchange mechanism is developed in this research article by presenting the concepts of cloud computing. Totally different from the traditional database system that are full of anomalies. This cloud-based architecture makes the doctors convenient to store an individual patient's clinical records in a standard format that are easy to extract, update, and process for future purposes
4	[80,115]	Hybrid model	This paper embeds multiple integrated applications such as IoT-based systems, sensors and arrays of sensors, big data analytics for emerging technologies. This hybrid model ultimately provides a smart application in low cost, low data storage requirements, with high communication capabilities in between the patient and practioners
5	[84,85,107,118]	Secure communication framework for doctor and patient	On one side the available of smart IoT-based applications made our life easier by presenting an easy access interface. But on other side it threatens our life by giving an opportunity to hackers to steal our information from distant offices. Keeping these drawbacks in mind this research work has reviewed different techniques proposed for efficient and secure communication in between patient and doctor
6	[87]	SOA-FOG based architecture	A three-tier architecture is developed for healthcare big data management using Fog computing (SOA-FOG architecture). It displays security information in fog layer, client layer, and in cloud layer. This framework presents a secured communication framework between the doctor and the patients and can address the man-in-the-middle network attack problems
7	[101]	Data mining techniques	Smart clinical intelligence model is proposed in this research article using data mining techniques. This model aims to assist the doctors to remotely process the clinical health records of a particular patient and validate his health conditions
8	[198]	mHealth 4.0	This paper surveyed different industry 4.0 applications by identifying the gaps in the available mHealth applications and suggested a new enhanced healthcare mobile smart application (mHealth 4.0). This advanced version of the mHealth application provides an easy communication facility for doctor and patient where they can exchange information easily
9	[20]	Machine learning based model	This paper has developed a hybrid ML-based model using mask encryption, monitoring patient activities, granularly access control, real time data classification techniques. This hybrid healthcare technique facilitates both doctors by presenting a more accurate disease diagnosing model
10	[199]	CrowdHEALTH	A real-time communicating framework is developed for practitioners and doctors using healthcare big data analytics and Holistic Health Records (HHR) in the context of the CrowdHEALTH. Big data analytics is executed on the database to retrieve stored records (i.e. HHRs) for clinicians and doctors to advice medications

Table 9. List of techniques proposed for efficient communication and information exchange.

and shallow architectures are proposed for the automatic diagnosing purposes. Keeping these gaps in mind, an optimum eHealth application can be developed by applying these hybrid techniques.

RQ5. What are different applications of big data analytics in healthcare domain? Big data analytics has revolutionized our lives by presenting many state of the art applications in various domains ranging from eHealth to mHealth, weather forecasting to climate changes, traffic management to object detection, and many others. This research question mainly focusing on enlisting different applications of big data analytics in Table 10.

Limitations

This article has a number of limitations. Some of these limitations are listed below.

- For this systematic analysis articles are only accumulated from six different peer-reviewed libraries (ACM, SpringerLink, Taylor & Francis, Science Direct = IEEE Xplore, and Wiley online library), but there exist a number of multi-disciplinary databases for articles accumulation purposes.
- This systematic analysis covers a specific range of years (2011 –2021), while a number of articles are reporting on daily basis.
- Articles are accumulated from online libraries using search queries, so if a paper has no matching words to the query, then it was skipped during search process.
- Google Scholar is skipped during the articles accumulation phase to shorten the searching time. Also, it gives access to both peer-reviewed and non-peer-reviewed journals and we only focused on peer-reviewed journals for the relevant articles.
- Being a systematic literature work it can be broadened to grab the knowledge about other varying topics such as healthcare data commercialization, health sociology etc.

Figure 14. Multiple disease diagnosing techniques proposed in the literature.

Besides these limitations we hope that this systematic research work will be an inspiration for future research in the recommended fields and will open gates for both industrialists and policymakers.

Conclusion and future work

In this research article, the existing research reported during 2011 to 2021 is thoroughly analysed for the efforts made by researchers to help caretakers and clinicians to make authentic decisions in disease diagnosing and suggest medicines accordingly. Based on the research problem and underlying requirements, the researchers proposed several feature extraction, identification, and remote communication frameworks to develop doctor and patient communication in a timely fashion. These real-time or nearer to real-time applications mostly use big data analytics and computational devices. This research work identified several key features and optimum management designs proposed in healthcare big data analytical domain to achieve effective outcomes in disease diagnosing. The results of this systematic work suggests that advanced hybrid machine learning-based models and cloud computing application should be adapted to reduce treatment cost, simulation time, and achieve improved quality of care. The findings of this research work will not only help the policymakers to encourage the researchers and practitioners to develop advanced disease diagnosing models, but it will also assist in presenting an improved quality of treatment mechanism for patients.

Advanced hybrid machine learning architectures for cognitive computing are considered as the future toolbox for the data-driven analysis of healthcare big data. Also, geometric-based features must be considered for feature extraction purposes instead of semantic and structural-based features. These geometric-based feature extraction techniques will not only reduce the simulation time, but it will also improve the identification and disease diagnosing capabilities of smart health devices. Additionally, these features can help in accurate identification of Alzheimer, tumours in PET or MRI images using upgraded machine learning and big data analytics. Cluster-based mechanism should be considered for data organization purposes to improve big data timely-access and easy-management capabilities. Promoting research in these areas will be crucial for future innovation in healthcare domain.

Data availability

The data used and/or analyzed during the current study available from the corresponding author on reasonable request.

S. no	Citations	Architectures	Description
1	193	Proper treatment	Healthcare big data analytics has enhanced the capabilities of nursing staff by practicing data science and analysis for accurate treatment using patients' clinical-relevant historical information
2	59	Lupus disease diagnosing	This research work developed lupus disease diagnosing system with the help of big data analytics
3	64	Silico medicine solutions	Integrating VPH technologies with healthcare big data analytics a robust and effective solution is developed in this research work that provides a silico medicine solution for individuals
4	200	Real-time health analysis	A smart real-time eHealth application is presented in this paper by embedding Amazon and Microsoft Azure services to solve health issues in a limited time
6	71	Alzheimer's disease detection	A statistical data analysis mechanism is proposed in this research work for the identification of Alzheimer's disease using Chronic nervous system and healthcare big data analytics
7	88	health risk assessment	Multiple eHealth applications are overviewed in this research work to develop an optimal technique for patient's health stratification and risk assessment purposes
9	93	eHearing loss application	An embedded solution is proposed in this research work for hearing loss (HL) patients. This application uses holistic management scheme to the HL-patients by facilitating public health services. This integrated solution also keeps the users away from this type of diseases in future
11	110,201	High storage capabilities	Big data has a large amount of data to be processed, particularly a plethora of types of data to process and extract enriched information regarding a problem of interest. Furthermore, storing this vast amount of data is another big challenge for the research community. Apache Pig and Apache HBase are the repository facilitates the research community to store this huge amount of information easily and provides an easy access interface for the retrieval purposes
12	114	Health status detection	This research work presents a machine learning-based model for checking the patient's health status. Apache Spark-based cloud computing and data processing engine is deployed to process tweets (related to a specific patient's health status)
13	121	Hadoop framework	Using healthcare big data analytics, this study work proposes a statistical diagnosing mechanism for diabetic analysis. Hadoop framework is used for the validation and implementation purposes
14	128,131,142	Indoor treatment system	In this over-populous and disease burden environment accessing hospitals and doctors for basic health purposes is more difficult. But this problem become more severe for old age patients. To address this problem, these research articles propose an IoT-based smart eHelath application that facilitates aged personnel at doorstep
15	7	Intelligent hospital appointment system	A smart hospital appointment system was developed in this research article by using the concept of first come-first serve (FIFO) concept. This model not only vanished the concept of traditional doctor-patient appointment system but also helps in fixing the appointment with an expert doctor using healthcare data bank
16	139	Digital HIV/AIDS detection system	This research article presents an electronic HIV/AIDS detection system using healthcare big data analytics. Then a regional based monitoring system is developed to monitor these patients. This model not only helps in identifying the HIV/AIDS infected patient in a specific region, but it also helps the practitioners, researchers, and government agencies to perform proper treatment of these patients
17	145	Parkinson disease diagnosing model	This study has developed a sensor-based robot model for Parkinson disease diagnosing purposes. Sensors are integrated in the infected patient's body. Based on the sensed data the robot moves with the patient and accumulate data. This accumulated data was then provided to artificial neural network for classification and recognition of Parkinson disease

Table 10. Applications of big data analytics in healthcare domain.

References

1. Rahman, F. & Slepian, M. J. Application of big-data in healthcare analytics—Prospects and challenges. In *2016 IEEE-EMBS International Conference on Biomedical and Health Informatics (BHI)* 13–16 (2016).
2. Khan, N. *et al.* Big data: Survey, technologies, opportunities, and challenges. *Sci. World J.* **2014**, 1–18 (2014).
3. Groves, P., Kayyali, B., Knott, D. & Van Kuiken, S. The 'big data 'revolution in healthcare. In *McKinsey Quarterly* (2013).
4. Andreu-Perez, J., Poon, C. C., Merrifield, R. D., Wong, S. T. & Yang, G.-Z. Big data for health. *IEEE J. Biomed. Health Inform.* **19**, 1193–1208 (2015).
5. Kumar, M. A., Vimala, R. & Britto, K. A. A cognitive technology based healthcare monitoring system and medical data transmission. *Measurement* **146**, 322–332 (2019).
6. Chen, H., Khan, S., Kou, B., Nazir, S., Liu, W. & Hussain, A. A smart machine learning model for the detection of brain hemorrhage diagnosis based internet of things in smart cities. *Complexity* **2020** (2020).
7. Liang, Y. & Zhao, L. Intelligent hospital appointment system based on health data bank. *Procedia Comput. Sci.* **159**, 1880–1889 (2019).
8. Galetsi, P. & Katsaliaki, K. A review of the literature on big data analytics in healthcare. J. Oper. Res. Soc. 1–19 (2019).
9. Lindell, J. What are big data and analytics?. In *Analytics and Big Data for Accountants* (2018).
10. Alharthi, H. Healthcare predictive analytics: An overview with a focus on Saudi Arabia. *J. Infect. Public Health* **11**, 749–756 (2018).
11. Lee, C. *et al.* "Big healthcare data analytics: Challenges and applications. In *Handbook of Large-Scale Distributed Computing in Smart Healthcare* 11–41 (Springer, 2017).
12. Hussain, A., Nazir, S., Khan, S. & Ullah, A. Analysis of PMIPv6 extensions for identifying and assessing the efforts made for solving the issues in the PMIPv6 domain: A systematic review. *Comput. Netw.* **179**, 107366 (2020).
13. Khan, H.-U. *et al.* Systematic analysis of safety and security risks in smart homes. *Comput. Mater. Contin.* **68**, 1409–1428 (2021).
14. Khan, S., Nazir, S. & Khan, H.-U. Analysis of navigation assistants for blind and visually impaired people: A systematic review. *IEEE Access* **9**, 26712–26734 (2021).
15. Nazir, S. *et al.* A comprehensive analysis of healthcare big data management, analytics and scientific programming. *IEEE Access* **8**, 95714–95733 (2020).
16. Kitchin, R. Big Data, new epistemologies and paradigm shifts. *Big Data Soc.* **1**, 2053951714528481 (2014).

17. Cox, M. & Ellsworth, D. Application-controlled demand paging for out-of-core visualization. In *Proceedings. Visualization'97 (Cat. No. 97CB36155)* 235–244 (1997).
18. Syed, L., Jabeen, S., Manimala, S. & Elsayed, H. A. Data science algorithms and techniques for smart healthcare using IoT and big data analytics. In *Smart Techniques for a Smarter Planet* 211–241 (Springer, 2019).
19. Venkatesh, R., Balasubramanian, C. & Kaliappan, M. Development of big data predictive analytics model for disease prediction using machine learning technique. *J. Med. Syst.* **43**, 272 (2019).
20. Kaur, P., Sharma, M. & Mittal, M. Big data and machine learning based secure healthcare framework. *Procedia Comput. Sci.* **132**, 1049–1059 (2018).
21. Patel, H. B. & Gandhi, S. A review on big data analytics in healthcare using machine learning approaches. In *2018 2nd International Conference on Trends in Electronics and Informatics (ICOEI)* 84–90 (2018).
22. Rumbold, J. M. M., O'Kane, M., Philip, N. & Pierscionek, B. K. Big Data and diabetes: The applications of Big Data for diabetes care now and in the future. *Diabetic Med.* (2019).
23. Oxman, A. D. *et al.* Users' guides to the medical literature: VI. How to use an overview. *JAMA* **272**, 1367–1371 (1994).
24. Swingler, G. H., Volmink, J. & Ioannidis, J. P. Number of published systematic reviews and global burden of disease: database analysis. *BMJ* **327**, 1083–1084 (2003).
25. Research, C. I. O. H. Randomized controlled trials registration/application checklist (12/2006). Available at: http://www.cihr-irsc.gc.ca/e/documents/rct_reg_e.pdf. Accessed 22 June 2009.
26. Young, C. & Horton, R. Putting clinical trials into context. *Lancet* **366**, 107–107 (2005).
27. P. Group, Moher, D., Liberati, A., Tetzlaff, J. & Altman, D. G. Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. *PLoS Med.* **6**, e1000097 (2009).
28. Kitchenham, B. & Charters, S. Guidelines for performing systematic literature reviews in software engineering (2007).
29. Van Solingen, R., Basili, V., Caldiera, G. & Rombach, H. D. Goal question metric (gqm) approach. *Encycl. Softw. Eng.* (2002).
30. Brereton, P., Kitchenham, B. A., Budgen, D., Turner, M. & Khalil, M. Lessons from applying the systematic literature review process within the software engineering domain. *J. Syst. Softw.* **80**, 571–583 (2007).
31. Achimugu, P., Selamat, A., Ibrahim, R. & Mahrin, M. N. R. A systematic literature review of software requirements prioritization research. *Inf. Softw. Technol.* **56**, 568–585 (2014).
32. Nazir, S., Ali, Y., Ullah, N. & García-Magariño, I. Internet of things for healthcare using effects of mobile computing: A systematic literature review. *Wirel. Commun. Mobile Comput.* **109**, 5931315 (2019).
33. Wohlin, C. Guidelines for snowballing in systematic literature studies and a replication in software engineering. In *Proceedings of the 18th International Conference on Evaluation and Assessment in Software Engineering* 1–10 (2014).
34. Kable, A. K., Pich, J. & Maslin-Prothero, S. E. A structured approach to documenting a search strategy for publication: A 12 step guideline for authors. *Nurse Educ. Today* **32**, 878–886 (2012).
35. Helmer, A., Kretschmer, F., Müller, F., Eichelberg, M., Deparade, R., Tegtbur, U. *et al.* Integration of medical models in personal health records using the example of rehabilitation training for cardiopulmonary patients. In *2011 4th International Conference on Biomedical Engineering and Informatics (BMEI)* 1887–1892 (2011).
36. Tian, M. Integrated feature based medical image retrieval. In *2011 International Conference on Control, Automation and Systems Engineering (CASE)* 1–3 (2011).
37. Chaves, R., Ramírez, J., Górriz, J. M., Illán, I. A. & Salas-Gonzalez, D. FDG and PIB biomarker PET analysis for the Alzheimer's disease detection using Association Rules. In *2012 IEEE Nuclear Science Symposium and Medical Imaging Conference Record (NSS/MIC)* 2576–2579 (2012).
38. Chute, C. G. Obstacles and options for big-data applications in biomedicine: The role of standards and normalizations. In *2012 IEEE International Conference on Bioinformatics and Biomedicine* (2012).
39. Goel, A. & Chandra, N. A prototype model for secure storage of medical images and method for detail analysis of patient records with PACS. In *2012 International Conference on Communication Systems and Network Technologies* 167–170 (2012).
40. Huang, H. & Hsiao, I. Use of anatomical information in a Bayesian reconstruction with an edge-preserving median prior. In *2012 IEEE Nuclear Science Symposium and Medical Imaging Conference Record (NSS/MIC)* 3321–3323 (2012).
41. López, C. M., Welkenhuysen, M., Musa, S., Eberle, W., Bartic, C., Puers, R. *et al.* Towards a noise prediction model for in vivo neural recording. In *2012 Annual International Conference of the IEEE Engineering in Medicine and Biology Society* 759–762 (2012).
42. Ng, H., Chuang, C. & Hsu, C. Extraction and analysis of structural features of lateral ventricle in brain medical images. In *2012 Sixth International Conference on Genetic and Evolutionary Computing* 35–38 (2012).
43. Patel, A. B., Birla, M. & Nair, U. Addressing big data problem using Hadoop and Map Reduce. In *2012 Nirma University International Conference on Engineering (NUiCONE)* 1–5 (2012).
44. Zheng, G., Yu, L., Feng, Y., Han, Z., Chen, L., Zhang, S. *et al.* Seizure prediction model based on method of common spatial patterns and support vector machine. In *2012 IEEE International Conference on Information Science and Technology* 29–34 (2012).
45. Li, L., Bagheri, S., Goote, H., Hasan, A. & Hazard, G. Risk adjustment of patient expenditures: A big data analytics approach. In *2013 IEEE International Conference on Big Data* 12–14 (2013).
46. Loshin, D. Chapter 8—Developing big data applications. In *Big Data Analytics* (ed. Loshin, D.) 73–81 (Morgan Kaufmann, 2013).
47. Loshin, D. Chapter 9—NoSQL data management for big data. In *Big Data Analytics* (ed. Loshin, D.) 83–90 (Morgan Kaufmann, 2013).
48. Loshin, D. Chapter 1—Market and business drivers for big data analytics. In *Big Data Analytics* (ed. Loshin, D.) 1–9 (Morgan Kaufmann, 2013).
49. Purkayastha, S. & Braa, J. Big data analytics for developing countries–Using the cloud for operational BI in health. *Electron. J. Inf. Syst. Dev. Ctries.* **59**, 1–17 (2013).
50. Lin, C.-H., Huang, L.-C., Chou, S.-C. T., Liu, C.-H., Cheng, H.-F. & Chiang, I. J. Temporal event tracing on big healthcare data analytics. In *2014 IEEE International Congress on Big Data* 281–287 (2014)
51. Martínez, J. G., Ramos-Becerril, F. J., Leija, L., López, F., García, U., Vera, A. *et al.* Development of an electronic equipment for the pre medical diagnose in the progress of diabetic foot disease. In *2014 International Conference on Control, Decision and Information Technologies (CoDIT)* 679–683 (2014).
52. Mian, M., Teredesai, A., Hazel, D., Pokuri, S. & Uppala, K. Work in progress—In-memory analysis for healthcare big data. In *2014 IEEE International Congress on Big Data* 778–779 (2014).
53. Panahiazar, M., Taslimitehrani, V., Jadhav, A. & Pathak, J. Empowering personalized medicine with big data and semantic web technology: Promises, challenges, and use cases. In *2014 IEEE International Conference on Big Data (Big Data)* 790–795 (2014).
54. Vargheese, R. Dynamic protection for critical health care systems using cisco CWS: Unleashing the power of big data analytics. In *2014 Fifth International Conference on Computing for Geospatial Research and Application* 77–81 (2014).
55. Archenaa, J. & Anita, E. A. M. A survey of big data analytics in healthcare and government. *Procedia Comput. Sci.* **50**, 408–413 (2015).
56. Boman, M. & Sanches, P. Sensemaking in intelligent health data analytics. *KI Künstliche Intell.* **29**, 143–152 (2015).
57. Chong, D. & Shi, H. Big data analytics: A literature review. *J. Manag. Anal.* **2**, 175–201 (2015).
58. Dantanarayana, G., Sahama, T. & Wikramanayake, G. Quality of information for quality of life: Healthcare big data analytics. In *2015 Fifteenth International Conference on Advances in ICT for Emerging Regions (ICTer)* 281–281 (2015).

59. Gomathi, S. & Narayani, V. Implementing big data analytics to predict systemic lupus erythematosus. In *2015 International Conference on Innovations in Information, Embedded and Communication Systems (ICIIECS)* 1–5 (2015).

60. Hussain, S. & Lee, S. Semantic transformation model for clinical documents in big data to support healthcare analytics. In *2015 Tenth International Conference on Digital Information Management (ICDIM)* 99–102 (2015).

61. Kuo, M., Chrimes, D., Moa, B. & Hu, W. Design and construction of a big data analytics framework for health applications. In *2015 IEEE International Conference on Smart City/SocialCom/SustainCom (SmartCity)* 631–636 (2015).

62. Mehmood, R. & Graham, G. Big data logistics: A health-care transport capacity sharing model. *Procedia Comput. Sci.* **64**, 1107–1114 (2015).

63. Raj, P., Raman, A., Nagaraj, D. & Duggirala, S. Big data analytics for healthcare. In *High-Performance Big-Data Analytics Computer Communications and Networks* 1525–1525 (Springer, Cham, 2015).

64. Viceconti, M., Hunter, P. & Hose, R. Big data, big knowledge: Big data for personalized healthcare. *IEEE J. Biomed. Health Inform.* **19**, 1209–1215 (2015).

65. Wang, M. D. Biomedical big data analytics for patient-centric and outcome-driven precision health. In *2015 IEEE 39th Annual Computer Software and Applications Conference* 1–2 (2015).

66. Batarseh, F. A. & Latif, E. A. Assessing the quality of service using big data analytics: With application to healthcare. *Big Data Res.* **4**, 13–24 (2016).

67. Buzzi, M. C. *et al.* Facebook: A new tool for collecting health data?. *Multimed. Tools Appl.* **76**, 10677–10700 (2016).

68. Chauhan, R., Jangade, R. & Mudunuru, V. K. A cloud based environment for big data analytics in healthcare. In *International Conference on Soft Computing and Pattern Recognition* 315–321 (2016).

69. Stefano, A. D., Corte, A. L., Lió, P. & Scatá, M. Bio-inspired ICT for big data management in healthcare. In *Intelligent Agents in Data-intensive Computing* 1–26 (Springer, 2016).

70. Gupta, S. & Tripathi, P. An emerging trend of big data analytics with health insurance in India. In *2016 International Conference on Innovation and Challenges in Cyber Security (ICICCS-INBUSH)* 64–69 (2016).

71. Haas, M. *et al.* Big data to smart data in Alzheimer's disease: Real-world examples of advanced modeling and simulation. *Alzheimers Dement.* **12**, 1022–1030 (2016).

72. Jiang, P. *et al.* An intelligent information forwarder for healthcare big data systems with distributed wearable sensors. *IEEE Syst. J.* **10**, 1147–1159 (2016).

73. Kankanhalli, A., Hahn, J., Tan, S. & Gao, G. Big data and analytics in healthcare: Introduction to the special section. *Inf. Syst. Front.* **18**, 233–235 (2016).

74. Kashyap, H., Ahmed, H. A., Hoque, N., Roy, S. & Bhattacharyya, D. K. Big data analytics in bioinformatics: Architectures, techniques, tools and issues. *Netw. Model. Anal. Health Inform. Bioinform.* **5**, 28 (2016).

75. Lv, Z., Chirivella, J. & Gagliardo, P. Bigdata oriented multimedia mobile health applications. *J. Med. Syst.* **40**, 120 (2016).

76. Pandey, M. K. & Subbiah, K. A novel storage architecture for facilitating efficient analytics of health informatics big data in cloud. In *2016 IEEE International Conference on Computer and Information Technology (CIT)* 578–585 (2016).

77. Plachkinova, M., Vo, A., Bhaskar, R. & Hilton, B. A conceptual framework for quality healthcare accessibility: A scalable approach for big data technologies. *Inf. Syst. Front.* **20**, 289–302 (2016).

78. Rahman, F. & Slepian, M. J. Application of big-data in healthcare analytics—Prospects and challenges. In *2016 IEEE-EMBS International Conference on Biomedical and Health Informatics (BHI)* 13–16 (2016).

79. Rallapalli, S., Gondkar, R. R. & Ketavarapu, U. P. K. Impact of processing and analyzing healthcare big data on cloud computing environment by implementing hadoop cluster. *Procedia Comput. Sci.* **85**, 16–22 (2016).

80. Sakr, S. & Elgammal, A. Towards a comprehensive data analytics framework for smart healthcare services. *Big Data Res.* **4**, 44–58 (2016).

81. Xu, B. *et al.* Healthcare data analytics: Using a metadata annotation approach for integrating electronic hospital records. *J. Manag. Anal.* **3**, 136–151 (2016).

82. Tresp, V. *et al.* Going digital: A survey on digitalization and large-scale data analytics in healthcare. *Proc. IEEE* **104**, 2180–2206 (2016).

83. Straton, N., Hansen, K., Mukkamala, R. R., Hussain, A., Gronli, T., Langberg, H. *et al.* Big social data analytics for public health: Facebook engagement and performance. In *2016 IEEE 18th International Conference on e-Health Networking, Applications and Services (Healthcom)* 1–6 (2016).

84. Abouelmehdi, K., Beni-Hssane, A., Khaloufi, H. & Saadi, M. Big data security and privacy in healthcare: A review. *Procedia Comput. Sci.* **113**, 73–80 (2017).

85. Alonso, S. G., de la Torre, Diez I., Rodrigues, J. J., Hamrioui, S. & Lopez-Coronado, M. A systematic review of techniques and sources of big data in the healthcare sector. *J. Med. Syst.* **41**, 183 (2017).

86. Anjum, A. *et al.* Big data analytics in healthcare: A cloud-based framework for generating insights. In *Cloud Computing* 153–170 (Springer, 2017).

87. Barik, R. K., Dubey, H. & Mankodiya, K. SOA-FOG: Secure service-oriented edge computing architecture for smart health big data analytics. In *2017 IEEE Global Conference on Signal and Information Processing (GlobalSIP)* 477–481 (2017).

88. Cano, I., Tenyi, A., Vela, E., Miralles, F. & Roca, J. Perspectives on big data applications of health information. *Curr. Opin. Syst. Biol.* **3**, 36–42 (2017).

89. A. Di Meglio and M. Manca, "From Big Data to Big Insights: The Role of Platforms in Healthcare IT," in New Perspectives in Medical Records, ed: Springer, 2017, pp. 33–47.

90. Manogaran, G. *et al.* Big data analytics in healthcare Internet of Things. In *Innovative Healthcare Systems for the 21st Century* 263–284 (Springer, 2017).

91. Plageras, A. P., Stergiou, C., Kokkonis, G., Psannis, K. E., Ishibashi, Y., Kim, B. *et al.* Efficient large-scale medical data (eHealth Big Data) analytics in Internet of Things. In *2017 IEEE 19th Conference on Business Informatics (CBI)* 21–27 (2017).

92. Pramanik, M. I., Lau, R. Y. K., Demirkan, H. & Azad, M. A. K. Smart health: Big data enabled health paradigm within smart cities. *Expert Syst. Appl.* **87**, 370–383 (2017).

93. Spanoudakis, G., Katrakazas, P., Koutsouris, D., Kikidis, D., Bibas, A. & Pontopidan, N. H. Public health policy for management of hearing impairments based on big data analytics: EVOTION at genesis. In *2017 IEEE 17th International Conference on Bioinformatics and Bioengineering (BIBE)* 525–530 (2017).

94. Wu, J., Li, H., Liu, L. & Zheng, H. Adoption of big data and analytics in mobile healthcare market: An economic perspective. *Electron. Commer. Res. Appl.* **22**, 24–41 (2017).

95. Aceto, G., Persico, V. & Pescape, A. The role of Information and Communication Technologies in healthcare: Taxonomies, perspectives, and challenges. *J. Netw. Comput. Appl.* **107**, 125–154 (2018).

96. Antoniou, C., Dimitriou, L. & Pereira, F. *Mobility Patterns, Big Data and Transport Analytics: Tools and Applications for Modeling* (Elsevier, 2018).

97. Bates, D. W., Heitmueller, A., Kakad, M. & Saria, S. Why policymakers should care about "big data" in healthcare. *Health Policy Technol.* **7**, 211–216 (2018).

98. Choi, T.-M., Wallace, S. W. & Wang, Y. Big data analytics in operations management. *Prod. Oper. Manag.* **27**, 1868–1883 (2018).

99. Forestiero, A. & Papuzzo, G. Distributed algorithm for big data analytics in healthcare. In *2018 IEEE/WIC/ACM International Conference on Web Intelligence (WI)* 776–779 (2018).

100. Ganesh, S. & Talukder, A. K. Formal methods, artificial intelligence, big-data analytics, and knowledge engineering in medical care to reduce disease burden and health disparities. In *International Conference on Big Data Analytics* 307–321 (2018).

101. Giacalone, M., Cusatelli, C. & Santarcangelo, V. Big data compliance for innovative clinical models. *Big Data Res.* **12**, 35–40 (2018).

102. Guha, S. & Kumar, S. Emergence of big data research in operations management, information systems, and healthcare: Past contributions and future roadmap. *Prod. Oper. Manag.* **27**, 1724–1735 (2018).

103. Gupta, V., Singh Gill, H., Singh, P. & Kaur, R. An energy efficient fog-cloud based architecture for healthcare. *J. Stat. Manag. Syst.* **21**, 529–537 (2018).

104. Hopp, W. J., Li, J. & Wang, G. Big data and the precision medicine revolution. *Prod. Oper. Manag.* **27**, 1647–1664 (2018).

105. Huang, H. K. Big data in PACS-based multimedia medical imaging informatics. In *PACS Based Multimedia Imaging Informatics* (ed Huang, H.) 575–589 (2018).

106. Istepanian, R. S. H. & Al-Anzi, T. m-Health 2.0: New perspectives on mobile health, machine learning and big data analytics. *Methods* **151**, 34–40 (2018).

107. Khaloufi, H., Abouelmehdi, K., Beni-hssane, A. & Saadi, M. Security model for big healthcare data lifecycle. *Procedia Comput. Sci.* **141**, 294–301 (2018).

108. Krittanawong, C., Johnson, K. W., Hershman, S. G. & Tang, W. H. W. Big data, artificial intelligence, and cardiovascular precision medicine. *Expert Rev. Precis. Med. Drug Dev.* **3**, 305–317 (2018).

109. Ma, X., Wang, Z., Zhou, S., Wen, H. & Zhang, Y. Intelligent healthcare systems assisted by data analytics and mobile computing. In *2018 14th International Wireless Communications & Mobile Computing Conference (IWCMC)* 1317–1322 (2018).

110. Manogaran, G. *et al.* A new architecture of Internet of Things and big data ecosystem for secured smart healthcare monitoring and alerting system. *Future Gener. Comput. Syst.* **82**, 375–387 (2018).

111. Mehta, N. & Pandit, A. Concurrence of big data analytics and healthcare: A systematic review. *Int. J. Med. Inform.* **114**, 57–65 (2018).

112. Miller, J. B. Big data and biomedical informatics: Preparing for the modernization of clinical neuropsychology. *Clin. Neuropsychol.* **33**, 287–304 (2018).

113. Moutselos, K., Kyriazis, D. & Maglogiannis, I. A web based modular environment for assisting health policy making utilizing big data analytics. In *2018 9th International Conference on Information, Intelligence, Systems and Applications (IISA)* 1–5 (2018).

114. Nair, L. R., Shetty, S. D. & Shetty, S. D. Applying spark based machine learning model on streaming big data for health status prediction. *Comput. Electr. Eng.* **65**, 393–399 (2018).

115. Pashazadeh, A. & Navimipour, N. J. Big data handling mechanisms in the healthcare applications: A comprehensive and systematic literature review. *J. Biomed. Inform.* **82**, 47–62 (2018).

116. Ravishankar Rao, A., Clarke, D. & Vargas, M. Building an open health data analytics platform: A case study examining relationships and trends in seniority and performance in healthcare providers. *J. Healthc. Inform. Res.* **2**, 44–70 (2018).

117. Sahoo, P. K., Mohapatra, S. K. & Wu, S.-L. SLA based healthcare big data analysis and computing in cloud network. *J. Parallel Distrib. Comput.* **119**, 121–135 (2018).

118. Sarkar, B. K. & Sana, S. S. A conceptual distributed framework for improved and secured healthcare system. *Int. J. Healthc. Manag.* 1–13 (2018).

119. Sebaa, A., Chikh, F., Nouicer, A. & Tari, A. Medical big data warehouse: architecture and system design, a case study: Improving healthcare resources distribution. *J. Med. Syst.* **42**, 59 (2018).

120. Shafqat, S., Kishwer, S., Rasool, R. U., Qadir, J., Amjad, T. & Ahmad, H. F. Big data analytics enhanced healthcare systems: A review. *J. Supercomput.*

121. Sivaparthipan, C. B., Karthikeyan, N. & Karthik, S. Designing statistical assessment healthcare information system for diabetics analysis using big data. *Multimed. Tools Appl.*

122. Tang, V. *et al.* An adaptive clinical decision support system for serving the elderly with chronic diseases in healthcare industry. *Expert. Syst.* **36**, e12369 (2018).

123. Wang, Y., Kung, L. & Byrd, T. A. Big data analytics: Understanding its capabilities and potential benefits for healthcare organizations. *Technol. Forecast. Soc. Change* **126**, 3–13 (2018).

124. Agrawal, A. & Choudhary, A. Health services data: Big data analytics for deriving predictive healthcare insights. In *Health Services Evaluation* 3–18 (2019).

125. Ahmed, M., Choudhury, S. & Al-Turjman, F. Big data analytics for intelligent internet of things. In *Artificial Intelligence in IoT* 107–127 (Springer, 2019).

126. Ahmed, Z. & Liang, B. T. Systematically dealing practical issues associated to healthcare data analytics. In *Future of Information and Communication Conference* 599–613 (2019).

127. Bora, D. J. Chapter 3—Big data analytics in healthcare: A critical analysis. In *Big Data Analytics for Intelligent Healthcare Management* (eds Dey, N. *et al.*) 43–57 (Academic Press, 2019).

128. Chanchaichujit, J., Tan, A., Meng, F. & Eaimkhong, S. Internet of Things (IoT) and big data analytics in healthcare. In *Healthcare 4.0* 17–36 (Springer, 2019).

129. Cirillo, D. & Valencia, A. Big data analytics for personalized medicine. *Curr. Opin. Biotechnol.* **58**, 161–167 (2019).

130. Dey, N., Das, H., Naik, B. & Behera, H. S. *Big Data Analytics for Intelligent Healthcare Management* (Academic Press, 2019).

131. Din, S. & Paul, A. Smart health monitoring and management system: Toward autonomous wearable sensing for Internet of Things using big data analytics. *Future Gener. Comput. Syst.* **91**, 611–619 (2019).

132. Galetsi, P., Katsaliaki, K. & Kumar, S. Values, challenges and future directions of big data analytics in healthcare: A systematic review. *Soc. Sci. Med.* **241**, 112533 (2019).

133. Guo, C. & Chen, J. Big data analytics in healthcare: data-driven methods for typical treatment pattern mining. *J. Syst. Sci. Syst. Eng.* **28**, 694–714 (2019).

134. Hussain, S. *et al.* Semantic preservation of standardized healthcare documents in big data. *Int. J. Med. Inform.* **129**, 133–145 (2019).

135. Mehta, N., Pandit, A. & Shukla, S. Transforming healthcare with big data analytics and artificial intelligence: A systematic mapping study. *J. Biomed. Inform.* **100**, 103311 (2019).

136. Muniasamy, A., Tabassam, S., Hussain, M. A., Sultana, H., Muniasamy, V. & Bhatnagar, R. Deep learning for predictive analytics in healthcare. In *International Conference on Advanced Machine Learning Technologies and Applications* 32–42 (2019).

137. Palanisamy, V. & Thirunavukarasu, R. Implications of big data analytics in developing healthcare frameworks–A review. *J. King Saud Univ. Comput. Inf. Sci.* **31**, 415–425 (2019).

138. Rajabion, L., Shaltooki, A. A., Taghikhah, M., Ghasemi, A. & Badfar, A. Healthcare big data processing mechanisms: The role of cloud computing. *Int. J. Inf. Manag.* **49**, 271–289 (2019).

139. Ramasamy, V., Gomathy, B. & Verma, R. K. Smart HIV/AIDS digital system using big data analytics. In *Progress in Advanced Computing and Intelligent Engineering* 415–421 (Springer, 2019).

140. Razzak, M. I., Imran, M. & Xu, G. Big data analytics for preventive medicine. *Neural Comput. Appl.*

141. Reiz, A. N., de la Hoz, M. A. & García, M. S. Big data analysis and machine learning in intensive care units. *Med. Intensiva* **43**, 416–426 (2019).

142. Saheb, T. & Izadi, L. Paradigm of IoT big data analytics in the healthcare industry: A review of scientific literature and mapping of research trends. *Telematics Inform.* **41**, 70–85 (2019).

143. Sahoo, A. K. *et al.* Chapter 9—Intelligence-based health recommendation system using big data analytics. In *Big Data Analytics for Intelligent Healthcare Management* (eds Dey, N. *et al.*) 227–246 (Academic Press, 2019).

144. Shahbaz, M., Gao, C., Zhai, L., Shahzad, F. & Hu, Y. Investigating the adoption of big data analytics in healthcare: The moderating role of resistance to change. *J. Big Data* **6**, 6 (2019).

145. Sivaparthipan, C. B. *et al.* Innovative and efficient method of robotics for helping the Parkinson's disease patient using IoT in big data analytics. *Trans. Emerg. Telecommun. Technol.* **31**, e3838 (2019).

146. Sousa, M. J., Pesqueira, A. N. M., Lemos, C., Sousa, M. & Rocha, Ãl. Decision-making based on big data analytics for people management in healthcare organizations. *J. Med. Syst.* **43**, 290 (2019).

147. Strang, K. D. Problems with research methods in medical device big data analytics. *Int. J. Data Sci. Anal.*

148. Thomas, J., Kneale, D., McKenzie, J. E., Brennan, S. E. & Bhaumik, S. Determining the scope of the review and the questions it will address. In *Cochrane Handbook for Systematic Reviews of Interventions* 13–31 (2019).

149. Wang, Y., Kung, L., Gupta, S. & Ozdemir, S. Leveraging big data analytics to improve quality of care in healthcare organizations: A configurational perspective. *Br. J. Manag.* **30**, 362–388 (2019).

150. Zetino, J. & Mendoza, N. Big data and its utility in social work: Learning from the big data revolution in business and healthcare. *Soc. Work Public Health* **34**, 409–417 (2019).

151. Nazir, S., Nawaz, M., Adnan, A., Shahzad, S. & Asadi, S. Big data features, applications, and analytics in cardiology—A systematic literature review. *IEEE Access* **7**, 143742–143771 (2019).

152. Shah, G., Shah, A. & Shah, M. Panacea of challenges in real-world application of big data analytics in healthcare sector. *J. Data Inf. Manag.* **1**, 107–116 (2019).

153. Galetsi, P., Katsaliaki, K. & Kumar, S. Big data analytics in health sector: Theoretical framework, techniques and prospects. *Int. J. Inf. Manag.* **50**, 206–216 (2020).

154. Iyengar, S. P., Acharya, H. & Kadam, M. Big data analytics in healthcare using spreadsheets. In *Big Data Analytics in Healthcare* 155–187 (Springer, 2002).

155. Kumar, S. A. & Venkatesulu, M. BrownBoost classifier-based bloom hash data storage for healthcare big data analytics. In *Information and Communication Technology for Sustainable Development* 53–69 (Springer, 2020).

156. Kumar, Y., Sood, K., Kaul, S. & Vasuja, R. Big data analytics and its benefits in healthcare. In *Big Data Analytics in Healthcare* 3–21 (Springer, 2020).

157. Naqishbandi, T. A. & Ayyanathan, N. Clinical big data predictive analytics transforming healthcare:-An integrated framework for promise towards value based healthcare. In *Advances in Decision Sciences* 545–561 (Springer, 2020).

158. Lambay, M. A. & Mohideen, S. P. Big data analytics for healthcare recommendation systems. In *2020 International Conference on System, Computation, Automation and Networking (ICSCAN)* 1–6 (2020).

159. Katarya, R. & Jain, S. Exploration of big data analytics in healthcare analytics. *In 2020 4th International Conference on Computer, Communication and Signal Processing (ICCCSP)* 1–4 (2020).

160. Javid, T., Faris, M., Beenish, H. & Fahad, M. Cybersecurity and data privacy in the cloudlet for preliminary healthcare big data analytics. In *2020 International Conference on Computing and Information Technology (ICCIT-1441)* 1–4 (2020).

161. Leung, C. K., Chen, Y., Hoi, C. S. H., Shang, S. & Cuzzocrea, A. Machine learning and OLAP on big COVID-19 data. In *2020 IEEE International Conference on Big Data (Big Data)* 5118–5127 (2020).

162. Akhtar, U., Lee, J. W., Bilal, H. S. M., Ali, T., Khan, W. A. & Lee, S. The impact of big data in healthcare analytics. In *2020 International Conference on Information Networking (ICOIN)* 61–63 (2020).

163. Mung, P. S. & Phyu, S. Effective analytics on healthcare big data using ensemble learning. In *2020 IEEE Conference on Computer Applications (ICCA)* 1–4 (2002).

164. Georgakopoulos, S. V., Gallos, P. & Plagianakos, V. P. Using big data analytics to detect fraud in healthcare provision. In *2020 IEEE 5th Middle East and Africa Conference on Biomedical Engineering (MECBME)* 1–3 (2020).

165. Leung, C. K., Chen, Y., Shang, S. & Deng, D. Big data science on COVID-19 Data. In *2020 IEEE 14th International Conference on Big Data Science and Engineering (BigDataSE)* 14–21 (2020).

166. Juddoo, S. & George, C. A Qualitative assessment of machine learning support for detecting data completeness and accuracy issues to improve data analytics in big data for the healthcare industry. In *2020 3rd International Conference on Emerging Trends in Electrical, Electronic and Communications Engineering (ELECOM)* 58–66 (2020).

167. Chauhan, R. & Yafi, E. Big data analytics for prediction modelling in healthcare databases. In *2021 15th International Conference on Ubiquitous Information Management and Communication (IMCOM)* 1–5 (2021).

168. Islam, M., Karim, R., Khatun, M. A. & Reza, S. A research on big data analytics in healthcare industry. In *2020 International Conference on Information Science and Communications Technologies (ICISCT)* 1–5 (2020).

169. Leung, C. K., Chen, Y., Hoi, C. S. H., Shang, S., Wen, Y. & Cuzzocrea, A. Big data visualization and visual analytics of COVID-19 data. In *2020 24th International Conference Information Visualisation (IV)* 415–420 (2020).

170. Balaji, S. & Prasathkumar, V. Dynamic changes by big data in health care. In *2020 International Conference on Computer Communication and Informatics (ICCCI)* 1–4 (2020).

171. Alahmar, A. & Benlamri, R. Optimizing hospital resources using big data analytics with standardized e-clinical pathways. In *2020 IEEE Intl Conf on Dependable, Autonomic and Secure Computing, Intl Conf on Pervasive Intelligence and Computing, Intl Conf on Cloud and Big Data Computing, Intl Conf on Cyber Science and Technology Congress (DASC/PiCom/CBDCom/CyberSciTech)* 650–657 (2020).

172. Sadineni, P. K. Developing a model to enhance the quality of health informatics using big data. In *2020 Fourth International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC)* 1267–1272 (2020).

173. Pramanik, M. I. *et al.* Healthcare informatics and analytics in big data. *Expert Syst. Appl.* **152**, 113388 (2020).

174. Ravikumaran, P., Vimala Devi, K., Kartheeban, K. & Narayanan Prasanth, N. Health data analytics: Framework & review on tool & technology. *Mater. Today Proc.* (2020).

175. Ramesh, T. & Santhi, V. Exploring big data analytics in health care. *Int. J. Intell. Netw.* **1**, 135–140 (2020).

176. Galetsi, P. & Katsaliaki, K. A review of the literature on big data analytics in healthcare. *J. Oper. Res. Soc.* **71**, 1511–1529 (2020).

177. Mehta, N., Pandit, A. & Kulkarni, M. Elements of healthcare big data analytics. In *Big Data Analytics in Healthcare* 23–43 (Springer, 2020).

178. Ehwerhemuepha, L. *et al.* HealtheDataLab–a cloud computing solution for data science and advanced analytics in healthcare with application to predicting multi-center pediatric readmissions. *BMC Med. Inform. Decis. Mak.* **20**, 1–12 (2020).

179. Sivasangari, A., Lakshmanan, L., Ajitha, P., Deepa, D. & Jabez, J. Big data analytics for 5G-enabled IoT healthcare. In *Blockchain for 5G-Enabled IoT* 261.

180. Ma, S. & Huai, J. Approximate computation for big data analytics. *SIGWEB Newsl.* (2021).

181. Uzunbaz, S. & Aref, W. G. Shared execution techniques for business data analytics over big data streams. In *Presented at the 32nd International Conference on Scientific and Statistical Database Management, Vienna, Austria* (2020).

182. Chalumporn, G. & Hewett, R. Health data analytics with an opportunistic big data algorithm. In *Presented at the Proceedings of the 11th International Conference on Advances in Information Technology, Bangkok, Thailand* (2020).

183. Minami, T. & Ohura, Y. Small data analysis for bigger data analysis. In *Presented at the 2021 Workshop on Algorithm and Big Data, Fuzhou, China* (2021).
184. Chakraborty, C. & Rathi, M. Chapter 2—Smart healthcare systems using big data. In *Demystifying Big Data, Machine Learning, and Deep Learning for Healthcare Analytics* (eds Kautish, P. N. S. & Peng, S.-L.) 17–32 (Academic Press, 2021).
185. Ilmudeen, A. Chapter 3—Big data-based frameworks for healthcare systems. In *Demystifying Big Data, Machine Learning, and Deep Learning for Healthcare Analytics* (eds Kautish, P. N. S. & Peng, S.-L.) 33–56 (Academic Press, 2021).
186. Mendhe, C. H., Henderson, N., Srivastava, G. & Mago, V. A scalable platform to collect, store, visualize, and analyze big data in real time. *IEEE Trans. Comput. Soc. Syst.* **8**, 260–269 (2021).
187. Sivabalaselvamani, D., Selvakarthi, D., Yogapriya, J., Thiruvenkatasuresh, M. P., Maruthappa, M. & Chandra, A. S. Artificial Intelligence in data-driven analytics for the personalized healthcare. In *2021 International Conference on Computer Communication and Informatics (ICCCI)* 1–5 (2021).
188. Harb, H., Mansour, A., Nasser, A., Cruz, E. M. & de la Torre Diez, I. A sensor-based data analytics for patient monitoring in connected healthcare applications. *IEEE Sens. J.* **21**, 974–984 (2021).
189. Jones, J. & Jones, J. Optimizing healthcare. In *2020 IEEE International Conference on E-health Networking, Application & Services (HEALTHCOM)* 1–6 (2021).
190. Hassan, S., Dhali, M., Zaman, F. & Tanveer, M. Big data and predictive analytics in healthcare in Bangladesh: Regulatory challenges. *Heliyon* **7**, e07179 (2021).
191. Khan, S. *et al.* KNN and ANN-based recognition of handwritten pashto letters using zoning features. *Mach. Learn.* **9**, 570–577 (2018).
192. Pant, D., Kumar, V., Kishore, J. & Pal, R. Healthcare data modeling in R. In *2017 1st International Conference on Intelligent Systems and Information Management (ICISIM)* 230–233 (2017).
193. Brennan, P. F. & Bakken, S. Nursing needs big data and big data needs nursing. *J. Nurs. Scholarsh.* **47**, 477–484 (2015).
194. Sreedevi, A. G., Nitya Harshitha, T., Sugumaran, V. & Shankar, P. Application of cognitive computing in healthcare, cybersecurity, big data and IoT: A literature review. *Inform. Process. Manag.* **59**, 102888 (2022).
195. Sinha, A., Hripcsak, G. & Markatou, M. Large datasets in biomedicine: A discussion of salient analytic issues. *J. Am. Med. Inform. Assoc. JAMIA* **16**, 759–767 (2009).
196. Alonso-Betanzos, A. & Bolón-Canedo, V. Big-Data analysis, cluster analysis, and machine-learning approaches (2018).
197. Dayal, M. & Singh, N. Indian health care analysis using big data programming tool. *Procedia Comput. Sci.* **89**, 521–527 (2016).
198. Jayaraman, P. P., Forkan, A. R. M., Morshed, A., Haghighi, P. D. & Kang, Y.-B. Healthcare 4.0: A review of frontiers in digital health. *WIREs Data Min. Knowl. Discov.* **10**, e1350 (2018).
199. Gallos, P. *et al.* CrowdHEALTH: Big data analytics and holistic health records. *Stud. Health Technol. Inform.* **258**, 255–256 (2019).
200. Wang, L., Ranjan, R., Kołodziej, J., Zomaya, A. & Alem, L. Software tools and techniques for big data computing in healthcare clouds. *Future Gener. Comput. Syst.* **43–44**, 38–39 (2015).
201. Kiourtis, A. *et al.* An autoscaling platform supporting graph data modelling big data analytics. *Stud. Health Technol. Inform.* **295**, 376–379 (2022).

Acknowledgements

Author contributions

S.K. wrote the original draft of the paper. He also revised the draft based on the reviewers suggestions. Dr. H.U.K. developed the experimental setup for the proposed systematic research work. Dr. S.N. performed articles accumulation and database development process.

Funding

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to H.U.K.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.