IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received April 30, 2020, accepted May 16, 2020, date of publication May 19, 2020, date of current version June 3, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.2995572

A Comprehensive Analysis of Healthcare Big Data Management, Analytics and Scientific Programming

SHAH NAZIR[1], SULAIMAN KHAN[1], HABIB ULLAH KHAN[2], SHAUKAT ALI[3], IVÁN GARCÍA-MAGARIÑO[4], RODZIAH BINTI ATAN[5], AND MUHAMMAD NAWAZ[6]

[1]Department of Computer Science, University of Swabi, Swabi 23430, Pakistan
[2]Department of Accounting and Information Systems, College of Business and Economics, Qatar University, Doha, Qatar
[3]Department of Computer Science, Islamia College University, Peshawar 25120, Pakistan
[4]Department of Software Engineering and Artificial Intelligence, Complutense University of Madrid, 28040 Madrid, Spain
[5]Department of Software Engineering and Information Systems, Faculty of Computer Science and Information Technology, University Putra Malaysia, Serdang 43400, Malaysia
[6]Department of Computer Science, Institute of Management Sciences Peshawar, Peshawar 25000, Pakistan

Corresponding author: Habib Ullah Khan (habib.khan@qu.edu.qa)

The work was supported by the Qatar National Library, Doha, Qatar.

ABSTRACT Healthcare systems are transformed digitally with the help of medical technology, information systems, electronic medical records, wearable and smart devices, and handheld devices. The advancement in the medical big data, along with the availability of new computational models in the field of healthcare, has enabled the caretakers and researchers to extract relevant information and visualize the healthcare big data in a new spectrum. The role of medical big data becomes a challenging task in the form of storage, required information retrieval within a limited time, cost efficient solutions in terms care, and many others. Early decision making based healthcare system has massive potential for dropping the cost of care, refining quality of care, and reducing waste and error. Scientific programming play a significant role to overcome the existing issues and future problems involved in the management of large scale data in healthcare, such as by assisting in the processing of huge data volumes, complex system modelling, and sourcing derivations from healthcare data and simulations. Therefore, to address this problem efficiently a detailed study and analysis of the available literature work is required to facilitate the doctors and practitioners for making the decisions in identifying the disease and suggest treatment accordingly. The peer reviewed reputed journals are selected for the accumulated of published research work during the period ranges from 2015 – 2019 (a portion of 2020 is also included). A total of 127 relevant articles (conference papers, journal papers, book section, and survey papers) are selected for the assessment and analysis purposes. The proposed research work organizes and summarizes the existing published research work based on the research questions defined and keywords identified for the search process. This analysis on the existence research work will help the doctors and practitioners to make more authentic decisions, which ultimately will help to use the study as evidence for treating patients and suggest medicines accordingly.

INDEX TERMS Healthcare, big data, big data management, big data analytics.

I. INTRODUCTION

Healthcare systems are being digitally transformed by technological enhancements in medical information systems, electronic medical records, wearable and smart devices, and handheld devices. This increase in medical big data, alongside the development of computational techniques in the field of healthcare, has enabled researchers and practitioners to extract and visualize medical big data in a new spectrum. In this modern technological age the information increases exponentially. Wearable devices continuously produce a vast amount of data that ultimately known as big data in layman's terms. Modification is required for the big data in the form of analytic based technique for proper management, visualization and extracting the hidden information within the big data. these modification are required due to scale, diversity, and

The associate editor coordinating the review of this manuscript and approving it for publication was Dian Tjondronegoro.

complexity in the data [1]. Big data has gained a significant attention in several fields such as; healthcare applications, banking, internet of things (IoT) based applications, imaging, smart cities, smart transportation system and many others [2]. The data is stored in a standardized manner to easily store, access, and retrieve the required relevant information.

Smart IoT based applications such as wearable devices, electronic medical report (EMR) generators, smart mobile phone healthcare systems has transformed the conventional healthcare system to digital healthcare system. A vast amount of data is generated from these devices on daily basis. Exponential increase in the medical big data has attracted the research community to extract and visualise the new insights from the healthcare big data. Many big data sources are available in the healthcare market such as; registration data, biometric data, electronic health records, imaging, patient reports, the internet data, biomarker data, clinical data, and administrative data [3].

Davarzani *et al.* [4] performed analysis on 499 elderly patients with congestive heart failure. The reported ages were greater than or equal to 60 years. The samples are collected for these patients after a continues follow-up for the 19 months in his clinic. Relationship between frequently measurements of biomarkers and treatment effects of loop diuretics, spironolactone, β-blockers, and renin-angiotensin system inhibitors on risk of HF hospitalization was examined in the generation of hypothesis. A generic mathematical model is applied to find out the correlation in between the patient's recurrences of events. Kinkorová and Topolčan [5] descibed the key Horizon 2020 of biobanking projects, financing schemes and the future perspective. The EPMA announced the enduring strategies for the useful endorsement of predictive, preventive, and personalized (PPP) medicine. Over 45 countries is contributing toward the implementation and improvement of the promotion of PPP medicine [6]. Golubnitschaja *et al.* [7], [8] reported several issues associated with the healthcare services of pandemic scenario in the progression of frequent non-communicable diseases, poor economy of health care, lack of specialised education, overdue interventional approaches of reactive medicine, challenging ethical characteristics of some treatments along with insufficient communications among policymakers and professional groups.

Distinctive nature of the healthcare big data makes it unique from other types of data sources. The researchers face many challenges in the processing of big data. Some of the significant challenges are; an efficient recognition model is required to process the huge amount of healthcare data to identify feature and perform classification for the identification of disease, lack of data-sharing incentives, and illegal use (blackmailing) of healthcare big data [9]. Some structural mechanism is required to accumulate healthcare big data [10]. Due to involvement of expensive hardware components (instruments), personal and discomfort of the patient makes the healthcare big data is more expensive. Healthcare big data has numerous applications in the field of public health, disease and safety surveillance, predictive modelling and clinical decision support, and many other research areas.

The researchers face many hurdles in extracting the enriched information from healthcare big data and applying for the diseases identification purposes. Some research work has been reported for reducing the cost of care, the time consumption, the quality of treatment, and the providence of the healthcare facilities at the door steps, but still no detailed research work in the healthcare big data is reported for feature identification, applications, and healthcare big data analytics.

A. PROBLEM DEFINITION

The literature cited reflects that a lot of work has been reported by many researcher around the world for healthcare big data analysis. This research focuses on extracting significant features of big data in the healthcare, applications of healthcare big data, and state of the art techniques proposed in the healthcare big data field which can ultimately be used for decision making in healthcare. The proposed research work organizes and summarizes the existing published research work based on the research questions defined and keywords identified for the search process. This analysis on the existence research work will help the doctors and practitioners to make more authentic decisions, which ultimately will help to use the study as evidence for treating patients and suggest medicines accordingly.

B. RESEARCH CONTRIBUTIONS

A total of 127 most relevant primary studies were included based on the inclusion or exclusion criteria, and quality assessment criteria. Main goal of the proposed systematic literature research work is to counter the given research questions;

- What are the most significant big data features in the field of healthcare to be analysed?
- How many state of the art techniques developed in the field of big data in healthcare during the census 2015 – 2019?
- For efficient big data management in the healthcare; the researchers proposed how many optimum solutions?
- What are the applications of the big data in scientific programming?
- What are the big data analytics in healthcare?

This paper is organized as follows; proposed methodology for this SLR process is outlined in section 2. Section 3 explains the overall research process based on the guidelines provided by Kitchenham *et al.* [11]. The results and discussions are given briefly in section 4. Some threats to validity are in section 5 for the proposed research work followed by the conclusion in section 6.

II. METHODOLOGY

In this modern technological age the information increases exponentially. Wearable devices continuously produce a vast amount of data that ultimately known as big data in layman's

terms. The big data is the data whose diversity, scale, and complexity need new structure, algorithm, technique and analytics for the management, visualization of the new insights from healthcare big data to further provide easy solution for it. For efficient and accurate disease detection this huge data must be processed and to be examined by the caretakers and by practitioners. Analysing this vast amount of data for disease detection, and prevention is a big challenge using the traditional techniques. To address this problem data mining techniques are applied to extract patterns and useful information for the investigation of diseases and its prevention [12].

Big data analytics has improved the quality of treatment by processing the healthcare big data and provides personalized medicines based on the analytical descriptions. The big data in healthcare can be classified into three major classes such as; large n and small p, small n and large p, and large n and large p, (n is the number of samples while p is the number of parameters) [13]. Big data has numerous features to be analysed especially in the field of healthcare. Some researchers used the features of blood sugar, pulse rate, Glasgow coma scale, respiratory rate, chronic nervous scale (CNS) are studied and analysed [14], [15]. Significant work has been reported for identifying different techniques to select big data features in the healthcare. A feature may uniquely identifies data and hence used to detect a specific disease. SLR was followed for the conduction of the proposed research due to the reason that it systematically collect, analyse and derive managing insights from the published material for the defined range of time and defined criteria while other methods (like survey) don't do so.

A. BIG DATA ROLE IN MEDICAL EDUCATION

In a layman terminologies "big data" means a huge amount of data. The researchers and teachers used big data in the medical curriculum for designing, analysis, planning, estimating, and ensure delivery of the teacher activities by enhancing healthcare education. Big data is commonly used in the medical education for improving the medical field. The visual analytic has many applications in combing the analysis of data and exploitation technique, human cognitive strength to perceive, information and knowledge representation, and identify visual patterns [16]. Medical data feature encapsulates all the basic information required for diagnosing a certain disease. Features are the static astute values presents in any big data systems. These data features are in conjunction with each other for a certain purpose. These data analytics help in identifying a disease for a specific patient based on its healthcare history. These data analytical values ultimately help the doctors and practitioners suggesting the accurate medicine for a certain disease for a particular patient.

The following sections discuss the proposed methodology followed by the conclusion section at the end.

B. RESEARCH PROCESS

Significant work has been reported in many research domains using SLR [17]. For objectively analysing a specific problem SLR is the alternate way to explore it. Several features exists for the healthcare big data to be analysed critically. These analyses include regression, classification, association, data mining, clustering, and many others. Semantic analysis results in semantically see, deliver, and assess all the available and published material relevant to the research questions formulated to provide a broader knowledge to the research community about a specific domain [17]. Based on the protocol selected for conducting the SLR [11], the activities are classified into three major classes includes; design of protocol, conduct the SLR, and the report evidence.

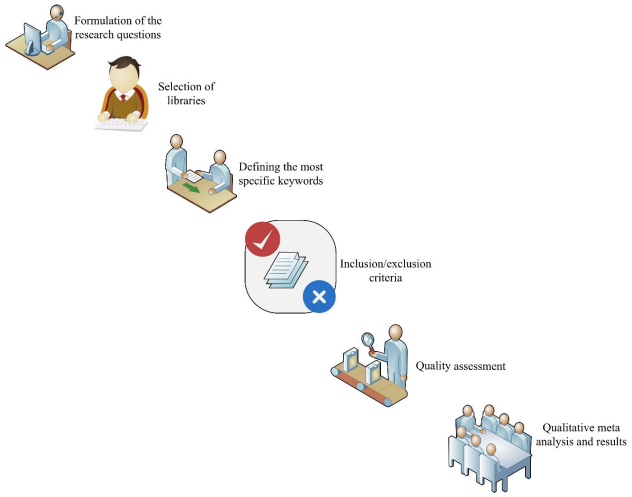

FIGURE 1. Steps of the SLR.

The next section provides details about the data accumulation, conducting SLR process, and quality assessment of the final selected papers.

C. RESEARCH DEFINITION

The main goal of the proposed SLR work is to perform in-depth study of the available literature in the field of healthcare. In-depth knowledge includes the study of big data features, and data analytics for the patient based on the history recorded. This data analytics and features helps the doctors and practitioners to identify the disease a particular patient based on the features (symptoms) recorded and suggest medicines accordingly. The systematic literature review used in this research work aims to perform a systematic and concise analysis of the healthcare big data to provide a simple and descriptive measures for the dives tools used in the medical and healthcare industries.

A sequence of steps required to perform a fruitful SLR work with objectives. Figure 1 depicts the number of steps taken in performing an SLR work [17].

D. RESEARCH PLAN AND METHOD

For the proposed research work an SLR protocol is followed based on the guidelines suggested by Kitchenham [11]. Figure 2 depicts the number of steps carried out to perform the proposed SLR process. The first step is the formulation

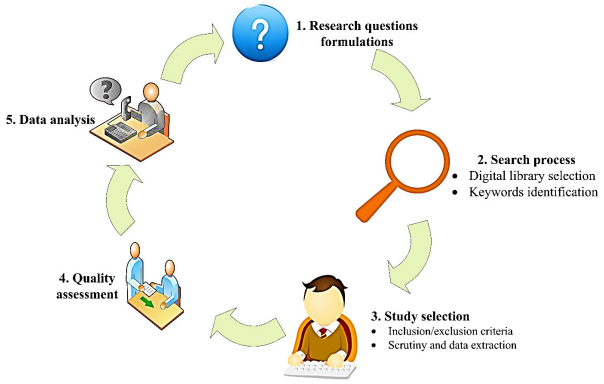

FIGURE 2. Research protocol followed the proposed SLR.

of research questions (5 research questions are selected), the selection of keywords for downloading the relevant papers from the selected digital libraries, the inclusion or exclusion criteria of the research articles (based on the contents provided in the paper), the quality assessment criteria based on the weighted numerical values assigned to each research question, and finally, the analysis of the data extracted from the included primary studies.

All these steps are discussed in detail below.

E. RESEARCH QUESTIONS

The research questions (RQ) selected for the proposed systematic literature review process are given below:

RQ1. What are the most significant big data features in the field of healthcare to be analysed? Big data has numerous features to be analysed especially in the field of healthcare. Researchers used the features blood sugar, pulse rate, Glasgow coma scale, respiratory rate, chronic nervous scale are studied and analysed. In depth analysis of the healthcare features will be studied.

RQ2. How many state-of-the-art techniques developed in the field of big data in healthcare during the census 2015 – 2019? This research questions focuses on describing how many available research based approaches are developed for healthcare big data for the retrieval of relevant information.

RQ3. For efficient big data management in the healthcare; the researchers proposed how many optimum solutions? The effective solutions will be studied as part of this research question to.

RQ4. What are the applications of the big data in scientific programming? Applications in the area o healthcare will be find from the existing literature to show the areas for further exploration.

RQ5. What are the big data analytics in healthcare? Healthcare big data analytics goal is to model, predict and inference, classification, clustering, regression, and other generic approaches which will be exploited in this research question.

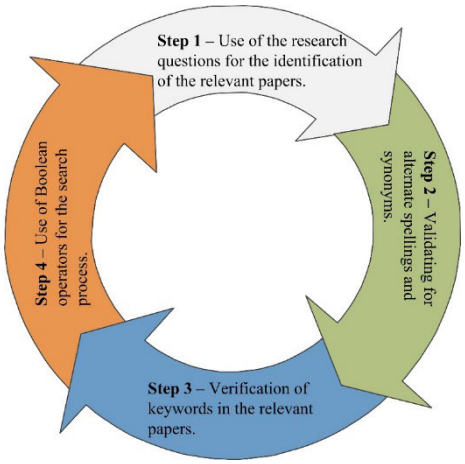

FIGURE 3. Steps involved in the search process.

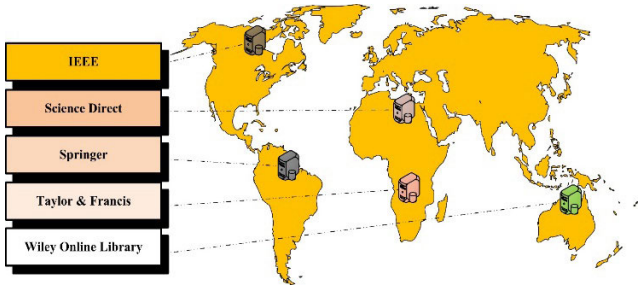

FIGURE 4. Libraries for the proposed research.

F. SEARCH PROCESS

For conducting an SLR process, a proper mechanism must be followed to ensure the accumulation of relevant studies from the selected digital libraries. A systematic mechanism is followed for accumulating the most relevant articles to the proposed field by formulating a set of most specific keywords. That was used for the search purposes in the selected peer review digital libraries for the retrieval of research articles (conference papers, journal papers, book section, survey papers, and many others). In the proposed SLR work, several keywords are formulated associated with big data features identification and analytics in the field of healthcare according to the research questions (given in section 2.3) were searched in the mentioned libraries shown in Figure 4. Figure 3 depicts the steps followed for the proposed study.

Figure 4 represents the digital libraries followed for collection of the relevant primary studies based on the keywords formulated. These libraries were selected due to the reasons that it is the most commonly used libraries and are publishing quality materials.

The authors defined the keywords for searching the relevant articles in the libraries. These keywords were kept specific, and short words are chosen for the said task. Using a combination of words instead of using shorten keywords results in huge bulk of articles such as using (big data in healthcare or healthcare big data) results in 2042 links in

TABLE 1. List of keywords selected for searching.

("Healthcare" OR "Medical") AND ("big data" OR "Big Data" OR "Medical Big Data") AND ("Management" OR "efficient management") AND ("Analytics" OR "Big data analytics") AND ("Programming" OR "Scientific programming")

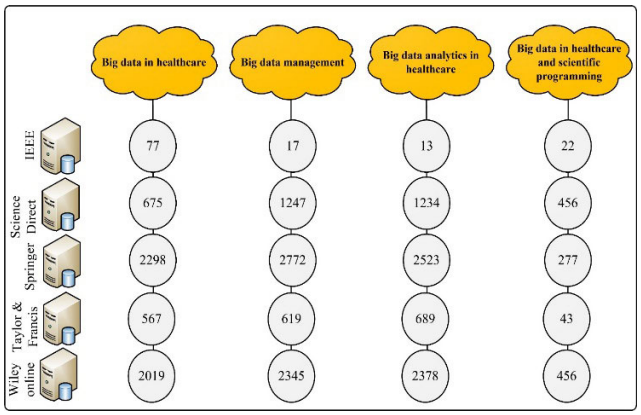

FIGURE 5. Search process in the given libraries based on keywords.

IEEE, 50775 in ScienceDirect, 198238 in Springer, 13852 in Taylor & Francis, and 37289 in Wiley Online Library. This huge information shows that the analysis and assessment, inclusion or exclusion process of every individual paper is a hectic job. To address this problem and accumulate only the relevant papers for the inclusion/exclusion process, and for the analysis and assessment process, the keywords are kept very specific as shown in table 1. These keywords were selected by the authors based on the title of the proposed study and research questions.

The searching process limited to the years ranging from 2015 to 2019 (a section of 2020 is included). The search process results research articles in the form journal publications, books, conference, workshops papers, and many other available materials. Based on the keywords formulated all the selected digital repositories were search manually for the accumulation of the relevant articles. Bibliographic details are stored in the EndNote tool [18]. Overall search process is depicted in figure 5.

For accumulating the relevant articles from the selected libraries a separate folder was created in the root directory and a total of 19789 relevant titles were found. Firstly, each folder is sorted out manually and all the downloaded articles were named based on their titles. This process helped in removing duplicate articles that ultimately helps in saving time during the quality assessment phase. After manual filtering, a total of 527 articles were finalized based on the title. Then the articles were manually checked based on the abstract provided and a total of 339 relevant articles were finalized. Then these papers are finalized for the quality assessment process based on the contents provided in the research articles. A total

TABLE 2. Inclusion/exclusion criteria of the relevant articles.

Inclusion
• Ensure the paper is published in the years ranging from 2015-2019 (a part of 2020 included) • Whether the contents of the paper provide satisfactory information for the selected topic? • The article is written in the English language • Select only the primary studies from the relevant studies • The selected articles provides a sound knowledge from the research questions formulated • The article must from the selected digital repositories
Exclusion
• The papers not in the range of January 2015- 2019 (a part of 2020 included) • Gray papers • Research papers less than three pages are excluded • Duplicate versions • Provides no information for the research questions selected • Written in other than English language

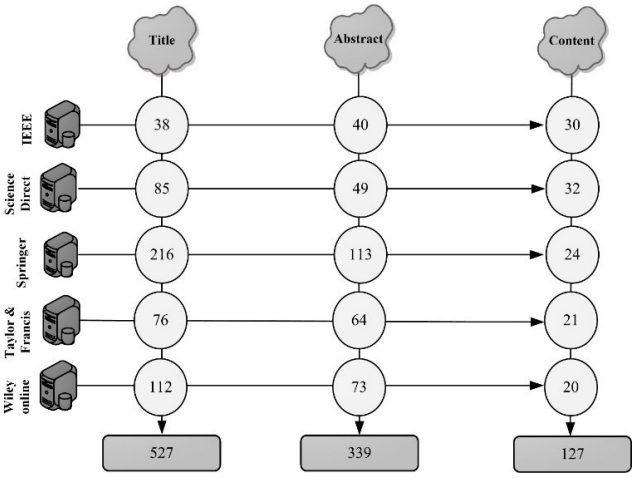

FIGURE 6. Filtering of papers by title, abstract, and contents.

of 127 paper were finalized that were the most relevant papers to the proposed topic. The process of inclusion/exclusion of papers is a hectic job because all these steps were performed manually. Whole this process retackled is depicted in figure 6.

EndNote tool is used for managing the bibliographic details for the final selected papers (127 papers).

G. STUDY SELECTION

The search process was performed on the selected peer reviewed digital libraries to extract the most relevant papers to the research questions formulated. A large number of papers (19789 peepers) were accumulated from these libraries that need further refinement to make a final pool of the most relevant papers for the assessment phase. An inclusion and exclusion process is performed to the accumulated articles. The authors the following inclusion/exclusion criteria for the inclusion of the papers in the final pool:

TABLE 3. Selection of primary studies.

Online libraries	Keywords based search results	Filtered by title	Filtered by abstract	Filtered by contents
IEEE	65	38	40	30
ScienceDirect	3833	85	49	32
Springer	7785	216	113	24
Taylor & Francis	1907	76	64	21
Wiley Online	7505	112	73	20
Total	21095	527	339	127

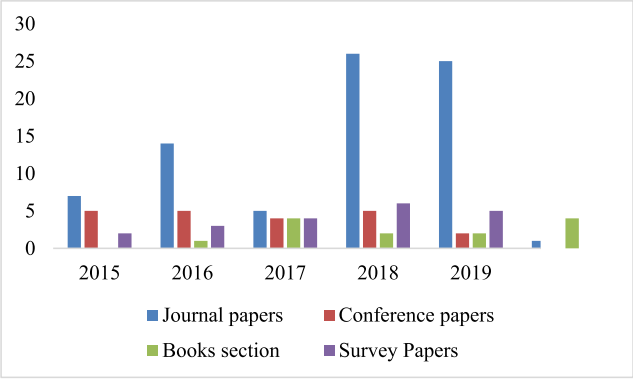

FIGURE 7. Details of selected papers.

- The papers provide a sound knowledge for the healthcare big data in terms of management, application, and scientific programming.
- The papers provide comprehensible details and context need to cover and answers the research questions defined in the research.

H. STUDY SELECTION PROCESS

In the systematic literature review works article selection is a crucial task, because at every stage the authors are confused whether an article should be skipped or included in the final pool of the selected relevant articles. This is the main step in SLR process, as it results in prominent assessment analysis. to perform the article selection phase we divided this step into three phases. Firstly, the relevant articles are selected based on the titles. A total of 527 relevant articles were finalized based on its titles. Secondly, these titled based selected papers were filtered again based on the abstract provided in the finalized articles. It results in 339 relevant articles. Third and the last stage is to filter the resultant articles (selected based on title and abstract) based on the contents provided in the relevant articles. It results in a total of 127 most relevant articles to the proposed field. All this process is done manually by the authors. The inclusion and exclusion criteria are shown in table 2.

Table 3 represents the process of filtering the primary studies for the proposed SLR process.

Figure 7 depicts the types of paper (conference paper, book section, journal paper, survey paper) after applying the exclusion and inclusion criteria [19]. Total number of papers by type and by year is depicted in Figure 7.

TABLE 4. Year-wise representations of selected papers.

2015	2016	2017	2018	2019	2020
[20]	[21]	[22]	[23]	[24]	[25]
[26]	[27]	[28]	[29]	[30]	[31]
[32]	[33]	[34]	[35]	[36]	[37]
[38]	[39]	[40]	[41]	[42]	[43]
[44]	[45]	[46]	[47]	[48]	[49]
[50]	[51]	[52]	[53]	[54]	
[55]	[56]	[57]	[58]	[59]	
[60]	[61]	[62]	[63]	[64]	
[65]	[66]	[14]	[67]	[68]	
[69]	[70]	[71]	[72]	[73]	
[74]	[75]	[76]	[77]	[78]	
[79]	[80]	[81]	[82]	[83]	
	[84]	[85]	[86]	[87]	
	[88]		[89]	[90]	
	[91]		[92]	[15]	
	[93]		[94]	[95]	
	[96]		[97]	[98]	
	[99]		[100]	[101]	
	[102]		[103]	[104]	
	[105]		[106]	[107]	
			[108]	[109]	
			[110]	[111]	
			[112]	[113]	
			[114]	[115]	
			[116]	[117]	
			[118]	[119]	
			[120]	[121]	
			[122]	[123]	
			[124]	[125]	
			[126]	[127]	
			[128]	[129]	
			[130]		
			[131]		

Year-wise distribution of the selected primary studies is shown in table 4.

Figure 8 represents the total number of papers per years. It is depicted from figure 8 that after 2015 the number of paper increases gradually that reflects the applicability of the big data in the fields of healthcare and medical applications.

Figure 9 represents the percentage contribution of every online library.

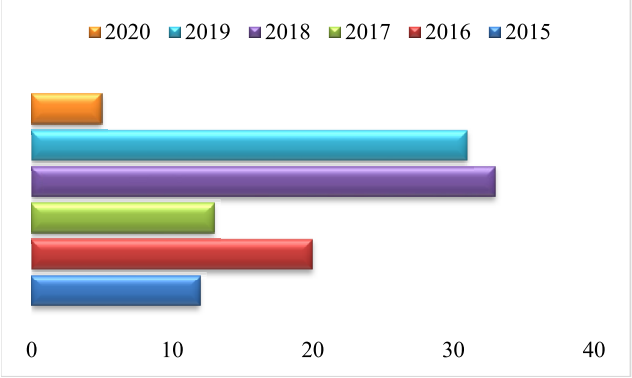

FIGURE 8. Trend of healthcare big data research from 2015–2019 (a part of 2020 included).

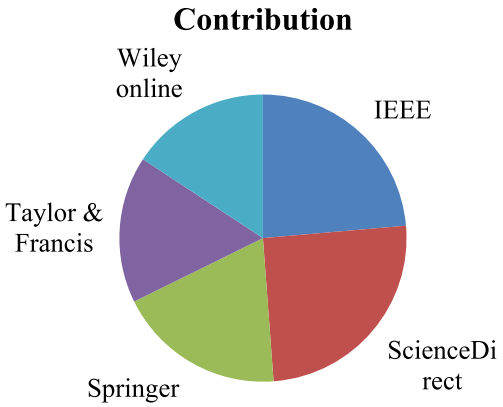

FIGURE 9. Percentage contribution.

I. QUALITY ASSESSMENT

After the inclusion and exclusion phase the quality assessment process is performed on the final selected articles based on the criteria defined as below. This process is iteratively implemented for all the selected papers for all the five research questions.

QR1. The papers provide a sound knowledge about the big data features in the healthcare.
QR2. The papers provide a detail of work reported during the period 2015 2019 (a portion of 2020 is included).
QR3. The papers provide state of the art technique for efficient data management and controlling.
QR4. The papers emphasize the applications of scientific programming in the field of healthcare big data.
QR5. The papers describe medical big data analytics in healthcare.

The authors analysed every papers manually and assessed each paper based on the criteria given below:

- 0 – if the paper fails in answering a particular research question.
- 0.5 – if a question moderately/partially satisfied in the paper.
- 1 – if a question satisfactory discussed in the paper.

All the selected primary studies are evaluated based on the assessment criteria defined for each question. Figure 10 shows the assessment results for every article.

After performing the assessment process and assigning weighted values to each article based on the research questions, all the assessment values are added to find out the most relevant papers to the proposed field. After performing this operation it was find out that the summed value of a paper is greater than or equal to 4 shows the most relevancy of the paper to the targeted field. The list of the most relevant papers is shown in figure 11 below.

J. DATA EXTRACTION

After performing the search process on the targeted online libraries, performing the quality assessment phase, and accumulation of the most relevant papers all the analysis are stored and inspected. The significant information retrieved from the assessment phase and inclusion/exclusion phase in the form of table or figures explained below;

- Figure 7 depicts the total included research articles along with required information.
- Table 3 depicts the annual distribution of the articles ranges from 2015 – 2019 (a part of 2020 is also included)
- Figure 8 gives results for total number of papers on yearly basis.
- Table 5 gives information for the significant healthcare big data features.
- Table 6 provides information about the research work reported since 2015 in the field of healthcare big data.
- Table 7 shows state of the art techniques developed for healthcare big data processing and information retrieval.
- Table 8 depicts the applications of healthcare big data, and scientific programming.
- Table 9 shows the healthcare big data analytics and its applications.

III. RESULTS AND DISCUSSION

The EPMA made a joint venture of 45 countries to work for the implementation and improvement of the promotion of PPP medicine [6]. The European programme 'Horizon 2020' made a commitment to combine the status of the professionals of PPPM toward the new lifelong instruments for the progress of technology and science in health related medical services.

In the upcoming sub-sections each research question is discussed and the relevant articles are classified based on the research questions formulated. A description of every research article is provided based on the research question. A total of 127 relevant primary studies are selected according to the inclusion and exclusion criteria as depicted in Table 4.

A. CONCEPT AND DEFINITION OF MEDICAL BIG DATA

There is no context name suggested for the "healthcare big data", but for ease of use and interpretation purposes classified into 5V structure. Figure 11 shows a 5V structure of the big data.

During the last few years a drastic change has been recorded in the field of healthcare due to the occurrence of smart IoT based devices and global communication systems.

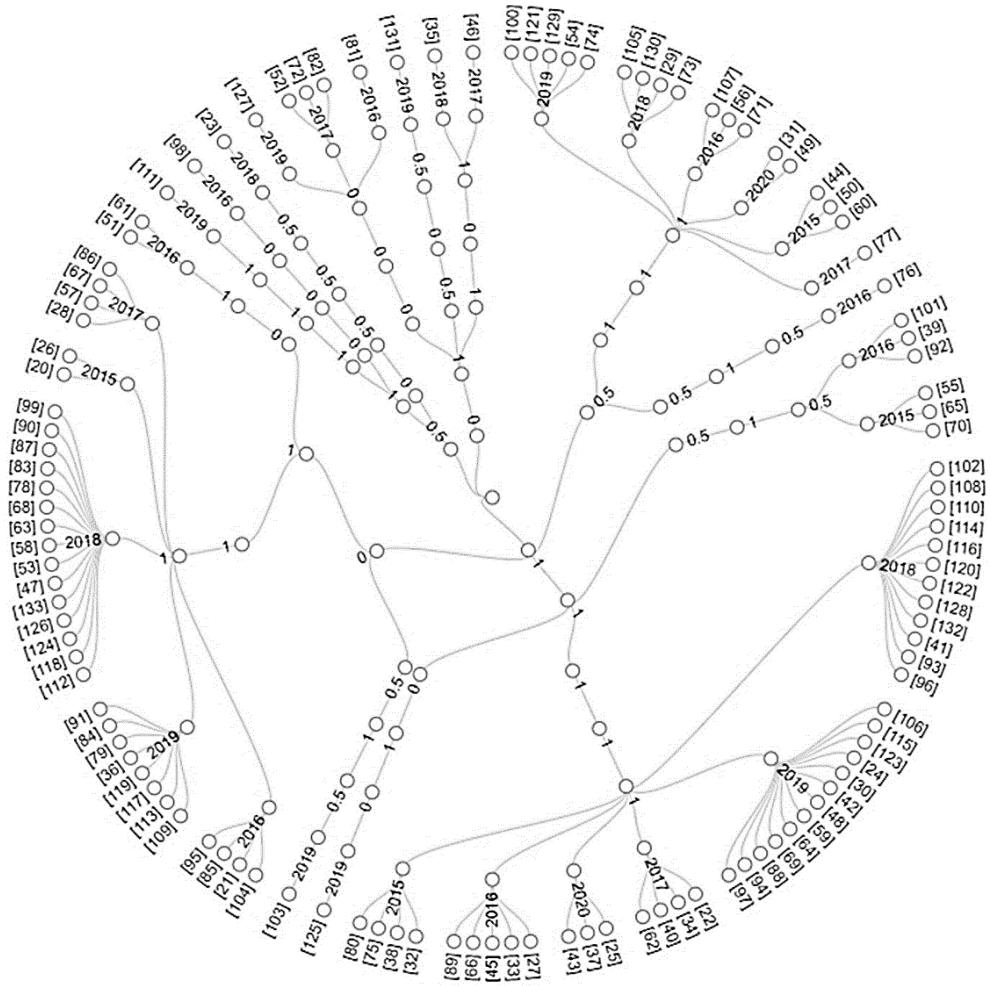

FIGURE 10. Quality evaluation of the selected papers.

In this digital age the information increases exponentially. Wearable devices continuously produce a vast amount of data that ultimately known as big data in layman's terms. The big data is the data whose diversity, scale, and complexity need new structure, algorithm, technique and analytics for the management, visualization and to pull-out hidden information. The idea of implementing big data analytics in the field of healthcare suggested for the retrieval of optimum information from complex data by applying different data mining and machine leering and neural network techniques [132]. These techniques ensure the quality of care, reducing care costs, and ensure minimal error rates in the care. McKinsey Global Institute [133] stated that, if the data is utilized efficiently, then the US health care can make an addition than $300 billion every year, of which 2/3 would be reducing expenses by about 8% in the healthcare. Multiple healthcare big data sources are available globally such as; health registration data, EMR data, patient reports of imaging, CT Scans, MRI, and X-Rays, biometrics, biomarkers data, clinical, and other administrative data [3]. The complication in healthcare results in misleading the doctor and practitioners in the prediction, prevention, and processing the healthcare big data for accurate purposes [6].

B. RQ1. WHAT ARE THE MOST SIGNIFICANT BIG DATA FEATURES IN THE FIELD OF HEALTHCARE TO BE ANALYSED?

Big data has numerous features to be analysed especially in the field of healthcare. Some researchers used the features of blood sugar, pulse rate, Glasgow coma scale, respiratory rate, chronic nervous scale (CNS) are studied and analysed [14], [15]. This analysis process consists of classification, features mining, association, clustering, and regression. Table 5 summarizes the work done reported for the question 1.

C. RQ2. HOW MANY STATE-OF-THE-ART TECHNIQUES DEVELOPED IN THE FIELD OF BIG DATA IN HEALTHCARE DURING THE CENSUS 2015 – 2019?

This research questions focuses on describing how many novel based approaches are developed for healthcare big data classification and for the retrieval of relevant information. Several diverse approaches are suggested by the researchers around the world for healthcare big data analysis, ensuring security and risk, storage requirements, and many others. Table 6 depicts the work reported during the years since 2015 – 2019 (a section of 2020 is included).

TABLE 5. Representation of the work done in the area of big data features.

S. No	Citation	Technique	Description
1	[32]	Big Data for Nursing staff training	Focuses on the emergence of federal big data initiatives, and exploration of exemplars from nursing informatics re-search to benchmark where nursing is already poised to participate in the big data revolution.
2	[55] [124]	Semantic transformation model	Globally, standardized healthcare documents are being used and process exponentially. The companies face several challenges in its comprehension and large scale analysis. To tackle this issue Hadoop based semantic transformation model is developed and clinical documents architecture (CAD) standards for the case studies.
3	[69]	Silico medicine solutions	Several big data features are considered for the proposed solution such as; Such requirements are: working with sensitive data; analytics of complex and heterogeneous data spaces, including nontextual information; distributed data management under security and performance constraints; specialized analytics to integrate bioinformatics and systems biology information with clinical observations at tissue, organ and organisms scales; and specialized analytics to define the "physiological envelope" during the daily life of each patient.
4	[27] [99]	Social network for data collection	Social network is an alternate solution for accumulating data about different problems especially for healthcare. Social media gives an easy access to all the users without any interruption and wait compared to the situations where the younger or the elders wait at the conventional social health poll stations. A Facebook application is developed where more 1400 users were recruited to collect data about adherence to the Mediterranean diet, considering the significant risk of cardiovascular and neurological degenerative diseases in subjects with poor adherence to a healthy diet. This data was collected in a short time without any promotional action.
5	[77]	Big data features	The book chapter presents two other features to 5V (volume, velocity, veracity, value, and variety) features. These features include vulnerability and complexity.

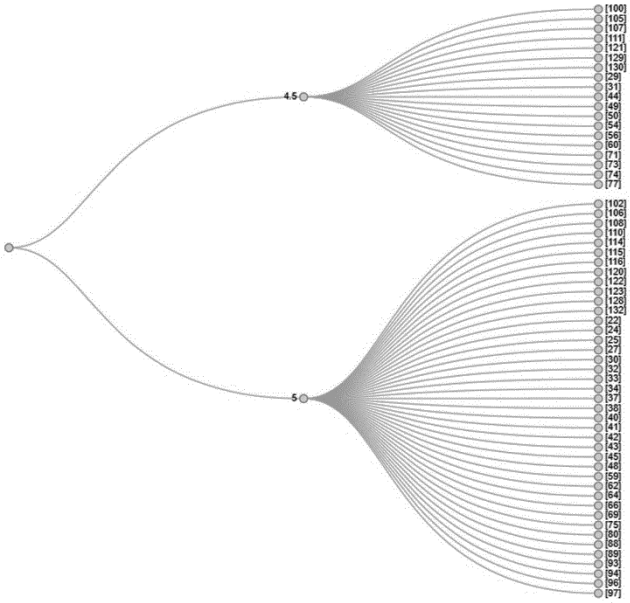

FIGURE 11. Most relevant papers to the proposed field.

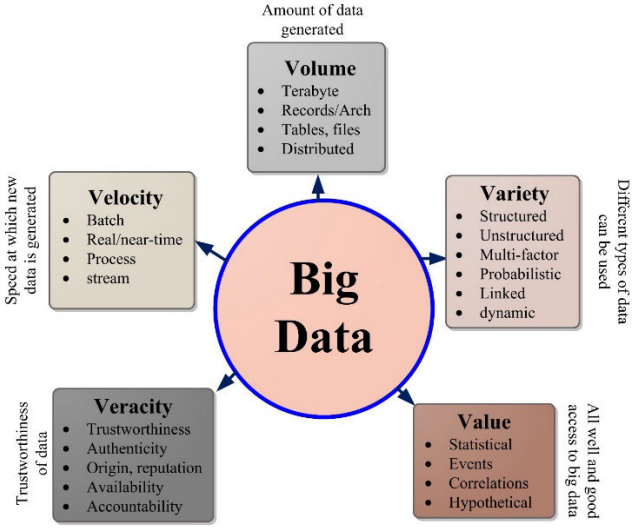

FIGURE 12. 5 Vs of big data.

D. RQ3. FOR EFFICIENT BIG DATA MANAGEMENT IN THE HEALTHCARE; THE RESEARCHERS PROPOSED HOW MANY OPTIMUM SOLUTIONS?

In layman terminologies the ''Bid Data'' means a huge amount of data. This huge data comes from many sources like industries, companies, aerospace services and research works, healthcare, smart cities, environmental changes, forecasting and many others, but in this research work the focus of research is the healthcare big data. This big data comes from imaging devices like (X-ray machine, MRI, CT-Scan machines and so on), IoT based applications, Electronic health reporting devices (EHR) devices and many other ICT based systems. These devices provide a huge amount of data on daily basis. Dealing with such a vast amount of data for extracting the most relevant information, and its

TABLE 6. Representation of the work done in the area of healthcare big data from 2015–2019.

S. No	Citation	Technique	Description
1	[44] [39]	Embedding data science to make the healthcare green	The emergence of the smart IoT devices, and the new communication technology challenges the world by facing a vast amount of data (big data). Dealing with such a huge amount of data using traditional mathematical formulas for retrieving the required results within a limited time span is about impossible. So to resolve this problem data science is applied in healthcare applications to make it green and efficient.
2	[60]	Hidden Markova model	The occurrence of new cities during the 21st century pressurized the utilization of resources and the standards of urban life. New ideas and techniques are needed to handle network traffic and basic healthcare needs in the populous cities. Additionally, a huge amount of data is available on daily basis for investigation. To handle this problem this research paper uses the concept of hidden Markova model for the big data mining purposes.
3	[33] [93] [34] [120]	Cloud computing	A vast amount of data is generated on the daily basis from sensor based technologies, imaging, electronic health report (EHR) and many others. Dealing with such a vast amount of data for extracting the rich information is a hectic job. To address this issue this research work focuses on by applying the cloud computing services that will ultimately acts as a benchmark to substantiate big data which may lead to discover of hidden patterns and trends to enhance knowledge for progression of disease.
4	[61] [14]	Statistical data analysis techniques	This research work focuses on applying statistical data analysis techniques for the detection of Alzheimer's disease based on Chronic nervous system using big data. This technique also capable of converting big data to smart data for efficient processing and management.
5	[66]	Distributed wearable sensors based model	In this populous environment the elders want to spend their lives independently, rather than relying on the intrusive care programs. Distributed wearable systems are used to record the behaviour of the humans using hidden Markova model. Locality Hashing mechanism is used to record the patterns of the sensors.
6	[75]	Machine learning techniques for bioinformatics	Bioinformatics in classified into incremental databases, voluminous, and complex data analytics techniques. These datasets such as; DNA, RNA and many others contains a huge information in other words termed as big data. For extracting the enriched information from these datasets machine learning techniques are used for fast and efficient processing.
7	[80]	Multimedia based mobile health application	This paper focuses on proposing two mHealth applications, can be considered as the terminals of big data based mHelath system for accumulating the information for electronic medical records (EMR) systems. The first mHealth application is a hybrid model for improving the user experience in hyperbaric oxygen chamber, while the second one is the voice interactive serious game for the therapists.
8	[96]	Hybrid model	This paper combines the internet of things (IoT), sensing technologies, and cloud of sensing, big data analytic systems for emerging technologies that ultimately results in cost efficient solution, low data storage requirements, network communication speed, and computation requirements.
9	[99]	Unsupervised machine learning techniques	This paper has analysed data from Facebook walls of 153 public organizations to get knowledge about user engagement and performances based on unsupervised machine learning techniques.
10	[105]	Phrase sense disambiguation	Electronic medical records (EMR) contain complex information and are difficult to access for the future use. To access this data and use it for future purposes annotated conversion is required. This paper proposes a model that converts the EMR data to annotated formats indirectly without interrupting the semantics of the data. To achieve high accuracy rats the phrase sense disambiguation (PSD) is applied.

TABLE 6. *(Continued.)* **Representation of the work done in the area of healthcare big data from 2015–2019.**

11	[62] [103]	MF-R and GC architectures	In this modern technological age wearable medical devices play a key role in many daily life applications such as weather forecasting, individual health monitoring on continuous basis, traffic flow monitoring, and smart home applications. The sensor from these wearable devices generates a huge amount of data that stores on the cloud networks. To process this huge amount of data for extracting enriched information this chapter proposed an IoT based model that uses Metafog-Redirection (MF-R) and Grouping and Choosing (GC) models for efficient data processing.
12	[47]	Distributing and self-organizing algorithm	During the last few years the exponential development and implementation of social media applications making available a continuous enormous data stream of individual information. Several disease surveillance systems, clinical trials and other healthcare applications are used to analyse data, but these applications are good in small amount of data and fails for a huge amount of data. To address this issue content delivery networks (CDNs) are proposed, but these models are failed in the dynamic and continuous data. To overcome this issue the authors presented a distributed and self-organizing algorithm for dynamic data analytics for the individual health perspectives.
13	[53] [94] [117]	Formal Methods, Artificial Intelligence	A novel approach is developed by using formal methods to develop medical care framework. This framework contains no programming phase. Encapsulate the medical intelligence they converted the medical pathways to multipartite directed weighted graph (MDWG). The autonomous interpreters in the server presents ***natural language generator*** (NLG) pathophysiology questions a doctor would normally ask a patient to understand the signs and symptoms of a disease. These models are also suggested for cardiovascular disease (CVDs) detection and medicine suggesting by the practitioners.
14	[67]	Fog-Cloud based architecture	This paper proposed a cloud based architecture named Fog-Cloud based model for energy efficiency in healthcare. This model reduces the energy consumption, latency and delays in the IoT applications.
15	[82] [112] [116] [90] [107] [125]	Machine learning and deep learning techniques	This paper presents a twofold model for healthcare big data. In the first section they provide the issues in the current mobile healthcare system, while in the second phase mHelath 2.0 they proposed deep neural network and machine learning techniques for big data processing and information retrieving purposes that ensures the data integrity, time saving, efficient, and manageable solution to make the healthcare green. These techniques are also applicable in intensive care units for critical patients health monitoring.
16	[110]	Policy development toolkit	Healthcare big data analytics plays a vital role for the healthcare big data and for healthcare policymakers. This research has presented a web based application known as policy development toolkit (PDT) that aims to promote the exploitation of heterogeneous holistic health data and pre-existing knowledge on public health policy models. The proposed approach incorporates mechanisms for forecasting, causal and risk analysis, as well for the compilation and visualization of predictions.
17	[130]	Adaptive clinical decision support system	This paper integrates cloud computing technology and the case-based reasoning (CBR) technique, medical records and documents pertaining to the elderly can be captured in real time, whereas appropriate treatment plans based on past similar treatment records can be formulated.
18	[78]	Cognitive data transmission model	This paper presents the cognitive data transmission model (CDTM) for data transmission and patient health relevant data.
19	[83]	Multi-objective programming and prospect theory	Instead of the conventional patient-doctor appointment system they presented a new state of the art model named intelligent hospital appointment system where the patient fix an appointment with the doctor based on expertise using healthcare big data (data bank). This system is verified by the universal first-come-first-serve approach.
20	[37]	BrownBoost Classifier	Big data analysis in the field of healthcare means to accumulate data from various sources, and process patient's history for useful information. But the researchers face many challenges in the form of processing, storing and retrieving useful information. Especially the storage requirement for such a vast amount of data is a big challenge for the researchers. To address this problem the proposed research work presents a BrownBoost Classifier-Based Bloom Hash Data Storage (BBC-BHDS) mechanism for storing and accessing the healthcare data from various locations in distributed environment with minimum space usage and in less timefor Healthcare Big Data Analytics.

TABLE 7. Research work done reported for efficient big data management in the healthcare.

S. No	Citation	Technique	Description
1	[44] [39]	Applications of data science in healthcare	Basics and potential of data science are discussed. The challenges and evaluations of the needs of data science techniques in the healthcare field to make it prosper and green.
2	[50]	Analytical framework of big data for Lupus disease	Analytical framework of big data has been suggested for the prediction of Lupus disease.
3	[33] [93] [34] [98]	Cloud computing	This paper focuses on applying various aspects of cloud based services to enable big data analytic in healthcare big data management system.
4	[51]	Bio-Inspired ICT	In the near future the life and death will be mostly depends upon the data and its organization. While the organization of a huge amount of data that is generated from the imaging, IoT devices and other electronic health repots devices and many others is a crucial task. To overcome this issue this paper has presented a Bio-inspired ICT based application for providing a smart healthier system for the caretaker and the patient.
5	[61]	Big data to smart data conversion	This research work focuses on applying statistical data analysis techniques for the detection of Alzheimer's disease based on Chronic nervous system using big data. This technique also capable of converting big data to smart data for efficient processing and management.
6	[84]	Big data in cloud	This research focuses on developing a novel storage architecture capable of providing ease in read, write or update functions contrary to the conventional database models that are full of anomalies. For the proposed work they have used application cluster mechanism for storing and retrieval of data form the user.
7	[96] [114]	Hybrid model	This paper combines the internet of things (IoT), sensing technologies, and cloud of sensing, big data analytic systems for emerging technologies that ultimately results in cost efficient solution, low data storage requirements, network communication speed, and computation requirements.
8	[22] [28] [92] [122]	Big data security and privacy	This paper provides a review of different techniques developed for big data privacy and security.
9	[40]	SOA-FOG based data privacy and management	This paper presents three tier architecture for big data management using SOA-FOG architecture (fog devices). It gives security information in the client layer, fog layer and cloud layer.
10	[71]	Data analytics in IoT applications	This paper provides review study of different primary studies for resolving the security, management and efficient use of healthcare big data using cloud computing, internet of things, and large scale data processing system.
11	[41] [63] [72] [119]	Big data analytics in operation management	In this research article the authors investigate the big data analytical techniques for their strengths, weaknesses, and major functionalities. Then they provide a literature study by discussing how many different operation management techniques are followed for different big data applications such as weather forecasting, urban traffic management, marketing, supply chain management, risk analysis and avoidance, revenue management, and inventory management.
12	[58]	Clinical intelligence approach	Clinical intelligence approaches based on data mining and process data mining for retrieving astute information.
13	[86]	mHealth 4.0	This paper presents a survey of the industry 4.0 to address the issues in the current mHealth system and proposed a new healthcare mobile smart system (mHealth 4.0) to provide efficient big data healthcare solution for patients.
14	[89]	Machine learning based healthcare big data security model	This paper proposed a hybrid techniques based on masking encryption, activity monitoring, granular access control, dynamic data encryption and end point validation models. The proposed hybrid healthcare model results in providing more effective disease diagnostic healthcare big data system.

proper management is a difficult job. Different techniques are suggested by multiple researchers around the world for extracting the enriched information from this huge amount of data and its efficient management. This research question

TABLE 8. Representation of the applications of medical big data in the healthcare.

S. No	Citation	Technique	Description
1	[32]	Applications of Big data in Nursing	Focuses on revolutionizing the nursing staff by applying big data analysis and data science for patient treatment based on the previous history.
2	[50]	Lupus disease detection system	They have implemented big data analytics for the Lupus disease detection.
3	[69]	Silico medicine solutions	This research work focuses on integrating big data analytics with VPH technologies to produce robust and effective in silico medicine solutions for individuals.
4	[74]	Software tools and techniques	Real time applications are developed by combining Microsoft Azure and Amazon with new communication models for solving healthcare problems in a short span of time.
5	[45]	Big data programming tools	This paper focuses on analysing the healthcare dataset against varying queries using Pig Latin Script, to improve the quality of health in India by analysing the medical big data.
6	[61]	Alzheimer's disease detection	This research work focuses on applying statistical data analysis techniques for the detection of Alzheimer's disease based on Chronic nervous system using big data. This technique also capable of converting big data to smart data for efficient processing and management.
7	[46]	Patient health risk assessment and stratification models	This paper provides an overview of different eHealth applications and provides a model for patient health risk assessment and stratification.
8	[14]	R programming	This paper presents the R programming tool for big data analytics and uses the features of blood sugar, pulse rate, Glasgow coma rate, respiratory rate, and CNS. A classification model is developed using the R programming tools statistical and visual tools.
9	[81]	EVOTION	A new European research journal aims to provide an integrated solution to the hearing loss (HL) patients. This journal also aims to provide a holistic management to the HL patients by providing public health policies and prevent the user from future this type diseases.
10	[53]	DocDx	A novel approach is developed by using formal methods to develop medical care framework. This framework contains no programming phase. Encapsulate the medical intelligence they converted the medical pathways to multipartite directed weighted graph (MDWG). The autonomous interpreters in the server presents ***natural language generator*** (NLG) pathophysiology questions a doctor would normally ask a patient to understand the signs and symptoms of a disease.
11	[103]	Apache Pig and Apache HBase	During this digital technological age wearable devices provides big data in the form of transportation, smart healthcare, smart cities, and many others. Dealing with such a huge amount of data for storage and processing is a big challenge for the researchers. Apache Pig and Apache HBase are the databases provide an opportunity to the researchers to save their records for future purposes.
12	[112]	Apache Spark	In this research work data processing engine, the Apache Spark deployed in cloud mainly focuses on applying machine learning technique for predicting patient's health status based on the tweets (related to patient's health).
13	[128]	Hadoop framework	This research work presents a statistical assessment process for diabetic's analysis using big data. Difference performance measures such as accuracy and F-measure values are evaluated from the proposed statistical assessment suing Hadoop framework.
14	[42] [54] [111]	Internet of things	In this modern technological age and over-populous environment accessing the basic caretakers for the treatment purposes is a big challenge for the patient and especially for the old age patient. In such case the smart interment of things (IoT) based applications provide basic treatment facilities at door steps.
15	[83]	Multi-objective programming and prospect theory	Instead of the conventional patient-doctor appointment system they presented a new state of the art model named intelligent hospital appointment system where the patient fix an appointment with the doctor based on expertise using healthcare big data (data bank). This system is verified by the universal first-come-first-serve approach.
16	[101]	Smart HIV/AIDS digital system	This paper aims to develop a smart HIV/AIDS digital system by accumulating data from various sources about the said disease. Then applying this data to a specific location to monitor the HIV/AIDS patients in that specific region. This system as well as data ultimately helps the researchers, doctors, government, patients and many others for proper treatment.

TABLE 8. *(Continued.)* **Representation of the applications of medical big data in the healthcare.**

17	[117]	Parkinson disease detection	Big data has accumulated a huge amount of data varying from smart transportation systems to healthcare system and robotics. In the proposed system the sensors are fixed in the Parkinson's affected patient. The robot is designed to move with the patient based on the sensor data for communication. Based on the sensor data the artificial neural network is applied for the detection purposes.

focuses on to extract the articles that have worked on the big data management (efficient use, control and retrieving abilities). Table 7 depicts the tools and techniques developed during the period ranges from 2015 – 2019 (a part of 2020 is also included).

E. RQ4. WHAT ARE THE APPLICATIONS OF THE BIG DATA IN HEALTHCARE, AND SCIENTIFIC PROGRAMMING?

Big data mining and analysis has unlimited applications in many fields of daily life such as; natural language processing, medical and healthcare applications, whether forecasting, climate change, aerospace research and many others. This research work mainly focuses on the healthcare big data and its applications in the healthcare. Table 8 represents some of the applications of the big data that are developed during the years ranging from 2015 – 2019 (a part of 2020 is also included).

F. RQ5. WHAT ARE THE BIG DATA ANALYTICS IN HEALTHCARE?

In the field of healthcare the major goal of the big data analytics is to model, predict and inference, classification, clustering, regression, and other generic approaches to be exploited [13]. Supervised learning mechanism is followed for healthcare big data analytical modelling. For healthcare improvements the Rumsfeld *et al.* [3] classified the big data analytics into eight major classes that are; (i) mathematical model for risk avoidance and usage of resources; (ii) disease and treatment heterogeneity; (iii) surveillance of drug and device safety in healthcare; (iv) management of population; (v) measurement of performance and quality care; (vi) clinical decision support and precision medicine; (vii) public health; and (viii) application of research.

Machine learning techniques and big data analytics are applied in the healthcare fields for risk management, clinical decision support, chronic disease detection, and for precision medicine using genomic information [134]. The authors presented Lean Six Sigma application for improvements of patients' satisfaction, processing errors in healthcare payer firm using Six Sigma, and services of Lean Six Sigma for morphological analysis of research literature [135]–[137]. Table 9 depicts the research work reported using big data analytics in the fields of healthcare.

IV. LIMITATIONS OF THE RESEARCH

A few threats to the validity for the proposed research work are listed below:

- A number of digital libraries exists for accumulating the research articles, but the proposed SLR work is limited to only five most extensively used libraries. This was decided to focus only on high quality peer reviewed articles.
- This research work is limited to specific number of years ranges from 2015 – 2019 (a part of 2020 is also included), but the papers are published on daily basis continuously.
- Google Scholar is skipped. Main reason behind this decision was; it gives access to every journal and paper, and to save time and access only the reputed journal papers.
- A research article may be skipped that contains the word healthcare, but have no concern with this word for the implementation purposes.

V. CONCLUSION

Healthcare big data analysis and management has becomes a challenging task for the researchers. On-time decision making in the healthcare systems requires massive potential for refining the care quality, minimize the cost of care, and decrease error and waste. Therefore, to address this issue an ideal way is to study and report the existing literature that will help the doctors and practitioners to take better decisions in early healthcare. The detailed study will eventually summarise the results of the available literature published related to big data in cardiology.

The current research is an endeavour toward comprehensive report on healthcare big data. The proposed study uses systematic literature protocol and guidelines as presented by Kitchenham *et al.* [11]. Data was collected from the work published during the year 2015 to 2019 (a portion of 2020 is also included) in the form of conference, journals, books, magazines and other online sources. A total of 127 most relevant papers to the healthcare big data are selected in the final pool for the quality assessment purposes.

This research work gives year-wise distribution of the included relevant articles ranging from 2015 to 2019 (a part of 2020 is also included). This paper outlines the significant features of the healthcare big data, state of the art techniques developed by multiple researchers for the identification of certain disease based on the feature map calculated for the healthcare big data, the applications of the healthcare big data, the proper management of the healthcare big data (management in the form of efficient storing, retrieving of certain information, updating and insertion of

TABLE 9. Research work reported for the healthcare big data analytics.

S. No	Citation	Technique	Description
1	[26]	Sensemaking in intelligent health data analytics	A systematic model is developed by for performing sense of health data. This model uses networked foresight technique for identifying the future medicine approach based on the current and the past data.
2	[38] [65] [70] [91] [102] [34] [57] [76] [23] [29] [35] [97] [106] [108] [110] [118] [126] [131] [24] [30] [36] [42] [48] [54] [59] [64] [68] [87] [15] [95] [101] [104] [107] [113] [115] [117] [119] [121] [125] [127] [129] [25] [31] [43] [49]	Big data analytics	A survey paper is presented in order to present an overview of big data analytics' content, scope and findings as well as opportunities provided by the application of big data analytics.
3	[44] [39] [123]	Data science and big data analytics for Healthcare	With rapid growth of information and communication technology and smart IoT based systems in the field of healthcare it is more difficult for the companies to retrieve the required information from a vast of data (big data). In order to address this issue Godbole and Lamb [44] combined data science with big data analytics to make healthcare green.
4	[50]	Big Data Analytics for Predicting Systemic Lupus Erythematosus	This paper deals with the possibilities of predicting the Lube disease by applying big data analytics.
5	[55] [124] [73]	Semantic transformation model	Around the world standardized healthcare documents are being used and process at an increasing order. The companies face several challenges in its comprehension and large scale analysis. To tackle this issue Hadoop based semantic transformation model is developed and clinical documents architecture (CAD) standards for the case studies.
6	[79]	Patient-Centric and Outcome-Driven Precision Health	Around the world many countries are interested in developing a novel effective ways to make sense of "big data" for evidence-based, outcome-driven, and affordable 5P (Patient-centric, Predictive, Preventive, Personalized, and Precise) healthcare. This research work focuses on multi-modal and multi-scale (i.e. molecular, cellular, whole body, individual, and population) biomedical data analytics for discovery, development, and delivery, including translational bioinformatics in biomarker discovery for personalized care; imaging informatics in histopathology for clinical diagnosis decision support; bionanoinformatics for minimally-invasive image-guided surgery; critical care informatics in ICU for real-time evidence-based decision making; and chronic care informatics for patient-centric health.
7	[21]	Applying big data analytics for assessing the quality of service in healthcare	This paper focuses on applying big data analytics to healthcare data accumulated from several sources to gain quality insights and grasp best practices of the field (using new healthcare specific big data tools).
8	[45] [56]	Big data programming for Indian healthcare analysis	Focuses on analysing the healthcare dataset against varying queries using Pig Latin Script, to improve the quality of health in India by analysing the medical big data.
9	[84]	Efficient Analytics of Health Informatics	This research work proposed an application cluster mechanism for data processing and efficient information retrieval from the database. This mechanism enables the read, written and update functionalities in the database with high efficiencies.
10	[88]	Quality healthcare accessibility	This research work classifies prior studies into Floating Catchment Area methodologies, a generic combination of methodologies that measure healthcare accessibility, and presents a framework that conceptualizes accessibility computation.
11	[99]	Big social data analytics for public health	This paper has analysed data from Facebook walls of 153 public organizations to get knowledge about user engagement and performances based on unsupervised machine learning techniques.
12	[105]	metadata annotation approach	Electronic medical records (EMR) contain complex information and are difficult to access for the future use. To access this data and use it for future purposes annotated conversion is required. This paper proposes a model that converts the EMR data to annotated formats indirectly without interrupting the semantics of the data. To achieve high accuracy rats the phrase sense disambiguation (PSD) is applied.

TABLE 9. *(Continued.)* **Research work reported for the healthcare big data analytics.**

13	[40]	SOA-FOG based data analytics	This paper presents three tier architecture for big data management using SOA-FOG architecture (fog devices). It gives security information in the client layer, fog layer and cloud layer.
14	[52]	Platform in healthcare IT	The book section focuses on describing state of the art of data analytics platforms and suggests possible applications and benefits in healthcare while cautioning against excessively utopian scenarios.
15	[14]	Big data analytic in healthcare using R programming tool	This paper presents the R programming tool for big data analytics and uses the features of blood sugar, pulse rate, Glasgow coma rate, respiratory rate, and CNS. A classification model is developed using the R programming tools statistical and visual tools.
16	[71]	Data analytics in IoT applications	This paper provides review study of different primary studies for resolving the security, management and efficient use of healthcare big data using cloud computing, internet of things, and large scale data processing system.
17	[85] [100]	Big data analytics in mobile healthcare market for economic purposes	In this over-populous world and still increasing exponentially results in restricting almost every patient to access their basic caretakers on time. This problem is then more critical in the developing countries. To address this issue smart phone based eHealth systems are developed to help the patient to access their basic health needs at door step. This not only helps the patient in accessing their caretakers guideline at doorstep, can get their treatment, but it also helps in increasing countries economy exponentially.
18	[41] [63] [72]	Big data analytics in operation management	In this research article the authors investigate the big data analytical techniques for their strengths, weaknesses, and major functionalities. Then they provide a literature study by discussing how many different operation management techniques are followed for different big data applications such as weather forecasting, urban traffic management, marketing, supply chain management, risk analysis and avoidance, revenue management, and inventory management.
19	[47]	Distributing and self-organizing algorithm	The social media applications making available a continuous enormous data stream of individual information. Several disease surveillance systems, clinical trials and other healthcare applications are used to analyse data, but these applications are good in small amount of data and fails for a huge amount of data. To address this issue content delivery networks (CDNs) are proposed, but these models are failed in the dynamic and continuous data. To overcome this issue the authors presented a distributed and self-organizing algorithm for dynamic data analytics for the individual health perspectives.
20	[53]	Big data analytics for reducing disease burden and health disparities	Medical errors and overtreatment are responsible for the exponential increase in the disease burden and health disparities when combined with non-communicable disease population. To address this issue critical safety system are used by applying formal methods to develop a system with the automation of zero defects.
21	[72]	Big data based medicine revolution	Accumulating information from daily life patients history and perform analytical operations to provide and suggest a revolution in the medicine field.
22	[82]	Machine learning and data analytics	This paper outlines the problem in the available mHealth system and proposed a new mHealth2.0 model for big data processing in the healthcare using machine learning model and deep learning models for data analytics.
23	[128] [109]	Statistical assessment for diabetics analysis	This research work presents a statistical assessment process for diabetic's analysis using big data. Difference performance measures such as accuracy and F-measure values are evaluated from the proposed statistical assessment suing Hadoop framework.

new data) and scientific programming, and the healthcare big data analytics. The exponential increase of the published research article on yearly basis concludes that the researchers show a keen interest in the field of healthcare big data analytics. This analysis on the existence research work will help the doctors and practitioners to make more authentic decisions, which ultimately will help to use the study as evidence for treating patients and suggest medicines accordingly.

IMPLICATIONS AND FUTURE DIRECTIONS

The proposed study have highlighted the most significant areas of research in healthcare big data and answered some research questions. Further, the implications and future directions for researchers and practitioners are to explore the area further to extract meaningful insights and information from the data for the researchers and practitioners to use in effective way in healthcare. This will require the effort of in-depth analysis of big data in healthcare.

CONFLICT OF INTEREST

The authors declared that there are no potential conflicts of interests regarding the article.

REFERENCES

[1] R. Bellazzi, "Big data and biomedical informatics: A challenging opportunity," *IMIA Yearbook Med. Informat.*, vol. 23, no.1, pp. 8–13, 2014.
[2] N. Khan, I. Yaqoob, I. A. T. Hashem, Z. Inayat, W. K. M. Ali, and M. Alam, "Big data: Survey, technologies, opportunities, and challenges," *Sci. World J.*, vol. 2014, pp. 1–18, Jul. 2014.
[3] J. S. Rumsfeld, K. E. Joynt, and T. M. Maddox, "Big data analytics to improve cardiovascular care: Promise and challenges," *Nature Rev. Cardiol.*, vol. 13, no. 6, pp. 350–359, Jun. 2016.
[4] N. Davarzani, S. S.-V. Wijk, M. T. Maeder, P. Rickenbacher, E. Smirnov, J. Karel, T. Suter, R. A. de Boer, D. Block, V. Rolny, C. Zaugg, M. E. Pfisterer, R. Peeters, and H.-P. B.-L. Rocca, "Novel concept to guide systolic heart failure medication by repeated biomarker testing—Results from TIME-CHF in context of predictive, preventive, and personalized medicine," *EPMA J.*, vol. 9, no. 2, pp. 161–173, Jun. 2018.
[5] J. Kinkorová and O. Topolčan, "Biobanks in horizon 2020: Sustainability and attractive perspectives," *EPMA J.*, vol. 9, no. 4, pp. 345–353, Dec. 2018.
[6] O. Golubnitschaja and V. Costigliola, "EPMA summit 2014 under the auspices of the presidency of Italy in the EU: Professional statements," *EPMA J.*, vol. 6, p. 4, Feb. 2015.
[7] O. Golubnitschaja, J. Kinkorova, and V. Costigliola, "Predictive, preventive and personalised medicine as the hardcore of 'Horizon 2020': EPMA position paper," *EPMA J.*, vol. 5, no. 1, pp. 1–29, Dec. 2014.
[8] O. Golubnitschaja, I. D. Watson, E. Topic, S. Sandberg, M. Ferrari, and V. Costigliola, "Position paper of the EPMA and EFLM: A global vision of the consolidated promotion of an integrative medical approach to advance health care," *EPMA J.*, vol. 4, no. 1, p. 12, Dec. 2013.
[9] S. B. Scruggs, K. Watson, A. I. Su, H. Hermjakob, J. R. Yates, III, and M. L. Lindsey, "Harnessing the heart of big data," *Circulat. Res.*, vol. 116, no. 7, pp. 1115–1119, 2015.
[10] W. Wang and E. Krishnan, "Big data and clinicians: A review on the state of the science," *JMIR Med. Informat.*, vol. 2, no. 1, p. e1, Jan. 2014.
[11] B. Kitchenham, O. P. Brereton, D. Budgen, M. Turner, J. Bailey, and S. Linkman, "Systematic literature reviews in software engineering–a systematic literature review," *Inf. Softw. Technol.*, vol. 51, no. 1, pp. 7–15, Jan. 2009.
[12] A. Lavecchia, "Machine-learning approaches in drug discovery: Methods and applications," *Drug Discovery Today*, vol. 20, no. 3, pp. 318–331, Mar. 2015.
[13] A. Sinha, G. Hripcsak, and M. Markatou, "Large datasets in biomedicine: A discussion of salient analytic issues," *J. Amer. Med. Inform. Assoc.*, vol. 16, no. 6, pp. 759–767, Nov. 2009.
[14] D. Pant, V. Kumar, J. Kishore, and R. Pal, "Healthcare data modeling in R," in *Proc. 1st Int. Conf. Intell. Syst. Inf. Manage. (ICISIM)*, 2017, pp. 230–233.
[15] S. Nazir, M. Nawaz, A. Adnan, S. Shahzad, and S. Asadi, "Big data features, applications, and analytics in cardiology—A systematic literature review," *IEEE Access*, vol. 7, pp. 143742–143771, 2019.
[16] C. Vaitsis, G. Nilsson, and N. Zary, "Visual analytics in healthcare education: Exploring novel ways to analyze and represent big data in undergraduate medical education," *PeerJ*, vol. 2, p. e683, Nov. 2014.
[17] B. Kitchenham, "Guidelines for performing systematic literature reviews in software engineering," Softw. Eng. Group School Comput. Sci. Math. Keele Univ., Keele, U.K. and Dept. of Comput. Sci. Univ. Durham, Durham, U.K., Tech. Rep. EBSE-2007-01, 2007.
[18] (2016). *Endnote*. [Online]. Available: https://endnote.com/
[19] T. Dybå and T. Dingsøyr, "Empirical studies of agile software development: A systematic review," *Inf. Softw. Technol.*, vol. 50, nos. 9–10, pp. 833–859, Aug. 2008.
[20] J. Archenaa and E. A. M. Anita, "A survey of big data analytics in healthcare and government," *Procedia Comput. Sci.*, vol. 50, no. 2015, pp. 408–413, 2015.
[21] F. A. Batarseh and E. A. Latif, "Assessing the quality of service using big data analytics: With application to healthcare," *Big Data Research*, vol. 4, pp. 13–24, Jun. 2016.
[22] K. Abouelmehdi, A. Beni-Hssane, H. Khaloufi, and M. Saadi, "Big data security and privacy in healthcare: A review," *Procedia Comput. Sci.*, vol. 113, pp. 73–80, Jan. 2017.
[23] G. Aceto, V. Persico, and A. Pescapé, "The role of information and communication technologies in healthcare: Taxonomies, perspectives, and challenges," *J. Netw. Comput. Appl.*, vol. 107, pp. 125–154, Apr. 2018.
[24] A. Agrawal and A. Choudhary, "Health services data: Big data analytics for deriving predictive healthcare insights," in *Data and Measures in Health Services Research* (Part of the Health Services Research book series (HEALTHSR)). Springer, 2019, pp. 3–18.
[25] P. Galetsi, K. Katsaliaki, and S. Kumar, "Big data analytics in health sector: Theoretical framework, techniques and prospects," *Int. J. Inf. Manage.*, vol. 50, pp. 206–216, Feb. 2020.
[26] M. Boman and P. Sanches, "Sensemaking in intelligent health data analytics," *KI-Künstliche Intelligenz*, vol. 29, no. 2, pp. 143–152, Jun. 2015.
[27] M. C. Buzzi, M. Buzzi, D. Franchi, D. Gazzè, G. Iervasi, A. Marchetti, A. Pingitore, and M. Tesconi, "Facebook: A new tool for collecting health data?" *Multimedia Tools Appl.*, vol. 76, no. 8, pp. 10677–10700, Apr. 2017.
[28] S. G. Alonso, I. de la Torre Díez, J. J. P. C. Rodrigues, S. Hamrioui, and M. López-Coronado, "A systematic review of techniques and sources of big data in the healthcare sector," *J. Med. Syst.*, vol. 41, no. 11, p. 183, Oct. 2017.
[29] H. Alharthi, "Healthcare predictive analytics: An overview with a focus on Saudi Arabia," *J. Infection Public Health*, vol. 11, no. 6, pp. 749–756, Nov. 2018.
[30] M. Ahmed, S. Choudhury, and F. Al-Turjman, "Big data analytics for intelligent Internet of Things," in *Artificial Intelligence in IoT*. Cham, Switzerland: Springer, 2019, pp. 107–127.
[31] S. P. Iyengar, H. Acharya, and M. Kadam, "Big data analytics in healthcare using spreadsheets," in *Big Data Analytics in Healthcare*. Cham, Switzerland: Springer, 2020, pp. 155–187.
[32] P. F. Brennan and S. Bakken, "Nursing needs big data and big data needs nursing," *J. Nursing Scholarship*, vol. 47, no. 5, pp. 477–484, Sep. 2015.
[33] R. Chauhan, R. Jangade, and V. K. Mudunuru, "A cloud based environment for big data analytics in healthcare," in *Proc. Int. Conf. Soft Comput. Pattern Recognit.*, 2016, pp. 315–321.
[34] A. Anjum, S. Aizad, B. Arshad, M. Subhani, D. Davies-Tagg, and T. Abdullah, "Big data analytics in healthcare: A cloud-based framework for generating insights," in *Cloud Computing* (Part of the Computer Communications and Networks book series (CCN)). Cham, Switzerland: Springer, 2017, pp. 153–170.
[35] D. W. Bates, A. Heitmueller, M. Kakad, and S. Saria, "Why policymakers should care about 'big data' in healthcare," *Health Policy Technol.*, vol. 7, pp. 211–216, Jun. 2018.
[36] Z. Ahmed and B. T. Liang, "Systematically dealing practical issues associated to healthcare data analytics," in *Proc. Future Inf. Commun. Conf.*, 2019, pp. 599–613.
[37] S. A. Kumar and M. Venkatesulu, "Brownboost classifier-based Bloom hash data storage for healthcare big data analytics," in *Information and Communication Technology for Sustainable Development*. Singapore: Springer, 2020, pp. 53–69.
[38] D. Chong and H. Shi, "Big data analytics: A literature review," *J. Manage. Analytics*, vol. 2, no. 3, pp. 175–201, Jul. 2015.
[39] Y. Chen, H. Chen, A. Gorkhali, Y. Lu, Y. Ma, and L. Li, "Big data analytics and big data science: A survey," *J. Manage. Analytics*, vol. 3, no. 1, pp. 1–42, Jan. 2016.
[40] R. K. Barik, H. Dubey, and K. Mankodiya, "SOA-FOG: Secure service-oriented edge computing architecture for smart health big data analytics," in *Proc. IEEE Global Conf. Signal Inf. Process. (GlobalSIP)*, Nov. 2017, pp. 477–481.
[41] T.-M. Choi, S. W. Wallace, and Y. Wang, "Big data analytics in operations management," *Prod. Oper. Manage.*, vol. 27, no. 10, pp. 1868–1883, 2018.
[42] J. Chanchaichujit, A. Tan, F. Meng, and S. Eaimkhong, "Internet of Things (IoT) and big data analytics in healthcare," in *Healthcare 4.0*. Singapore: Palgrave Pivot, 2019, pp. 17–36.
[43] Y. Kumar, K. Sood, S. Kaul, and R. Vasuja, "Big data analytics and its benefits in healthcare," in *Big Data Analytics in Healthcare*. Cham, Switzerland: Springer, 2020, pp. 3–21.
[44] N. S. Godbole and J. Lamb, "Using data science & big data analytics to make healthcare green," in *Proc. 12th Int. Conf. Expo Emerg. Technol. Smarter World (CEWIT)*, Oct. 2015, pp. 1–6.

[45] M. Dayal and N. Singh, "Indian health care analysis using big data programming tool," *Procedia Comput. Sci.*, vol. 89, no. 2016, pp. 521–527, 2016.

[46] I. Cano, A. Tenyi, E. Vela, F. Miralles, and J. Roca, "Perspectives on big data applications of health information," *Current Opinion Syst. Biol.*, vol. 3, pp. 36–42, Jun. 2017.

[47] A. Forestiero and G. Papuzzo, "Distributed algorithm for big data analytics in healthcare," in *Proc. IEEE/WIC/ACM Int. Conf. Web Intell. (WI)*, Dec. 2018, pp. 776–779.

[48] D. Cirillo and A. Valencia, "Big data analytics for personalized medicine," *Current Opinion Biotechnol.*, vol. 58, pp. 161–167, Aug. 2019.

[49] T. A. Naqishbandi and N. Ayyanathan, "Clinical big data predictive analytics transforming healthcare: An integrated framework for promise towards value based healthcare," in *Advances in Decision Sciences, Image Processing, Security and Computer Vision*. Cham, Switzerland: Springer, 2020, pp. 545–561.

[50] S. Gomathi and V. Narayani, "Implementing big data analytics to predict systemic lupus erythematosus," in *Proc. Int. Conf. Innov. Inf., Embedded Commun. Syst. (ICIIECS)*, Mar. 2015, pp. 1–5.

[51] A. Di Stefano, A. La Corte, P. Lió, and M. Scatá, "Bio-inspired ICT for big data management in healthcare," in *Intelligent Agents in Data-intensive Computing*. Cham, Switzerland: Springer, 2016, pp. 1–26.

[52] 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*. Cham, Switzerland: Springer, 2017, pp. 33–47.

[53] S. Ganesh and A. K. Talukder, "Formal methods, artificial intelligence, big-data analytics, and knowledge engineering in medical care to reduce disease burden and health disparities," in *Proc. Int. Conf. Big Data Anal.*, 2018, pp. 307–321.

[54] S. Din and A. Paul, "Smart health monitoring and management system: Toward autonomous wearable sensing for Internet of Things using big data analytics," *Future Gener. Comput. Syst.*, vol. 91, pp. 611–619, Feb. 2019.

[55] S. Hussain and S. Lee, "Semantic transformation model for clinical documents in big data to support healthcare analytics," in *Proc. 10th Int. Conf. Digit. Inf. Manage. (ICDIM)*, Oct. 2015, pp. 99–102.

[56] S. Gupta and P. Tripathi, "An emerging trend of big data analytics with health insurance in india," in *Proc. Int. Conf. Innov. Challenges Cyber Secur. (ICICCS-INBUSH)*, Feb. 2016, pp. 64–69.

[57] C. Lee, Z. Luo, K. Y. Ngiam, M. Zhang, K. Zheng, and G. Chen, "Big healthcare data analytics: Challenges and applications," in *Handbook of Large-Scale Distributed Computing in Smart Healthcare*. Cham, Switzerland: Springer, 2017, pp. 11–41.

[58] M. Giacalone, C. Cusatelli, and V. Santarcangelo, "Big data compliance for innovative clinical models," *Big Data Res.*, vol. 12, pp. 35–40, Jul. 2018.

[59] P. Galetsi and K. Katsaliaki, "A review of the literature on big data analytics in healthcare," *J. Oper. Res. Soc.*, vol. 12, pp. 1–19, Jul. 2019.

[60] R. Mehmood and G. Graham, "Big data logistics: A health-care transport capacity sharing model," *Procedia Comput. Sci.*, vol. 64, no. 2015, pp. 1107–1114, 2015.

[61] M. Haas, D. Stephenson, K. Romero, M. F. Gordon, N. Zach, H. Geerts, and B. Health Modeling Initiative (BHMI), "Big data to smart data in Alzheimer's disease: Real-world examples of advanced modeling and simulation," *Alzheimer's Dementia*, vol. 12, no. 9, pp. 1022–1030, Sep. 2016.

[62] G. Manogaran, D. Lopez, C. Thota, K. M. Abbas, S. Pyne, and R. Sundarasekar, "Big data analytics in healthcare Internet of Things," in *Innovative Healthcare Systems for the 21st Century*. Cham, Switzerland: Springer, 2017, pp. 263–284.

[63] S. Guha and S. Kumar, "Emergence of big data research in operations management, information systems, and healthcare: Past contributions and future roadmap," *Prod. Oper. Manage.*, vol. 27, no. 9, pp. 1724–1735, 2018.

[64] P. Galetsi, K. Katsaliaki, and S. Kumar, "Values, challenges and future directions of big data analytics in healthcare: A systematic review," *Social Sci. Med.*, vol. 241, Nov. 2019, Art. no. 112533.

[65] P. Raj, A. Raman, D. Nagaraj, and S. Duggirala, "Big Data Analytics for Healthcare," in *High-Performance Big-Data Analytics: Computing Systems and Approaches*. Cham, Switzerland: Springer, 2015, p. 1525.

[66] P. Jiang, J. Winkley, C. Zhao, R. Munnoch, G. Min, and L. T. Yang, "An intelligent information forwarder for healthcare big data systems with distributed wearable sensors," *IEEE Syst. J.*, vol. 10, no. 3, pp. 1147–1159, Sep. 2016.

[67] V. Gupta, H. Singh Gill, P. Singh, and R. Kaur, "An energy efficient fog-cloud based architecture for healthcare," *J. Statist. Manage. Syst.*, vol. 21, no. 4, pp. 529–537, Jul. 2018.

[68] C. Guo and J. Chen, "Big data analytics in healthcare: Data-driven methods for typical treatment pattern mining," *J. Syst. Sci. Syst. Eng.*, vol. 28, no. 6, pp. 694–714, Dec. 2019.

[69] M. Viceconti, P. Hunter, and R. Hose, "Big data, big knowledge: Big data for personalized healthcare," *IEEE J. Biomed. Health Inform.*, vol. 19, no. 4, pp. 1209–1215, Jul. 2015.

[70] A. Kankanhalli, J. Hahn, S. Tan, and G. Gao, "Big data and analytics in healthcare: Introduction to the special section," *Inf. Syst. Frontiers*, vol. 18, no. 2, pp. 233–235, Apr. 2016.

[71] A. P. Plageras, C. Stergiou, G. Kokkonis, K. E. Psannis, Y. Ishibashi, B.-G. Kim, and B. B. Gupta, "Efficient large-scale medical data (eHealth big Data) analytics in Internet of Things," in *Proc. IEEE 19th Conf. Bus. Informat. (CBI)*, Jul. 2017, pp. 21–27.

[72] W. J. Hopp, J. Li, and G. Wang, "Big data and the precision medicine revolution," *Prod. Oper. Manage.*, vol. 27, no. 9, pp. 1647–1664, Sep. 2018.

[73] S. Hussain, M. Hussain, M. Afzal, J. Hussain, J. Bang, H. Seung, and S. Lee, "Semantic preservation of standardized healthcare documents in big data," *Int. J. Med. Informat.*, vol. 129, pp. 133–145, Sep. 2019.

[74] L. Wang, R. Ranjan, J. Kołodziej, A. Zomaya, and L. Alem, "Software tools and techniques for big data computing in healthcare clouds," *Future Gener. Comput. Syst.*, vols. 43–44, pp. 38–39, Feb. 2015.

[75] H. Kashyap, H. A. Ahmed, N. Hoque, S. Roy, and D. K. Bhattacharyya, "Big data analytics in bioinformatics: Architectures, techniques, tools and issues," *Netw. Model. Anal. Health Informat. Bioinf.*, vol. 5, no. 1, p. 28, Sep. 2016.

[76] M. I. Pramanik, R. Y. K. Lau, H. Demirkan, and M. A. K. Azad, "Smart health: Big data enabled health paradigm within smart cities," *Expert Syst. Appl.*, vol. 87, pp. 370–383, Nov. 2017.

[77] H. K. Huang, "Big data in PACS-based multimedia medical imaging informatics," in *PACS-Based Multimedia Imaging Informatics*, H. Huang, Ed. Hoboken, NJ, USA: Wiley, 2018, pp. 575–589.

[78] M. A. Kumar, R. Vimala, and K. R. A. Britto, "A cognitive technology based healthcare monitoring system and medical data transmission," *Measurement*, vol. 146, pp. 322–332, Nov. 2019.

[79] M. D. Wang, "Biomedical big data analytics for patient-centric and outcome-driven precision health," in *Proc. IEEE 39th Annu. Comput. Softw. Appl. Conf.*, Jul. 2015, pp. 1–2.

[80] Z. Lv, J. Chirivella, and P. Gagliardo, "Bigdata oriented multimedia mobile health applications," *J. Med. Syst.*, vol. 40, no. 5, p. 120, Mar. 2016.

[81] G. Spanoudakis, P. Katrakazas, D. Koutsouris, D. Kikidis, A. Bibas, and N. H. Pontopidan, "Public health policy for management of hearing impairments based on big data analytics: EVOTION at genesis," in *Proc. IEEE 17th Int. Conf. Bioinf. Bioeng. (BIBE)*, Oct. 2017, pp. 525–530.

[82] R. S. H. Istepanian and T. Al-Anzi, "M-health 2.0: New perspectives on mobile health, machine learning and big data analytics," *Methods*, vol. 151, pp. 34–40, Dec. 2018.

[83] Y. Liang and L. Zhao, "Intelligent hospital appointment system based on health data bank," *Procedia Comput. Sci.*, vol. 159, no. 2019, pp. 1880–1889, 2019.

[84] M. K. Pandey and K. Subbiah, "A novel storage architecture for facilitating efficient analytics of health informatics big data in cloud," in *Proc. IEEE Int. Conf. Comput. Inf. Technol. (CIT)*, Dec. 2016, pp. 578–585.

[85] J. Wu, H. Li, L. Liu, and H. Zheng, "Adoption of big data and analytics in mobile healthcare market: An economic perspective," *Electron. Commerce Res. Appl.*, vol. 22, pp. 24–41, Mar. 2017.

[86] P. P. Jayaraman, A. R. M. Forkan, A. Morshed, P. D. Haghighi, and Y.-B. Kang, "Healthcare 4.0: A review of frontiers in digital health," *WIREs Data Mining Knowl. Discovery*, vol. 10, no. 2, p. e1350, 2020.

[87] N. Mehta, A. Pandit, and S. Shukla, "Transforming healthcare with big data analytics and artificial intelligence: A systematic mapping study," *J. Biomed. Informat.*, vol. 100, Dec. 2019, Art. no. 103311.

[88] M. Plachkinova, A. Vo, R. Bhaskar, and B. Hilton, "A conceptual framework for quality healthcare accessibility: A scalable approach for big data technologies," *Inf. Syst. Frontiers*, vol. 20, no. 2, pp. 289–302, Apr. 2018.

[89] P. Kaur, M. Sharma, and M. Mittal, "Big data and machine learning based secure healthcare framework," *Procedia Comput. Sci.*, vol. 132, no. 2018, pp. 1049–1059, 2018.

[90] A. Muniasamy, S. Tabassam, M. A. Hussain, H. Sultana, V. Muniasamy, and R. Bhatnagar, "Deep learning for predictive analytics in healthcare," in *Proc. Int. Conf. Adv. Mach. Learn. Technol. Appl.*, 2019, pp. 32–42.

[91] F. Rahman and M. J. Slepian, "Application of big-data in healthcare analytics—Prospects and challenges," in *Proc. IEEE-EMBS Int. Conf. Biomed. Health Informat. (BHI)*, Feb. 2016, pp. 13–16.
[92] H. Khaloufi, K. Abouelmehdi, A. Beni-Hssane, and M. Saadi, "Security model for big healthcare data lifecycle," *Procedia Comput. Sci.*, vol. 141, no. 2018, pp. 294–301, 2018.
[93] S. Rallapalli, R. R. Gondkar, and U. P. K. Ketavarapu, "Impact of processing and analyzing healthcare big data on cloud computing environment by implementing Hadoop cluster," *Procedia Comput. Sci.*, vol. 85, pp. 16–22, May 2016.
[94] C. Krittanawong, K. W. Johnson, S. G. Hershman, and W. H. W. Tang, "Big data, artificial intelligence, and cardiovascular precision medicine," *Expert Rev. Precis. Med. Drug Develop.*, vol. 3, no. 5, pp. 305–317, Sep. 2018.
[95] V. Palanisamy and R. Thirunavukarasu, "Implications of big data analytics in developing healthcare frameworks—A review," *J. King Saud Univ.-Comput. Inf. Sci.*, vol. 31, no. 4, pp. 415–425, 2019.
[96] S. Sakr and A. Elgammal, "Towards a comprehensive data analytics framework for smart healthcare services," *Big Data Res.*, vol. 4, pp. 44–58, Jun. 2016.
[97] J. Lindell, "What are big data and analytics?" in *Analytics and Big Data for Accountants*. Hoboken, NJ, USA: Wiley, 2018, pp. 1-1–1-13.
[98] L. Rajabion, A. A. Shaltooki, M. Taghikhah, A. Ghasemi, and A. Badfar, "Healthcare big data processing mechanisms: The role of cloud computing," *Int. J. Inf. Manage.*, vol. 49, pp. 271–289, Dec. 2019.
[99] N. Straton, K. Hansen, R. R. Mukkamala, A. Hussain, T.-M. Gronli, H. Langberg, and R. Vatrapu, "Big social data analytics for public health: Facebook engagement and performance," in *Proc. IEEE 18th Int. Conf. E-Health Netw., Appl. Services (Healthcom)*, Sep. 2016, pp. 1–6.
[100] X. Ma, Z. Wang, S. Zhou, H. Wen, and Y. Zhang, "Intelligent healthcare systems assisted by data analytics and mobile computing," in *Proc. 14th Int. Wireless Commun. Mobile Comput. Conf. (IWCMC)*, Jun. 2018, pp. 1317–1322.
[101] V. Ramasamy, B. Gomathy, and R. K. Verma, "Smart HIV/AIDS digital system using big data analytics," in *Progress in Advanced Computing and Intelligent Engineering*. Singapore: Springer, 2019, pp. 415–421.
[102] V. Tresp, J. Marc Overhage, M. Bundschus, S. Rabizadeh, P. A. Fasching, and S. Yu, "Going digital: A survey on digitalization and large-scale data analytics in healthcare," *Proc. IEEE*, vol. 104, no. 11, pp. 2180–2206, Nov. 2016.
[103] G. Manogaran, R. Varatharajan, D. Lopez, P. M. Kumar, R. Sundarasekar, and C. Thota, "A new architecture of Internet of things and big data ecosystem for secured smart healthcare monitoring and alerting system," *Future Gener. Comput. Syst.*, vol. 82, pp. 375–387, May 2018.
[104] M. I. Razzak, M. Imran, and G. Xu, "Big data analytics for preventive medicine," *Neural Comput. Appl.*, vol. 32, pp. 4417–4451, Mar. 2019.
[105] B. Xu, K. Xu, L. Fu, L. Li, W. Xin, and H. Cai, "Healthcare data analytics: Using a metadata annotation approach for integrating electronic hospital records," *J. Manage. Anal.*, vol. 3, no. 2, pp. 136–151, Apr. 2016.
[106] N. Mehta and A. Pandit, "Concurrence of big data analytics and healthcare: A systematic review," *Int. J. Med. Informat.*, vol. 114, pp. 57–65, Jun. 2018.
[107] A. N. Reiz, M. A. Armengol de la Hoz, and M. Sánchez García, "Big data analysis and machine learning in intensive care units," *Medicina Intensiva (English Edition)*, vol. 43, no. 7, pp. 416–426, Oct. 2019.
[108] J. B. Miller, "Big data and biomedical informatics: Preparing for the modernization of clinical neuropsychology," *Clin. Neuropsychologist*, vol. 33, no. 2, pp. 287–304, Feb. 2019.
[109] J. M. M. Rumbold, M. O'Kane, N. Philip, and B. K. Pierscionek, "Big data and diabetes: The applications of big data for diabetes care now and in the future," *Diabetic Med.*, vol. 37, no. 2, pp. 187–193, Feb. 2020.
[110] K. Moutselos, D. Kyriazis, and I. Maglogiannis, "A Web based modular environment for assisting health policy making utilizing big data analytics," in *Proc. 9th Int. Conf. Inf., Intell., Syst. Appl. (IISA)*, Jul. 2018, pp. 1–5.
[111] T. Saheb and L. Izadi, "Paradigm of IoT big data analytics in the healthcare industry: A review of scientific literature and mapping of research trends," *Telematics Informat.*, vol. 41, pp. 70–85, Aug. 2019.
[112] L. R. Nair, S. D. Shetty, and S. D. Shetty, "Applying spark based machine learning model on streaming big data for health status prediction," *Comput. Electr. Eng.*, vol. 65, pp. 393–399, Jan. 2018.
[113] G. Shah, A. Shah, and M. Shah, "Panacea of challenges in real-world application of big data analytics in healthcare sector," *J. Data, Inf. Manage.*, vol. 1, nos. 3–4, pp. 107–116, Dec. 2019.
[114] A. Pashazadeh and N. J. Navimipour, "Big data handling mechanisms in the healthcare applications: A comprehensive and systematic literature review," *J. Biomed. Informat.*, vol. 82, pp. 47–62, Jun. 2018.
[115] M. Shahbaz, C. Gao, L. Zhai, F. Shahzad, and Y. Hu, "Investigating the adoption of big data analytics in healthcare: The moderating role of resistance to change," *J. Big Data*, vol. 6, no. 1, p. 6, Dec. 2019.
[116] H. B. Patel and S. Gandhi, "A review on big data analytics in healthcare using machine learning approaches," in *Proc. 2nd Int. Conf. Trends Electron. Informat. (ICOEI)*, May 2018, pp. 84–90.
[117] C. B. Sivaparthipan, B. A. Muthu, G. Manogaran, B. Maram, R. Sundarasekar, S. Krishnamoorthy, C. Hsu, and K. Chandran, "Innovative and efficient method of robotics for helping the Parkinson's disease patient using IoT in big data analytics," *Trans. Emerg. Telecommun. Technol.*, vol. e3838, Dec. 2019, Art. no. e3838.
[118] A. Ravishankar Rao, D. Clarke, and M. Vargas, "Building an open health data analytics platform: A case study examining relationships and trends in seniority and performance in healthcare providers," *J. Healthcare Informat. Res.*, vol. 2, nos. 1–2, pp. 44–70, Jun. 2018.
[119] M. J. Sousa, A. M. Pesqueira, C. Lemos, M. Sousa, and Á. Rocha, "Decision-making based on big data analytics for people management in healthcare organizations," *J. Med. Syst.*, vol. 43, no. 9, p. 290, Sep. 2019.
[120] P. K. Sahoo, S. K. Mohapatra, and S.-L. Wu, "SLA based healthcare big data analysis and computing in cloud network," *J. Parallel Distrib. Comput.*, vol. 119, pp. 121–135, Sep. 2018.
[121] K. D. Strang, "Problems with research methods in medical device big data analytics," *Int. J. Data Sci. Analytics*, vol. 9, no. 2, pp. 229–240, Feb. 2019.
[122] B. K. Sarkar and S. S. Sana, "A conceptual distributed framework for improved and secured healthcare system," *Int. J. Healthcare Manage.*, vol. 11, pp. 1–13, Jan. 2018.
[123] L. Syed, S. Jabeen, S. Manimala, and H. A. Elsayed, "Data science algorithms and techniques for smart healthcare using IoT and big data analytics," in *Smart Techniques for a Smarter Planet*. Cham, Switzerland: Springer, 2019, pp. 211–241.
[124] A. Sebaa, F. Chikh, A. Nouicer, and A. Tari, "Medical big data warehouse: Architecture and system design, a case study: Improving healthcare resources distribution," *J. Med. Syst.*, vol. 42, no. 4, p. 59, Apr. 2018.
[125] R. Venkatesh, C. Balasubramanian, and M. Kaliappan, "Development of big data predictive analytics model for disease prediction using machine learning technique," *J. Med. Syst.*, vol. 43, no. 8, p. 272, Aug. 2019.
[126] S. Shafqat, S. Kishwer, R. U. Rasool, J. Qadir, T. Amjad, and H. F. Ahmad, "Big data analytics enhanced healthcare systems: A review," *J. Supercomput.*, vol. 76, no. 3, pp. 1754–1799, Feb. 2018.
[127] Y. Wang, L. Kung, S. Gupta, and S. Ozdemir, "Leveraging big data analytics to improve quality of care in healthcare organizations: A configurational perspective," *Brit. J. Manage.*, vol. 30, no. 2, pp. 362–388, Apr. 2019.
[128] C. B. Sivaparthipan, N. Karthikeyan, and S. Karthik, "Designing statistical assessment healthcare information system for diabetics analysis using big data," *Multimedia Tools Appl.*, vol. 79, nos. 13–14, pp. 8431–8444, Nov. 2018.
[129] J. Zetino and N. Mendoza, "Big data and its utility in social work: Learning from the big data revolution in business and healthcare," *Social Work Public Health*, vol. 34, no. 5, pp. 409–417, Jul. 2019.
[130] V. Tang, P. K. Y. Siu, K. L. Choy, H. Y. Lam, G. T. S. Ho, C. K. M. Lee, and Y. P. Tsang, "An adaptive clinical decision support system for serving the elderly with chronic diseases in healthcare industry," *Expert Syst.*, vol. 36, no. 2, Apr. 2019, Art. no. e12369.
[131] Y. Wang, L. Kung, and T. A. Byrd, "Big data analytics: Understanding its capabilities and potential benefits for healthcare organizations," *Technological Forecasting Social Change*, vol. 126, pp. 3–13, Jan. 2018.
[132] S. Khan, H. Ali, Z. Ullah, N. Minallah, S. Maqsood, and A. Hafeez, "KNN and ANN-based recognition of handwritten Pashto letters using zoning features," *Mach. Learn.*, vol. 9, no. 10, pp. 570–577, 2018.
[133] J. Manyika, M. Chui, B. Brown, J. Bughin, R. Dobbs, and C. Roxburgh, *Big Data: The Next Frontier for Innovation, Competition & Productivity*. New York, NY, USA: McKinsey Global Institute, 2011.
[134] A. Alonso-Betanzos and V. Bolón-Canedo, "Big-data analysis, cluster analysis, and machine-learning approaches," in *Sex-Specific Analysis of Cardiovascular Function* (Advances in Experimental Medicine and Biology), vol. 1065, P. Kerkhof and V. Miller, Eds. Cham, Switzerland: Springer, 2018.
[135] S. M. Vijaya, S. Mahalingam, and N. K. M. Sai, "Improving patients' satisfaction in a mobile hospital using lean six sigma—A design-thinking intervention," *Prod. Planning Control*, vol. 31, no. 6, pp. 512–526, 2020, doi: 10.1080/09537287.2019.1654628.

[136] S. M. Vijaya and N. R. Kunnath, "Six sigma to reduce claims processing errors in a healthcare payer firm," *Prod. Planning Control*, vol. 31, no. 6, pp. 496–511, 2020, doi: 10.1080/09537287.2019.1652857.
[137] M. V. Sunder, L. S. Ganesh, and R. R. Marathe, "A morphological analysis of research literature on lean six sigma for services," *Int. J. Oper. Prod. Manage.*, vol. 38, no. 1, pp. 149–182, 2018, doi: 10.1108/IJOPM-05-2016-0273.

SHAH NAZIR received the Ph.D. degree in computer science with a specialization in software engineering from the University of Peshawar, in 2015. He has several research publications in well-reputed international journals and conference proceedings. He is currently serving as an Assistant Professor and the Head of the Department with the University of Swabi. Prior to this, he worked at the University of Peshawar. His research interests include component-based software engineering, software birthmarks, systematic literature review, and decision making. He is a Reviewer of several journals and conferences.

SULAIMAN KHAN received the M.S. degree in computer systems engineering from the University of Engineering and Technology at Peshawar, Peshawar. He has several research articles in international conferences and journals. He is currently working as a Lecturer with the University of Swabi. His research areas are in machine learning algorithms.

HABIB ULLAH KHAN received the Ph.D. degree in management information systems from Leeds Beckett University, U.K., in 2008. He has nearly 20 years of industry, teaching, and research experience. He is an Associate Professor of MIS with the Department of Accounting and Information Systems, College of Business and Economics, Qatar University, Qatar. His research interests are in the areas of IT adoption, social media, the Internet addiction, mobile commerce, computer mediated communication, IT outsourcing, bigdata, and IT security.

SHAUKAT ALI received the M.Sc. and M.S. degrees in computer science from the University of Peshawar, Peshawar, Pakistan, in 2007 and 2010, respectively, and the Ph.D. degree in computer science from the Department of Computer Science, University of Peshawar. Apart from this, he is also working as a Lecturer with the Department of Computer Science, Islamia College University, Peshawar. His areas of interest include big data, data analytics, information security, and data protection.

IVÁN GARCÍA-MAGARIÑO received the Ph.D. degree in computer science engineering from the Complutense University of Madrid, in 2009. He is currently a Lecturer and a contributor to the GRASIA Research Group with the Complutense University of Madrid. Prior to commencing this position in 2018, he was a Ph.D. Assistant Professor with the University of Zaragoza, from 2014 to 2018, and before he was a Lecturer with the Madrid Open University, from 2010 to 2014. His main research interests include agent-based simulators, multi-agent systems, and agent-oriented software engineering. Among journals, book chapters, conferences and workshops, he has over 135 publications (over 60 in journals with ISI Thomson JCR). He was a recipient of an FPI scholarship, from 2006 to 2010. He actively collaborates with the EduQTech research group from the University of Zaragoza on m-health projects, the HCI Lab from the University of Udine, Italy, on human-computer interaction projects, Institute of Technology Blanchardstown, Ireland, on datamining projects, and the Multicultural Alzheimer Prevention Program (MAPP) of Massachusetts General Hospital and Harvard University, USA, on m-health projects. He is Editor in several journals and guest editor in several special issues in journals with impact factor. http://grasia.fdi.ucm.es/ivan/.

RODZIAH BINTI ATAN received the Ph.D. degree in software engineering from University Putra Malaysia, in 2005. She currently works with the Faculty of Computer Science and Information Technology, University Putra Malaysia. She is currently a Lecturer and handles classes for more than 1000 undergraduate students, and has had eight master's and seven Ph.D. degrees students graduated under her supervision. She specializes in the areas of software engineering, software process modeling, and cloud computing services. She has also published academic books and chapters in books, and has more than 100 journal publications and conference papers. Her current research interest is in the commercialization of research products, where she works in collaboration with industry and other institutions of higher learning. She has also held positions as a keynote speaker for an international conference and a colloquium, a panel assessor in a workshop, a subject matter expert, a local and international journal and conference Reviewer, and an industrial training head coordinator.

MUHAMMAD NAWAZ received the Ph.D. degree from the University of Brunel, in 2013, which is quite a reputable Institution in the U.K. He has strong background in Computing discipline which is evident from the master's and bachelor's degrees in the relevant discipline. He has over 20 years of professional teaching and administrative experience and he has taught various computing subjects at the undergraduate and graduate levels. It is worthwhile to note that he has taught some of the most hardcore subjects in the computing field–Advance Multimedia Technologies, Advance Image Processing, Computer Vision, Compiler Construction, Automata Theory, and Artificial Intelligence. He has more than ten publications in recognized journals by the HEC Pakistan. Besides journal publications, he has presented in major computing conferences around the world which ultimately indicates his participation in the research activities and collaboration at national and international level.