IJCSE International Journal of Computer Sciences and Engineering Open Access
Survey Paper Vol.-7, Special Issue-13, May 2019 E-ISSN: 2347-2693

A Survey: Big Data Ethics and Challenges in Healthcare Division

Madhu H K[1]*, Prakash B R[2]

[1]SSAHE Research Center, SSIT, Tumkur, India
[2]Department of Master of Computer Applications, SSIT, Tumkur, India

*Corresponding Author: madhuhkbit@gmail.com, Tel.: +919980008835

DOI: https://doi.org/10.26438/ijcse/v7si13.1624 | Available online at: www.ijcseonline.org

Abstract - The era of automation, robotics, Internet of Things and collaborative networked systems generated an uniquely characterized data which is named as Big data due to the presence of sensor, camera and text generators poses challenges of storage, processing, and analysis. In this paper, the advanced health care system of future is discussed which generates the Big data along with the conventional methods of data analytics on medical Big data its trends, limitations and further research possibilities. In this paper, a general characteristic of Bigdata is discussed then a thorough study on the future healthcare process and management will be described. Then discuss how the data generated in the context of the Healthcare system will have big data characteristics. Finally, collect all the state of artwork for medical big data analytics and discuss the research gap and open issues.

Keywords **–** Analysis, Bigdata, Healthcare, Medical, Machine Learning.

I. INTRODUCTION

In the current scenario of the world, the population explosion demands mass production of every utilities, goods, and service required for humankind. These mass productions require technologies like automation, Internet of Things, advance communication and computing mechanisms to balance the demand and supply ratio. Recent trends of collaborative networks of people such as social network, a collaboration of networked control system, smart and intelligent systems, robotics, surveillance systems and IoT, etc. all generates continuous data. The data generation on the time series makes its reposit larger and larger over a period of time, and this assigns a tag of volume characteristic to the data[1][2]. These data can be in the form of structured, unstructured or in semi-structured. The processing of these gigantic data is a kind of the challenging task, it is not enough to store and process on a single machine, because this data may be in billions or few petabytes. The conventional file system for storage like Unix or NTFS file system ceases its capacity to handle volumes data and even a relational data management system has been proven inefficient to handle such a vast amount of data[3-4]. Hence, new tools and technologies are in demanding phase for effective analysis and management of these massive data. Despite this, the explosive growth and widespread availability of data have led researchers into the wave of more research activities in the era of digital healthcare due to which the trend of medicine is more shifting from treatment to prevention. Hospitals and medical associations generates huge amount of heterogeneous clinical data through medical imaging, sensors and patient records that can be viewed as a great source of big data repositories system that can be effectively utilized to make medical systems better [5-6]. The heterogeneous medical data refers to Electronic Health Record (EHR) where all information is related to every single patient such as clinical data, prescriptions, medical images, laboratory records, drugs documents are kept as a record to support medical practice and help to derive future healthcare insights[7]. The EHR includes two types of data information that are as 1) Electronic Medical Records (EMR): EMR is set of time series data which holds patients records such as patient medical history, patient hospital visiting time diagnoses, lab reports, medications, doctors notes, image data's such as MRI, CT-SCAN, etc. and 2) Sensors data: The advancement of wireless technology and IoT are now expected to provide big support for the improvement of healthcare industries. The small and energy- efficient sensors (such as wearable device, ECG sensors, etc.) are highly used to analyze the physical environment of hospitals and patients in real-time. These sensors generate numerous and variant types of data streams which can be useful to understand the psychological and physiological behavior of patients after monitoring and analyzing these data[8]. The effective use of EHR data analytics can be beneficial for the growth of digital healthcare system in efficient way. For instance, reducing of high cost by predicting patient's readmission & management; on-demand health service for lifestyle recommendation, health risk prediction, disease prediction, etc. Although, mining of EHR data is challenging because it contains unstructured data which has the following property: high-

dimensionality, noise, irregularity, missing data and sparsity [9] [10]. So, there is a need for an efficient analytical tool to understand the complexity of data and extracting useful knowledge from it. Therefore, this paper conducts survey for understanding big data analytics research trends and issues. The rest of the paper organization in the following manner: Section 2 contains background knowledge, Section3 carries literature review of existing studies, and finally Section 4 discloses the open research gap, conclusion and future work.

Background

This section presents the general concept of big data, analytical techniques and big data opportunities in healthcare department which are as follows:

Big Data

The term big data refers to the complex and huge amount of data that can both structured and non-structured. The data is continuously generated from the variety of resources such as research centers, GPS data, Medical data, data generated from sensors equipped in monitoring and surveillance applications, social data like Face book data, twitter data and customer's feedback streams from commercial sites, etc. The big data concept can be characterized by 4 V's properties that are as volume, variety, velocity, and veracity [11].

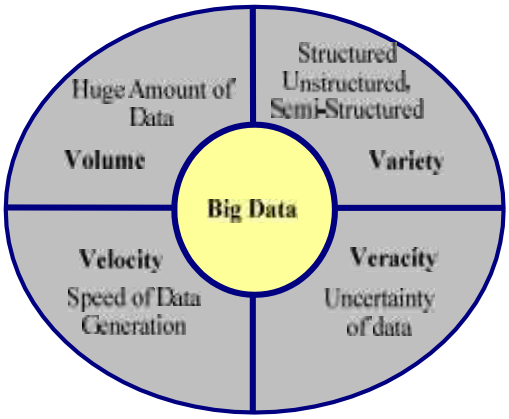

Figure 1. Characteristic of Big Data

A. Volume

Volume refers to the gigantic amount of data that has no fixed size and number. It can be term as big data, which is continuously generated by the various organizations and is difficult to process with traditional tool and technique. According to health data archive report, a health data volume is increasing 48 percent per year. In 2013, 153 exabytes of health data were estimated, and it was projected that 2314 Exabyte's of data would be generated by 2020 [12].

B. Variety

Variety implies the format of big data that is in both structured and non-structured data. Usually, in healthcare, data are classified as structure, non- structure and semi-structured. There are only about to 10% data are structured which can easily store and managed in databases. Similarly, semi- structured data is also about 8% to%10. These structured data and semi-structured data generates from the health record such as CT scan, MRI reports, blood test, diagnosis and laboratory reports. Most of the data is unstructured that is about 75% to
80%. The unstructured data generated from the varieties of resources that includes e-mails, feedback stream, sensor data, written documents, missing content, etc. The unstructured data is very complicated in handling and for the analytical process. Thus with effective processing platform based on efficient data mining, advanced machine learning technique, a high potential can be achieved to improve the services and system of healthcare departments [13].

C. Velocity

The explosive growth rate of data volume and data variety refers to the velocity of big data. In the healthcare

system day-by-day, newly data are generated continuously and in a constant manner. The data are generated in the form of both structured and unstructured such as the data generated from monitoring patient, patient admission, EHR progression, etc. The high velocity of healthcare running data creates another challenge for the big data analytics that it cannot process instantly and even it cannot store efficiently on the cloud system. To handle this situation, researchers bring new tools named Hadoop which is a platform that can handle massive running data for the analysis.

D. Veracity

Big data veracity refers to the trustworthy, reliability, noise and uncertainty of information received from the different datasets. Generally, unstructured data are highly variable and often incomplete. Healthcare data analytics should need to determine some valuable insights from unstructured data to have better treatment of patients and to make better decisions making system [14].

Big Data Analytics Outlook

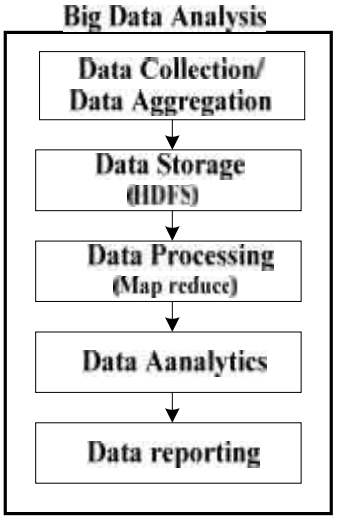

Figure 2. A General Architecture of Big Data Analytics

The above figure 2 shows the modeling of big data analytics. The big data is composed of structured and unstructured type data, and it becomes very tough to deal with such kinds of oversize data. Through this analytical model, a significant amount of data can be processed and further useful information can be extract to solve the hidden problem. The first layer of model initializes the data collection, and aggregation processes form the relevant health record resources. In the second layer, the data is stored in the Hadoop distributed file system storage (HDFS). HDFS is an open source of distribution platform provided by Apache Hadoop used to store such a massive amount of data. Further, the processing layer uses Map reduce model to process these data by implementing some machine learning algorithm that performs some classification and feature extraction operation. Then analysis is made based on this feature and providing essential facts for generating reports [15].

Big Data Challenges in Healthcare

Big data analytics in healthcare faces various challenges which include multiple factors such as storing data, appropriate data collection and summarization. The organization of resultant data after extraction from different layers is also a challenging task. It includes several sub- processes such as authentication, archiving, securing, information retrieval, management, and representation. Also, the quality information in each level should be verified by applying security protocols and preservation methods. In addition, following are the vital points of big data challenges in healthcare are represented below:

- Maintaining the correlation among the patient and their records in longitudinal ways.
- Understanding the unstructured data in a productive way.
- Analyzing the missing health data content
- High dimensionality problem of big health imaging data
- Maintaining privacy and security towards patient identity

- Analyzing genomic data is also a big challenging task that it requires more storage and processing capacity
- Capturing the patient behavior and their social interaction through several sensors

Openings of Big Data in Medical

Big data can formulate a high potential to contribute in many sectors of medical and healthcare industries. This section presents some excellent initiatives that used to enhance the quality feature of healthcare services. But also this is not enough to keep up with the demand of the healthcare services and the increasing cost. The following are some relevant key points that are mentioned below:

- Clinical notes and electronic health records are the primary sources of data. These can be efficiently served as the availability of patients and individual data.
- Structural data can efficiently use for service optimization.
- With the use of these big healthcare data and analysis can be made to reduce the mistakes and cost optimization
- Analysis of big medical data for clinical decision support, personalized care, future disease prediction.
- Analysis of big data can also enhance the healthcare policy, financial and administrative decisions
- Patient privacy maintenances
- Mobile healthcare facilities
- To Prevent Unnecessary emergency ward visits and increase customer satisfaction by E-health services.

II. RELATED WORK

The expansion of the analytical concept emphasizes the value of data collection, data integration, and data processing that how these data knowledge discovery acts to support better decision making property and helps to understand the behavior and trends of the organization. Therefore, this part of the paper presents the role of Big Data Analytics in the era of medical system and healthcare research associations through discussing various existing works that carried in the following sections.

Big Data Analytics for Disease prediction in the Healthcare System

Various researchers have explored that studies toward medical disease conditions concerning the advance analytics are still in developing phases. According to the findings of work carried out by Rahimi et al. [16], it was noticed that there is an evolving importance towards data quality. It was also found that ontological approaches could significantly assist in improving data quality. Biomedicals and health care system generates lots of data, based on these data a beneficial feature can be extracted to predict the disease more accurately. In the work of Chen et al. [17] convolution neural network (CNN) concept is used to handle the structured and unstructured data generated from the hospitals. The contribution of work is that it uses both structured and unstructured data which achieves good prediction rate of about 95 percent when compared to other existing methods. About 10 percent of the world population suffers from the metabolic disorder of diabetes due to improper lifestyle, and more critically, youth and children are becoming more subjected to a person with diabetes. The existing system is not sufficient as much that it faces real-time complexities, lacks data sharing and real-time monitoring. To cope the inefficiency of existing diabetic system the work of Chen et al. [18] have introduced an intelligent model where, diagnosis system is prepared using multiple layers, and smart analysis model is constructed using machine learning approach which includes smart clothing, intelligent data centers, and smart-phone. In the same way, the design of an efficient ECG monitoring system is expected to minimize the death rate of young adults that caused due to arrhythmia heart-attacks. So, Yin et al. [19] uses combined features of ECG signals and radio signals to mitigate the issues of motion artifacts occurs in ECG recordings and to accurately classify the arrhythmia. Furthermore, a hybrid monitoring framework is designed using the concept of CNN which extracts the features of both ECG signals and radio signals. The efficiency of the model is tested on caffe deep learning platform and obtains 85 to 90 percent accuracy rate for classifying abnormal and normal heartbeat. Different data mining techniques were adopted to analyze and predict the heart disease in [20]. The accuracy of each approach that is Naïve Bayes, Decision Tree and artificial neural network are verified and achieves 86.53%, 89% and 85.53% of accuracy rate respectively. Another work of Hsi-Jen Chiang et al. [21] have proposed a framework for analyzing analytical indicators in dental embed therapy. The author have studied1161 implants from 513 number of patients. Further, the Data on 23 samples are considered as impact factors on dental implants. The 1161 implants are investigated by using C5.0 technique. The experimental outcome of this model achieves the better performance with 97.67% accuracy rate.

Big Data Analytics towards the personalized healthcare system

In the effort of improving human lifetime, bioinformatics and computational analytic tool provide potential support in the field of personal healthcare and medical health decision- making system. The significant work carried out by Chawla, and Davis [22] have discussed a pre-prediction model for forecasting possible disease risk in the future. The author have proposed foundations of the presented model that derives essential aspects of big data and networks concept to bring the emergence of personalized medical care system. It was also noticed that similar work carried by Chen et al. [23], in which they have considered the problem of human unhealthy lifestyle and designed a smart system to suggest user's to avoid the unhealthy behavior. The presented model uses machine learning and deep learning technique embedded with the sensing device and analytical module for monitoring the social, mental and physiological functioning of human based on their daily life activity. The primary objective of the presented model is to guide human to fallow a healthier lifestyle. The effectiveness of this intelligent adviser system is evaluated on the smart testbed system which achieves efficient monitoring feature compared to some existing techniques. An IoT with big data concept in healthcare facilitates a remote service to medical experts so that they can analyze the action and behavior of the patients. By applying the same idea, Magarino et al. [24] suggested a computational model for recognizing the sleeping posture of patients in the hospital. The motive of the presented work is to provide a supportive tool in the research area for advancing the measurement process of human activity through smart beds and IoT. The simulation results display that it achieves a good accuracy rate of about 95 to 98 percent compared to existing works. Similarly, Viceconti et al. [25], have given the joint concept of big data and virtual physiological human towards improving the critical situation of computational medicine. The author have discussed some critical problem such as data confidentiality, data size, integrating bioinformatics, data localization, etc. in order minimize the shortcoming in the research zones so that an efficient and robust solution could be made in the cycle of big data healthcare. In the work of Yassine et al. [26] have constructed a framework utilizes the smart home electronic device to recognize and learn the human pattern and then extract knowledge to support health care system in respect to improving human living style. The presented framework uses the electric meter as a means of data mining tool for identifying human behavior and their routine changes through analyzing temporal electricity consumption by home appliances in multiple time frames. The author uses UK electricity data-sets between the years of 2012-2015 with 109 electric appliances from five houses. The experimental outcomes of this study achieve a better result in predicting short and long terms of human activity patterns.

Big Data Analytics towards privacy preservation in the healthcare system

However, big data analytics offers many beneficial features to the medical health care department, but also it raises some security breaches that related to patient privacy and organizations confidentiality. Therefore, there is crucial need of efficient security system which serves to preserve the security and privacy of the big data medical assistance system. The work carried out by Abouelmehdi et al. [27] have investigated various research works that have been carried out in the domain of big data healthcare security based on encryption and Anonymization approaches. Furthermore, the author have reviewed and compared various privacy-preserving schemes along with its potential strength and limitations. It seems that the main objective of the presented work is to attract the researches attention to design a feasible privacy preserving algorithm in the era of big healthcare and computational medical system. Similarly, Jain et al. [28] and Shahi et al. [29] also presented security challenges and security requirements based on a study of various recent privacy protection techniques such as Anonymization, hiding a needle in haystack and differential privacy towards medical care in big data streams. Big data in healthcare often considered as a distributed system where transmission and exchange of data are processing from various locations in multiple resources. Thus, distributive nature of transmission system includes various security complications such as patient privacy leakage, data loss, etc. To ensure data integrity sarkar [30] has presented brief information about the concept of data mining in bio-medical and designed a secure distributed architecture for electronic health records operations. The performed study achieves good security constraints, low-cost implementation and also delivers a good overview that big data plays crucial role in era of bioinformatics for the non-expert readers. An incident was reported in Forbes magazine that shows concern about the patient privacy [31]. It is mentioned that Target Corporation forward baby care coupons to a teenage girl unknown to her fathers. This incident refers to the researchers and analytics to focus on important privacy and security in big data healthcare. In the study of Patil et al. [32], harsh et al.[33] and in [34] some traditional approaches are described for privacy preserving in big data. Although, these approaches are used traditionally to ensure the patient's privacy and their demerits led to the arrival of advanced techniques. Similarly, various analytical actions are proposed in the work of [35, 36] to compute information loss due to Anonymization. It has been observed that the presented work has not reflected in the actual value of data.

Big Data Analytics in healthcare system based on clinical data and machine learning approach

Medical industries generated a vast amount of clinical data such as imaging reports, diagnosis reports, billing report and medical prescriptions. These data's are collected from the various sources and distributed to different health departments. However, reusing of these clinical data's provides useful and essential supportive role to assist the medical professionals regarding event

prediction, patient health, quick decision making and helps to enhance biomedical care system and medical research. Various optimization and preprocessing technique have been applied to these big clinical data so that an active learning and knowledge discovery process could be made. The author Huddar et al. [37], have made their contribution towards investigating the patient at the risk of critical health. The author have used a preprocessing technique for feature extraction and matrix factorization learning approach for modeling and classifying unstructured clinical data. The performances of the study display that the presented feature extraction technique performs better than other existing technology. Also, from the experimental outcomes, it has been observed that the matrix factorization based classification approach suffers from the limitation of interpretability loss with respect to feature characteristic in final prediction. It is also becoming a necessity to understand the complexity, weakness, and requirements of the current diverse clinical data sets to enter the big data into computation medical. The work of Johnson et al. [38] carried a survey task and talked about the current trends of machine learning approach in the critical care. The main focused drawn towards the problem such as complexity, compartmentalization, and corruption that arises in the clinical data gathering and preprocessing techniques. The efficient and Innovative software is designed by the Luo [39] to support and facilitate the medical researchers with constructing an optimized predictive model based on clinical data. The software model contains multiple techniques such as fast-pivoting, visual query, hyper-parameter values, and automatic selection of learning algorithms are utilized with the machine learning approach. A current trend of cloud infrastructure for big data healthcare system has discussed in the study of Tawalbeh et al. [40], where the author have brought their effort to analyze the big data of medical such as patient records, diagnosis reports, etc., into real time. Based on the convolution neural network technology and natural language processing Yao et al. [41] developed an automated diagnosis framework which objective to provide human-like medical attention for guidance and consultation. The experiment task is carried out on the real data sets and found to be very effective than other existing methods. Similarly, in [42] Zhang et al. have presented a healthcare model based on cloud technique and big data which used to establish medical services according to patient requirements and preferences. The model exhibits multiple layers for data collection, storage and data distribution with parallel computing. In [43] Jiang et al. have focused to the lifestyle of elderly which live independent and proposed approach of a smart wearable system based on the Hidden Markov model which objective is to recognize the elderly physiological and mental behavior pattern. From the experimental result, it is observed that the proposed approach achieves good performances, that it transmits only important information to the big data server so that elderly safety to be maintained. In [44], the authors have mentioned some data analytics tools and methods based on data warehouse, cloud computing, and big data concept to improve healthcare services in many sectors such as prediction and prevention system. In the work of Ali et al. [45], introduced about the needs of data warehouse and Business Intelligence for healthcare department and discusses the difference role of Business Intelligence and its need in medical systems.

III. RESEARCH CHALLENGES IN BIG DATA HEALTHCARE

Big data concept can facilitate enormous benefits to many organizations and individuals. The application of big data includes information science, data modeling and analysis, distributed computing, machine learning, statistical learning, and pattern recognition. The practical and efficient integration of such technologies and analytical model helps to predict future deviations in advance. However, benefits of big data applications can also opens a number of security challenges and privacy risk because most of the big data tools are available online as open source and which may be an optional key for intrusions and attackers. From the state of artworks that have been carried in above section, it is found that most of the research works have been focused on the data extraction and knowledge discovery from the uncertain pattern of big data and very less effort has strained on the privacy and security concern in the healthcare era. The privacy and security related to patients and medical organization, there must be a system that ensures the patient personal and medical information to be protected and secure when it is using in the analytical modeling. Furthermore, it is also investigated that there is a need of skilled full analyst because there are big shortages of big data specialists. Therefore the main focus of this section is to discuss open research challenges in healthcare big data analytics and researchers needs to more focus on the following key points:

Privacy and Security Issues

The privacy and security are the primary concerns of individuals and organizations that hold valuable information about the persons. The security becomes a prime concern when working with big data analytics. The data generated from the healthcare system contains personal and confidential data, and it must be handled with the very careful responsibilities. In order to protect patient's privacy and organizational data, some contribution should be made just as personal information & identity should initially be backed up somewhere, and then deleted during the data collection process. This process provides a secure step for individual privacy protection and can be recovered as long as the healthcare organization needs it. Additionally, some security protocol can also be integrated into the user's data which serves as better security protection in the age of advanced technology.

Storage and Processing Issues

The healthcare system generates a massive amount of structured and unstructured continuously and constantly every day. So storing and processing such data becomes another challenge in the medical big data analytics. For instance, genome sequences of single human generate a massive amount of data that takes about three terabytes of storage memory. These enormous amounts of data generated and created around all over the world due to which processing of these data become another challenge. The processing issue can be solved by adopting software that transmits only the data which is essential to the analysis purpose. However, this may because another problem related to data integrity and data source. To overcome these situations and challenges, researchers should apply more effort in their research work and must introduce secure and efficient data storage and processing solutions.

Data-Sets Updating

Healthcare data is not fixed it grows continuously, so most of the elements require frequent updates to make analysis better. For instance, the detail such as individual home address, signature, etc. may be required to update. So understanding the volatility of big data is also one of the challenging tasks. It is very difficult to keep monitoring the degree of deviation in big data. To deal such types this situation, the system should be designed in which data provider must have explicit knowledge about the data set updating, what information to be updated and which data should be updated. The Organization must also ensure that they are not building unnecessary copying or duplication of records while attempting the update. They must also ensure that the update operation is being performed without affecting the quality and integrity of the data set.

A Requirement of Skillful and Rich Infrastructure

Many organizations still lacks the good technology infrastructure required to influence their researchers. More effort and investment is required, and also proper policies and legal agreements must apply to all entities to preserve privacy amongst the organizations and the patient's details. Form the recent study; it has been observed that there is a shortage of the skilled full data analyst. So our educational system must make the student aware of the current situation and demand of the technology, required skill in the big data analytics.
Hence, from the big healthcare analysis viewpoints majority of the research gap found and which seems completely as serious issues. Therefore, there is a need for a multi-objective solution for the healthcare data collection and big data analysis task. Consequently, it is very essential to realize the state of the big data analytics in healthcare under which many processes are carried for multiple purposes such as improving human life cycle and improving the healthcare system in more advance. If researchers and experts can overcome some of these complications, then the power of big data may be released that making the possibilities infinite.

IV. CONCLUSION AND FUTURE DIRECTIONS

Big data is reshaping the healthcare and the individual's life. It is evolving into a promising area where insight is made from extensive data sets which results in changing the level of outcomes at a tremendous level. Anyhow, its potential is great, but there remain challenges to overcome. This paper presents the general information and fundamental concepts about the big data in healthcare. This paper also survey some existing works related to the big data healthcare and discusses some challenges and the possible opportunities that bring effective and efficient services for the healthcare department in terms of flexible medical services, cost optimization, better decision making support, providing valuable services to patients and also improving the human life cycle by guiding their life style activity. In the last, this paper also discloses the some research challenges in the era of medical big data analytics and hence, it suggests that medical sector need to be more perfect in the way of data processing, discovering knowledge from data, and enhance their interaction with patient's, and customers through data- driven decisions. However, the current situations shows that medical industries have not yet achieved significant value from the data analytics and data mining and achieving these significance will little difficult. But it opens numbers of opportunities and their possibilities are so high that healthcare leaders, policymakers and administrators should not ignore these opportunities.

REFERENCES

[1] Big Data Fundamentals: Concepts Drivers: Concepts, Drivers and techniques Paperback – 25 Jul 2016
[2] John Walker, Saint. "Big data: A revolution that will transform how we live, work, and think." (2014): 181-183.
[3] Marjani M, Nasaruddin F, Gani A, Karim A, Hashem IA, Siddiqa A, Yaqoob I. Big IoT data analytics: architecture, opportunities, and open

researchchallenges. IEEE Access. 2017; 5:5247-61.
[4] Lee C, Luo Z, Ngiam K Y, Zhang M, Zheng K, Chen G, Ooi BC, Yip WL. Big healthcare data analytics: Challenges and applications. In Handbook of Large-Scale Distributed Computing in Smart Healthcare 2017 (pp. 11-41). Springer, Cham.
[5] Sacristán, José A., and Tatiana Dilla. "No big data without small data: learning health care systems begin and end with the individual patient." Journal of evaluation in clinical practice21.6 (2015): 1014- 1017.
[6] Raghupathi W, Raghupathi V. Big data analytics in healthcare: promise and potential. Health information science and systems. 2014 Dec 1; 2(1):3.
[7] Bates D W, Saria S, Ohno-Machado L, Shah A, Escobar G. Big data in health care: using analytics to identify and manage high-risk and high- cost patients. Health Affairs. 2014 Jul 1;33(7):1123-31.
[8] Palanisamy V, Thirunavukarasu R. Implications of Big Data Analytics in developing Healthcare Frameworks–A review. Journal of King Saud University-Computer and Information Sciences. 2017 Dec 9.
[9] https://healthitanalytics.com/news/using-the-ehr-to-dive-into-data- mining-clinical-analytics
[10] Jensen P B, Jensen L J, Brunak S. Mining electronic health records: towards better research applications and clinical care. Nature Reviews Genetics. 2012 Jun;13(6):395.
[11] Zikopoulos P, Eaton C. Understanding big data: Analytics for enterprise class hadoop and streaming data. McGraw-Hill Osborne Media; 2011 Oct 19.
[12] https://www.healthdataarchiver.com/health-data-volumes-skyrocket- legacy-data-archives-rise-hie/
[13] McAfee A, Brynjolfsson E, Davenport T H, Patil DJ, Barton D. Big data: the management revolution. Harvard business review. 2012 Oct 1;90(10):60-8.
[14] Tsai C W, Lai C F, Chao H C, Vasilakos A V. Big data analytics: a survey. Journal of Big Data. 2015 Dec;2(1):21.
[15] Demchenko Y, De Laat C, Membrey P. Defining architecture components of the Big Data Ecosystem. InCollaboration Technologies and Systems (CTS), 2014 International Conference on 2014 May 19 (pp. 104-112). IEEE.
[16] Rahimi A, Liaw S T, Ray P, Taggart J, Yu H. Ontological specification of quality of chronic disease data in EHRs to support decision analytics: a realist review. Decision Analytics. 2014 Dec 1;1(1):5.
[17] Chen M, Hao Y, Hwang K, Wang L, Wang L. Disease prediction by machine learning over big data from healthcare communities. IEEE Access. 2017;5:8869-79.
[18] Chen M, Yang J, Zhou J, Hao Y, Zhang J, Youn C H. 5G-Smart diabetes: towards personalized diabetes diagnosis with healthcare big data clouds. IEEE Commun. 2018 Apr 1.
[19] Yin W, Yang X, Zhang L, Oki E. ECG monitoring system integrated with IR-UWB radar based on CNN. IEEE Access. 2016;4:6344-51.
[20] Soni, J, U. Ansari, D. Sharma and S. Soni, "Predictive data mining for medical diagnosis: An overview of heart disease prediction," Int. J. Comput. Applic. 17: 43-48
[21] Hsi-Jen. "A retrospective analysis of prognostic indicators in dental implant therapy using the C5.0 decision tree algorithm", Journal of Dental Sciences, Volume 8, Issue 3 , 248-255, 2013
[22] Chawla N V, Davis DA. Bringing big data to personalized healthcare: a patient-centered framework. Journal of general internal medicine. 2013 Sep 1;28(3):660-5.
[23] Chen M, Zhang Y, Qiu M, Guizani N, Hao Y. SPHA: Smart personal health advisor based on deep analytics. IEEE Communications Magazine. 2018 Mar;56(3):164-9.
[24] García-Magarino I, Lacuesta R, Lloret J. Agent-based simulation of smart beds with Internet-of-Things for exploring Big Data analytics. IEEE Access. 2018; 6:366-79.
[25] Viceconti M, Hunter P J, Hose R D. Big data, big knowledge: big data for personalized healthcare. IEEE J. Biomedical and Health Informatics. 2015 Jul 1;19(4):1209-15.
[26] Yassine A, Singh S, Alamri A. Mining human activity patterns from smart home big data for health care applications. IEEE Access. 2017;5:13131-41.
[28] Jain P, Gyanchandani M, Khare N. Big data privacy: a technological perspective and review. Journal of Big Data. 2016 Dec 1;3(1):25.
[29] Sahi M A, Abbas H, Saleem K, Yang X, Derhab A, Orgun M, Iqbal W, Rashid I, Yaseen A. Privacy Preservation in e-Healthcare Environments: A Review. IEEE Access. 2017.
[30] Sarkar B K. Big data for secure healthcare system: a conceptual design. Complex & Intelligent Systems. 2017 Jun 1;3(2):133-51.
[31] Hill K. How target figured out a teen girl was pregnant before her father did. Forbes, Inc. 2012
[32] Patil P, Raul R, Shroff R, Maurya M. Big data in healthcare. 2014.
[33] Big Data security and privacy issues in healthcare—Harsh KupwadePatil, Ravi Seshadri. 2014
[34] Sectorial healthcare strategy 2012–2016 Moroccan healthcare ministry
[35] LeFevre K, Ramakrishnan R, DeWitt D J. Mondrian multidimensional k-anonymity. In: Proceedings of the ICDE. 2006. p. 25
[36] Iyenger V. Transforming data to satisfy privacy constraints. In: Proceedings of the ACM SIGKDD. 2002;279–88.
[37] Huddar V, Desiraju B K, Rajan V, Bhattacharya S, Roy S, Reddy C K. Predicting complications in critical care using heterogeneous clinical data. IEEE Access. 2016;4:7988-8001.
[38] Johnson A E, Ghassemi M M, Nemati S, Niehaus K E, Clifton D A, Clifford GD. Machine learning and decision support in critical care. Proceedings of the IEEE. 2016 Feb;104(2):444-66.
[39] Luo G. MLBCD: a machine learning tool for big clinical data. Health information science and systems. 2015 Dec 1;3(1):3.
[40] Lo'ai A T, Mehmood R, Benkhlifa E, Song H. Mobile cloud computing model and big data analysis for healthcare applications. IEEE Access. 2016;4:6171-80.
[41] Yao C, Qu Y, Jin B, Guo L, Li C, Cui W, Feng L. A convolutional neural network model for online medical guidance. IEEE Access. 2016;4:4094-103.
[42] Zhang Y, Qiu M, Tsai CW, Hassan M M, Alamri A. Health-CPS: Healthcare cyber-physical system assisted by cloud and big data. IEEE Systems Journal. 2017 Mar;11(1):88-95.
[43] Jiang P, Winkley J, Zhao C, Munnoch R, Min G, Yang LT. An intelligent information forwarder for healthcare big data systems with distributed wearable sensors. IEEE systems journal. 2016 Sep; 10(3):1147-59.
[44] Ahmad A M, Khoei A T, Ghapanchi A H, BAnalysis of Research in Healthcare Data Analytics.^, Australasian Conference on Information Systems.

AUTHORS PROFILE

Mr. H K Madhu is a research scholar at Sri Siddhartha Institute of Technology, Tumkur(SSAHE), having nineteen years of Teaching expereance at Department of M C A, in Bangalore Institute of Technology, Bengaluru. His research interest includes Data Mining and Big Data Analytics.

Dr B R Prakash is working as an Associate Proffessor, in the department of M C A, at Sri Siddhartha Institute of Technology, Tumkur (SSAHE), having fifteen years of experience in areas of teaching and research. Dr. Prakash complited his Ph.D in Computer Science from Bangalore University. He has published more than 55 papers in reputed journals and International Conferences.

