Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 132 (2018) 1049–1059

Procedia Computer Science

www.elsevier.com/locate/procedia

International Conference on Computational Intelligence and Data Science (ICCIDS 2018)

Big Data and Machine Learning Based Secure Healthcare Framework

Prableen Kaur[a], Manik Sharma[b*], Mamta Mittal[c]

[a]*Ph.D. Research Scholar, Department of Computer Science and Applications, DAV University, Jalandhar, India*
[b]*Department of Computer Science and Applications, DAV University, Jalandhar, India*
[c]*Department of Computer Science and Engineering, G. B. Pant Govt. Engineering College, New Delhi, India*

Abstract

The paper presents a brief introduction to big data and its role in healthcare applications. It is observed that the use of big data architecture and techniques are continuously assisting in managing the expeditious data growth in healthcare industry. Here, initially an empirical study is performed to analyze the role of big data in healthcare industry. It has been observed that significant work has been done using big data in healthcare sector. Nowadays, it is intricate to envision the way the machine learning and big data can influence the healthcare industries. It has been observed that most of the authors who implemented the use of machine learning and big data analytics in disease diagnosis have not given significant weightage to the privacy and security of the data. Here, a novel design of smart and secure healthcare information system using machine learning and advanced security mechanism has been proposed to handle big data of medical industry. The innovation lies in the incorporation of optimal storage and data security layer used to maintain security and privacy. Different techniques like masking encryption, activity monitoring, granular access control, dynamic data encryption and end point validation have been incorporated. The proposed hybrid four layer healthcare model seems to be more effective disease diagnostic big data system.

Peer-review under responsibility of the scientific committee of the International Conference on Computational Intelligence and Data Science (ICCIDS 2018).

* Manik Sharma
E-mail address: manik_sharma25@yahoo.com
Keywords: Big Data, Healthcare, Big data analytics, disease diagnosis, predictive analysis, security and privacy.

Peer-review under responsibility of the scientific committee of the International Conference on Computational Intelligence and Data Science (ICCIDS 2018).
10.1016/j.procs.2018.05.020

1. Introduction

Today, due to the expeditious development of internet and cloud computing, data grows rapidly at uncontrollable rate in every organization [1]. Wal-Marts import approximately 2.5 petabytes of data every hour into databases, Facebook handles more than 250 million photos and 900 million objects every day etc [2]. Due to this explosive growth of data, explications are essential in order to glean the valuable insights from datasets. The effective utilization of data is important as it is deemed as a building block for an organization. The effective data analysis can be very useful in sale forecasting, disease diagnosis, economic analysis, business management, social network analysis etc. Some organizations formerly use analytics on the organized data in the form of reports. The early idea of big data was introduced in paper "Visually Exploring Gigabyte Datasets in Real Time" and was published in 1999 [3]. In 2001, Doug Laney defines the big data characteristics in 3V's i.e. Velocity, Volume and Variety in his paper 3D Data Management: Controlling Data Volume, Velocity and Variety.Hadoop is one of dominant framework that is used to manage and analyze the unstructured big data [4].

In general, big data refers to the voluminous and complex amount of data collected from sources like web, enterprise applications, mobile devices and digital repositories which cannot be easily managed by using traditional tools. Big data is not only about the large data size, rather, it is an act of storing and managing data for eventual analysis [5]. As the humans are getting digitized, therefore, the computing embroils data with greater variety, volume and velocity [6]. As per Doug Laney, there are various V's involved in big data, however, 3V's model is the basic big data model. The significance of 3V's is briefly mentioned below [7][8]:

- Volume: Big data analyze comparatively a huge quantity of data such as in terabytes [9][10].
- Variety: Big data incorporates data from distinct sources that appears in numerous formats such as structured, unstructured, multifactor, probabilistic [9][10].
- Velocity: Big data handles the fast processing of data to promote the decision making process [9][10].

One more characteristic of big data is concatenated with 3V's i.e. veracity that represents the credibility and appropriateness of data for users [9][10]. Big data offer tremendous opportunities in healthcare sector to improve efficiency and quality in healthcare, detection of health threats, managing human health by diagnosis disease in early stage and in assisting better decisions.[11] However, data storage, data standardization, data transfer and privacy are the some of the major challenges in big data industry[12][13]. Several researches have tried to incorporate security aspect in healthcare system. Two different approaches have been presented below:

Ahmed E. Youssef (2014) proposed a framework based on big data analytics for healthcare information system (HIS) shown in fig. 1. This framework was used to manage and analyze huge amount of data to benefit the patients as well as healthcare professionals. HIS framework consists of five main components as mentioned below.

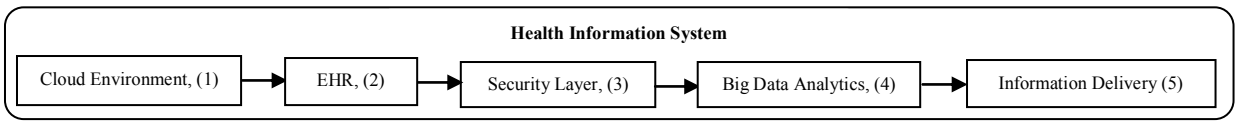

Fig.1. Health Information System framework proposed by Youssef, 2014[3]

First component is cloud environment used to provide various services and allow data sharing to authorized users. Second component is electronic health record used to integrate the patients' data from different locations. Third component is the security layer, where the various security issues such as, authentication, authorization, confidentiality of patients' data was managed effectively. This layer includes various encryption algorithms viz. AES, RC4 to protect the data. Also different authentication techniques viz. One Time Password (OTP), Two Factor Authentication (2FA) was used to restrict the unauthorized users. OTP is a method used to generate a different password every time to secure the data on network. This layer was also used to grant or revoke privileges to authorized users. Not only this layer was advantageous for patients and healthcare professionals, but it also benefits health insurance agencies to verify the decisions on the treatments of their customers. Fourth layer of HIS was used to deploy big data analytic tools get insight from raw data. Finally the care delivery organizations distribute the

healthcare information in different locations. This framework assured the confidentiality and privacy of health data to improve the security and quality of healthcare services [14].

To maintain the privacy of patients' health data Bouhriz Mounia et al., (2015) introduced the two way procedure. First procedure was focused to control the unauthorized access on medical records at the time of production, processing, sharing and storing of data. This can be performed by applying various cryptographic algorithms and data backup strategies. In the second half procedure, the patients' profiles have generated to protect the patients' medical records. This can be performed by applying some privacy constraints such as OTPs. The second procedure was focused on generating and protecting the patients' medical records [15].

Initially, big data and its role in different sectors have been briefly presented. Additionally, 19 different papers related to big data and healthcare has been reviewed and their outcomes are highlighted. The innovation lies in the proposed healthcare framework. A four layer framework has been proposed. Data security and privacy layer has been incorporated to preserve high degree of confidentiality and privacy of patients' data.

Section 2 presents the applications of big data in various sectors. In Section 3, analysis of big data applications based on Google Scholar database is presented. Section 4 shows the overview on big data analytics tools and techniques. Quality healthcare and big data analytics is discussed in section 5. Section 6 presents the article selection process. Design of Secure Big Data and Machine Learning based Healthcare Framework is presented in section 7. Conclusion is presented in section 8.

2. Big Data Applications

Since 2010, big data was and still remains in the spot. Big data has been increasingly used in different industrial, social and professional sectors [16]. Some of the major categories and applications where big data is used are shown in Table 1.

Table 1: Applications of big data in various categories

Category	Applications
Public Sector	Tax reduction, Social security, Energy exploration, Environmental protection, Power investigation, Public safety.
Healthcare industry	Cost reduction in medical treatments, Prediction of diseases, Eliminate the risk factors associate with diseases, Improves the preventive care , Analyzing drug efficiency.
Education and learning	Students' preferred learning mode, Track students' performance, Provide guidance, Gives real time feedbacks and updates, Improving the learning material, Cross checking of assignments, Digital students assessment.
Insurance industry	Predicting customer behavior, Evaluate the risk of insuring, Monitoring real time claims, Customer retention, Managing premium for the policies, Manage the fraudulent claims.
Transportation sector	Traffic control, Route planning, Intelligent transport systems, Congestion management, Revenue management in private sector, Technological enhancement, Forecasting routes to reduce cast on petroleum.
Industrial and natural resources	Integrating geospatial, temporal, graphical and text data, Analyze consumption of utilities.
Banking	Analyzing big businesses, Prognostic Analytics, Analyzing shopping patterns of customers, Analyzing CRM tactics of competitors, Customer statistics alteration.
Fraud detection	Detect misuse of credit cards, debit cards, Archival of inspection tracks, Treatment for venture credit hazard, Public analytics for business
Entertainment	Manage content for target audience, Measure content performance.

3. Analysis of Big Data Applications based on Google Scholar Database

To find the exact usage of big data in different applications, following queries have been executed on Google Scholar and the analysis of last eight years (2010 to 2017) have been accessed and presented in Fig. 2. "Applications of Big Data" in "Public Sector", "Applications of Big Data" in "Healthcare", "Applications of Big

Data" in "Transportation", "Applications of Big Data" in "Banking", "Applications of Big Data" in "Insurance", "Applications of Big Data" in "Fraud Detection", "Applications of Big Data" in "Entertainment", "Applications of Big Data" in "Learning".

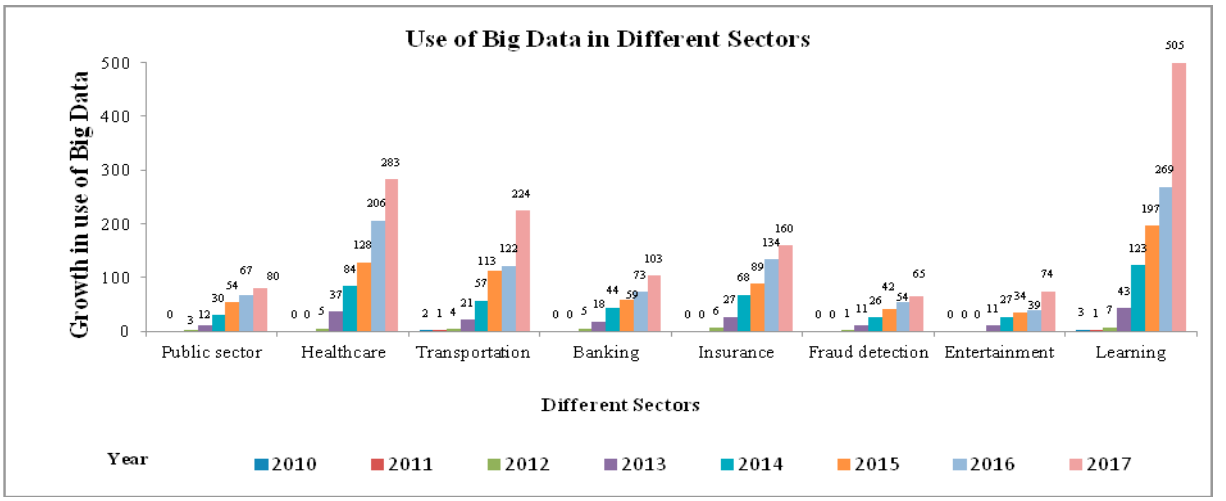

Fig 2. Comparison of big data applications in various sectors from 2010 to 2017

Fig. 2 shows the escalation in the utilization of big data from 2010 to 2017. It is observed that to get valuable insights of data, the big data has been successfully used in different sectors like public, healthcare, transportation, banking, insurance, entertainment, learning and fraud detection. Maximum work was done in healthcare and learning sector with big data in 2017. The highest rate of increase in the use of big data has been observed in learning sector. The rate of escalation in last eight years is 99.6%. Likewise, the rate of increase is 99% and 96% for transportation and insurance section respectively. However, entertainment and banking are the areas where minimum rate of increase in the usage of big data has been witnessed.

4. Tools and Techniques used in Big Data Analytics

Big data enables an organization to store and manage large amount of data at high speed to get the right insight from data. Some tools and techniques are required to explore heterogeneous and voluminous data to transform raw data into information to enhance the process of decisions makings. Big data analytics (BDA) refers to the tools and methods that transform huge data into useful information for analytical use. In general, BDA is blending technology comprises of IT, business professional and data scientist. It deals with getting deeper insights of the business organization so that the organization can be steering into right direction. With BDA, one may give competitive edge to their rival parties. The reason behind the emergence BDA is the unexpected tremendous growth of annual data. Nowadays, there are billion, trillion of data transactions over a day or sometimes in an hour. In twitter and Facebook, hundreds of millions, "tweets" and billions of likes are posted in a day. On average, 2.5 quintillions bytes are produced in a day i.e. a mammoth data volume. To manage big data various techniques are available, but these can be employed depending upon the type of data to be analyzed and the research question or problem to be solved [17][18]. Based upon problems, and the type of data to be analyzed, different big data tools can be used. Table 2 shows some tools and techniques in big data scenario.

Table 2. Tools and techniques for big data analytics

Techniques	Tools	Description
NoSQL	Cassandra	It is an open source and distributed database management system with fault tolerance facility [19].

	Hbase	It is a distributed database for storing large tables and provides real time read/write access and fault tolerance [19].
	ZooKeeper	It provides data synchronization services for distributed applications [19].
Map Reduce	Hive	It is a data warehouse software tool for applying ad hoc queries over huge datasets [19].
	Pig	It is a scripting language to write the Map Reduce transformations and provides parallel processing to handle huge datasets [19].
	Flume	It is a service used to stream huge logs of data from distinct sources [19].
	Oozie	It is an open source tool to manage the workflow [20].
Storage	HDFS	Hadoop Distributed File System is a java based file system used to manage huge quantity of data and follow the master/slave structure to manage the data with fault tolerance facility [19].

5. Quality Healthcare and Big Data Analytics

Healthcare and medical data is kind of a big data not because of its huge volume and size but also for its intricacy and timeliness. Health information includes patient related data viz. patients' Electronic Health Reports (EHR), diagnostic reports, doctor's prescription, medical images, pharmacy records, and research data from medicinal journals [21]. Today, it is imperative to digitize the data generated by healthcare industries to improve the quality of treatment, early diagnosis of diseases to avoid risk factors and for better management of information system in hospitals. There are prodigious opportunities to impact quality and productivity of healthcare using big data analytics. Analytical insights reaped from substantial analysis of medical data that can probably improve to clinical models and realize efficiencies through wise delivery of care [22]. Some challenging tasks for healthcare industry are:

- How to decide the most effective treatment for particular disease?
- How certain policies impact the outlay and behavior?
- How does the healthcare cost likely to rise for different aspects in future?
- How the claimed fraudulent can be identified?
- Does the healthcare outlay vary geographically?

There are four major pillars for quality healthcare. Fig. 3 represents these pillars.

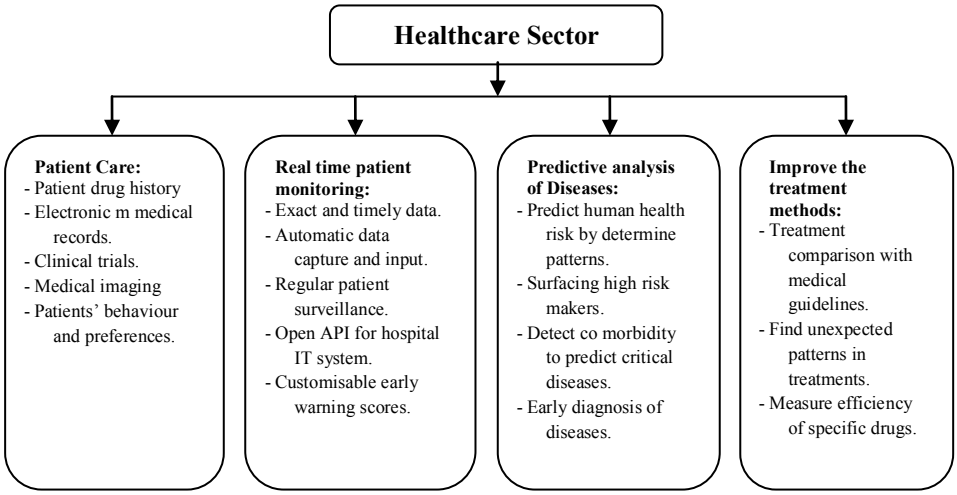

Fig.3. Big data in healthcare sector

All these four pillars of quality healthcare can be potently managed by using descriptive, predictive and prescriptive big data analytical techniques.

- **Patient Centric Care:** Provide relief to the patients by providing evidence based diseases diagnoses based on clinical data at early stage and minimizing drug dosage to avoid side effects. This helps to decrease readmission rates in hospitals and also reducing cost for the patients.
- **Predictive Analysis of Diseases:** Predicting the viral disorders in early stage before spreading based on the live analysis. This can be determined by analyzing the patients' social logs who are suffering from a disease in a particular location. This further helps the healthcare professionals to advise the victims by taking necessary preventive measures [23].
- **Real Time Patient Monitoring:** Monitoring whether the hospitals are setup according to the norms setup by Indian medical council. This periodical check-up helps government in taking necessary measures against disqualifying hospitals [23].
- **Improving the Treatment Methods:** Customized patient treatment monitoring depends on the analysis of medications dosages can be changed for rapid relief. Analysis of patients' data who are enduring from similar symptoms help doctors to provide effective medicines to new patients [23].

Some diseases viz. psychological, cardiovascular, liver disorder and diabetes are related with long term medical issues. Big data analysis helps to enhance the healthcare by early prediction of diseases which further improve the life expectancy of patients. Due to early predictions, healthcare experts can also suggest better treatment at early stage [21].

6. Big Data in Healthcare

The study includes a systematic review on big data applications in healthcare sector. There are 2357 number of articles published from 2010 to 2016 for big data applications in all sectors. Here, 460 articles represent the performance of big data in healthcare. After successful screening (six level), 19 articles are finally integrated in this study. These selected articles are divided in four categories under healthcare sector viz. patient centred care (4 articles), predicting diseases (6 articles), real time monitoring of patients (6 articles) and improving treatment methods (3 articles).

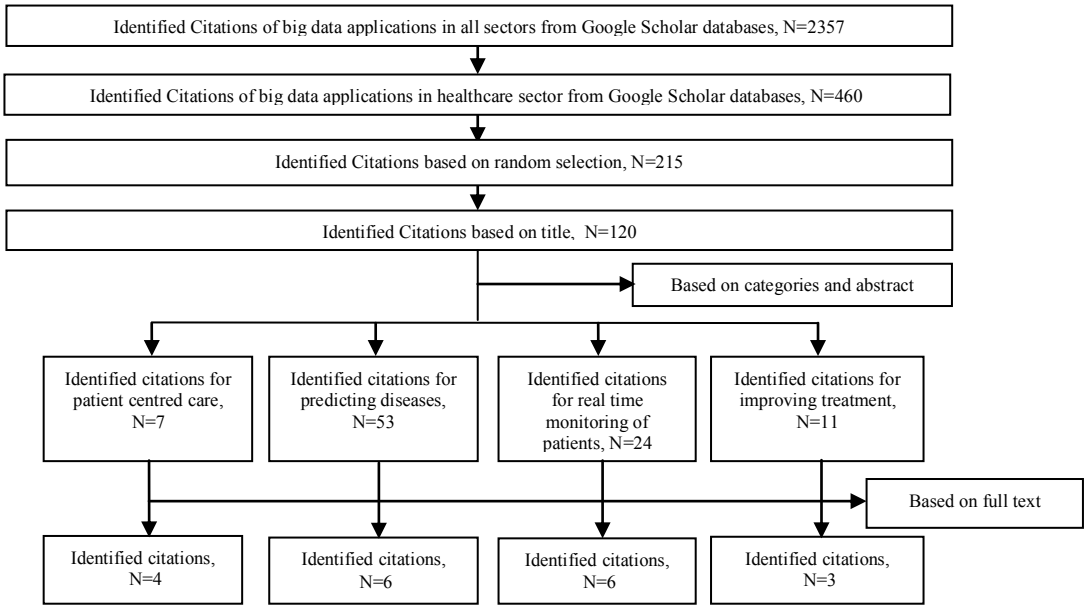

Fig.4. Article Selection Process

Table 3 demonstrates the research outcomes from 2013 to 2016 in healthcare industry using big data.

Table 3. Review of big data applications in healthcare industry

Category	Author & Year	Outcome
Providing patient centric Care	Raghupathi et al. (2014) [20]	The study introduced a conceptual architecture to present big data analytic outlines in healthcare. The framework is categorized into four levels to get insights from raw data by applying big data techniques viz. hadoop, map reduce, pig Cassandra, Hbase, zookeeper, oozie, avro, mahout etc. By introducing big data tools healthcare cost can be curtailed and get appropriate information from large datasets.
	Kim et al. (2015) [24]	Author used a Complex Event Processing (CEP) based technology to propose a real time big data analytics system to manage the ERP systems for medical institutions. The study provides the accurate information by analysing thousands of records in every second and applicable for medium sized hospitals.
	Patel et al. (2016) [25]	Authors discussed the different opportunities of big data analytics in healthcare sector such as reduction in healthcare cost, predict diseases to overcome risks, providing preventive and virtual care, identification of patients, drug discovery and efficacy. The study also point some challenges of big data viz. patient's privacy and security, cost for establishing big data architecture, data aggregation in healthcare industry.
	Dayal et al. (2016) [26]	Authors analysed the quality of healthcare services in India for a period of 1950 to 2015. Datasets was analyzed using Pig Latin Script to generate research queries. The purpose of study is to provide the better healthcare services to all the stratum of the society.
Predictive Analysis of diseases	Chawla et al. (2013) [27]	Authors proposed a patient centric healthcare framework called Collaborative Assessment and Recommendation Engine (CARE). The system helps physicians to manage the risk of diseases for patients. The study used filtering techniques to find the patient similarities to produce risk of diseases.
	Kuriyan et al. (2013) [28]	Authors proposed architecture for forecasting cancer and other chronic diseases viz. diabetes, hypertension, high cholesterol, osteoarthritis, asthma and morbid obesity in patients. Big data analytics was used for efficient diagnosis also helps to lessen the healthcare cost. The study confirmed that the diabetic patients are more prone to cancer disease..
	Saravana et al. (2015) [21]	Author used Hadoop and Map reduce environment to predict the type of diabetic patient. The outcomes of this study can further help the physicians to give better treatment to the patients.
	Abinaya (2015) [29]	Authors proposed the e-Health service application for diagnosing heart diseases. Hadoop map reduce with data mining methods and techniques were used to develop the architecture of system and Heterogeneous Autonomous Sources with Distributed and Decentralized Control and evolving the relationship among various data (HACE) theorem was used to manage the relationship between data.
	Wang et al. (2016) [30]	The study proposed to diagnose the undiagnosed diabetes from raw data by using the concept of mHealth application to track the data through cloud. Authors concluded that older adults and women are commonly affected by type 2 diabetes mellitus.
	Razavian et al. (2015) [31]	The study proposed a model by using machine learning and big data analytics to predict type2 diabetes as well as uncover the associated risk factors. The study performed the risk assessment on large data sets with 42000 variables collected from healthcare utilization, administrative, pharmacy and laboratory records. Authors concluded that the administrative data has shown the better prediction performance and enable to uncover the risk factors associated in different stages.
Real time monitoring of patients and hospital's quality	Shinde (2015) [32]	The study proposed a system for real time monitoring by using hadoop and HDFS. Authors analyze that the healthcare system can be enhanced by delivering information to right person on time.
	Sukumar et al. (2015) [22]	The study considered the data quality concept in healthcare industry by managing health data lifecycle, erroneous data, process of data collection source and pedigree of data.
	Luo et al. (2015) [33]	The study has shown the progress of big data applications in healthcare industry and discussed the threats and opportunities associated with it. Authors highlighted the progress of storage and retrieval of patients' electronic record in big data, data security of records and analysis of data for decision making process.
	Balladini et al. (2015) [34]	The study proposed real time architecture of big data for Francisco Lopez Lima Hospital at Argentinia to process physiological data. Authors analyze the current state of patients' data in intensive care unit by using public and local cloud computing platform that will be further useful for the hospital with same characteristics.
	Archenaa et al.	The study performed the real time analysis on the huge amount of healthcare and government data.

	(2015) [23]	Big data analytics was performed using hadoop/map reduce for predicting needs of citizen. The study targets on the management of citizen care by providing effective services to the citizens.
	Boukenze et al. (2016) [35]	Authors employ big data analysis and data mining to develop the architecture for monitoring diseases in real time to perform efficient decision making.
Improve the treatment methods	Lusher et al. (2013) [36]	The study focused on evolution of data-driven medicinal chemistry. The data driven research used to enhance decision making process for drug discovery. Here, big data was used to manage data from distinct sources. Authors concluded that the data driven approaches uncover meaningful patterns and relationships from large datasets.
	Herland et al. (2014) [37]	The study analyzes the data from the four levels including tissue, molecular, patient and population by using big data tools. Using these four levels study proposed to predict data using gene expressions, ICU readmission, using MRI data for further predictions.
	Belle et al. (2015) [38]	The study focused on the area of genomic data processing, image analytics and signal processing of psychological data.

Today, in healthcare sector data size is not only the one major issue, security and privacy of medical records is equally important. Generally, authors in their study only concentrates on the interoperability and integration of big data for analyzing informative patterns and hardly concerned with the security. Whereas, the security of medical records is the major challenge.

7. Design of Secure Big Data and Machine Learning based Healthcare Framework

The above deliberation motivated the need for using big data ecosystem in healthcare to address solutions related to the privacy and optimization of data. Fig. 5 represents the design for secure and optimal healthcare system. The system has four layers viz. Data source, Data storage, Security and Machine learning based application layer. The innovation lies in the data optimization and security features that have been incorporated in this united framework. The distributed environment and storage optimization helps to optimize storage resources that ultimately enhance the operation performance of the system. Additionally, the effort is made to maintain security and privacy of patients' data by incorporating different data protection techniques.

First layer deals with heterogeneous data source. Clinical records, operational data (flat files, relational data, and ASCII format), genome data, medical images (X-ray, MRI, CT Scan etc.) are main source of data layer. Second layer is responsible for storing heterogeneous data. Depending upon the type of data, it can be stored accordingly in the form of data files or in data warehouse. Distributed platform provide the facility of decomposing the data and to store it on different sites. Finally this layer takes care of storage optimization so that the memory resources can be effectively utilized. In healthcare sector, security is most important concern. In this model, to make patients' data secure, different techniques like masking encryption, dynamic data encryption, Granular access control and activity monitoring module have been incorporated in security and privacy layer. The machine learning based application layer is another important layer of proposed framework. It has five sub modules called early diagnosis, drug discovery, epidemic outbreak forecasting, data analytics and visualization. Here different machine learning methods like naive bayes, support vector machine, decision tree, genetic algorithm etc. may be used for the above mentioned sub modules. Depending upon the performance and accuracy, machine learning techniques may further be combined with other soft computing techniques to get better results. In Layer4, predictive models are developed using big data methods to improve the decision making process and to get insight from data. This layer encapsulates the healthcare information services and applications. Finally, the generated information can be presented using machine learning, SQL queries and medical imaging analytics.

This framework is proposed to provide integration and analysis of distributed data that further help patients to take better treatments and healthcare experts to take effective decisions. Privacy and security are considered as a critical issue in big data. Data privacy involves the appropriate use of data to protect it from unauthorized access. Whereas, security is the act of safeguarding data and data resources using technology and procedures preparing from inspections, unauthorized access, recordings etc. It is hard to have good data privacy without good security practices. Various methods viz. masking encryption, activity monitoring, granular access control and dynamic data encryption are used to manage the privacy and security of big data in healthcare sector.

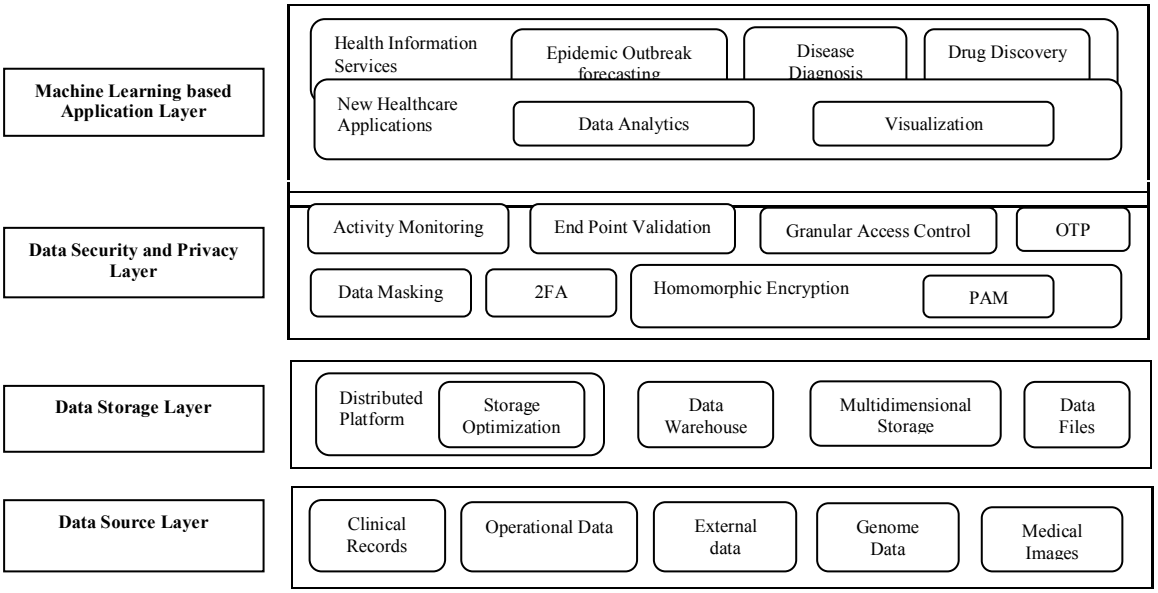

Fig. 5. Design of secure big data and machine learning based healthcare framework

Data masking ensures the confidentiality of data by encrypting the crucial segments of Personality Identifiable Information (PII) such as patients' personal information. It muddles or obscures the sensitive data by supplanting it with another data to protect to original data. In healthcare, data masking can be achieved in real time by using various EHR masking techniques such as data obfuscation, data perturbation, data exclusion, data hashing etc. Some algorithms viz. crypto algorithms, matrix encryption, MASK algorithm and cyber-secure protocols will be used to hide huge amount of patients' original data effectively. These algorithms can also used to encrypt the image data such as X-ray and MIR reports of patients [39].

Another way to protect the data is data encryption. It involves the transformation of original data into encoded patterns (i.e. cipher text) to restrict the unauthorized access. Once the data is encrypted there is a way called decryption to get original data from cipher text. Data encryption can be performed at different levels viz. column or field level, disk level, HDFS files, packet level in clusters to protect the data. Various symmetric and asymmetric algorithms are used to encrypt medical records of patient viz. Data Encryption Standards (DES), triple DES, Advanced Encryption Standard (AES) and blowfish. Among all the encryption algorithms blowfish encrypt huge data in less time with best security conditions [40]. One way to track the security of data is activity monitoring where the activities of privileged users are under observations. This method helps to monitor the access of data managed by privileged users, block the suspect activities, identify changes in the privileges, and revoke the unused rights. To manage the permission levels for authorized user, granular access control methods are used to restrict the actions of users [41]. Homomorphic encryption can be used to encrypt the result of a big data query. It decodes the results of the query so that on execution of the query unauthorized users are not able to get it. Authorized users can then use homomorphic decryption algorithm on the results to convert it into plain text. Moreover, to secure non relational data, the encryption techniques like AES, RSA, SHA-256 can be employed. In addition, the facilities like TLS and SSL can also be incorporated. The used of Pluggable Authentication Module (PAM) can further assist in protecting the data. The end point security can be effectively used to identify and connect only trusted wired or mobile devices [42].

8. Conclusion

This paper provides brief introduction to big data and its applications. The major intention of this study is to propose an optimized and secure big data healthcare framework. It is observed from the existing literature that the

applications of big data in healthcare industry revolutionize medical industry by providing better health and information to patients. Moreover, the use of information technology assist in reducing the costs associated with healthcare diagnosis. Here, initially the big data concept along with its major applications has been briefly introduced. From the publication statistical analysis, it is evident that in last eight years, a lot of research was done on big data application in distinct sectors like public, healthcare, transportation, banking, insurance, fraud detection, entertainment and learning. Furthermore, superlative research has been done using big data in healthcare and learning sector. The four major components of healthcare sector viz. patient care, real time patient monitoring, disease prediction and improvement treatment methods have also been highlighted.

Additionally, an inventive and secure four layers healthcare model has been designed and proposed. The bottom layer i.e. data source layer deals with heterogeneous data sources. This layer is supposed to manage heterogeneous data and transform heterogeneous to homogenous if required. The 2nd layer data storage layer is supposed to manage the storage optimization process. The data security and privacy layer provides several advance security and privacy features like data masking, activity monitoring, homomorphic encryption, PAM and granular access control. Finally, machine learning based application layer takes care of different tasks like disease diagnosis, drug discovery, data analytics and visualization support. Here, machine learning techniques like traditional data mining and emerging nature inspired computing can be employed for early disease diagnosis. The precision in disease diagnosis can be further enhanced by incorporating the concepts of fuzzy logic and information theory.

There are four layers of this framework. The uniqueness of this model lies in optimization and security of patients' data. To make optimal use of system resources, different techniques of optimizations like indexing, normalization has been used. Further, different security features like masking encryption, activity monitoring, granular access control and dynamic data encryption make patient data private and secure.

The proposed model can be practically implemented to verify the empirical results. The design can be enhanced in such a way that the framework can work on different mobile devices like mobile, tablet, phablet etc. The emerging nature inspired techniques and their hybridization can be implemented to further improve the precision of this diagnostic framework. Finally, the effort can be extended to examine the effect of preprocessing and different types of data over the overall efficiency of this system.

References

[1] Ozgur C., Kleckner M. and Li Y. (2015) "Selection of Statistical Software for Solving Big Data Problems: A Guide for Businesses, Students, and Universities." Sage Open: 1-12.

[2] Sagiroglu S. and Sinanc D. (2013) "Big Data: A Review. Presented in International Conference: Collaboration Technologies and Systems (CTS)." IEEE Xplore.

[3] Picciano A. G. (2012) "The Evolution of Big Data and Learning Analytics in American Higher Education." Journal of Asynchronous Learning Networks **16(3)**: 9-20.

[4] Gandomi A. and Haider M. (2015) "Beyond the hype: Big data concepts, methods, and analytics." International Journal of Information Management **35 (2)**: 137-144.

[5] M. M. Hansen1 , T. Miron-Shatz2 , A. Y. S. Lau3 , C. Paton (2014) "Big Data in Science and Healthcare: A Review of Recent Literature and Perspectives" Yearb Med Inform: 1-6.

[6] Lee I. (2017) "Big Data: Dimensions, Evolution, Impact, and Challenges" Business Horizons 60(3): 293-303.

[7] Zikopoulos P. and Eaton C. (2011) "Understanding Big Data: Analytics for Enterprise Class Hadoop and Streaming Data." McGraw-Hill Osborne Media.

[8] Chen M., mao S. and Liu Y. (2014) " Big Data: A Survey." Mobil Netw Appl **19(2):** 171-209.

[9] McAfee A. and Brynjolfsson E. (2012) "Big Data: The Management Revolution." Harvard Business Review.

[10] https://www.cognizant.com/industries-resources/healthcare/Big-Data-is-the-Future-of-Healthcare.pdf

[11] Kruse C. S., Goswami R. Raval Y., Marawi S. (2016) "Challenges and opportunities of big data in Healthcare: A systematic Review" JMIR Med Inform 4(4).

[12] Bhadani A. K., Jothimani D. (2016) "Big data: Challenges, opportunities and realities" In Singh, M.K., & Kumar, D.G. (Eds.), Effective Big Data Management and Opportunities for Implementation: 1-24.

[13] White S. E. (2014) "A review of Big Data in Healthcare: Challenges and Opportunities" Open Access Bioinformatics **4(6)**: 13-18.

[14] Youssef A. E.(2014) "A framework for secure healthcare systems based on big data analytics in mobile cloud computing environments." International Journal of Ambient Systems and Applications **2 (2)**: 1-11.

[15] Mounia B., and Habiba C.(2015) "Big Data Privacy in Healthcare Moroccan context." Paper presented at the 2nd International Workshop on Privacy and Security in HealthCare, Procedia Computer Science **63**: 575 – 580.

[16] Boinepelli H.(2015) "Applications of Big Data". In: Mohanty H., Bhuyan P., Chenthati D. (eds) Big Data. Studies in Big Data, Springer, **11**.

[17] Acharya. S. and Chellappan S. (2015) "Big Data and Analytics." Wiley

[18] Philip Chen C. L., and Zhang C. Y.(2014) "Data-intensive applications, challenges, techniques and technologies: A survey on Big Data." Information Sciences **275(10)**: 314-347.

[19] Ishwarappa and Anuradha J. (2015) "A Brief Introduction on Big Data 5V's Characteristics and Hadoop Technology." Procedia Computer Science **48**: 319 – 324.

[20]Raghupathi W., and Raghupathi V. (2014) "Big data analytics in healthcare: promise and Potential." Health Information Science and Systems **2(3)**: 5-10.

[21] Saravana N. M. K., Eswari T., Sampath P., and Lavanya S. (2015) "Predictive Methodology for Diabetic Data Analysis in Big Data." Procedia Computer Science **50**: 203-208.

[22] Sukumar R., Ramachandran N., and Ferrell R. K. (2015) "Big Data' in health care: How good is it?." International Journal of Health Care Quality Assurance: 2-9.

[23] Archenaa J. et al. (2015) "A Survey of Big Data Analytics in Healthcare and Government." Procedia Computer Science. **50**: 408 – 413.

[24] Kim M. J., and Yu Y. S. (2015) "Development of Real-time Big Data Analysis System and a Case Study on the Application of Information in a Medical Institution." International Journal of Software Engineering and Its Applications **9(7)**: 93-102.

[25] Patel S., and Patel A. (2016) "A Big Data Revolution in Health Care Sector: Opportunities, Challenges and Technological Advancements." International Journal of Information Sciences and Techniques **6**:155-162.

[26] Dayal M., and Singh N. (2016) "Indian Health Care Analysis using Big Data Programming Tool." Procedia Computer Science **89**: 521 – 527.

[27] Chawla N. V., and Davis D. A. (2013) "Bringing Big Data to Personalized Healthcare: A Patient-Centered Framework." Journal of General Internal Medicine **28(3)**: 660-665.

[28] Kuriyan J., and Cobb N. (2013) "Forecasts of Cancer and Chronic Patients: Big Data Metrics of Population Health." Cornell University library: 1-26.

[29] Abinaya K.(2015) "Data Mining with Big Data e-Health Service Using Map Reduce." International Journal of Advanced Research in Computer and Communication Engineering, **4(2)**: 123-127.

[30] Wang L., and Alexander C. A. (2016) "Big Data Analytics as Applied to Diabetes Management." European Journal of Clinical and Biomedical Sciences **2(5)**: 29-38.

[31] Razavian N., Blecker S., Schmidt A. M., McLallen A. S., Nigam S., and Sontag D. (2015) "Population-Level Prediction of Type 2 Diabetes from Claims Data and Analysis of Risk Factors." Big Data **3(4)**: 277-282.

[32] Shinde K. V. (2016) "A Real Time Monitoring System In Healthcare With Hadoop." 'Research Journey' International Multidisciplinary E-Research Journal, Special Issue-I : 15-19.

[33] Luo J., Wu M., Gopukumar D., and Zhao Y. (2016) "Big Data Application in Biomedical Research and Health Care: A Literature Review." Biomedical Informatics Insights **8**: 1-10.

[34] Balladini J., and Rozas C. (2015) "Big Data Analytics in Intensive Care Units: challenges and applicability in an Argentinian Hospital." Journal of Computer Science and Technology **15(2):** 61-67.

[35] Boukenze B., Mousannif H. and Haqiq A.(2016) "A Conception of a Predictive Analytics Platform in Healthcare Sector by Using Data Mining Techniques and Hadoop." International Journal of Advanced Research in Computer Science and Software Engineering **6(8):** 65-70.

[36] Lusher S. J., McGuire R., Schaik R. C. V., Nicholson C. D., and Vlieg J. D. (2014) "Data-driven medicinal chemistry in the era of big data." Drug Discovery Today **9(7)**: 859-868.

[37] Herland M., Khoshgoftaar T. M., and Wald R. (2014) "A review of data mining using big data in health informatics." Journal of Big Data **1(2)**:1-35.

[38] Belle A., Thiagarajan R., Soroushmehr S. M. R., Navidi F., Beard D. A., and Najarian K. (2015) "Big Data Analytics in Healthcare." BioMed Research International **2015**: 1-16.

[39] Wali H. G., Iyer N. C. (2016) "Modified MASK Algorithm for Image Encryption." Paper presented at the International Conference on Information Security and Privacy 2015, Nagpur. Procedia Computer Science **78:** 603 – 609.

[40] Arora R., and Sharma, S (2012) "Performance Analysis of Cryptography Algorithms." International Journal of Computer Applications **48 (21)**: 35-39.

[41] Gaddam A. (2015) "Securing Your Big Data Environment." Black Hat USA: 1-8.

[42] Potey M. M., Dhote C. H., Sharma D. H. (2015) "Homomorphic Encryption for Security of Cloud Data." Procedia Computer Science **79**: 175-81.