IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 3, March 2024

Big Data Analytics in Healthcare Systems

Ms. Fazilat Parkar[1], Mrs. Ashwini Sheth[2], Mrs. Akshata Chavan[3]
Student, M.Sc.IT.[1]
Assistant Professor, Department of I.T[2,3]
I.C.S. College, Khed, Ratnagiri

Abstract*: Big Data analytics has had a significant impact on the healthcare industry in recent years, allowing for improved patient outcomes, personalized care, and cost-reduction. This paper examines the role of healthcare data, its applications and benefits, and the technological advances made possible by big data, such as cloud computing and stream processing. Additionally, it highlights the challenges and opportunities that Big Data analytics can bring to health care systems. Ultimately, Big Data analytics can help to improve clinical decision-making, personalized medicine, and overall healthcare outcomes.*

This paper looks at how Big Data analytics impacts the healthcare system, the applications and benefits of analysing big data, the challenges of using big data in healthcare, and how Big Data can affect healthcare policy, resource allocation, and how to make healthcare delivery processes more efficient. It also looks at the ethical issues that Big Data can bring to the healthcare industry, like responsible data governance, patient consent, and transparency. Overall, it's a great read that will give you a better understanding of how Big Data can change healthcare.

Keywords: Big data, Big Data analytics, Healthcare, Personalized medicine, Precision medicine, Cloud computing, Stream processing

I. INTRODUCTION

Ensuring that patients are matched with the right medical treatments for their disease can reduce unwanted side-effects, enhance treatment quality, and prevent unnecessary treatment or waste in healthcare services. It can also lead to new medical treatments by exploring new drugs or by using existing drugs for novel or more targeted uses. Systems biology is a powerful approach that integrates multiple data sources and studies of biological processes. Many studies use network models to describe etiopathogenesis or immune responses, which can help discover new biomarkers for early detection. However, it is important to avoid bias of clinical data when using network models.

A wide range of medical equipment, particularly wearable devices, captures data continuously. The high speed of generated data often necessitates rapid processing in an emergency situation. Although the value hidden in a single data source may be small, the deep value of healthcare data (such as public health warning or personalized health guidance) could be maximized by combining data from electronic medical records (EMRs) with EHRs (see Zhang et al., 2017). Structural MRI is a high-resolution imaging technique that can be used to visualize a patient's brain. It is a powerful source of high-dimensional data and provides detailed brain maps in high spatial resolution. This technique is highly useful for research and clinical applications, as it can be used to reveal structural characteristics of the brain.

Mobile/web applications have been developed in the healthcare field. Patients can send a symptom query to providers via a server. The mobile applications may include first aid instructions. Patients may receive emergency assistance for further treatment or be directed to appropriate departments.

Mobile Cloud Computing (MCC) is a healthcare system that collects and analyzes real-time biomedical data (e.g. blood pressure, EKG) from users at various locations. A personalized healthcare app is installed on the user’s mobile device. The health data is synced to the healthcare system’s cloud computing service for storage and analysis.

With the help of cutting-edge information technology, big data can be captured in healthcare. This makes it possible to explore information to support policy-making. A life table is a useful tool for conducting research on population ageing and medical expenditure, which supports evidence-based policy-making. The cost of healthcare also increases as the population ages. Japan has started to use Big Data technologies in healthcare to enhance medical treatment and care for the elderly.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 3, March 2024

Big Data analytics can extract valuable information from big and complex datasets through data mining.

In this paper, literature research was conducted using the Scopus database and IEEE Xplore database. For the literature review, the combination of keywords, "Big data and healthcare", "Big Data and health care", and "big data and medical" were used to search for papers published from January 2015 to May 2018. The duplicate papers found from both databases were removed.

Healthcare Data

Healthcare data sources range from clinical text to biomedical images, electronic health records (EHRs) and genomic data to biomedical signals and sensing data. Analysis of genomic data allows people to have a much deeper understanding of the relationships between various genetic markers, mutations and disease conditions. Translating genetic discoveries into personalized medicine practice presents many challenges.

Clinical text mining transforms data from unstructured clinical notes into useful information.

Information retrieval and NLP are methods for extracting useful information from large amounts of clinical text.

Social network analysis helps uncover knowledge and new patterns that can be used to model and forecast global health trends (such as the outbreak of infectious epidemics).

This research is based on a variety of social media sources, including web logs, Twitter, Facebook, and other social networks, as well as search engines. To accurately diagnose diseases, it is necessary to employ appropriate diagnostic techniques prior to assessing their severity. Table 1 outlines five layers of personal information related to health.

Table 1. A scheme for diseases diagnosis in a system

Disease	**Diagnostic method**	**Health measures via IoT**
Hypertension	Frequency based, scale based	Blood pressure
Obesity	Scale based	Body weight, blood pressure
Heart diseases	Frequency matching, pattern-matching	ECG pattern
Water borne or infectious disease	Frequency based, scale based	ECG, temperature sensor, Camera pill (gastro intestinal tract)
Stress index	Frequency based, pattern-matching based	Emotiv EPOC sensor, other stress measuring sensors
Respiration index	Frequency matching, pattern-matching	Respiration sensor

Table 2. Datafication layers of personal health data

Layers	**Description**
Layer 1	Search, smartphone,buying,,healthcare servicessensors, medical and fitness device
Layer 2	Health-related clinical and non-clinical information in defined formats that is collected, stored and disseminated to third parties
Layer 3	In the majority of cases, the raw data is collected by private companies from Layer1 and Layer 2, as well as other private and public data sources. The resulting data is distributed or sold in a de-identified or identifiable form.
Layer 4	The data is re-processed, re-disseminated, or re-sold for a variety of uses by national governments, private companies, or public entities around the world.
Layer 5	There are laws and treaties around the world that protect people's privacy when it comes to their health data.

In a WBSN (Wireless Body Sensor Network), there are 6 wireless physiological sensors. These six sensors collect a patient's 6 vital signs (Body Temperature, Heart Rate/Pulse, Blood Glucose, Blood Pressure, ECG and Oximetry).

Healthcare data integration is an important topic, covering everything from Personal Health Information to Epigenomics. There are various ways to integrate healthcare data, such as data warehouse, data schema integration, link

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 3, March 2024

integration, webpage presentation, serviceoriented architectures, data dynamically at web in familiar format, view integration, data view integration, and data mash-ups.

Table 3 shows some of the key aspects or challenges of data fusion.

Table 3. Critical aspects in data fusion

Aspect	Description
1	No data processing
2	Balances information from diverse sources
3	Production with varying, denying and contradictoryinformation
4	Founding loss or impartialpurposes and regulation/penalty footings
5	Separating between delicate and difficult information joins, i.e. considering a arbitrary handle from which the information is created as subject to same parameters, or instep bookkeeping fair for conditions, covariations, similarity/dissimilarity, etc.

Big Data Analytics in Healthcare Systems

Big data, as shown in Table 4, has a high value in terms of volume, speed, and variety. It also has a low value in terms of value and complexity.

Big data has many applications in healthcare, such as disease surveillance and epidemic control. It can also be used to support clinical decision making and population health management.

In addition, big data can be used to detect diseases at an earlier stage.

By integrating Big Data analytics into smart healthcare systems, new forms of e/m and mobile health can be created. This can lead to improved efficiency and lower medical costs.

Table 4. Big data features

Aspects	Description	Examples in healthcare
Volume	Data size	Treatment Plans, Multiple Conditions, Cohorts of Patients
Velocity	Data generation rate (batches, streams, infrequent intervals)	Sensing & Diagnosis Transmitting Patient Status & Behaviors
Variety	Various formats and data types (numbers, text, images)	Clinical data and images, Medical data and images Omics data and images, Patients under different conditions
Variability	Data change with time	Wearable sensor health data
Veracity	Imprecise or untruthful data	Clinician's observations on patients' condition, feedback from patients
Value	Inherent value (often achieved through data mining)	Examining hundreds of patient testimonials and determining the adverse reactions of a drug
Complexity	Hierarchies, linkages between items and recurrent structure of data	The term “multi-pharma, multi-disorder” is often used interchangeably.
Sparseness	Low density of useful information (due to null values, missing data, etc.)	There are a lot of gaps in patient feedback about their progress and symptoms.

Precision medicine uses predictive analytics to:

Predict pharmaceutical outcomes. Identify patients who are most likely to benefit from pharmacist intervention. Help pharmacists better understand the risks associated with particular medication-related issues. Deliver interventions that are specifically tailored to meet the needs of patients Data collection and management, Data sharing and privacy, Data integration, Data mining, Visualization. As biotechnological advances continue, complex biomedical data with vast amounts of data are becoming available.Big Data analytics is needed to leverage this heterogeneous data. This data is used in various application areas, such as:Health informatics, Sensor informatics, Bioinformatics, Image informatics

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 3, March 2024

Veracity is a key factor in big data analytics. For example, personal health records may contain acronyms, typographical mistakes, and mysterious notes. Similarly, ambulatory measurements may be completed in unsupervised and unreliable environments compared to clinical data collected by trained clinicians in a clinical environment. Social media data from spontaneous unmanaged sources may lead to inaccurate predictions. Also, data sources can sometimes be biased. Noise data is a huge problem, especially when it is growing rapidly. Databases of varying completeness or quality can result in heterogeneous results. This increases the risk of false findings and ‘unreliable fact-finding trips’. The low quality of the data and the lack of randomization can also lead to biases. To increase the value of the big data, it is often used to link different databases and to analyze all existing data and related data. The process of data pre-processing includes data cleaning, data integration, data transformation, and data reduction. Data discretization is another important step in the data pre-processing process.

Systems based on big data streams, such as hospital discharge records (patient-level), electronic death certificates (electronic death certificates), and medical claims data (medical claims data using ICD coding), have been developed. Surveillance tactics have been proposed using big data streams (crowd sourcing, social media, internet search queries). Big Data technologies, such as NoSQL databases, have been used in the processing of healthcare information. Some features, such as local access and rational distribution of logical and physical data, are important to enhance parallel processing performance in distributed databases. A Big Data-driven approach and process that integrates clinical and molecular information has been proposed. Candidate biomarkers/therapeutic targets/drugs have been identified in the approach, followed by cross-species analysis, resulting in reduced costs and time required for biomarker /therapeutic development.

A clinical data warehouse has been developed for structured data. A set of modules have been developed for unstructured content analysis. The goal of the research was to build the first implementation of a big data framework. The framework manages the modules in a cluster of Hadoop and leverages the distributed computing power of Big Data.

A Hadoop based architecture has been developed to manage the big data of Twitter health. Twitter health big data analysis has the potential to revolutionize the way in which people and healthcare providers leverage advanced technologies to gain new clinical insights.

Open sources such as:Hadoop, Kafka, ApacheStorm, NoSQL, Cassandra, Apache Storm contains a set of generic primitives in which big data can be processed in real-time (Vanathi, 2017).Table 5 equivalences Apache Storm with Hadoop.

Table 5. A judgment of structures between Storm and Hadoop (Vanathi and Khadir, 2017)

Features	Storm	Hadoop
Data handling	Handled as topology	Handled as jobs
Data Processing	Real-time oriented	Batch oriented
Database Compatibility	Cassandra, NoSQL	HBase, SQL
Performance	Low latency	High latency

Studies on attribute reduction have been conducted using MapReduce, which is based on the RST. The procedures are as follows:

1) parallelize large-scale RST methods for feature acquisition.
2) implement these RST methods on a MapReduce runtime system like Twister, Phoenix, or Hadoop.
3) extract features from large datasets through data mining.
4) use the < key, < value > pair framework structure to accelerate the calculation of equivalence classes, and attribute significance.
5) parallelize the traditional attribute reduction process using MapReduce.
6) HPC stands for high-performance computing.
7) It is CPU-oriented with intensive computing for large-scale distributed data through internal or external networking.
8) It stands for cluster or grid computing with intensive computing for big data.
9) Hadoop HPC stands for Hadoop Hadoop Extended Computing.
10) Hadoop Efficient, reliable, and scalable.

IJARSCT

ISSN (Online) 2581-9429

International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)

International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal

Impact Factor: 7.53

Volume 4, Issue 3, March 2024

Table 6: Comparison of tools used for big data analysis in the healthcare system

Table 6. A comparison of tools used for analyzing big data

Tools	Type of Databases	Platforms	Advantages	Limitations
Hadoop	Non-relational database	Open source and cloudbased platform	Stores data with any structure such as Web logs	Lacks technical support and security
MapReduce	Non-relational database	Open source and cloudbased platform	Works well with semistructured and unstructured data such as visual and audio data.	Lacks indexing capabilities of modern database systems.
Google Big Query	Columnar database	Open source and cloudbased platform	Allows data to be replicated across diverse data centers.	Does not support indexes.
Microsoft Windows Azure	Relational database	Public cloud based platform	Allows users to make relational queries against structured, semistructured and unstructured files.	The size of the database is limited; it cannot handle huge databases.
Jaql	It is a query language for JavaScript object notation.	It is a proprietary query language.	Supports both structured and semi-structured data.	No user defined types; schema information only for possible values of a domain

Industry 4 and Industry 5 are strategic plans for manufacturing and custom-made medical devices, drugs, etc. Precision medicine in health is a type of big data application that takes advantage of multi-omics and IoT. Industry 5 is a plan that uses AI, IoT and next-gen tech policies to make sense of big data. For example, an intelligent healthcare framework was created using IoT technology to give everyone access to healthcare during their workout. It was also able to predict a person's health related vulnerability with the help of an AI neural network model. Verma and Sood (2018) list Data Management, Model Development, Visualization, and Business Models as four main areas of big data analytics. To get a better understanding of how to mine big data from an EHR, check out Table 7 from Wu (2017).

Table 7.Some methods for EHR data mining

Methods	Advantages	Limitations
Hidden Markov models	Simultaneous segmentation, detection, and classification in a waveform	Sensitive to the design of trained Markov model
Logistic regression with LASSO regularization	Reduces feature space	Prone to over-fitting
Logistic regression, local regression, cox regression	Direct estimates of relevant hazards for Cox regression; simple to interpret and implement	Sensitive to an outlier
Rule mining, directed acyclic graph, Allen's interval algebra	Temporal modeling/mining capabilities	Requires specific design of experiment
Conditional random fields	Resistant to differences in class prevalence; supporting temporal analysis	Sensitive to feature space size and regularization
Windowing, episode rule mining, relational subgroup discovery	Valid sequential methods for some clinical applications	Tradeoffs between complexity, simplicity and temporal resolution

Challenges of Big Data in Healthcare Systems

Healthcare big data comes with a lot of challenges, like how to capture, store, share, search, and analyse health data. It's also tricky to organize data after it's extracted from different sources and how to make it work together. Plus, it's hard to integrate physical data with high-tech methods for making clinical decisions. With more and more genomic data available, it's even harder to figure out what to do with it. Finally, there's the question of how to get consent to use healthcare data, like genetic data. There have been academic debates about the legality of this, and there are even arguments that Big Data can be used to improve healthcare systems. Here are some of the general challenges that Big Data in healthcare has to face.

Privacy and security measures are effective on small data sets; however, the ability to apply the same measures to large and streaming data sets may be a challenge, especially when dealing with patients' health data.

This can affect the quality of the data and the decision-making process for patients' healthcare.

Additionally, delays in real-time data processing can lead to lower quality patient care.

The mixing of dissimilar data bases can be difficult due to the data being dispersed across hospitals, laboratories, EHRs, and financial information systems.

Finally, there are no uniform standards for the types of healthcare data collected, which can make it difficult to further process the data.

II. CONCLUSION

The ability of traditional data processing techniques to process vastquantities of information in health careorganizations is being challenged by the emergence of Big Data analytics. This technology has the potential to revolutionize the healthcare industry, as it has the potential to be used in a variety of areas such as disease surveillance, outbreak control, clinical decision making, and population health management. The use of data analytics, which is based on the Hadoop architecture, offers a range of advantages, such as increased efficiency, increased reliability, and increased scalability. However, there are a variety of challenges associated with the usage of Big Data in health care systems, such as the need to capture, store, share, search, and analyse data, as well as data security and privacy issues, as well as the integration of disparate or heterogeneous data, and the need for standards for healthcare data.

The benefits are significant, but challenges remain, including data security, privacy, and standardized data formats. To ensure trust between patients and healthcare providers, robust cybersecurity measures and interoperable data standards must be implemented.

Looking forward, the future of big data analytics in healthcare promises exciting opportunities. Machine learning, AI, and data science technologies will continue to improve diagnostic accuracy and speed. Collaboration between researchers, healthcare professionals, and technology professionals will be necessary to overcome current barriers and unlock the potential of big data for global health.

To sum up, this research highlights the transformative potential of Big Data Analytics for healthcare, highlighting its potential to transform patient care, enhance operational efficiency, and improve the health of individuals and communities around the world. As we look ahead, a coordinated effort from all stakeholders in the healthcare ecosystem is needed to overcome challenges and unlock the full power of Big Data for global health.

REFERENCES

[1]. 1.Andreu-Perez, J., Poon, C. C., Merrifield, R. D., Wong, S. T. kaj Yang, G. Z. (2015). Terveydelle on big data. IEEE J Biomed Health Inform, 19(4), 1193-1208.

[2]. Belle , A. , Thiagarajan , R. , Soroushmehr , S. M. , Navidi , F. , Beard , D. A. and Najarian , K. (2015). Suurandmete analüüs tervishoius. BioMed Research International, 2015.

[3]. Capobianco, E. (2017). A systems and precision medicine approach to diabetes heterogeneity: a big data perspective. Clinical and Translational Medicine, 6(1), 23.

[4]. Cunha, J., Silva, C. and Antunes, M. (2015). Managing Health Twitter Big Bata with Hadoop Framework. Procedural Computing, 64, 425-431.

[5]. De Silva, D., Burstein, F., Jelinek, H.F. and Stranieri, A. (2015). Addressing the Complexity of Big Data Analytics in Healthcare: The Case of Diabetes Crime. Australasian Journal of Information Systems, 19, S99-S115.

[6]. Ding, W., Lin, C.T., Chen, S., Zhang, X., and Hu, B. (2018). Multiagent consensus-MapReduce-based feature reduction using co-evolutionary quantum PSO for big data applications. Neurocomputing, 272, 136-153.

[7]. Farid, D. M., Nowe, A., & Manderick, B. (2016, December). A feature clustering method for clustering high-dimensional genomic big data. In Future Technologies Conference (FTC) (pp. 260268). IEEE.

[8]. Hernandez, I. and Zhang, Y. (2017). Using predictive analytics and big data to optimize drug outcomes. American Journal of Health-System Pharmacy, 74 (18), 1494-1500.

[9]. Istephan, S. ja Siadat, M. R. (2015, novembris). An Extensible Query Framework for Unstructured Medical Data - A Grand Data Approach. En Data Mining Workshop (ICDMW), 2015 IEEE Internacia Konferenco pri (pp 455-462). IEEE.

[10]. Knoppers, B. M., and Thorogood, A. M. (2017). Ethics and big data in health. Nuna Opinion in Systematics, 4, 53-57. http://dx.doi.org/10.1037/0021-843X.111.1

[11]. Loai, A. T., Mehmood, R., Benkhlifa, E. and Song, H. (2016). Mobile Cloud Computing Model and Big Data Analytics for Healthcare Applications. IEEE Access, 4, 6171-6180.

[12]. Mathew, P. S. and Pillai, A. S. (2015, March). Big data solutions in healthcare: issues and perspectives. Innovations in Information, Embedded and Communication Systems (ICIIECS), 2015 International Conference (pp. 1-6). IEEE.

[13]. Mendelson, D. (2017). Legal protection of personal health data in the age of big data - a proposal for a regulatory framework. Ethics, Medicine and Public Health, 3(1), 37-55.

[14]. Murphy , S.N. , Avillach , P. , Bellazzi , R. , Phillips , L. , Gabetta , M. , Eran , A. and Kohane , I.S. (2017). Combining clinical and genomic questions using i2b2-Tri methods. PloS one, 12(4), e0172187.

[15]. Ni, J., Chen, Y., Sha, J. and Zhang, M. (2015, November). Hadoop-based distributed computing algorithms for healthcare and clinical data processing. In Online Computing for Science and Engineering (ICICSE), 2015 Eighth International Conference (pp. 188-193). IEEE.

[16]. Olaronke, I. and Oluwaseun, O. (2016, December). Big data in healthcare: future perspectives, challenges and solutions. In Future Technologies Conference (FTC) (pp. 1152-1157). IEEE.

[17]. Özdemir, V. ja Hekim, N. (2018). The Emergence of Industry 5.0: Making Sense of Big Data with Artificial Intelligence, the Internet of Things and Next Generation Technology Policy. Omics: Journal of Integrative Biology, 22(1), 65-76.

[18]. Panda, M., Ali, S.M. and Panda, S. K. (2017, March). Big data in healthcare: a mobile solution. In Big Data Analytics and Computational Intelligence (ICBDAC), 2017 International Conference (pp. 149-152). ieee.

[19]. Pramanik, M.I., Lau, R.Y., Demirkan, H. kaj Azad, M.S. A.K. (2017). Smart Health: Granda Datumo Ebligita Smart Cities Health Paradigm. Expert Systems with Applications, 87, 370-383. and ends with a specific patient. Journal of Evaluation in Clinical Practice, 21(6)