Chapter 32
Big Data Analytics and Internet of Things in Health Informatics

Pawan Singh Gangwar and Yasha Hasija

Abstract The increasing amount of data in healthcare industry has necessitated the adoption of big data techniques to deliver quality health services. As the healthcare and technology industries are intensely entwined, an accelerated expansion is seen in the area of Internet of Things (IoT) and biomedical big data. Hence, a number of technologies like health devices and mobile applications are being integrated with telehealth and telemedicine via the biomedical IoT which constantly monitors auto-administer therapy-based devices, health indicators, or devices which keep real-time track of patient's data of a self-administered therapy. Nowadays, due to increased Internet and smartphone access, patients have started using wearable biosensors, mobile apps for personalized mHealth and eHealth technologies managing their daily health needs. This paper reviews healthcare big data analytics and biomedical IoT and analyzes growing concerns in IoT technology pertaining to smarter ways of healthcare applications underlining the big data privacy and security challenges.

32.1 Introduction

In past few years, the association between technology and healthcare has seen a big rise all over the world. Since then, big data analytics and Internet of Things (IoT) have progressively gained attention for eHealth and mHealth next-generation services. Data amount and its generation speed in different fields have taken a big leap in recent years. Science and Nature have published special issues to uncover big data opportunities and to overcome its challenges. Nowadays, healthcare data producing sources are improved quickly, like wearable medical devices, high-throughput instruments, sensor systems, which produce large amounts of data. Big data has been playing a crucial role in a variety of sectors, viz. scientific researches, health care,

---

P. S. Gangwar · Y. Hasija (✉)
Department of Biotechnology, Delhi Technological University, Delhi 110042, India
e-mail: yashahasija06@gmail.com

P. S. Gangwar
e-mail: pawansgangwar@gmail.com

A. N. R. Reddy et al. (eds.), *Intelligent Manufacturing and Energy Sustainability*,
Smart Innovation, Systems and Technologies 213,
https://doi.org/10.1007/978-981-33-4443-3_32

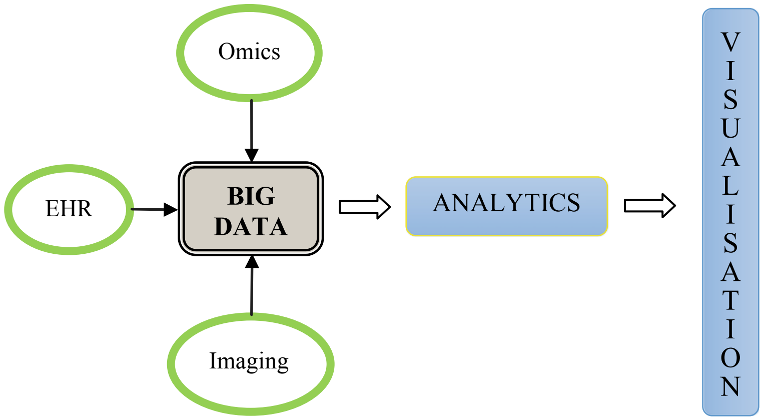

Fig. 32.1 Big data flow from its sources to storage, analytics, and visualization

industry, social networking, natural resource management, and public administration. Big data consists of large unstructured data which becomes tricky for analyses using traditional data processing techniques and platforms and requires advanced real-time analysis [1]. Characteristics of big data are defined by 6 V's of volume (lots of data), variety (data obtained from different sources exists in different forms), velocity (data is accumulated speedily), variability (consistency of data over time), veracity (uncertainty of the data), and value (data relevance). It is more interesting to see big data as it relates to sources, repositories, and its analysis (Fig. 32.1).

Internet of Things (IoT) is a system of devices and different equipments, integrated with sensor, software, electronics, and network connection, enabling such devices to gather and perform data exchange [2]. Its effect on medication would be perhaps more personalized and most significant. The amalgamation of healthcare and information technology, like biomedical informatics, would surely transform healthcare in several ways as reducing costs, minimizing inefficiencies and saving lives.

32.2 Biomedical Big Data

Data types in biomedical sciences are majorly clinical and scientific data. Patient data and data pertaining to patient care like health surveys and epidemiological data are clinical data, whereas scientific data consists of bench science data. The data sources are broadly categorized into primary and secondary data source. In primary data analysis, an individual or group of researchers design, collect, and analyze the data. Primary data comes with the advantages of guaranteed data quality with minimum number of missing values and consistency of the instrument. The secondary data source depends on the existing data or the data collected already for some other purpose. Secondary data sources have several advantages as low cost, less time-consuming data collection and large data samples. Several famous big data biomedical projects which generated large-scale sequencing data include 1,000 Genome Project, The Cancer Genome Atlas (TCGA), etc. The types of healthcare data sources on which big data analytics is applied are Electronic Health Records (EHRs), Omics, medical imaging, clinical texts, etc.

32.2.1 Big EHR Data

The principal source of big data in human health is the electronic health records (EHRs), since conversion from handwritten charts to EHRs. EHR stores each patient's information, like research facility tests and results, statistic data, analysis, medications, clinical notes, and radiological images [3]. To analyze such data, it needs to be converted from text to a much-structured form, with or without natural language processing (NLP). EHR data types are unstructured (clinical notes) or structured (clinical data, imaging data, administrative data, charts, and medication).

32.2.2 Medical Imaging Data

The image data consists of X-rays and CT scans. Picture Archival and Communication Systems (PACS) is a medical imaging data storage system which is used for image storage and retrieval [4]. Medical images are stored in database of biomedical images. These images are much complex and occupy more space. It is a two-phase system, i.e., training and testing. During the training phase, features are extracted by applying several algorithms using 80% data. In testing phase, rest 20% data is tested to check if the system gives true results of the input query image.

32.2.3 Clinical Text Mining Data

Healthcare data are structured, unstructured and might have textual records. Text/Data mining is a method by which high-quality information is extracted from unstructured data. Text mining when applied electronic medical records (EMRs) helps to find patient stratification, unknown diseases, better drug targeting, and distinct side effects of the drug. Natural language processing (NLP), a machine learning method, is generally used to extract information from clinical texts. Several data mining tools are used to analyze textual records of health domain, e.g., clinical Text Analysis and Knowledge Extraction System (cTAKES).

32.2.4 Big OMICs Data

Omics data contains catalog of molecular profiles (e.g., genomics, proteomics, transcriptomics, metabolomics, and epigenomics) which provides the base for personalized medicine. The genomes, transcriptomes, and epigenomes are in upstream as compared to the metabolome and proteome. *Genomics*. Complete set of DNA of an organism. A genome information is contained in frameshift mutation (insertion/deletion), single nucleotide polymorphism (SNP) and four copy number variations (CNVs); *Transcriptomic*. All the gene present in a cell. Transcriptomic knowledge is carried in gene expression, transcripts expression and alternative splicing; *Epigenomics*. A large number of chemical compounds which direct the genome; Proteomics. The total protein which are encoded by the genome; *Metabolomics*. A well-rounded catalog of metabolites in the cell of an organism.

32.3 Healthcare Internet of Things (IoT)

A decade prior just individuals were associated with the Internet, however, now things/devices are additionally associated with the Internet which is due to the sensors embedded in them. Hence, the core of IoT is the wireless sensor network, which senses the events and shares the data [5]. It can be understood by an example of a typical IoT hospital in practice where a diabetic patient has an ID card, which when scanned, links to a secure cloud that stores his/her EHR information, lab results, medical and prescription history. These records can be then used on smartphone, tablet, or computer by physician and nurses [6].

32.3.1 IoT Architecture

IoT produces huge amounts of data for real-time processing; however, there is a delay in data transfer between cloud and end user. Hence, to diminish this delay, a fog assisted real-time alert generation and remote monitoring system architecture was proposed [7]. It has three layers: device layer (where user information is collected by IoT devices and medical sensors), fog layer (here patient information is analyzed using classification rules), and cloud layer (informs cautioning alerts to the relatives enabling them to check the patients critical signs from anyplace anytime). Figure 32.2 shows the generalized healthcare IoT system architecture. Using machine learning and advanced inference algorithms, the healthcare IoT system itself learns from patient history and sensor data to give feedback about the present and predicted health in future of the patient, and could even generate cautions if vital.

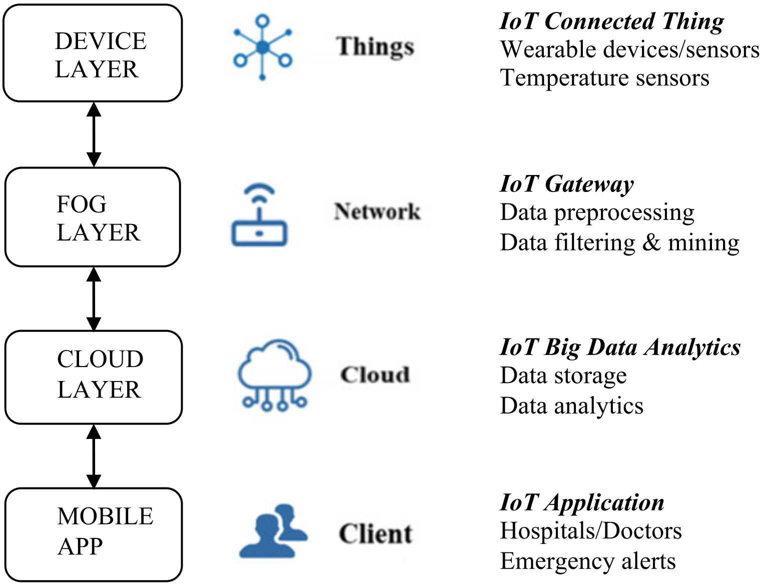

Fig. 32.2 Healthcare IoT system architecture [8]

32.3.2 IoT Data Source

Health and behavioral data of patients in IoT framework is recorded by distinct types of sensors (wearable or implanted) equipped on patient for personal monitoring on multiple parameters and to transfer real-time data to medical personnel.

Source of IoT data depends on the type of IoT application. The data is broadly branched in three classes: passive, active, and dynamic data. Passive data is from active response lacking device. Active data is collected from the IoT device which is actively reacting and sending the response. Dynamic data is collected and helps making the self-decision to do a better performance [9]. These types of data are sourced from Industrial Control Systems (Cortana Analytics and IBM Watson), Business Applications, Wearables (wearable devices are embedded with the sensors, e.g., Mi Smart Band 4), Agriculture Assistant (agri analytics), Open and Web Data (publicly available social network data like Facebook, Twitter to deduce the current trends among a group of people), GPS data (Global Positioning System).

32.4 Studies Related to Big Data Analytics in Healthcare IoT

IoT and big data in healthcare has been a very hot topic recently. Some of the recent trends and developments in light of big data and IoT eHealth are mentioned.

Rahmani et al. [10] proposed a smart eHealth Gateway in a fog computing-assisted system architecture. The gateway offers many features like local storage, real-time data processing, and integrated data mining. Additionally, the described architecture is powerful enough to deal several rising issues in omnipresent healthcare frameworks such as scalability, mobility, reliability, and energy efficiency. At last, using a prototype, authors have described some of the high-level qualities of the gateway like IoT-based Early Warning Score (EWS) patient health monitoring.

Woo et al. [11] focused on very crucial matter of fault-tolerant healthcare data services. Therefore, authors presented a fault-tolerant algorithm reliable for the IoT network. Gateways are linked forming a daisy chain for fault tolerance in this architecture. Additionally, one gateway stores the backup copy of the previous gateway located immediately before of the gateway in the daisy chain. Using this approach, two gateway faults can be recovered which occurred simultaneously.

Ammae et al. [12] described body movements sensing method by measuring changes in signal strength of Wi-Fi among two Wi-Fi enabled devices which allows an unnoticeable method to measure quality sleep. Based on maximum likelihood linear regression (MLLR), they adapted a person's model of body movement detection by using training data of other users. This allowed to adapt to a movement detection model which is user independent. Their method was assessed by using real data of 60 sessions collected from six participants and the model achieved high detection accuracy.

Farahani et al. [13] provided eHealth and mHealth IoT architecture systematic review. The ongoing present healthcare challenges all over the world are thoroughly described by the authors. The review suggests to move health services from hospital centric model to individual centric by IoT support. A multi-layer, holistic IoT environment is put forward which is driven by three layers of health devices, fog gateway/computing and cloud computing. It points IoT environment present challenges and suggests possible solutions. Smart textiles and smart eyeglasses case studies are also described demonstrating the competence of the proposed eHealth IoT ecosystem.

Several studies have been performed on biomedical IoT in health informatics. Table 32.1 illustrates some studies with description, methodology, or algorithm of

Table 32.1 Reviewed IoT healthcare system

Author	Description	Methodology
Kalid et al. [14]	Prioritization of telemedicine patient remote health monitoring using big data analytics	Remote monitoring real-time system
Prajapati et al. [15]	Proposed for continuous ICU patients monitoring	Intelligent real-time IoT-based system
Rani et al. [16]	Early diagnosis and preventive measures to control chikungunya virus	Fuzzy K-nearest neighbor algorithm
Sood et al. [17]	Identify and control chikungunya virus by wearable IoT device	Fuzzy C-means algorithm

study.

Health Devices and Mobile Apps In the current era of information, enriched with data and knowledge, apps and devices are seen as a "health buddy." *Myo*, a game controller, is presently being used in the orthopedic patients advised for post-fracture exercise. Patients monitor self-progress and doctors could measure angle of movement with Myo's help. *Zio Patch* is for measuring heart/pulse rate and it is approved by FDA (US Food and Drug Administration) [18]. DarioHealth delivers evidence-based personalized digital therapeutic interventions for diabetes [19]. SleepBot is a mobile and web application, which has a "smart alarm" feature and tracks sleep [20]. RANKED Health is a project which critically evaluates and ranks health apps and connected devices is run by the Hacking Medicine Institute [21].

32.5 Challenges for Medical IoT and Big Data in Healthcare

IoT leading platforms have to provide powerful, yet simple application access for IoT data and devices for fast development of mIoT apps and analytics application. Leading IoT platforms must enable: (1) Easy device management: Enables improved resource accessibility, expanded throughput and decreased maintenance costs. (2) Simple connectivity: It is simple for devices to connect and carryout management functions in a better IoT platform. (3) Information ingestion: Store and transform IoT data intelligently. APIs connect the gap between cloud and data, making it simple to draw the data which is required. (4) Informative analytics: Gain IoT platform from large bulk of IoT data for making better decisions and real-time analytics application to monitor present status and react accordingly. Leverage cognitive analytics to realize situations and learn as conditions change. (5) Reduced risk: Act on notifications and isolate incidents created anywhere from a single console.

Security and privacy of the patient/individual are the two major issues of big data in healthcare [22]. Medical data is very sensitive and several countries consider it as legally possessed by the patient. Big data analytics software solutions should use advanced encryption algorithms and pseudo-anonymization of personal data to address the security and privacy challenges. These software solutions must give security on system level and authenticate for every associated user, ensure privacy and security, and also set up better governance norms and practices.

32.6 Conclusion

Technology and health have consistently been associated, and this relation has fundamentally transformed because of fast development of IoT and the dominance of wearable devices lately. Smart IoT sensor devices, fog and cloud computing and smartphones collectively function as an efficient and flexible architecture for IoT in health care. Wearable health devices like fitness bands, smartwatch, smartphone, smart shirts, and headband detect user's body temperature, heart/pulse rate, blood pressure, and different activities leading to personalized healthcare by increased healthcare IoT access. Today, the best quality of IoTs in health industry is remote health monitoring system and real-time location feature, in which patients can be guided and monitored in real-time from anyplace. As medical IoT is depended on patient's data for quality health monitoring, healthcare IoT devices require efficient security algorithms to provide secure communication between server and devices, so that individual privacy is maintained along with data security to prevent misuse of data by an unauthorized entity. This review addresses major aspects of eHealth and mHealth IoT technologies which include wearable smart sensors, pervasive advanced health systems, and big healthcare data analytics furnishing better eHealth services leading to healthy society.

References

1. Y. Wang, L.A. Kung, T.A. Byrd, Big data analytics: understanding its capabilities and potential benefits for healthcare organizations. Technol. Forecast. Soc. Change (2018). https://doi.org/10.1016/j.techfore.2015.12.019
2. A. Zanella, N. Bui, A. Castellani, L. Vangelista, M. Zorzi, Internet of things for smart cities. IEEE Internet Things J. (2014). https://doi.org/10.1109/JIOT.2014.2306328
3. P.S. Gangwar, Y. Hasija, *Deep Learning for Analysis of Electronic Health Records (EHR)* (2020). https://doi.org/10.1007/978-3-030-33966-1_8
4. J. Sun, C.K. Reddy, Big data analytics for healthcare (2013). https://doi.org/10.1145/2487575.2506178
5. N. Nalini, P. Suvithavani, A study on data analytics: internet of things & health-care. Int. J. Comput. Sci. Mob. Comput. **6**, 20–27 (2017)
6. D.V. Dimitrov, Medical internet of things and big data in healthcare. Healthc. Inform. Res. (2016). https://doi.org/10.4258/hir.2016.22.3.156
7. P. Verma, S.K. Sood, Fog assisted-IoT enabled patient health monitoring in smart homes. IEEE Internet Things J. (2018). https://doi.org/10.1109/JIOT.2018.2803201
8. V. Jagadeeswari, V. Subramaniyaswamy, R. Logesh et al., A study on medical Internet of Things and Big Data in personalized healthcare system. Health Inf. Sci. Syst. **6**, 14 (2018). https://doi.org/10.1007/s13755-018-0049-x
9. S. Patil, A.R. Kokate, D.D. Kadam, Precision agriculture: a survey. Int. J. Sci. Res. (2013). https://doi.org/10.4172/2157
10. A.M. Rahmani et al., Exploiting smart e-Health gateways at the edge of healthcare Internet-of-Things: a fog computing approach. Futur. Gener. Comput. Syst. (2018). https://doi.org/10.1016/j.future.2017.02.014
11. M.W. Woo, J.W. Lee, K.H. Park, A reliable IoT system for personal healthcare devices. Futur. Gener. Comput. Syst. (2018). https://doi.org/10.1016/j.future.2017.04.004
12. O. Ammae, J. Korpela, T. Maekawa, Unobtrusive detection of body movements during sleep using Wi-Fi received signal strength with model adaptation technique. Futur. Gener. Comput. Syst. (2018). https://doi.org/10.1016/j.future.2017.03.022
13. B. Farahani et al., Towards fog-driven IoT eHealth: promises and challenges of IoT in medicine and healthcare. Futur. Gener. Comput. Syst. (2018). https://doi.org/10.1016/j.future.2017.04.036

14. N. Kalid et al., Based real time remote health monitoring systems: a review on patients prioritization and related 'big data' using body sensors information and communication technology. J. Med. Syst. (2018). https://doi.org/10.1007/s10916-017-0883-4
15. B. Prajapati, S. Parikh, J. Patel, An intelligent real time IoT based system (IRTBS) for monitoring ICU patient, in *Smart Innovation, Systems and Technologies* (2018). https://doi.org/10.1007/978-3-319-63645-0_44
16. S. Rani, S.H. Ahmed, S.C. Shah, Smart health: a novel paradigm to control the chickungunya virus. IEEE Internet Things J. (2019). https://doi.org/10.1109/JIOT.2018.2802898
17. S.K. Sood, I. Mahajan, Wearable IoT sensor based healthcare system for identifying and controlling chikungunya virus. Comput. Ind. (2017). https://doi.org/10.1016/j.compind.2017.05.006
18. C.E. Tung, D. Su, M.P. Turakhia, M.G. Lansberg, Diagnostic yield of extended cardiac patch monitoring in patients with stroke or TIA. Front. Neurol. (2015). https://doi.org/10.3389/fneur.2014.00266
19. Dario Health, *Blood Sugar Meter & Diabetes Tracker by Dario (Version 4.5.0.0.18) [Mobile app]* (2020). Retrieved from https://play.google.com
20. Insomaniapps, *SleepBot. [Sleep Tracking Device]* (2010). Retrieved from: https://mysleepbot.com
21. RANKED Health, Retrieved from: http://www.rankedhealth.com (2016)
22. J. Archenaa, E.A.M. Anita, A survey of big data analytics in healthcare and government, in *Procedia Computer Science* (2015). https://doi.org/10.1016/j.procs.2015.04.021