Bulletin of Environment, Pharmacology and Life Sciences
Bull. Env. Pharmacol. Life Sci., Special Issue [1]2022 : 1111-1119

Online ISSN 2277-1808
Journal's URL:http://www.bepls.com
CODEN: BEPLAD

ORIGINAL ARTICLE **OPEN ACCESS**

The future prospects and challenges of implementing big data in healthcare management using Structural equation model analysis

Balaji Ramkumar Rajagopal[1], E.Kannapiran[2], Ayan Das Gupta[3,*], Mohammadali Momin[4], DSK Chakravarthy[5]

[1]Data Architect, Cognizant Technology solutions, Chennai, Tamilnadu, India
[2]Assistant Professor, Department of Electrical and Electronics Engineering, Dr.N.G.P. Institute of Technology, Coimbatore, Tamilnadu, India
[3]WBES, Assistant Professor, Postgraduate Department of Geography, Chandernagore Government College, Hooghly, West Bengal, Chandernagore Government College affiliated to the University of Burdwan.
[4]Research Scholar, Department of Commerce, Veer Narmad South Gujarat University Surat
[5]Agile Coach, Virtusa Consulting Services Private Limited
*Email : dasguptaayan11111@gmail.com

ABSTRACT

Big data is a massive and complex dataset that is impossible for conventional computer systems to manage effectively and extract meaningful output. Healthcare systems have observed several benefits after using big data. For example, big data helped to save billions of dollars by effective financial planning; helped in tracking medical records; providing remote care to the patients and many more. Besides the advantages, certain drawbacks and future challenges have been found as well. These include privacy and security concerns, data reliability issues, inaccuracy, costs, scalability, technical limitations and lack of skilled personnel. These concerns ultimately do not allow every healthcare sector to implement this technology. To address these issues, several studies are available which can bring potential solutions for the future. This study has conducted a primary quantitative survey on 56 healthcare professionals to understand the practical challenges, solutions and possible future prospects of big data analytics. The 56 healthcare professionals work in their respective healthcare sectors and use big data analytics for administrative, medical care and remote care purposes. An online survey has been conducted on them via google survey form and they responded to the major challenges and possible solutions. Findings showed that privacy and security is the major issue in big data analytics which can be addressed by installing advanced software and establishing government policies. Scalability, inaccuracy and reliability issues can be addressed by data governance and information sharing.

Keywords: *Healthcare, Google survey, Internet of Things, Data privacy and protection, big data*

Received 13.02.2022 Revised 22.03.2022 Accepted 21.04.2022

INTRODUCTION

Big data defines a huge and complex dataset that cannot be controlled by the conventional data management system to obtain meaningful outcomes. Although big data can be accessed and processed by traditional management systems; however, it cannot process that amount of data promptly [1]. The data often requires petabytes of storage system which is not possible to be installed in traditional windows or apple computers [2]. Implementation of big data analytics is still a matter of concern for many healthcare sectors; however, the installation and analysis has shown potential benefits.

It has been reported that Web 2.0 company has become benefitted by integrating big data. Moreover, they invented several big data features. Those innovations provide today's industries access to open-source information and simultaneously extract meaningful information for business development [3]. Financial reports of industries show that big data provides a potential improvement in the growth of financial values in healthcare and other sectors [4]. Below fig. 1 suggests the estimated big data revenue by the year 2025. Currently, big data is providing revenue of more than 15 billion USD. The estimated total storage system required for storing big data is 181 zettabytes. To simplify, 181 zettabytes refer to 181 trillion gigabytes.

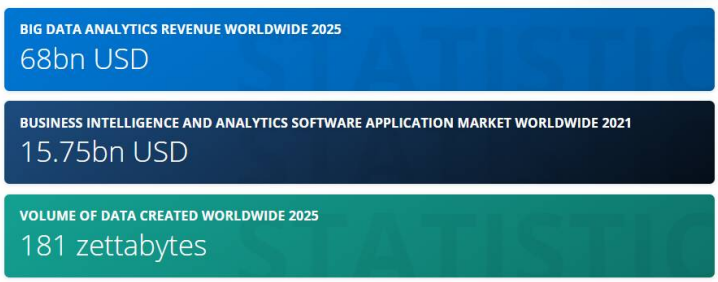

Figure 1: Big data revenue in 2021 and estimated revenue in 2025 with the volume of data (Source: [5])

In the healthcare system, big data is used for providing help and promising care to the patients; and to monitor the patients through several models. Nowadays, Smart models utilise big data to provide patient care remotely; to elder patients and patients with chronic illnesses [6]. These models are connected via ***Wireless Networks***, "***Radio-Frequency Identification***" or **RFID** and "***Remote patient Monitoring***" or RPM systems. These systems collect patient data with the help of different sensors. The patient data is then stored in the cloud environment for further processing and analysis. After analysing the data in the cloud and in the open-source, effective treatment and care strategies are extracted for patients.
Apparently, several future prospects and estimated revenues have been identified from the available studies. This study will understand the aspects of clinicians regarding the big data in the healthcare system via an online survey. Initially, past literature will be examined; followed by research methodology, analysis, discussion, and finally, the study will be concluded with the major findings.

LITERATURE REVIEW

Big data is being used by the healthcare industries to access the huge amount of data available online. This process is accomplished by the *"**Big Data revolution in Healthcare**"* where available research data, electronic records, pharmaceutical data, medical research data, medicine data, trial data and other available information are accessed to extract meaningful information for the healthcare system [7]. Reports suggest that the advancement of software and hardware allowed the healthcare sectors to readily access big data. Still, "***Data protection and privacy***" is a major concern in the field of big data analytics which needs to be addressed for its effective implementation [8]. According to Castaneda and colleagues, Big data analytics can save more than 390 billion USD in the healthcare systems [9]. Many healthcare sectors have already implemented these analytics systems and obtained potential benefits. For example, "Kaiser Permanente", a healthcare system saved around 1 billion USD by using big data analytics. Other healthcare systems, such as ***NantHealth*** and ***Blue Shield*** saved billions of monies by using big data.
Concerning today's challenge to access and analyse the huge amount of big data, Shilo and co-workers explained on the contrary that, still the computer systems of healthcare are not well-developed to process the huge amount of big data. As the data are highly complex and too large, the traditional computer systems cannot access it efficiently to extract valuable information [10].

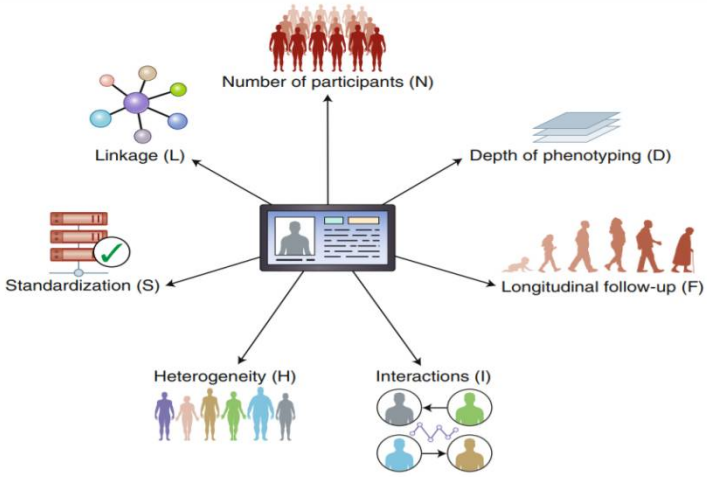

Figure 2: Axes of healthcare big data consists of interactive, standardisation, phenotyping, longitudinal and numerical data[10]

The authors also stated that medical data comprises some unique features and categories which is not applicable for other industry's big data. The data include scan imaging data, administrative information, registry information, electronic medical data, biobank data and many more. Thus, the data is too diverse and moreover, unstructured data is available which is more complex to process. Unstructured data are compared with each other which include the comparison of biobank research data with the patient-care data [11]. Therefore, the healthcare data are highly complex and consist of variant diversities and taxonomies.

Shilo and co-workers defined these medical big data as *axes* which have been represented in figure 2. Those data are obtained from either mega-biobank, electronic records or cohorts.

According to Agrawal and Prabakaran, big data is generated through research data, biobank data and patient medical record data. The patient medical records are mostly generated by the healthcare sector. **Biobanks** collect the specimens and bio-materials for research purposes and subsequently, they upload the data on the internet. The data of the biobank is too diverse and contains different characteristics which are suitable for research purposes [12]. Those data are available for big data analytics and it has been reported that those data can address several medical problems. This is a definition of biobanks; however, a mega biobank contains a huge volume of data consisting of the axes of data (figure 2).

Big data in the healthcare system has shown several benefits; however, concerning challenges are also observed. Pastorino and co-researchers stated that big data in healthcare has legal and ethical concerns. These challenges include the security and privacy of medical data; autonomy of a person; transparency; fairness and trust [13]. The authors explained that heterogeneity of data, data protection, infrastructure lacking, and inefficient data storage capabilities in healthcare sectors will bring an endangerment in big data analytics. Other ethical issues include the commercialisation of patient and medical data, privacy and security issues of medical data without understanding the people's interest. Thus, *the "**General Data Protection Regulation**"* or GDPR is ensuring sharing of data inside the healthcare sector and the data can be used for research purposes. Concerning this, the government is also required to establish policies to protect the confidentiality of patient data whenever big data analytics will be implemented.

Mehta and Pandit identified other challenges of big data analytics apart from privacy issues. They explained that big data contains unstructured, structured and semi-structured data which require technical advancements. Thus, it becomes a challenging factor for most of the healthcare sectors without advanced computers. Moreover, as big data is huge and too complex, skilled personnel and advanced computers are required for validation. Other activities include observation, understanding data reliability, identifying missing data, inaccuracy, scalability and standardisation [14]. Apart from this, ***string network bandwidth*** is essential to manage real-time data.

MATERIAL AND METHODS

A mixed research approach has been selected for accomplishing the major findings for this research. Initially, primary research has been carried out with **56 individuals** to collect the data from the aspect of healthcare professionals. On the social media platforms, the researchers posted about the requirement of survey responses and asked the social media users to drop comments on who will be interested to participate in the survey. Around 154 comments were obtained that included individuals from different fields. From that, the researchers selected 56 respondents who practically used the Internet of Things (IoT) and big data analytics in their respective healthcare sectors.

Those individuals were asked to provide their email addresses, and then, an online Google survey form link has been sent via those emails. The google form contains both open and close-ended questions which will help the researchers to identify the prevalence of a particular challenge and the possible future challenges. The **56 respondents** provided their answers in both open and close-ended scenarios. Those response data were then collected in Microsoft Excel for further analysis.

Some of the data were obtained in textual form; which was converted to numerical form for effective graphical representation and percentile calculation. The researchers used Microsoft Excel for both percentile calculation and diagrammatic representation. After that, the data were analysed and interpreted in the following section. ***The purposive sampling*** **technique** has been chosen for this analysis where the researchers have drawn their own judgments from the data of primary research. As most of the survey questions were close-ended and open-ended questions did not provide an effective explanation, the researchers decided to draw judgements based on their own understanding. Apparently, other available journal articles have been examined for effective validation.

To carry out further analysis and discussion, the researchers considered a ***deductive research approach***by collecting online journal articles belonging to the last five years (2018-2022). These articles will provide recent research findings that will either validate the primary research data or will address

some challenges. It was previously mentioned, the researchers obtained close-ended responses (mostly), and purposive sampling and deductive approaches are the only options. Thus, the available journal articles will be matched with the primary research data and the major and minor challenges will be observed.

After the entire analysis and discussion, a *"**Structural Equation Model**"* will be represented where the significance (p) values of those challenges will be shown. Those p values will help future researchers to address the issues based on their prevalence and potential.

Below, the research questions and flowchart have been provided.

Research Questions

What are the future prospects of big data analytics in the healthcare system?

What are the challenges of big data analytics in the healthcare system?

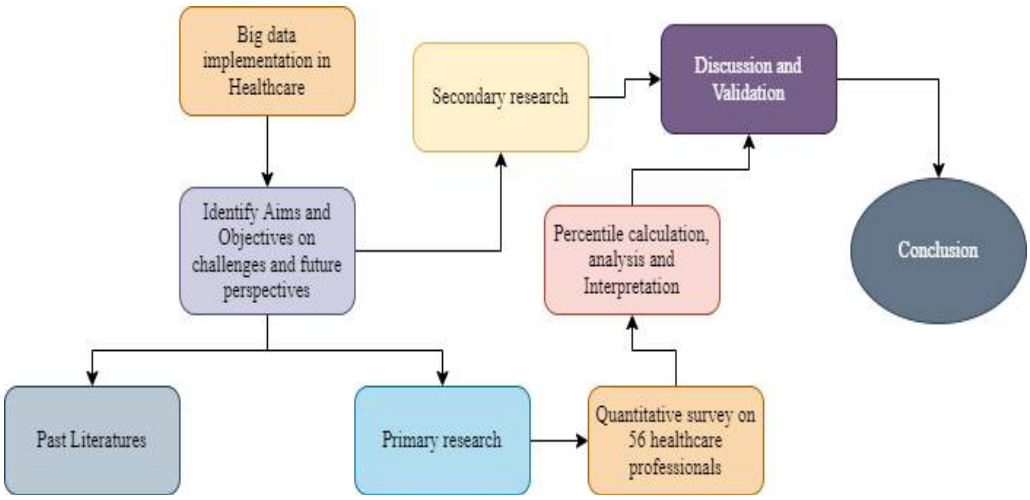

Figure 3: Research Flowchart

RESULTS

A survey has been carried out with 56 healthcare professionals who used big data analytics using IoT systems. The survey questions, responses, percentile calculations and graphical representations are shown below.

Q1. Where is big data mostly used in your healthcare sector?

TABLE I. SURVEY QUESTION 1 RESPONSE AND PERCENTILE VALUES

Options provided in survey form	Total Participants	Collected responses	Percentile
Health record tracking and analysing	56	23	41%
Providing remote care	56	19	34%
Administrative and finance department	56	8	14%
Predicting health conditions and diseases	56	6	11%

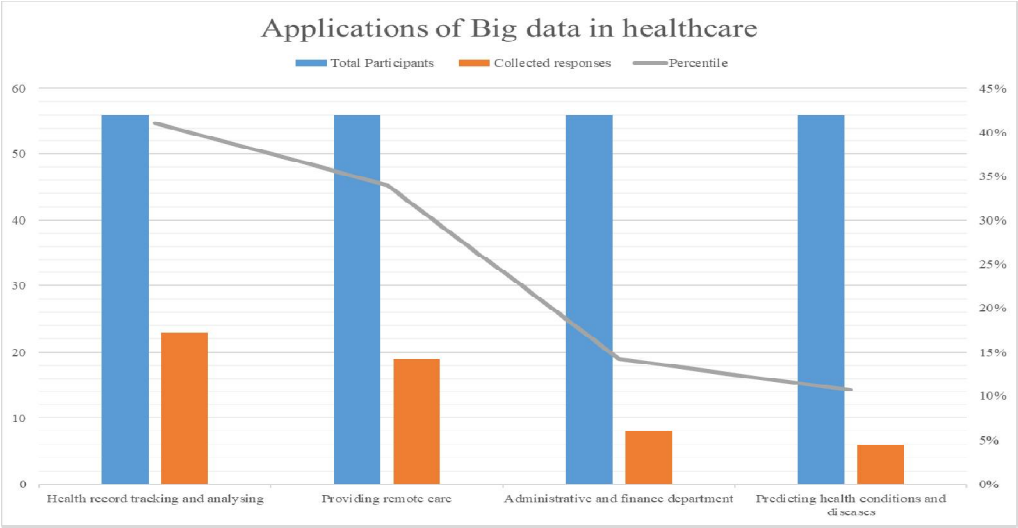

Figure 4: Graphical representation of survey question 1 response

Survey question 1 has been provided for understanding the field where big data analytics are used mostly (Table I, Figure 4). The responses found that big data is used in tracking health records mostly (41%). 34% of the respondents stated that big data is used in providing patient care remotely. 14% of respondents stated that big data is used in administrative work for financial reporting and managing financial balance. Lastly, 11% of them stated that big data is used for predicting health conditions and diseases. Big data, therefore, is mostly used in health record tracking and remote care.

Q2. What do you think is the major challenge of big data analytics for implementation in healthcare sectors?

TABLE II. RESPONSES OF SURVEY QUESTION 2

Options provided in survey form	Total Participants	Collected responses	Percentile
Lack of Data protection policy by the government	56	12	21%
Limited connectivity	56	9	16%
Halting	56	6	11%
Bandwidth	56	1	2%
Technical limitations and challenges	56	13	23%
Lack of skilled IT infrastructure	56	15	27%

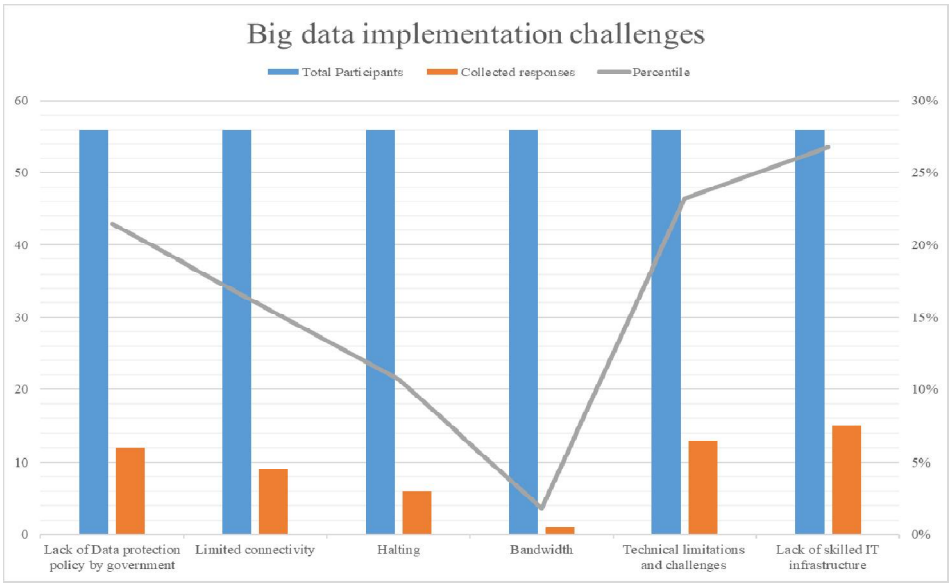

Figure 5: Graphical representation of third survey question responses

The second survey question (Table II, figure 5) has been asked to understand the big data analytics implementation challenges. More specifically, the challenges that do not allow the healthcare sectors to implement this technology. The responses showed that lack of technicians and skilled IT infrastructure does not allow the healthcare sectors to implement the technology (27%). Other challenges include the technical limitations (23%) which suggest that the conventional computer systems cannot access the huge big data. 21% of the respondents agreed that big data brings questions on data protection and policies. Few of them stated other challenges; such as the limited connectivity (16%); halting (11%); and bandwidth (2%). In this case, halting refers to the unstable internet connection and an interruption will result in technical glitches.

Q3. What do you think is the major challenge of big data analytics after implementation?

TABLE III. RESPONSES OF SURVEY QUESTION 3

Options provided in survey form	Total Participants	Collected responses	Percentile
Scalability	56	3	5%
Data privacy and security	56	13	23%
Inaccuracy of big data	56	6	11%
Cost	56	11	20%
Reliability	56	6	11%
Data fragmentation	56	7	13%
Lack of knowledge	56	10	18%

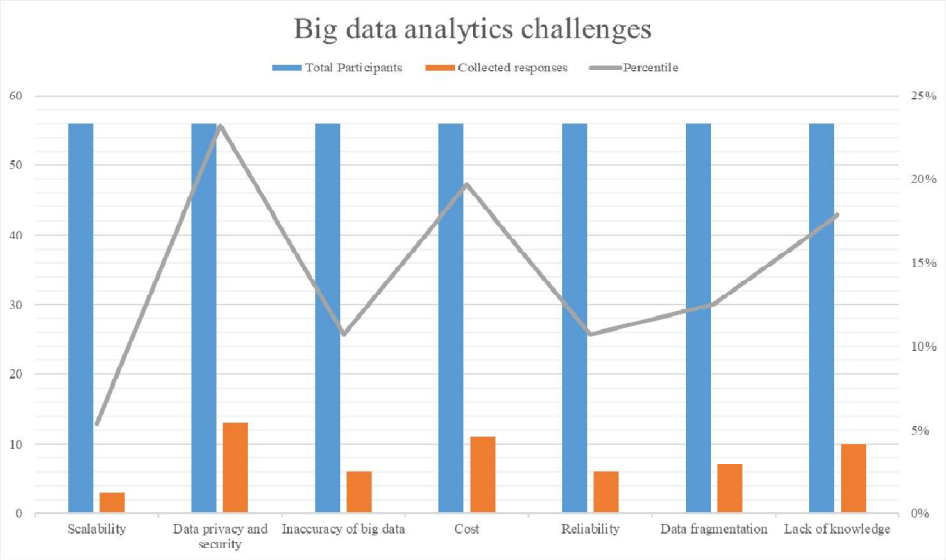

Figure 6: Graphical representation of third survey question responses

The third question (Table III, Figure 6) is based on the challenges that have been observed after the implementation of the big data analytics system. 23% of the respondents agreed that data privacy and security is the major issue they have faced. 18% of the respondents stated that they lack skilled technicians and proper knowledge on this topic which result in negative effects. 20% of respondents stated that cost is a major concern in big data analytics. 22% responded that reliability and inaccuracy are two challenges. Apart from this, data fragmentation and scalability are two other problems (13% and 5% respectively).

Q4. How do you think the big data challenges can be overcome?

TABLE IV. RESPONSES OF SURVEY QUESTION 4

Options provided in survey form	**Total Participants**	**Collected responses**	**Percentile**
By implementing information sharing system	56	9	16%
By strengthening security and privacy	56	16	29%
By data governance	56	10	18%
Training on cloud computing and big data analytics	56	11	20%
Funds from the government	56	10	18%

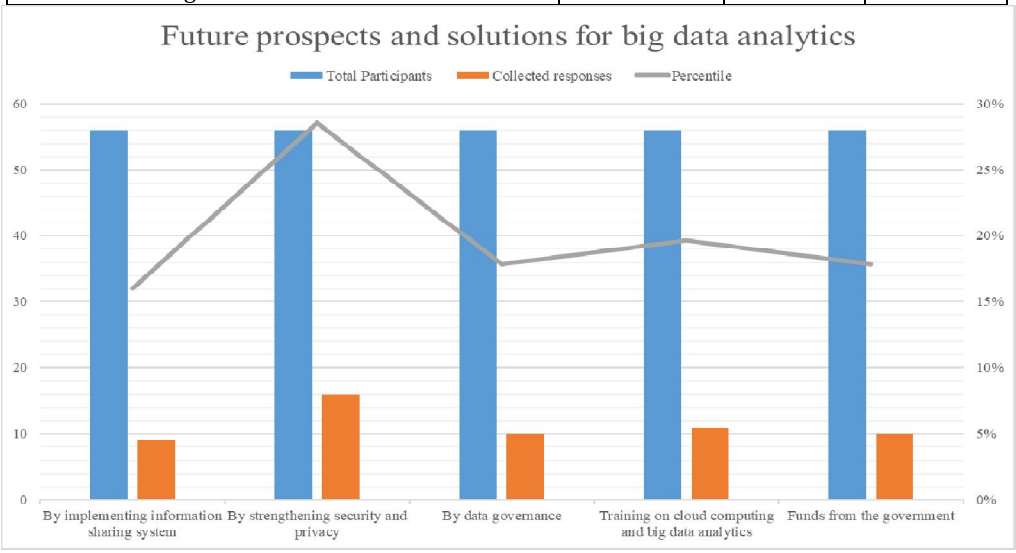

Figure 7: Graphical representation of fourth survey question responses

The final survey question (Table IV, figure 7) is related to the possible solution and future prospects of big data analytics. As the major challenges have been found, possible solutions are required as well to address the problems and develop this technology for receiving advantages in future. 29% of respondents

stated that strengthening the security and privacy features will allow the healthcare sectors to implement this technology. Patient data and other sensitive data are shared online which are prone to hijacking by hackers. Therefore, strengthening the security system is the first priority. 20% responded that training is required before and after the implementation of big data analytics. 18% stated data governance can help to mitigate the inaccuracy, scalability and reliability issues. 16% of respondents stated that information sharing systems can help mitigate those issues as well. Lastly, 18% of respondents agreed that government funds are required, as big data analytics require advanced technology and large storage systems.

DISCUSSION

The current analysis upheld the relevant responses identified from the survey and excluded the irrelevant responses. Demographic data and data related to the experiences of respondents have been excluded. Although this data could be useful for in-depth analysis; however, in this research, only major challenges, issues and future prospects have been analysed.

The first survey responses stated that, mostly, big data is being used in patient health record tracking and remote care. A study by Kalid and colleagues also found that big data analytics are used in telemedicine, remote patient care and chronic disease identification. Information related to administrations has not been found [15].

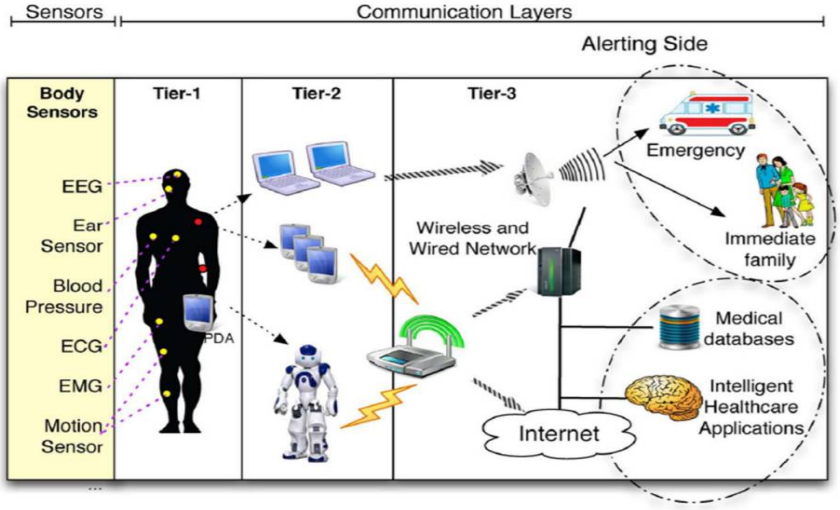

Figure 8: Health monitoring architecture[15]

The current research has also found that big data is being analysed for remote patient care (figure 8) and record tracking. Another aspect is predicting health conditions. Big data contains axes of data that are structured or unstructured. Therefore, with the help of IoT devices, the patient data are analysed along with the available big research data. This provides effective identification of health conditions when analysed through big data. Moreover, big data contains possible care strategies and medication recommendations which are then provided to the respective patients (remotely) [16]. The survey responses state that data protection, privacy and security are the major concerns. Other available studies also describe that data privacy and security is major concern. Ristevski and Chen explained that medical data are sensitive and need to be accessed and analysed by considering legal and ethical considerations [17]. Therefore, advanced software and government policies can address these problems.

Other issues have also been found; such as the poor bandwidth, halting, technical limitations which can be addressed by applying for government funds. The third survey question mentioned that cost and data protection are two major issues [18]. The reason behind cost is providing training to the personnel and implementation of advanced software and hardware. These issues and solutions did not mention the future applications and only provided ideas on the prospects of big data for the future [19]. Strengthening security and privacy systems, applying for government funds, establishing policies, technical advancement and effective training will help the healthcare sector to practice IoT big data analytics. Future applications include the "***Massive Parallel Processing***" and medicine recommendations which can be easily accomplished by using big data analytics [20].

As big data is huge and complex, scalability, inaccuracy, data fragmentation and reliability are the concerns. Many of the respondents agreed that these challenges are being faced after big data implementation. Concerning this, Mirza and co-researchers stated that big data requires the use of ***high-throughput omics*** which ultimately brings the issues. Therefore, multiple computers are required to

address these issues [21]. Survey responses stated that, by data governance and information sharing system, the inaccuracy, scalability and reliability issues can be solved [22]. According to Brous and colleagues, data governance is a solution for success in big data analytics. Therefore, the available findings validate the findings of the current research [23]. Other studies also stated that information sharing systems can improve big data analytics in the healthcare system [24]. Below, the "*Structural Equation Model" or SEM* shows the probable solutions for mitigating big data challenges that can bring future prospects (Figure 9).

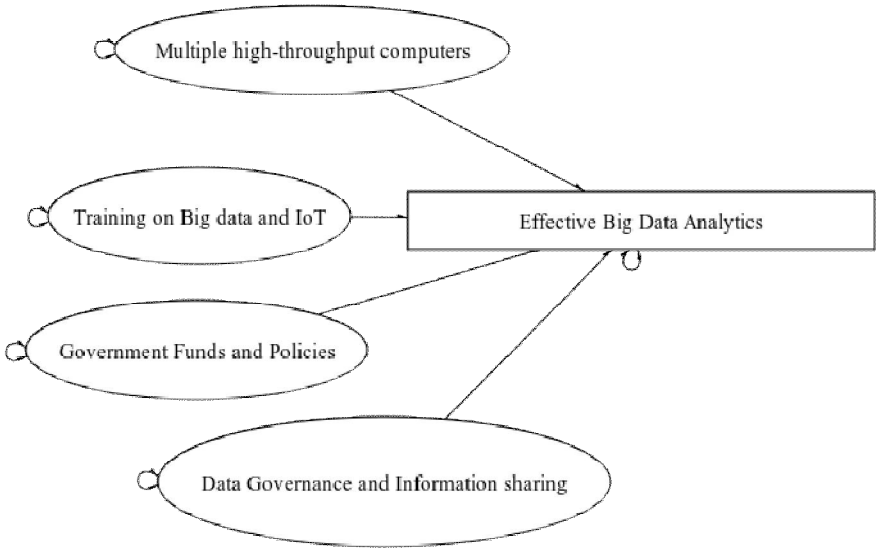

Figure 9: SEM for solutions of the big data analytics

CONCLUSION

The study has been carried out with 56 participants from different healthcare sectors who have used the big data analytics system. Mostly, close-ended and some open-ended responses have been obtained which have been discussed in the previous sections. It has been observed that, due to the issues of privacy and security, healthcare sectors are unable to rely on big data analytics. Several studies are available which say about the security and privacy issues of big data analytics which allow the healthcare sectors to rethink its implementation. However, the survey data and other available journal articles addressed the major challenges. For example, challenges in complex and high throughput data management can be managed by *data governance, multiple computer systems* and *information sharing systems*. Privacy and Security concerns can be addressed by using advanced software and establishing government policies. When these challenges will be addressed effectively, big data analytics will bring profitable future aspects to healthcare sectors.

REFERENCES

1. Hancock, J.T. and Khoshgoftaar, T.M., (2020). CatBoost for big data: an interdisciplinary review. *Journal of big data*, *7*(1), pp.1-45.
2. Qiao, S., Nicoara, A., Sun, J., Friedman, M., Patel, H. and Ekanayake, J., (2019). Hyper dimension shuffle: Efficient data repartition at petabyte scale in scope. *Proceedings of the VLDB Endowment*, *12*(10), pp.1113-1125.
3. Anshari, M., Almunawar, M.N., Lim, S.A. and Al-Mudimigh, A., (2019). Customer relationship management and big data enabled: Personalization & customization of services. *Applied Computing and Informatics*, *15*(2), pp.94-101.
4. Pejić Bach, M., Krstić, Ž., Seljan, S. and Turulja, L., (2019). Text mining for big data analysis in financial sector: A literature review. *Sustainability*, *11*(5), p.1277.
5. Statista.com. (2022). *Big data - Statistics & Facts.* Available at: https://www.statista.com/topics/1464/big-data/#topicHeader__wrapper [accessed 12 February 2022]
6. Hassan, M.K., El Desouky, A.I., Elghamrawy, S.M. and Sarhan, A.M., (2019). Big data challenges and opportunities in healthcare informatics and smart hospitals. In *Security in smart cities: Models, applications, and challenges* (pp. 3-26). Springer, Cham.
7. Dash, S., Shakyawar, S.K., Sharma, M. and Kaushik, S., (2019). Big data in healthcare: management, analysis and future prospects. *Journal of Big Data*, *6*(1), pp.1-25.
8. Gruschka, N., Mavroeidis, V., Vishi, K. and Jensen, M., (2018), December. Privacy issues and data protection in big data: a case study analysis under GDPR. In *2018 IEEE International Conference on Big Data (Big Data)* (pp. 5027-5033). IEEE.

9. Castaneda, G., Morris, P. and Khoshgoftaar, T.M., (2019), April. Maxout neural network for big data medical fraud detection. In *2019 IEEE Fifth International Conference on Big Data Computing Service and Applications (BigDataService)* (pp. 357-362). IEEE.
10. International.kaiserpermanente.org. (2022). *Kaiser Permanente: The Electronic Health Record Journey.* Available at: https://international.kaiserpermanente.org/wp-content/uploads/2019/07/ EHR_Journey_final.pdf [accessed 12 February 2022]
11. Shilo, S., Rossman, H. and Segal, E., (2020). Axes of a revolution: challenges and promises of big data in healthcare. *Nature medicine*, *26*(1), pp.29-38.
12. Keyes, K.M. and Westreich, D., (2019). UK Biobank, big data, and the consequences of non-representativeness. *Lancet (London, England)*, *393*(10178), p.1297.
13. Agrawal, R. and Prabakaran, S., (2020). Big data in digital healthcare: lessons learnt and recommendations for general practice. *Heredity*, *124*(4), pp.525-534.
14. Pastorino, R., De Vito, C., Migliara, G., Glocker, K., Binenbaum, I., Ricciardi, W. and Boccia, S., (2019). Benefits and challenges of Big Data in healthcare: an overview of the European initiatives. *European journal of public health*, *29*(Supplement_3), pp.23-27.
15. Mehta, N. and Pandit, A., (2018). Concurrence of big data analytics and healthcare: A systematic review. *International journal of medical informatics*, *114*, pp.57-65.
16. Kalid, N., Zaidan, A.A., Zaidan, B.B., Salman, O.H., Hashim, M. and Muzammil, H.J.J.O.M.S., (2018). Based real time remote health monitoring systems: A review on patients prioritization and related" big data" using body sensors information and communication technology. *Journal of medical systems*, *42*(2), pp.1-30.
17. Li, W., Chai, Y., Khan, F., Jan, S.R.U., Verma, S., Menon, V.G. and Li, X., (2021). A comprehensive survey on machine learning-based big data analytics for IoT-enabled smart healthcare system. *Mobile Networks and Applications*, *26*(1), pp.234-252.
18. Ristevski, B. and Chen, M., (2018). Big data analytics in medicine and healthcare. *Journal of integrative bioinformatics*, *15*(3).
19. A. Jain, A. K. Yadav & Y. Shrivastava (2019), "modelling and optimization of different quality characteristics in electric discharge drilling of titanium alloy sheet" material today proceedings, 21, 1680-1684
20. A. Jain, A. k. pandey, (2019), "Modeling and optimizing of different quality characteristics in electrical discharge drilling of titanium alloy (grade-5) sheet" material today proceedings, 18, 182-191
21. A. Jain, A. k. Pandey, (2019), "Multiple quality optimizations in electrical discharge drilling of mild steel sheet" material today proceedings, 8, 7252-7261
22. Mirza, B., Wang, W., Wang, J., Choi, H., Chung, N.C. and Ping, P., (2019). Machine learning and integrative analysis of biomedical big data. *Genes*, *10*(2), p.87.
23. Brous, P., Janssen, M. and Krans, R., (2020), April. Data governance as success factor for data science. In *Conference on e-Business, e-Services and e-Society* (pp. 431-442). Springer, Cham.
24. V. Panwar, D. K. Sharma, K.V.P.kumar, A. Jain & C. Thakar, (2021), "experimental investigations and optimization of surface roughness in turning of en 36 alloy steel using response surface methodology and genetic algorithm" materials today: proceedings, https://doi.org/10.1016/j.matpr.2021.03.642

CITATION OF THIS ARTICLE

B R Rajagopal, E.Kannapiran, A D Gupta, M Momin, DSK Chakravarthy. The future prospects and challenges of implementing big data in healthcare management using Structural equation model analysis. Bull. Env. Pharmacol. Life Sci., Spl Issue [1] 2022 : 1111-1119