

# The Role of Big Data Analytics in Healthcare

Nek Dil Khan, Muhammad Younas, Muhammad Taimoor Khan, Duaa, Adil Zaman



**Abstract:** Recently a rapid increase has been seen in the technology and health care became an unavoidable sector because an enormous amount of data is collected day to day. The Big data analytics has many uses and it is associated to big data. It gives an important judgment in the area of health care. This paper aims to present a review of big data in health care that how big data can aid the healthcare. In this paper multiple research paper has been reviewed and the limitations and research gaps of these papers has been discussed accordingly.

**Keywords:** Big data, Artificial intelligence, Deep neural networks, Healthcare

## I. INTRODUCTION

Healthcare is an important sector as it contains large amount of data which is gathered from day to day. This data is useful for healthcare sector for further use. Big data analytics is important to mine comprehensive data i.e., structured, unstructured or semi structured. Big data analytics has many uses and it is associated to big data. Big data analytics gives useful judgment in healthcare sector. Development of distributed computing huge information examination turn out to be a reality. Disseminated programming systems like Hadoop and Spark are accessible. Huge information investigation apparatuses to foster an astute clinical stage. The results show that the planned stage is versatile with the medical care information. The main problem in healthcare platform is that to choose the precise huge information apparatus to handle the unstructured information. Because recently the research area is transferred to huge efficient analysis of information. They also explain huge amount of data produced in medical data centers, which is a big challenge to handle even for cloud data centers [1]. They explain the main characteristics of the huge information i.e., volume (proportion of data increasing), velocity (beginning of data), variety (structured or unstructured data), value (capability of converting information in expressive

method). Facts are obtained from different ways i.e., patient clinical or health records. So, the healthcare industry looking solution in machine learning which is useful for huge information storing stage. The fortitude of ensemble learning is sorting out the unstructured data into meaningful and actionable knowledge. They explain currently some models are used i.e., Apache Mahout, Skytree, Karmasphere They proposed IMP (Intelligent medical platform) big data storage as a solution to deal with heterogeneous medical data. Their results show viability of the planned stage. Examination of the versatility of the planned structure with other system is not discussed in this paper, which is a future work.

## II. LITERATURE REVIEW

The authors in [2] describes big data in healthcare is a collection of records i.e., patient, doctor, hospital, medical treatment. It's difficult to manage heterogeneous data and obviously it's a big challenge to deal with the reliability of data and the normal data set frameworks are sufficiently not to maintain information. This article relates with the previous in such a way that, in this article, authors recommended machine learning tools to manage huge measure of information utilizing enormous information analytical context. The machine learning platforms uses Computer based intelligence to deliver advance information investigation on unstructured information and provide intuitive dashboard to perform bid information examination and perception. They presented consequences of large information in making medical services structure. And emphasized shortcomings of the current huge information investigation. In this paper, one AI approach is used to get better accuracy which is ensemble learning. EHR furthermore, enormous measure of complex information is utilized to get customized, prescient and prescient medication. EHR data may be structured, semi structured, unstructured. While the health care sector needs modified and modernized huge information investigation while the huge information investigation is hard to investigate and deal with usual calculations. They describe using ensemble learning methods i.e., KNN, Decision tree, huge information investigation performance can be increased. Evaluation of the accuracies on ensemble learning is not discussed in this paper which is a future work. In [3], the authors provides the brief overview of different healthcare Recommendation systems based on huge information investigation. Healthcare Industry is a source of huge information as well as one of the ways to manage that data Distributed file System like Hadoop and spark. Another way to manage huge size of data to develop the algorithms based on the machine learning.

Manuscript received on August 09, 2021.

Revised Manuscript received on August 14, 2021.

Manuscript published on September 30, 2021.

\* Correspondence Author

**Nek Dil Khan\***, Master Student, Software College, Northeastern University, Shenyang 110169, China. Email: [naikdil2002@gmail.com](mailto:naikdil2002@gmail.com)

**Muhammad Younas**, Master Student, Software College, Northeastern University, Shenyang 110169, China.. Email: [younas.cs@hotmail.com](mailto:younas.cs@hotmail.com)

**Muhammad Taimoor Khan**, Riphah Institute of Science and Engineering, Islamabad 44000, Pakistan. Email: [taimourkhan86@gmail.com](mailto:taimourkhan86@gmail.com)

**Duaa**, Riphah Institute of Science and Engineering, Islamabad 44000, Pakistan. Email: [duaam.shoukat@gmail.com](mailto:duaam.shoukat@gmail.com)

**Adil Zaman**, Bachelors degree, Riphah Abdul Wali Khan University, Mardan 23200, Pakistan. Email: [deviladie48@gmail.com](mailto:deviladie48@gmail.com)

© The Authors. Published by Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP). This is an [open access](https://creativecommons.org/licenses/by-nc-nd/4.0/) article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>)

Large information investigation is related with AI models to achieve better decisions rather than traditional methods. Writers inspect some important recommendation systems and explore the separating procedures like content-based sifting, cooperative based sifting and mixture-based separating to generate the recommendations and their system in light of the characterization and highlight choice. The main purpose of their system is collection of data, data processing, learning and prediction. The results discovered that Mean Absolute Error (MAE) fixes the presentation of recommender framework. On the off chance that more MAE is recorded, the precision of the expectation prototype will low and vice versa. The authors of [4] explained that with the advancement in the data of medical field, health care records also increased. As a result, to manage medical big data on the way of storing, cost efficiency, data needed with in a partial period as well as decreasing errors is becoming a big task. It can be handled with the help of modern technologies as the big data management used in different fields like banking, healthcare app etc. Traditional methods of medical are replaced with modern skill i.e., electronic clinical report generators, cell phone medical care applications. are utilized for the treatment of the sickness for compelling yield. This must be conceivable with the assistance of large information investigation in medical care. Numerous enormous information sources are open in the medical care for example enrollment information, biometric information, electronic wellbeing records, imaging, patient reports, the web information, biomarker information, clinical information, and regulatory information. Unmistakable nature of the medical care large information makes it remarkable from different kinds of information sources. The scientists face numerous difficulties in the preparing of large information. A portion of the difficulties are: a proficient acknowledgment model is needed to handle the enormous measure of medical services information to distinguish highlight and perform arrangement for the recognizable proof of infection, absence of information sharing and illicit utilization of medical care large information. In [5], The authors explain CPs are used in emergency clinics as paper-based reports and outlines or slabs that healthcare plan used to treatment of patient from the side of doctor and paramedical staff. In this paper authors are focused on reducing missing data in CPs because CP data is the primary source in healthcare data. Traditionally, missing values are the big problem in CP data. AyMEN and others present in this exploration a philosophy-based CP computerization model that model with the help of SNOMED CT (Systematized Nomenclature of Medicine Clinical Terms) and ensemble learning algorithms applied on CP based dataset. CPs automated model use for reducing missing values in datasets, provide complete measurable examinations on CP information and also used to improve the consequences of information examination calculations. The experimental results using ensemble learning algorithm shows LOS (length of stay) consequences of information examination calculations and outflanks the expectation utilizing the Electronic Health Record constructed dataset. Currently CP based data is the big challenge to use because CP based data most of the paper-based data. [6] In this research paper, Hadi et al., 2020 presents the concept of e-healthcare system utilizing the enormous information

examination and AI calculations in multi-tier 5G organization. Writers utilize three machine calculations in particular guileless Bayesian classifier, strategic relapse and choice tree as an outfit structure to examine unquestionable clinical records of stroke out patients and readings from body associated IoT sensors to expect the likelihood of an exceptional stroke. Then, at that point convert the stroke likelihood into a risk factor filling in as a need in a blended whole number straight programming advancement model. The undertaking is to allot actual asset squares to HetNet consumers though focusing on out patients (Ops) by allowing them high addition blocks as per seriousness of their clinical state. Hence, enabling the OPs to send their basic information to their medical services supplier with limited rescheduling. With that in mind, two advancement approaches are proposed, weighted aggregate rate amplification approach and a corresponding decency method. This work used to two wireless techniques with alteration traditional HetNets in 6G by contributing them with a client driven AI measurement. In light of that, huge data assessment fueled framework was proposed to have sway in uplink radio asset apportioning headway of HetNets. Big data analytics with the help of ML algorithms in healthcare system is used to detect disease at initial stage so that it can be cured properly without affecting more dangerously. [7] Dash et al., 2019 explains big data means vast amount of information. In this article, they used different framework to deal with the large and complex data. One of them is Hadoop to manage huge measure of information. Map Reduce utilizes guide and decrease to record each intelligent record in the contribution to a bunch of key esteem contests, and diminish activity consolidates every one of the qualities that common a similar key. It productively matches the calculation and handles disappointments. Hadoop Distributed File System (HDFS) is the document framework module that gives an adaptable, proficient and imitation based stockpiling of information at different hubs that structure a piece of a group. Along these lines, the dread of losing the significant information of patients, clinical records and so on can be overseen. [8] Chen & Islam, 2019 in this article, authors focus on big data in such a way that it's difficult to find useful data from huge volume of healthcare data. Big data analytics give benefits in such a way that it's easier to find disease at early stage to cure it. Because, Big data is a modern technology instead of conventional healthcare systems, it uses electronic health records, medical imaging, clinical records, clinical database to overcome the huge and unstructured data problems in specific area or to detect specific disease. In this way, it not only provides information about the disease symptoms, illness, treatments but also it also helps to cope with disease at earlier stage to make right decisions. This article studies the unique characteristics of big data, but it ignores comprehensive review of several big data analytical techniques which are available for healthcare techniques. [9] Shatnawi et al., 2020 in this paper presents the benefits and difficulties of large information investigation in medical services frameworks and applications.

Distributed computing and stream handling innovations is created late years to advance the large information examination in medical services. Some Mobile application give medical aid therapy of patient and might be crisis therapy admittance to the individual divisions. Versatile distributed computing in medical services was made for gathering and breaking down constant biomedical signs for example ECG from clients in a few spots. A customized medical care application is introduced on the cell phone and wellbeing information is synchronized into the distributed computing administration of the medical care framework for capacity and investigation. For huge scope dispersed information utilizing Hadoop-based Big Data enjoys benefits in effectiveness, unwavering quality, and versatility. There are some large difficulties of Big Data investigation in medical care frameworks to catching, putting away, sharing, looking, and dissecting information. Furthermore, information security and protection, information quality, ongoing preparing, combination of heterogeneous or unique information, and guidelines for medical services information are likewise difficulties of Big Data examination in medical care frameworks. [10] In this article, Zhang et al., 2019 describes how we can improve the electronic health record from adaptation of the big data analytics tools and can be get meaningful information using big data analytics with EHR. Over past years, EHRs has been recommended to upgrade the medical care administration productivity and adequacy, yet it doesn't imply that basically embracing the EHRs framework could prompt those advantages. Because managing and dealing of huge amount of data is very difficult task and most of the organization failed to adopted the EHR systems. TO overcome that challenges, effective use of big data analytics tools within EHR not only manage the huge amount of data but also generate the high-quality information and help the organizations to overcome the cost, time and resources. The proposed model methodology includes collection of data, storage of data and utilization of data. According to the results, proposed research model is a decent indicator of significant utilization of EHRs with regards to nursing division as the R2 represents 60.70% of the difference. this examination just gathers information from a huge emergency clinic as the exploration test. Future exploration may evaluate possible contrast among age gatherings, among working experience gatherings, and among various clinical division gatherings, with a more agent test. [11] In this paper, author describes the different features used huge information analytics in gastric medication as well as general healthcare managing, medical gastroenterology linked study that provides better decision making in clinical practices and also clinical research using electronic healthcare record. In 2012 USA government started plan of "Big data research" and the determination of exploitation huge information analytics to discover the patterns in medical research. In 2015 China started to promote the projects of huge information in diverse areas include medical. The medical services enormous information stage is upheld for complex investigation of medical services information, the helping of the stomach related framework illness early admonition and avoidance, early recognition and treated, which has significant clinical importance and esteem and is adding to work on the soundness of individuals. The reasonable model of stomach

related illness proposed approach incorporate trial information, information preprocessing, include extraction, build up anticipated model and model approval. The outcomes show the positive pace of intestinal illness was higher in the sound individuals with expanding age. More steady and exact clinical data can be accommodated clinical reference by man-made brainpower PC quantitative information investigation model for the logical exploration stage. [12] In this article, Panayides et al., 2019 describes how precision medicine improve the clinical practices in healthcare system using big data analytics and deep learning. author has discussed the cancer and Alzheimer case studies and highlights the challenges in this field with big data analytics perspective. Deep learning models and big data analytics tools are very helpful in radio genomics research to deal with new algorithms, approaches, and system progresses, through resourceful data storage, data sharing and data indexing. In this paper Author has focused on the image segmentation methods and biostatistical methods to analyze and visualize the better results in the clinical field and using big data analytics to overcome the complexity of clinical data. The proposed methodology includes data acquisition, preprocessing, image segmentation, feature extraction and predict disease using deep learning algorithms. the results show the deep learning models provide the different results to predicting the disease behalf of the image segmentation methods. in Future research Deep learning models are very important in the image analysis solutions and we can also increase the model performance using the hybrid methods that combine the information from different medical imaging resources.

### III. EXPLANATION OF THE TABLE

In this article, author used Intelligent Medical Platform (IMP) for storage of big data. This proposed methodology can manage heterogeneous data of health i.e., patient appointments, prescriptions, medication etc. The results are effective due to proposed methodology in such a way that its performance as well as scalability is good in distributed systems. The research gap in this paper is that the scalability of proposed framework is used in this article that can't be compared with other platforms to enhance performance as only a single IMP platform is used which can't give the desired results i.e., IMP can be compared with other platforms to increase the performance [1]. In this article,[2] soft voting method is used. It includes gather the data, feature selection and data cleaning and developing the model using ensemble learning method which has more accuracy than the individual classifiers. The research gap of this paper is that it can't compare the accuracies of machine learning with other classifiers as well as combining methods. As there are numerous arrangement calculations too as consolidating strategies that can measure up in this situation.

As there are numerous arrangement calculations too as consolidating strategies that can measure up in this situation. In this paper, in view of enormous information investigation recommender systems are observed.



## The Role of Big Data Analytics in Healthcare

There are various strategies for proposals i.e., content-based separating, community sifting and crossover sifting. Shared separating will be memory based. Prototypical based filtering is used to achieve recommender system results using machine learning algorithms. The results show that the higher MAE recorded the less exactness of the model. The

recommendation system has different drawbacks, this will be the future work of this paper. When number of users increases CF (Collaborative Filtering) algorithm creates scalability problem. Second problem is that when there is less data about a doctor or patient then the system creates problem.

Ref	Methodology	Research Gap	Future improvement
1	Intelligent Medical Platform	Salability of planned framework is used in this article that can't be compared with other platforms	IMP can be compared with other platforms to increase the performance.
2	Soft voting method	Can't compare the accuracies of machine learning with other algorithms	There are many classification algorithms as well as combining methods that can be compared in this scenario.
3	Content-based filtering, collaborative filtering and hybrid filtering	Recommender system has limitations	It can be improved with the help of deep learning or machine learning methods.
4	EHR (Electronic Health Records)	Adoption, cleaning and storage related issues	Possible to make better healthcare system by using big data analytics
5	Systematic literature protocol	Incomplete to specific number of years	Papers are distributed on everyday schedule constantly which is unnoticed and can't be examined in this paper.
6	Artificial intelligence	New strategies and technologies didn't be discussed	New procedures and innovations can be utilized to build execution
7	SNOMED CT all CP information, metaphysics based displaying and HL 7 correspondence	Only CP base datasets are presented	CP base dataset can be collected with the collaboration of different hospitals.
8	WSRMax and PF	Challenge that is related with 6G communication	6G communication improve with latest technology
9	Hadoop	Possible to enhance the performance using machine or deep learning	It is possible to enhance the performance using machine
10	Collection of data, storage of data and utilization of data.	Potential difference among age groups	Potential differences can be ignored
11	Experimental data, data preprocessing, feature extraction, establish predicted model and model validation	Standard terminology knowledge	Ceaseless work to further develop the AI procedures and calculation to offer better computational outcomes for investigation.
12	Includes data acquisition, preprocessing, image segmentation, feature extraction and predict disease using deep learning algorithms	Deep learning needs specific environment and it's difficult to handle that affects the accuracy	Increase the model performance using the hybrid methods that combine the information from different medical imaging resources.

There will be the issue when there are similar things with various sections. Along these lines, the recommender framework has limits. It can be improved with the help of deep learning or machine learning methods [3]. [8] In this article, EHR (Electronic Health Records) is used to maintain the patient data as every patient has data. Maintaining data through paper is an old technique. It may be lost. By using EHR a patient can save and check all data online or electronically. EHR is the useful application as the patient can maintain test report etc. and send it to doctor easily. This application can help doctor to make better decisions. The

research gap of this study is that it can be possible to make better healthcare system by using big data analytics.

As there are many challenges in future regarding healthcare data. Because medical sector is complex. The most important challenge while using big data is adoption, cleaning and storage related issues.

The proposed study utilizes deliberate writing convention, it incorporates relapse, characterization, affiliation, information mining and bunching. The SLR utilized in this paper purposes to play out a deliberate and compact examination of the medical service huge information. This exploration work is restricted to explicit number of years goes from 2015 to 2019 and some piece of 2020 is additionally included, yet the papers are distributed on regular routine persistently which is overlooked and can't be talked about in this paper [4] The proposed methodology focuses on artificial intelligence (AI) to make logic from large amount of data. There are many advantages of EHR to handle healthcare data. The pros of EHR are to improve the record of the patients. In this methodology, big amount of data can be stored. The research gap for this paper is that new methodologies and innovations didn't be examined to comprehend the idea of organized, unstructured and semi organized information. Since it is hard to deal with intricacy of information i.e., measurements and qualities [7]. [5] In this paper, CP automation framework is used which consist of SNOMED CT all CP information, cosmology based displaying and HL 7 correspondence. This system comprises of complete measurable examinations on CP information which help in improving the clinic assets and gives better medical care administrations to patients. The datasets are imbalanced which will cause unstable results like LOS problem. The research gap for this paper is that only CP base datasets are presented. One important point is to get more and large CP based dataset with the collaboration of multiple hospitals which is a challenging task. It may be in the form of charts or prescription and it is possible there will be missing values in this data. So, it is difficult to get more accuracy.

[6] In this article, WSRMax and PF are compared to check the fairness in terms of SINR. The approach WSRMax increment the OP normal SINR by up to 57%. Then again, PF approach further developed SINR by 95%. This work introduced two interdisciplinary techniques to change ordinary HetNets in 6G by giving them with a client driven AI measurement. Future work for this paper is the challenge that is related with 6G communication i.e., network security, privacy implementation, getting a cost-effective approach to fast network disposition and expansion with importance on remote and outside areas. But it is difficult to fulfill all the features. This study only highlighted the challenges and possible solutions for 6G communications. The proposed model is Hadoop [9] based and have potential help in generating actual analysis with support of machine learning in the ECG filed and its apps. The results show that proper planning and decision making in emergency situation can help in a good way. The research gap for future work is that while using big data analytics it is possible to enhance the performance using machine or deep learning to produce fundamental innovation in the field of medical. [10] The proposed model methodology includes collection of data, storage of data and utilization of data. According to the results, proposed research model is a decent indicator of significant utilization of EHRs with regards to nursing office as the R2 represents 60.70% of the change. this investigation just gathers information from an enormous emergency clinic as the exploration test. Future examination may evaluate likely distinction among age gatherings, among working experience gatherings, and among various clinical division gatherings, with a more delegate test. The applied model of stomach related illness proposed system incorporate test

information, information preprocessing, include extraction, set up anticipated model and model approval. The outcomes show the positive pace of intestinal sickness was higher in the solid individuals with expanding age. More steady and exact clinical data can be accommodated clinical reference by man-made consciousness PC quantitative information examination model for the logical exploration stage. Ceaseless attempting to build the standard phrasing information base for various illness with most recent updates from clinical aides. Nonstop work to expand the information assortment for work for stomach related medication large information for more extensive information inclusion including wellbeing, climate, climate. Persistent work to further develop the AI procedures and calculation to offer better computational outcomes for examination [11] In this paper Author [12] has focused on the image segmentation methods and biostatistical methods to analyze and visualize the better results in the clinical field and using big data analytics to overcome the complexity of clinical data. The proposed methodology includes data acquisition, preprocessing, image segmentation, feature extraction and predict disease using deep learning algorithms. the results show the deep learning models provide the different results to predicting the disease behalf of the image segmentation methods. in Future research Deep learning models are very important in the image analysis solutions and we can also increase the model performance using the hybrid methods that combine the information from different medical imaging resources.

#### IV. CRITICAL ANALYSIS AND FUTURE WORK

The articles that we examined the accompanying stages are utilized in huge information investigation viz Intelligent Medical Platform, content-based sifting, EHR (Electronic Health Records), methodical writing convention, SNOMED CT all CP information, metaphysics based demonstrating and HL 7 correspondence, WSRMax, PF and Hadoop [1], [2], [11], [12], [3]–[10]. As per our perspective after steps have the accompanying downsides which are as per the following; intelligent clinical stage can charge a big amount of time and money to construct, modify, and fix. Automated fix can happen to diminish time and people vital for fixing it however it will charge more assets and money. Volume is costly yet recovery and access may not swift relations in memory just as people. The substance based separating prototype can just make suggestions that are dependent upon present welfares of the client.

Electronic Health Records including certain burdens and failures just as likely protection and network safety concerns, the possibility to unnecessarily alarm patients, expanded misbehavior responsibility concerns, and matters of cost as far as both time and cash. Eventually, policymakers and specialists concur that once EHRS are received and genuinely utilized on an inescapable premise, they can impressively profit patients, suppliers, and society overall. Precise writing convention audits can be viewed as a solid wellspring of data, especially when they are excellent surveys that incorporate every pertinent examination and in this way limit inclination.

Dim writing, (for example, government reports and strategy archives) isn't normally included, which may predisposition points of view. The two fundamental benefits are to advance straightforwardness of orderly audit technique and to stay away from duplication of deliberate surveys. Web Ontology Language (W3C) impediment comes from the manner in which OWL utilizes limitations. They serve to indicate how information ought to be organized and forestall adding information conflicting with these limitations. This isn't generally helpful. Frequently, information imported from another source into the RDF triple store would be primarily conflicting with the imperatives set utilizing OWL. Therefore, this new information would need to be changed prior to being coordinated with what is as of now stacked in the triple store. One such limit is the accessible property develops. For instance, while giving amazing class develops, the latest form of the Web Ontology Language OWL2 has a fairly restricted arrangement of property builds. Another limit comes from the manner in which OWL utilizes requirements. HL 7 correspondence its trouble of articulation of clinical discoveries, signs, and manifestations, which involve electronic clinical records, all in light of absence of sufficient code of these things, and its being a depiction of fixed planning of assessment results, not being a grouping of undertaking of the patient. Hadoop upholds cluster preparing just, it doesn't handle streamed information, and consequently in general execution is slower. The MapReduce system of Hadoop doesn't use the memory of the Hadoop bunch to the greatest. Learning the clinical phrasing frameworks requires persistence and constancy albeit this is relevant just to framework creators and design. Making and question articulations is a specialized test. SNOMED CT likewise has a few information irregularity issue SNOMED CT additionally have a slower update plan than would need for the upkeep of a creation undertaking reference wording [13]–[16]

To settle restrictions of Hadoop flash is utilized that works on the exhibition, yet Spark stream preparing isn't pretty much as effective as Flink as it utilizes miniature clump handling. Flink works on the general execution as it gives single run-time to the spilling just as cluster preparing. Flink utilizes local shut circle emphasis administrators which make AI and diagram handling quicker. Quick headways in Artificial Intelligence keep on acquainting the innovation with a wide scope of mechanical applications with the clinical gadget industry being no exemption. Advances in scientific and figuring strategies, fixed with the blast of information in medical services associations, achieve numerous potential use instances of man-made consciousness to the fields of medication and medical care, uncovering driving clinical works on, lessening research disclosure time, smoothing out organization and offering new customized patient therapy. Included at a mechanical scale, AI can possibly adjust patients and guardian's choices in manners that further develop treatment results and add esteem. [16] Rather than content-based displaying, communitarian separating System needn't bother with the highlights of the things to be given. Each client and thing are portrayed by a component vector or inserting. It makes inserting for the two clients and things all alone. It installs the two clients and things in the equivalent inserting space. It thinks about other clients' responses while suggesting a specific client. It notes which things a specific client likes and

furthermore the things that the clients with conduct and likings like client likes, to prescribe things to that client. It gathers client criticisms on various things and utilizations them for suggestions. EHR give may assist with giving improved deliberation to patients and computerize a variety of undertakings for the training. In addition, they are profoundly adaptable and can be intended by the necessities of your experimental practice.

## V. CONCLUSION

This paper presents a systematic review of big data in health care in order to find the research gaps and future improvements. The huge information examination is ceaselessly assisting with building a superior insightful structure. Recently, medical care area has understood the significance of large information and subsequently applied huge information investigation in medical care and clinical practices. Innovation is helping in separating significant data from huge information in diminished time-frames. The authors conclude that Big Data and machine learning techniques can aid the Healthcare industry in the upcoming years.

## REFERENCES

1. U. Akhtar, J. W. Lee, H. S. Muhammad Bilal, T. Ali, W. A. Khan, and S. Lee, "The Impact of Big Data In Healthcare Analytics," *Int. Conf. Inf. Netw.*, vol. 2020-January, pp. 61–63, 2020, doi: 10.1109/ICOIN48656.2020.9016588.
2. P. S. Mung and S. Phyu, "Effective Analytics on Healthcare Big Data Using Ensemble Learning," *2020 IEEE Conf. Comput. Appl. ICCA 2020*, pp. 1–4, 2020, doi: 10.1109/ICCA49400.2020.9022853.
3. M. A. Lambay and S. Pakkir Mohideen, "Big Data Analytics for Healthcare Recommendation Systems," *2020 Int. Conf. Syst. Comput. Autom. Networking, ICSCAN 2020*, pp. 3–8, 2020, doi: 10.1109/ICSCAN49426.2020.9262304.
4. S. Nazir et al., "A comprehensive analysis of healthcare big data management, analytics and scientific programming," *IEEE Access*, vol. 8, pp. 95714–95733, 2020, doi: 10.1109/ACCESS.2020.2995572.
5. A. D. Alahmar and R. Benlamri, "SNOMED CT-Based Standardized e-Clinical Pathways for Enabling Big Data Analytics in Healthcare," *IEEE Access*, vol. 8, pp. 92765–92775, 2020, doi: 10.1109/ACCESS.2020.2994286.
6. M. S. Hadi, A. Q. Lawey, T. E. H. El-Gorashi, and J. M. H. Elmirghani, "Patient-Centric HetNets Powered by Machine Learning and Big Data Analytics for 6G Networks," *IEEE Access*, vol. 8, pp. 85639–85655, 2020, doi: 10.1109/ACCESS.2020.2992555.
7. S. Dash, S. K. Shakyawar, M. Sharma, and S. Kaushik, "Big data in healthcare: management, analysis and future prospects," *J. Big Data*, vol. 6, no. 1, 2019, doi: 10.1186/s40537-019-0217-0.
8. G. Chen and M. Islam, "Big Data Analytics in Healthcare," *Proc. - 2019 2nd Int. Conf. Saf. Prod. Informatiz. IICSPI 2019*, pp. 227–230, 2019, doi: 10.1109/IICSPI48186.2019.9095872.
9. M. Shatnawi, M. B. Yassein, and I. Jalabneh, "The Use of Big Data Analytics in Medical Images: A Survey," *2020 11th Int. Conf. Inf. Commun. Syst. ICICS 2020*, pp. 294–298, 2020, doi: 10.1109/ICICS49469.2020.239514.
10. C. Zhang, R. Ma, S. Sun, Y. Li, Y. Wang, and Z. Yan, "Optimizing the electronic health records through big data analytics: A knowledge-based view," *IEEE Access*, vol. 7, pp. 136223–136231, 2019, doi: 10.1109/ACCESS.2019.2939158.
11. J. P. Lu Yan, Weihong Huang, Liming Wang, Song Feng, Yonghong Peng, "Data-enabled Digestive Medicine: A New Big Data Analytics Platform."

12. A. S. Panayides, M. S. Pattichis, S. Leandrou, C. Pitris, A. Constantinidou, and C. S. Pattichis, "Radiogenomics for Precision Medicine with a Big Data Analytics Perspective," IEEE J. Biomed. Heal. Informatics, vol. 23, no. 5, pp. 2063–2079, 2019, doi: 10.1109/JBHI.2018.2879381.
13. J. I. Sakai, T. Mori, S. Saito, Y. Tanaka, and H. Aurass, "Simulated enhancement of solar type II radio bursts during the collision of two shocks associated with coronal mass ejections," Astron. Astrophys., vol. 454, no. 3, pp. 983–988, 2006, doi: 10.1051/0004-6361:20065130.
14. M. F. Furukawa, J. King, V. Patel, C. J. Hsiao, J. Adler-Milstein, and A. K. Jha, "Despite substantial progress in EHR adoption, health information exchange and patient engagement remain low in office settings," Health Aff., vol. 33, no. 9, pp. 1672–1679, 2014, doi: 10.1377/hlthaff.2014.0445.
15. A. Kankanhalli, J. Hahn, S. Tan, and G. Gao, "Big data and analytics in healthcare: Introduction to the special section," Inf. Syst. Front., vol. 18, no. 2, pp. 233–235, 2016, doi: 10.1007/s10796-016-9641-2.
16. P. Galetsi, K. Katsaliaki, and S. Kumar, "Values, challenges and future directions of big data analytics in healthcare: A systematic review," Soc. Sci. Med., vol. 241, p. 112533, 2019, doi: 10.1016/j.socscimed.2019.112533.
17. Javeed, D., Gao, T., & Khan, M. T. (2021). SDN-Enabled Hybrid DL-Driven Framework for the Detection of Emerging Cyber Threats in IoT. Electronics, 10(8), 918.
18. Javeed, D., Gao, T., Khan, M. T., & Ahmad, I. (2021). A Hybrid Deep Learning-Driven SDN Enabled Mechanism for Secure Communication in Internet of Things (IoT). Sensors, 21(14), 4884.

### AUTHORS PROFILE



**Nek Dil Khan**, is currently a Master student at the Software College at Northeastern University, China. He received his Bachelor of Software Engineering (BSc. SE) in 2016 from University of Science and Technology Bannu. His research

interests include artificial intelligence, machine learning, natural language processing, big data, and text data mining.



**Muhammad Younas** is currently a Master student at the Software College at the Northeastern University, China. He received his Bachelor of Computer Science degree in 2019 from (IBMS) Institute of Business and Management Sciences at The University of Agriculture, Peshawar,

Pakistan. His research interests include blockchain, artificial intelligence, machine learning, and big data.



**Muhammad Taimoor Khan** received the M.E degree in Information security from Riphah Institute of Science and Engineering, Islamabad 44000, Pakistan in 2020. He is currently working as a SOC Analyst in Trillium Information Securities, Pakistan. He has several research

publications to his credit as main author and co-author. His research interests include deep learning, cybersecurity, network security, software-defined networking, Internet of Things, Intrusion detection, and prevention system.



**Duaa** received the M.E degree in Information security from Riphah Institute of Science and Engineering, Islamabad 44000, Pakistan in 2020. She is currently working as an AM IS Audit at PTCL, Islamabad, identifying and handling

risks by providing treatment recommendations. Her research interests includes cybersecurity, network security, Big Data and IDS.



**Adil Zaman** received his Bachelors degree in computer science from Abdul Wali Khan University Mardan, Pakistan in 2016. He is currently working as a Unity Game developer in Tansals Technologies. His research interests include Big Data, Network Security, Deep learning.