Global Trends in Science and Technology

ISSN : 3079-5761
Volume 1, Number 3, 2025
https://doi.org/10.70445/gtst.1.3.2025.29-47

Big Data and Health Informatics: Managing Privacy, Accuracy, and Scalability

Mohammed Javeedullah[1*]

[1]New England College 98 Bridge Street, Henniker, NH 03242

[1]JMohammed3GPS@nec.edu

open access

Corresponding Author
Mohammed Javeedullah
JMohammed3GPS@nec.edu

Article History:
Submitted: 23-05-2025
Accepted: 28-06-2025
Published: 03-07-2025

Keywords

Big Data, Health Informatics, Data Privacy, Interoperability, Data Accuracy, Scalability, Ethical Implications, Artificial Intelligence.

Global Trends in Science and Technology is licensed under a Creative Commons Attribution-Noncommercial 4.0 International (CC BY-NC 4.0).

ABSTRACT

Big data has become a revolutionary element in the field of health informatics, as they provide the ability to achieve new clinical decision-making, precision medicine, and population health management. In the scope of this review, privacy, data accuracy, scalability, interoperability, and ethical, legal, and social implications (Elsi) of making use of big data in the healthcare sector will be discussed as pivotal issues of concern. As health data sources become more varied and larger in volume, e.g. the use of wearable devices in addition to electronic health records, assuring the quality of health data, safeguarding patient privacy, and the ability to easily integrate data into the system, becomes more complicated. The article was also focusing on new technologies including cloud computing, artificial intelligence and block chain as these solutions are scalable and secure. Also, it suggests that regulatory frameworks and ethical standards are critical in order to control responsible use of data. Finally, this review demonstrates that interdisciplinary cooperation is needed to create an adaptable, just, and evidence-based infrastructural context of health informatics enabling innovation and preserving the trust of the people.

INTRODUCTION

Big data in health informatics has transformed the face of contemporary healthcare as it allows making more accurate, timely, and data-driven decisions. Health informatics that has long concerned itself with the management of electronic health records (EHRs) and information systems in clinical settings, has evolved swifty as big data technologies have appeared [1]. The current development has extended the capabilities of informatics beyond mere record-keeping to sophisticated analytics,

This is an Creative Commons License This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

ISSN : 3079-5761
Volume 1, Number 3, 2025
https://doi.org/10.70445/gtst.1.3.2025.29-47

predictive modeling, and population health monitoring and real-time clinical decision support [2].

In the medical sector, big data is described using the five Vs, including Volume, Velocity, Variety, Veracity, and Value. The amount of data produced by healthcare systems has grown tremendously as it is being collected in a variety of sources such as EHRs, wearables, imaging equipment, genomic sequencing as well as patient-reported outcomes through mobile applications and social networks [3]. The rate at which this data is generated and must be managed either in real-time or otherwise poses an opportunity or challenge on how to handle data. The type of data that is involved (structured, semi-structured and unstructured) requires sophisticated ways of integration and analysis, and the quality of the data (i.e., whether it is correct and reliable) is fundamental not to derive misleading and potentially harmful conclusions [4]. When talking about the worth of big data, the capacity to transform complicated data into valuable insights, which will help better the outcomes of patients and streamline the system, is essential.

The role of big data in health informatics, however, has created some major issues regarding data privacy, security, and ethics governance due to ever-growing reliance on big data. Among the most sensitive types of personal information, it is possible to mark the health-related information, incorrect handling or leakage of which can cause a considerable damage [5]. Meanwhile, the incentives to be accurate due to the volume and complexity of data collected warrant considerations regarding the validity of the data as well as bias and uniformity thereof. In addition to this, infrastructure scalability, algorithm scalability and human knowledge are crucial contributors to the sustainability of systems as datasets become exponential [6].

The rest of the paper seeks to understand big data and health informatics by examining three fundamental issues namely privacy, accuracy, and scalability. It gives an overview of the existing environment, lists technological and ethical challenges, and talks about new solutions. This paper is a contribution to a better comprehension of the fact that big data can be responsibly and efficiently utilized in the field of healthcare delivery, research, and policy, by providing answers to these challenges [7].

AN OVERVIEW OF BIG DATA IN HEALTH INFORMATICS

Big data has become a cornerstone in the current health informatics history, changing the way healthcare systems capture, handle and exploit information. In essence, big data in health informatics implies large and complicated sets of data that are produced as a result of diverse mechanisms within the healthcare ecosystem [8]. The volume, velocity, and heterogeneity demand the use of new technologies and methods because these systems cannot be effectively handled using outdated data-processing programs.

This is an Creative Commons License This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

ISSN : 3079-5761
Volume 1, Number 3, 2025
https://doi.org/10.70445/gtst.1.3.2025.29-47

The origin of big data in healthcare is numerous and it keeps on growing. They encompass structured data retrieved in electronic health records (EHRs), lab systems, pharmacy, and billing systems, and unstructured or semi-structured data found in clinical notes, image techniques, genomic sequences, mobile health (mo health) applications, wearable devices, and social media networks [9]. All these sources will provide dimensions of information about patients and populations that are unique and may allow a more comprehensive view of health determinants, patterns of diseases, and outcomes of treatment efforts [10].

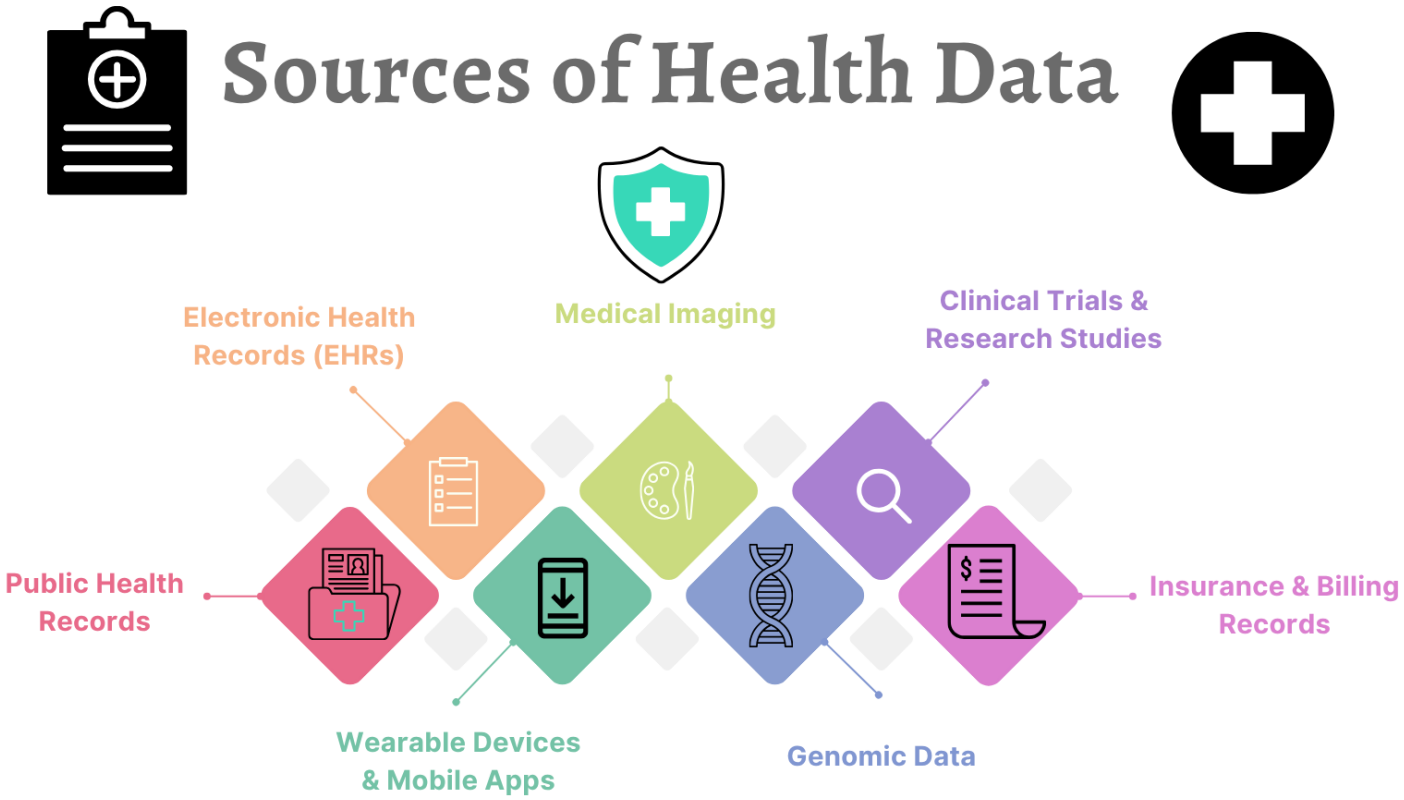


Figure: 1 showing sources of health data

Big data in health informatics promises to bring together these various differences indata. It facilitates the large number of applications such as data-driven clinical decision support systems, predictive models, personalized and population health management, and population health surveillance [11]. To illustrate, the application of machine learning algorithms analyzed on large patient databases can determine a group of patients at the highest risk of adverse events, streamline the treatment process, or predict the onset of an infectious disease. In a similar manner, using genomic data and EHRs allows precision medicine by making interventions according to the genetic profile of individuals [12].

Nevertheless using big data in health informatics is not a walk in the park either. The data may be of different types capture anything form an image, free text to sensor data inputs, the integration processes and data harmonization strategies are complex [13]. The rate of data generation, particularly

This is an Creative Commons License This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

ISSN : 3079-5761
Volume 1, Number 3, 2025
https://doi.org/10.70445/gtst.1.3.2025.29-47

in real time such as in intensive care or in tele-medicine, driven applications necessitates the existence of powerful computing disturbances that can easily process such data at a rapid rate. What is more, the accuracy, consistency, and credibility of the data, i.e., its veracity, is an ongoing issue of concerns, especially where data is generated through various often incompatible sources [14].

Nevertheless, the utility of big data is enormous with these barriers. Big data is on the cutting edge of healthcare innovation by providing the ability to make predictions more accurately, streamline operations, and make evidence-based policymaking. With the current development of this field, it is difficult to underestimate the need to learn about its elementary aspects and practical implications in order to fully exploit its possibilities and cope with the relevant ethical, technical, and regulatory challenges [15].

PRIVACY PROBLEMS AND SECURITY IN HEALTH INFORMATICS

With the increasing data volume and complexity on health issues, privacy and security are now in the focal point of health informatics. Healthcare big data analytics frequently entails collecting very sensitive personal data collected by a variety of devices and applications including electronic health records (EHRs), wearable devices, genetic testing tools, and mobile health apps. As much as these data are very valuable towards enhancing care delivery and outcomes, they are a very big risk in case they are not handled in a responsible manner [16].

One of the main concerns is how to keep the patients confidential and still be able to share data to be used during research, clinical decision making and to monitor the state of health of the general population. Breach of health information may lead to loss of money, theft of identity, discrimination and mistrust of healthcare organization [17]. Other large-scale affairs with unauthorized access, ransomware attacks, or inadvertently released data have been revealing the frailty of even the long-established healthcare system.

This is an Creative Commons License This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

ISSN : 3079-5761
Volume 1, Number 3, 2025
https://doi.org/10.70445/gtst.1.3.2025.29-47

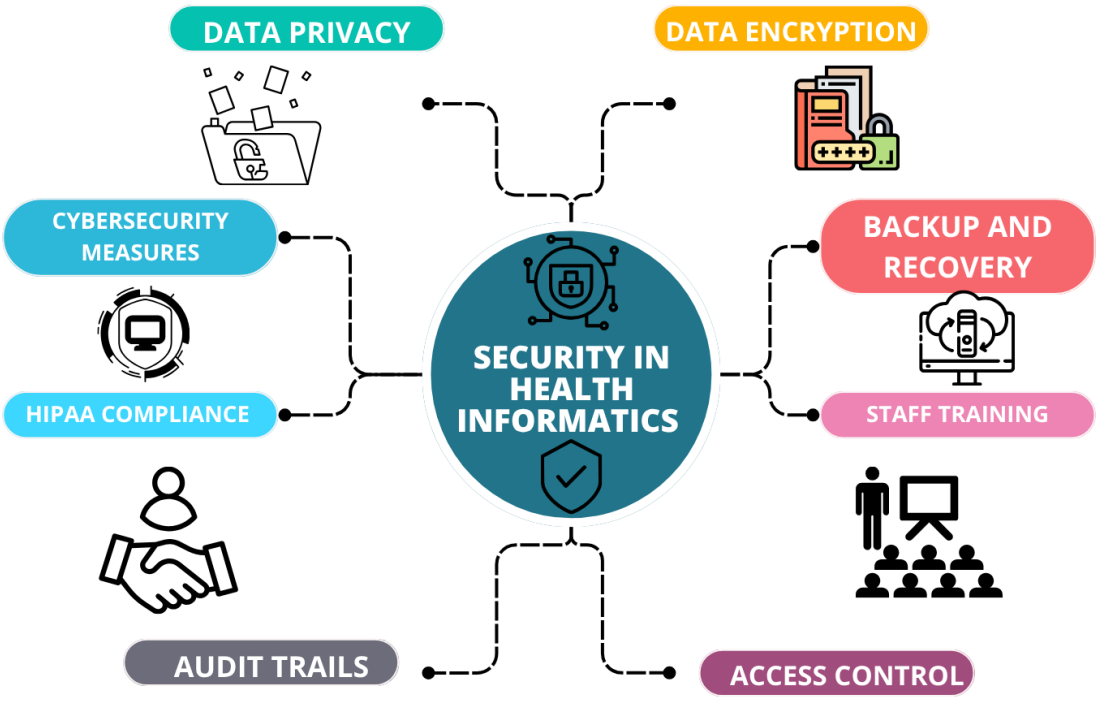


Figure: 2 showing security in health informatics

To deal with such issues a number of privacy laws and frameworks have been introduced world over. As an example, Health Insurance Portability and Accountability Act (HIPAA) in the United States, and the General Data Protection Regulation (GDPR) in the European Union stipulate severe provisions regarding the collection, storage, access, and sharing of data [18]. These laws require people to sign an informed consent, use data minimally, and have the right to access and remove their personal data. Nonetheless, regulation compliance may turn out to be complicated, particularly, when it comes to cross-border collaboration on research or even integrating the technologies of third-party vendors, whether cloud-based storage or machine learning algorithms [19].

Individual identities are usually safeguarded by techniques called anonymization and de-identification. These approaches are however ineffective as health datasets are becoming richer and have more granularity. An issue that continues to rage on is the attempt of re-identification attacks, whereby anonymously generated data can be linked to other databases in order to rediscover identities [20]. The newer methods of algorithmic privacy like differential privacy and homomorphic encryption provide good hints towards providing a successfully enhanced data security without compromising analytical performance, but these methods are also at their beginning phases of employment in real-life healthcare context [21].

Cybersecurity is also a significant issue, especially with more healthcare systems being tied via IoT,

This is an Creative Commons License This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

ISSN : 3079-5761
Volume 1, Number 3, 2025
https://doi.org/10.70445/gtst.1.3.2025.29-47

telemedicine, and cloud networks. These technologies increase an attack surface to the malicious actors. Consequently, cybersecurity measures are becoming more and more required, thanks especially to multi-factor authentication, network segmentation, and constant monitoring of threats [22].

To sum it up, the process of privacy and security management of health informatics is a trade between patient rights and innovation enabler. With the increasing volumes of data and the abilities in data analysis, a pro-active, multi-dimensional strategy of data governance is needed: data that is governed legally, ethically, and also by technology, in a multi-layered form of governance to guarantee trust and accountability in digital healthcare systems [23].

QUALITY AND ACCURACY OF DATA IN HEALTH INFORMATICS

Reliability of insights of big data in health informatics is directly related to accuracy, and the quality of the used data. In healthcare, the quality of data may be an issue of life or death, and ineffective quality may trigger misdiagnosis, delayed treatment, unreliable research results, in-service resource management. In spite of the hope of big data, the discipline still has a long way to go concerning data integrity, completeness, consistency and standardization [24].

Heterogeneous nature of health data is one of the major problems. Clinical data can exist in structured (i.e. laboratory results and prescription), semi-structured (i.e. HL7 messages) and unstructured text (i.e. doctor notes) formats. Millions of lines of non-standardized data are constantly produced by wearable devices and mobile health applications [25]. The integration of these heterogeneous sources creates multifold chances of introducing data errors in entry, internal copies, elimination of values and differences in terminologies and coding.

Preprocessing and cleansing of data are important to enhance quality of data. These involve error correction, elimination of duplicates, management of missing values as well as alignment of data to standard formats or ontologies. Valid coding systems like ICD-10, SNOMED CF, LOINC, RxNorm are essential to have a continuous consistency and interoperability among different health care systems and systems [26]. Yet, they are not evenly applied in institutions and countries making integration effort even harder.

This is an Creative Commons License This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

ISSN : 3079-5761
Volume 1, Number 3, 2025
https://doi.org/10.70445/gtst.1.3.2025.29-47

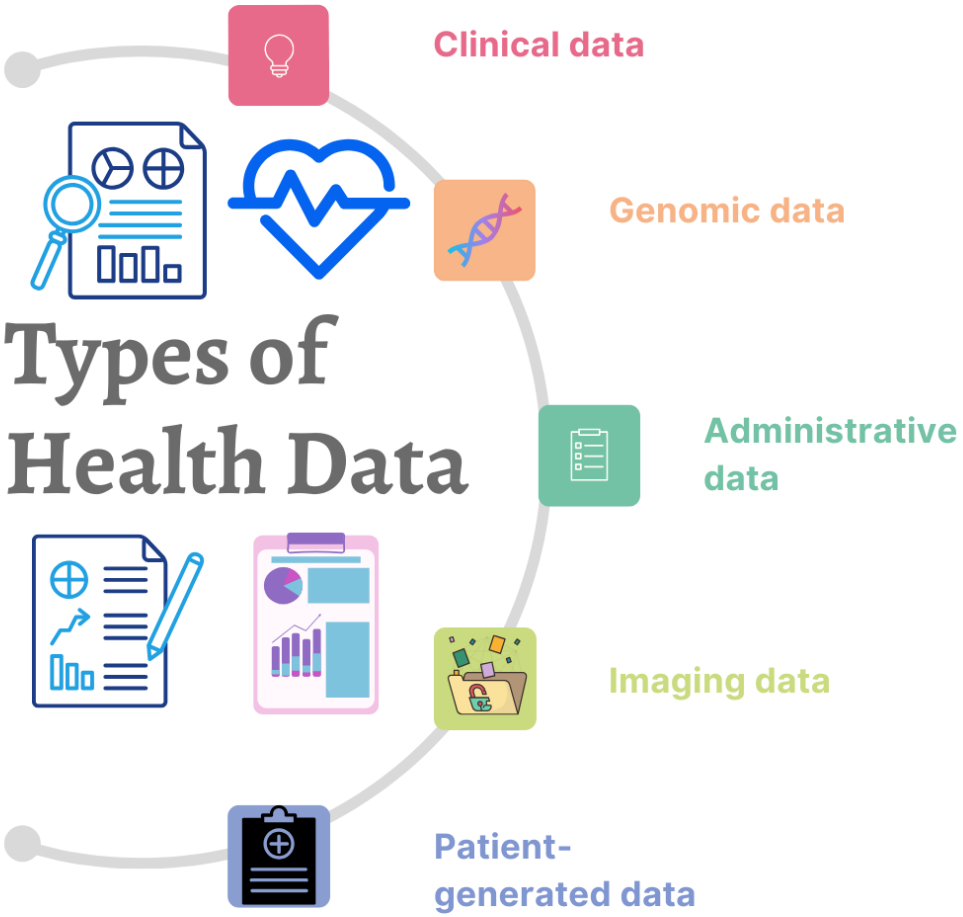


Figure: 3 showing types of health data

The other issue is bias in datasets that may be based on disparity in the representation of patient groups or on missing medical history or even on subjective clinical notes. When left unmitigated, these biases have the potential to transfer into machine learning models and analytical tools to create the existence of systemic disparities in care delivery and health outcomes [27]. To achieve fairness within data-driven healthcare, it is necessary to continuously mind creating audits of datasets with regard to representativeness and modify methodologies on an ongoing basis.

On-the-fly data confirmation is likewise turning into significant consideration, particularly when there are arrangements that depend on constant streaming information sent through IoT equipment or distant observing devices. There will be no use of real-time analytics in the absence of tools to check the correctness and exactness of such data. Furthermore, legacy solutions and manual work assignment can cause increased chances of errors on the part of humans, which supports the necessity of verification-related automation [28].

Finally, high quality data is not only a technical issue, but a strategic issue as well. It demands a

This is an Creative Commons License This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

ISSN : 3079-5761
Volume 1, Number 3, 2025
https://doi.org/10.70445/gtst.1.3.2025.29-47

concerted approach by healthcare providers, information technology specialists, data scientists and policy makers. The best way of putting big data and data assets into meaningful use, to improve patient care and research, and to make health care a more efficient, data-asset-driven system, is to establish effective data governance structures, and to make investments in data stewardship, and to help build a culture of accountability [29].

SCALABILITY INFRASTRUCTURE

With large healthcare systems turning more and more to big data to support final decision-making, research, and efficiency in the process of operation, the necessity to have a scalable and robust infrastructure has never been more evident than now. Scalability in health informatics is how a system can deal with rapidly increasing amount of data, data variety and data velocity [30]. As electronic health records (EHRs), imaging systems, genomic sequencing, mobile health applications, and the IoT devices continue to generate excessive amounts of data, modern health analytics often require infrastructural support that conventional systems simply cannot promote [31].

One way to deal with this is by the migration of databases in use by many healthcare organisations to the cloud environment. Flexibility, scalability and cost-efficiency are some of the advantages of cloud computing enabling the institutions to process and store large volumes of data without incurring the actual cost of hardware [32]. Document the likes of Amazon Web Services (AWS), Google cloud, and Microsoft Azures are platforms that offer on-demand computing and storage facility that scales up with the demand. But this transition brings in new issues of data privacy, regulatory compliance, and lock-in vendor in the case of sensitive environments like healthcare [33].

One more important feature of scalable infrastructure is applying distributed frameworks of computing, e.g., Apache Hadoop and Apache Spark. These technologies allow large data to be processed in parallel using several servers which has a great effect on fast and efficient processing of data [34]. As an example, in-memory processing offered by the Spark can be especially valuable when dealing with real-time analytics (a patient under critical care placement or the web of a disease outbreak).

Healthcare systems also require investment in high-performance computing (HPC) infrastructure to enable advanced analytics, with applications such as machine learning and deep learning being other high on the agenda. These are GPUs, TPUs, and other dedicated heavy-duty hardware to perform the numerous calculations needed by artificial intelligence algorithms. Most advanced models cannot work without scalable infrastructure, and the lack of that may be the reason why they will get stuck with processing delays, data throughput, or a memory bottleneck [35].

Also important to scalable health informatics are data integration platforms and interoperability

This is an Creative Commons License This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

ISSN : 3079-5761
Volume 1, Number 3, 2025
https://doi.org/10.70445/gtst.1.3.2025.29-47

standards. In many real-time applications, systems are required to combine well-defined and loosely defined data provided within a variety of sources with ease. This demands embracement of the common data exchange formats like HL7 FHIR (Fast Healthcare Interoperability Resources), which guarantees compatibility and effective communication amid various systems [36].

To sum up, scalable infrastructure should be built to help unleash the potential of big data in health informatics. It demands a just investment of software, hardware, cloud solutions and information management. Healthcare institutions can only handle the increasing complexity of data and provide timely and data-driven insights to enhance patient outcomes and the efficiency of operations by building on scalable, resilient, and secure systems [37].

INTEGRATION AND INTEROPERABILITY IN HEALTH INFORMATICS

The integration and interoperability are the key concepts of achieving the potential of big data in health informatics. Since health data is being created in many systems (electronic health records (EHRs) and laboratory information system (LIS), radiology systems, pharmacy systems, wearable devices, and telehealth apps) the ease of sharing and understanding that data is essential [38]. These data silos prevent well-coordinated care, restrict research capabilities, as well as decrease efficiency of the entire system in absence of interoperability.

Integration simply means a technical practice of amalgamating data in diverse sources through a single system. It does not just need to accumulate structured and unstructured data but also align them to common ontologies and terminologies so that they will be consistent. Effective integration allows complete patient records, information on cross-institutional data sharing, and analytics and decision-support systems [39].

Interoperability on the other hand would entail the capacity of various health information systems, devices and applications to read, retrieve, interpret and use data in a consistent manner. It works at four levels: Though it is vital, the issue of interoperability has been of major concern in healthcare. Lack of standardization is one of the fundamental problems. The fact is that health systems are prone to proprietary formatting and vendor-specific interface, which are incompatible with industry-wide data standards. The difference in the application of medical terminologies (e.g. ICD, SNOMED CT, and LOINC) also makes data harmonization difficult [40].

To solve such problems, universal attempts have occurred to introduce interoperability standards, the most prominent of which are Health Level Seven (HL7), Clinical Document Architecture (CDA), and, of course, the Fast Healthcare Interoperability Resources (FHIR). FHIR is a contemporary and web-based strategy in healthcare information exchange that supports less complex integration and sharing of information across platforms and organizations [41].

This is an Creative Commons License This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

ISSN : 3079-5761
Volume 1, Number 3, 2025
https://doi.org/10.70445/gtst.1.3.2025.29-47

There is also the emergence of new technologies like Application Programming Interfaces (APIs) where the connectivity between systems has been enhanced through modular or real-time access of data between platforms. In addition, block chain has been suggested as an approach to safe, decentralized health information trading; such an approach is yet under consideration in regard to its scalability and regulation [42]. Not enough is technical solutions, however. Aligning policies, cooperation between institutions, and governance structures are also essential to allow safe and ethical data sharing. Use of regulatory requirements, including the U.S. 21 st Century Cures Act, are nudging vendors and healthcare providers to more open and interoperable systems [43].

To sum up, interoperability and integration are not only technic missions but strategic needs with patient-centered, data-driven, and scalable healthcare. The further development of such capabilities will open access to a larger version of high-quality data, provide the conditions of evidence-based care, and innovate in the health field of informatics [44].

ETHICAL, LEGAL AND SOCIAL IMPLICATIONS (ELSI) IN HEALTHCARE

Big data is not only posing technical and operation issues as it continues to be integrated in health informatics, but also has huge ethical, legal and social consequences (ELSI). This entails the collection of huge volumes of sensitive and at times personal health information, which begs such questions as privacy, autonomy, equity, consent, accountability, and public trust [45]. The concern of ELSI is important in order to guarantee that technological developments do not conflict with basic conflicts of humanity and values of the society.

One primary ethical issue is autonomy and informed consent by a patient. More conventional models of consent, namely when individuals consent to particular uses of their data before care is provided, tend to be insufficient when it comes to big data, wherein data might be repurposed to research, learning AI models, or other more commercial uses many orders of magnitude removed [46]. The dynamic consent or broad consent models were suggested to enable management of permission on an ongoing basis, although the earlier approaches should be well formulated that the patient should not be overburdened or disoriented [47].

This is an Creative Commons License This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

ISSN : 3079-5761
Volume 1, Number 3, 2025
https://doi.org/10.70445/gtst.1.3.2025.29-47

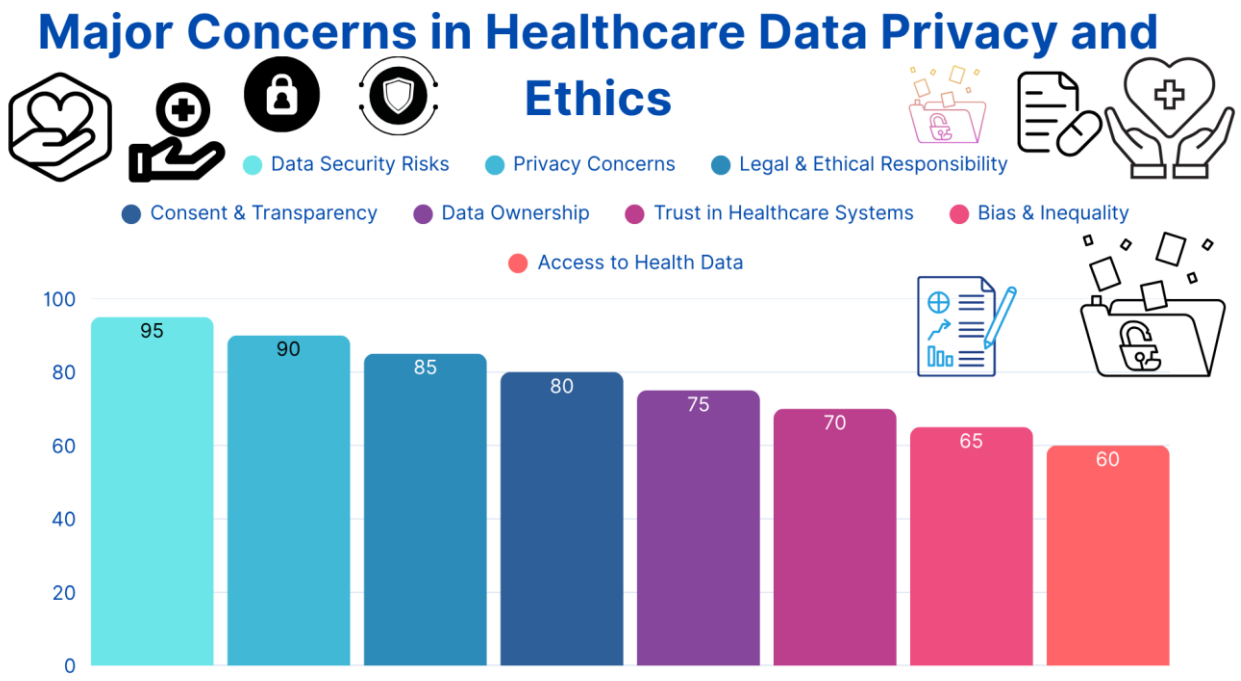


Figure: 4 showing major concerns in healthcare data privacy and ethics

Other disputable issues are data ownership and control. Although people create health data, the latter can be stored and manipulated by health professionals, insurance companies, or third-party systems. Who owning this data and who is entitled to use, share, or make money at its expense is a question left unascertained on the legally and ethically pertinent grounds in many jurisdiction? Increasing the disease victims with their health data such as access and portability rights is an ethical priority that is gaining a lot of concern [48].

Legally speaking, there is a number of regulations that are designed to protect health data, including the Health Insurance Portability and Accountability Act (HIPAA) in the United States and the General Data Protection Regulation (GDPR) in Europe. These legislations create schemes of data utilization, disclosure, and notification of breaches and fines in case of non-conformance [49]. Nevertheless, at times, there is a mismatch in technological advancements with effort to match according to regulation, thus giving way to a cloudy law in most cases- notably in new technologies such as AI, cloud and cross-border data flow [50].

Socially, equity and algorithmic fairness are becoming increasingly matters of concern. The trained algorithms that become the machine learning models could recreate health disparities because they make biased predictions or treatment suggestions. Case in point, the lack of equality in data pertaining to minority populations may amount to wrong diagnoses or futile treatments. It is important not to exclude information about data inclusivity and to install bias audits to avoid unintended harms [51].

Lastly, it is crucial to have the faith of the people. As long as people do not trust their data management

This is an Creative Commons License This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

ISSN : 3079-5761
Volume 1, Number 3, 2025
https://doi.org/10.70445/gtst.1.3.2025.29-47

applications, they will be less likely to engage in digital health initiatives, which will undermine innovation based on data. The principles of transparency, stakeholder involvement, and ethical monitoring are the main aspects in building a responsible data culture. ELSI considerations do not lie on the periphery of big data in health informatics; however, they are the core of responsible development and implementation. Such a combination of ethical thinking, legal standards and social responsibilities is the only way to make certain that technological advancement will contribute to fair and just healthcare results [52].

FUTURE DIRECTIONS AND TRENDS

Big data is changing the face of the future of health informatics by offering a revolutionary change in how care, delivery of care, and personalization of care takes place. With the ever-faster technological changes, a number of new trends as well as future paths are emerging and redefining the area, with new opportunities but also new challenges which need a careful approach [53].

The increased adoption of artificial intelligence (AI) and machine learning (ML) in clinical practice can be listed among the most important ones. Such predictive analytics is being created so that it will anticipate the outcomes of patients, diagnose them faster, and provide individualized methods of treatment [54]. These models should get more precise with increased data volume, enhanced computational capacity, and due to the availability of diverse datasets of high quality. They however should be integrated into standard care with explain ability, clinical verification as well as ethical supervision [55].

The second key trend is developing a precision medicine, which is supported by the usage of big data sources of genomic, proteomic and metabolomics data. With the addition of these along with EHRs and real-time sensor data, the healthcare providers will be able to target patient interventions based on both a genetic profile and lifestyle as opposed to a one-size-fits-all solution [56].

The emergence of the digital health ecosystems, such as telemedicine and the use of mobile health apps and wearable sensors, is producing constant, real-time health data that is not captured in conventional clinical scenarios. These platforms do not only enable patients to have greater control of their health status, but the data they have access to are also very informative about population-level health patterns [57].

Another approach that is getting traction as a possible means of solving secure and decentralized health data sharing is block chain, allowing improved data traceability and greater control by the patient. It is still very experimental, but may have the potential of enhancing trust and transparency of data transactions. In the prospect, the interdisciplinary cooperation of clinicians, data scientists, engineers, ethicists, and policymakers will be needed. Similarly, the creation of international data-

This is an Creative Commons License This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

ISSN : 3079-5761
Volume 1, Number 3, 2025
https://doi.org/10.70445/gtst.1.3.2025.29-47

sharing networks based on international privacy regulations and other moral principles will be the key to conducting research on a larger scale and greater influence [58]. Future impact of health informatics is about making healthcare system smarter, responsive and equitable with the leverage of big data. It will be essential to embrace new technologies and consider ethical, legal, and social issues as a prerequisite to the real-life benefits of the emerging technologies as far as their impact on patient care and population health is concerned.

CONCLUSION

The emergence of big data as an addition to health informatics remains a paradigm shift in the health care that is delivered, administered and conceptualized. The possible uses of big data range widely in terms of accuracy of diagnosis, customizability of treatment plans, forecasting of trends in the population of a given region and increased efficiency of operations. But all these gains are not without their major setbacks. Data privacy, data accuracy and quality, interoperability, and scalability of infrastructures are all important factors that need to be taken into account so that all the potential of data-driven healthcare could be utilized.

During this review, we have investigated the complex environment of big data in health informatics. The overarching theme that arises is that although the technology is progressing at an alarming rate, ethical, legal and social standards have to keep in tandem with the technology. The sheer amount of health information generated each day improper stewardship of health data, including, but not limited to, clinical records, genomics data, and mobile apps and wearable device data, in itself creates a challenge; and the sheer complexity of the data cannot be ignored. These convert health information that is fragmented, underutilized and can be hazardous as a result of glitches or misappropriation.

In this digital age, the concern is mainly on privacy and security. The higher the amount of personal health information gathered and transferred through the networks, the higher the opportunities of breach, unauthorized access, and re-identification. The legal frameworks offer the necessary protection, but there are still gaps, particularly when health data leaves the traditional healthcare setting and enters in the consumer-focused platforms. The use of technological protection mechanisms, e.g. encryption, anonymization, and privacy-preserving computation should be spread wider and improved.

Similarly, accuracy, thoroughness and timeliness of health data are important issues. The low quality of the data may result in making inaccurate conclusions, biased algorithms, and unsafe decisions in the clinical setting. This demands organized methods of data correction, validation and monitoring of constant quality together with introduction of common terminologies as well as codifications.

Scalability and interoperability are not technical dreams, they are a precondition to responsive and

This is an Creative Commons License This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

ISSN : 3079-5761
Volume 1, Number 3, 2025
https://doi.org/10.70445/gtst.1.3.2025.29-47

resilient health informatics infrastructure. There needs to be universal adoption of cloud-based solutions, APIs and interoperability's standards such as FHIR in order to overcome data silos and provide the ability to easily share between systems. It will lend a hand to coordinated care, eliminate duplication and improve decision-making no matter what level of healthcare one is relating to.

Ethical, legal, and social implications (ELSI) of big data cannot be overestimated. The future of health informatics is left to the responsibility of what can be done with data as well as how accountably it is done. Problems of consent, equity, transparency, and accountability should be discussed as much urgently as technical innovations. The key factors to support equitable benefiting of all groups by health informatics practices are inclusive data practices, fairness-aware algorithms, and patient-centered models of governance.

Moving into the future and the roughly foreseeable future, the next logical step in big data will be integrated with artificial intelligence, precision medicine, and live-time data monitoring to further extend the means of which healthcare operations are possible. But, the potential is large and with it large responsibility. A health informatics ecosystem needs to be innovative, ethical, secure, and sustainable and can only be achieved by stakeholders, such as researchers, clinicians, technologists, policymakers, and patients, collaborating to achieve its building.

Big data has amazing possibilities to change health informatics and advance the well-being of people. Its happy accomplishment will be, however, largely determined by the way in which we meet all its complex difficulties, and in shaping systems so thoughtful, as far as possible, and so enlightened and inclusive, as well as powerful.

REFERENCES

[1]. Wang X, Williams C, Liu ZH, Croghan J. Big data management challenges in health research—a literature review. Briefings in bioinformatics. 2019 Jan;20(1):156-67.

[2]. Hai Guang Li, Xindong Wu, Zhao Li, and Wei Ding. 2013. Online group feature selection from feature streams. In 27th AAAI Conference on Artificial Intelligence. Citeseer, 1627–1628.

[3]. Rongjian Li, Wenlu Zhang, Heung-Il Suk, Li Wang, Jiang Li, Dinggang Shen, and Shuiwang Ji. 2014. Deep learning based imaging data completion for improved brain disease diagnosis. In Medical Image Computing and Computer-Assisted Intervention (MICCAI'14). Springer, 305–312.

[4]. Shutao Li, Haitao Yin, and Leyuan Fang. 2012. Group-sparse representation with dictionary learning for medical image denoising and fusion. IEEE Transactions on Biomedical Engineering 59, 12 (2012), 3450– 3459.

This is an Creative Commons License This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

ISSN : 3079-5761
Volume 1, Number 3, 2025
https://doi.org/10.70445/gtst.1.3.2025.29-47

[5]. Znaonui Liang, Gang Zhang, Jimmy Xiangji Huang, and Qmming Vivian Hu. 2014. Deep learning for healthcare decision making with EMRs. In 2014 IEEE International Conference on Bioinformatics and Biomedicine (BIBM'14). 556–559.

[6]. Manhua Liu, Daoqiang Zhang, and Dinggang Shen. 2012. Ensemble sparse classification of Alzheimer's disease. NeuroImage 60, 2 (2012), 1106–1116.

[7]. Tao Meng and Mei-Ling Shyu. 2013. Biological image temporal stage classification via multi-layer model collaboration. In 2013 IEEE International Symposium on Multimedia (ISM'13). 30–37.

[8]. Andrew Pavlo, Erik Paulson, Alexander Rasin, Daniel J. Abadi, David J. DeWitt, Samuel Madden, and Michael Stonebraker. 2009. A comparison of approaches to large-scale data analysis. In Proceedings of the 2009 ACM SIGMOD International Conference on Management of Data. 165–178.

[9]. Mykola Pechenizkiy, Alexey Tsymbal, and Seppo Puuronen. 2004. PCA-based feature transformation for classification: Issues in medical diagnostics. In Proceedings of the 17th IEEE Symposium on ComputerBased Medical Systems (CBMS'04). IEEE, 535–540.

[10]. Saouabi M, Ezzati A. A comparative between hadoop mapreduce and apache Spark on HDFS. In: Proceedings of the 1st international conference on internet of things and machine learning. Liverpool: ACM; 2017. p. 1–4. 22.

[11]. Strickland NH. PACS (picture archiving and communication systems): flmless radiology. Arch Dis Child. 2000;83(1):82–6.

[12]. FLi L, et al. Identifcation of type 2 diabetes subgroups through topological analysis of patient similarity. Sci Transl Med. 2015;7(311):311ra174.

[13]. Valikodath NG, et al. Agreement of ocular symptom reporting between patient-reported outcomes and medical records. JAMA Ophthalmol. 2017;135(3):225–31.

[14]. Fromme EK, et al. How accurate is clinician reporting of chemotherapy adverse efects? A comparison with patientreported symptoms from the Quality-of-Life Questionnaire C30. J Clin Oncol. 2004;22(17):3485–90.

[15]. Patil HK, Seshadri R. Big data security and privacy issues in healthcare. In2014 IEEE international congress on big data 2014 Jun 27 (pp. 762-765). IEEE.

[16]. Merelli I, Pérez-Sánchez H, Gesing S, D'Agostino D. Managing, analysing, and integrating big data in medical bioinformatics: open problems and future perspectives. BioMed research international. 2014;2014(1):134023.

This is an Creative Commons License This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

ISSN : 3079-5761
Volume 1, Number 3, 2025
https://doi.org/10.70445/gtst.1.3.2025.29-47

[17]. Mir AA. Optimizing mobile cloud computing architectures for real-time big data analytics in healthcare applications: Enhancing patient outcomes through scalable and efficient processing models. Integrated Journal of Science and Technology. 2024 Jun 29;1(2).

[18]. El Aboudi N, Benhlima L. Big data management for healthcare systems: architecture, requirements, and implementation. Advances in bioinformatics. 2018;2018(1):4059018.

[19]. Mazumder MS. The Transformative Impact of Big Data in Healthcare: Improving Outcomes, Safety, and Efficiencies. Global Mainstream Journal of Business, Economics, Development & Project Management. 2024;3(03):01-12.

[20]. Harerimana G, Jang B, Kim JW, Park HK. Health big data analytics: a technology survey. Ieee Access. 2018 Oct 26;6:65661-78.

[21]. Harerimana G, Jang B, Kim JW, Park HK. Health big data analytics: a technology survey. Ieee Access. 2018 Oct 26;6:65661-78.

[22]. Senthilkumar SA, Rai BK, Meshram AA, Gunasekaran A, Chandrakumarmangalam S. Big data in healthcare management: a review of literature. American Journal of Theoretical and Applied Business. 2018 Jul 3;4(2):57-69.

[23]. Mazumder MS. The Transformative Impact of Big Data in Healthcare: Improving Outcomes, Safety, and Efficiencies. Global Mainstream Journal of Business, Economics, Development & Project Management. 2024;3(03):01-12.

[24]. Beckles GL, et al. Agreement between self-reports and medical records was only fair in a cross-sectional study of performance of annual eye examinations among adults with diabetes in managed care. Med Care. 2007;45(9):876–83.

[25]. Echaiz JF, et al. Low correlation between self-report and medical record documentation of urinary tract infection symptoms. Am J Infect Control? 2015; 43(9):983–6.

[26]. Or-Bach, Z. A 1,000x improvement in computer systems by bridging the processor-memory gap. In: 2017 IEEE SOI3D-subthreshold microelectronics technology unifed conference (S3S). 2017.

[27]. Dollas, A. Big data processing with FPGA supercomputers: opportunities and challenges. In: 2014 IEEE computer society annual symposium on VLSI; 2014.

[28]. Safman M. Quantum computing with atomic qubits and Rydberg interactions: progress and challenges. J Phys B: At Mol Opt Phys. 2016;49(20):202001.

[29]. Nielsen MA, Chuang IL. Quantum computation and quantum information. 10th anniversary ed. Cambridge: Cambridge University Press; 2011. p. 708.

This is an Creative Commons License This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

ISSN : 3079-5761
Volume 1, Number 3, 2025
https://doi.org/10.70445/gtst.1.3.2025.29-47

[30]. Raychev N. Quantum computing models for algebraic applications. Int J Scientifc Eng Res. 2015; 6(8):1281–8.

[31]. Zeb S, Nizamullah FN, Abbasi N, Qayyum MU. Transforming Healthcare: Artificial Intelligence's Place in Contemporary Medicine. BULLET: Jurnal Multidisiplin Ilmu. 2024;3(4):592385.

[32]. Buchanan W, Woodward A. Will quantum computers be the end of public key encryption? J Cyber Secur Technol. 2017; 1(1):1–22.

[33]. Mott A, et al. Solving a Higgs optimization problem with quantum annealing for machine learning. Nature. 2017; 550:375.

[34]. Rebentrost P, Mohseni M, Lloyd S. Quantum support vector machine for big data classifcation. Phys Rev Lett. 2014; 113(13):130503.

[35]. Gandhi V, et al. Quantum neural network-based EEG fltering for a brain-computer interface. IEEE Trans Neural Netw Learn Syst. 2014; 25(2):278–88.

[36]. Jamaludin H, Achlison U, Rokhman N. Enhancing AI Model Accuracy and Scalability Through Big Data and Cloud Computing. Journal of Technology Informatics and Engineering. 2024 Dec 24;3(3):296-307.

[37]. Zeb S, Nizamullah FN, Abbasi N, Fahad M. AI in healthcare: revolutionizing diagnosis and therapy. International Journal of Multidisciplinary Sciences and Arts. 2024 Aug 17;3(3):118-28.

[38]. Bellazzi R. Big data and biomedical informatics: a challenging opportunity. Yearbook of medical informatics. 2014;23(01):08-13.

[39]. Raghupathi W, Raghupathi V. Big data analytics in healthcare: promise and potential. Health information science and systems. 2014 Dec;2:1-0.

[40]. Gollapalli VS. Enhancing disease strati fication using federated learning and big data analytics in healthcare systems. International Journal of Management Research and Business Strategy. 2020 Oct 26;10(4):19-38.

[41]. Das J. Leveraging Cloud Computing for Medical AI: Scalable Infrastructure and Data Security for Advanced Healthcare Solutions. INTERNATIONAL JOURNAL OF RESEARCH AND ANALYTICAL REVIEWS. 2020;7:504-14.

[42]. Mehta N, Pandit A. Concurrence of big data analytics and healthcare: A systematic review. International journal of medical informatics. 2018 Jun 1;114:57-65.

[43]. McPadden J, Durant TJ, Bunch DR, Coppi A, Price N, Rodgerson K, Torre Jr CJ, Byron W, Hsiao AL, Krumholz HM, Schulz WL. Health care and precision medicine research: analysis

This is an Creative Commons License This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

ISSN : 3079-5761
Volume 1, Number 3, 2025
https://doi.org/10.70445/gtst.1.3.2025.29-47

of a scalable data science platform. Journal of medical Internet research. 2019 Apr 9;21(4):e13043.

[44]. Mathew PS, Pillai AS. Big Data solutions in Healthcare: Problems and perspectives. In2015 International conference on innovations in information, embedded and communication systems (ICIIECS) 2015 Mar 19 (pp. 1-6). IEEE.

[45]. Varshney M, Bhushan B, Haque AB. Big data analytics and data mining for healthcare informatics (HCI). InMultimedia Technologies in the Internet of Things Environment, Volume 3 2022 Apr 5 (pp. 167-195). Singapore: Springer Singapore.

[46]. Imran S, Mahmood T, Morshed A, Sellis T. Big data analytics in healthcare− A systematic literature review and roadmap for practical implementation. IEEE/CAA Journal of Automatica Sinica. 2020 Sep 24;8(1):1-22.

[47]. Sakly H, Guetari R, Kraiem N, Abed M. Big Data and AI Solutions for Transforming Healthcare: Frameworks, Challenges, and Future Directions. InScalable Artificial Intelligence for Healthcare 2025 (pp. 58-76). CRC Press.

[48]. Chaluvadi A, Karthick M. Leveraging cloud computing and big data for enhanced healthcare decision-making: Integrating LSTM for predictive modelling. International Journal of Business Management and Economic Review. 2021;4(5):142.

[49]. Ahmed A, Xi R, Hou M, Shah SA, Hameed S. Harnessing big data analytics for healthcare: A comprehensive review of frameworks, implications, applications, and impacts. IEEE Access. 2023 Oct 10;11:112891-928.

[50]. Yang X, Huang K, Yang D, Zhao W, Zhou X. Biomedical big data technologies, applications, and challenges for precision medicine: a review. Global Challenges. 2024 Jan;8(1):2300163.

[51]. Belle A, Thiagarajan R, Soroushmehr SR, Navidi F, Beard DA, Najarian K. Big data analytics in healthcare. BioMed research international. 2015;2015(1):370194.

[52]. Sakly H, Guetari R, Kraiem N, Abed M. Big Data and AI Solutions for Transforming Healthcare: Frameworks, Challenges, and Future Directions. InScalable Artificial Intelligence for Healthcare 2025 (pp. 58-76). CRC Press.

[53]. Bates DW, Saria S, Ohno-Machado L, Shah A, Escobar G. Big data in health care: using analytics to identify and manage high-risk and high-cost patients. Health affairs. 2014 Jul 1;33(7):1123-31.

[54]. Pramanik MI, Lau RY, Demirkan H, Azad MA. Smart health: Big data enabled health paradigm within smart cities. Expert Systems with Applications. 2017 Nov 30;87:370-83.

This is an Creative Commons License This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

ISSN : 3079-5761
Volume 1, Number 3, 2025
https://doi.org/10.70445/gtst.1.3.2025.29-47

[55]. Mohammed EA, Far BH, Naugler C. Applications of the MapReduce programming framework to clinical big data analysis: current landscape and future trends. BioData mining. 2014 Dec;7:1-23.

[56]. Uddin MK, Hossan KM. A Review of Implementing AI-Powered Data Warehouse Solutions to Optimize Big Data Management and Utilization. Academic Journal on Business Administration, Innovation & Sustainability. 2024 Jul 28;4(3):10-69593.

[57]. Sakr S, Elgammal A. Towards a comprehensive data analytics framework for smart healthcare services. Big Data Research. 2016 Jun 1; 4:44-58.

[58]. Nizamullah FN, Zeb S, Abbasi N, Qayyum MU, Fahad M. AI in Healthcare: Breaking New Ground in the Management and Treatment of Cancer. Asian Journal of Engineering, Social and Health. 2024 Oct 18;3(10):2325-43.

This is an Creative Commons License This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.