Send Orders for Reprints to reprints@benthamscience.net

Current Medical Imaging, 2019, *Volume, Page Enation* 1

ARTICLE TYPE

Big Data for Treatment Planning: Pathways and Possibilities for Smart Healthcare Systems

Samiya Khan[*1,a], Shoaib Amin Banday[2,b], and Mansaf Alam[3,c]

[a]*School of Mathematics and Computer Science, University of Wolverhampton, United Kingdom;* [b]*Department of Electronics & Communication, Islamic University of Science & Technology, India;* [c]*Department of Computer Science, Jamia Millia Islamia, India*

[1]*samiya.khan@wlv.ac.uk;* [2]*bandaysa@ieee.org;* [3]*malam2@jmi.ac.in*

ARTICLE HISTORY

Received:
Revised:
Accepted:

DOI:

Abstract:

Background: Treatment planning is one of the crucial stages of healthcare assessment and delivery. Moreover, it also has a significant impact on patient outcomes and system efficiency. With the evolution of transformative healthcare technologies, most areas of healthcare have started collecting data at different levels, as a result of which there is a splurge in the size and complexity of health data being generated every minute.

Introduction: This paper explores the different characteristics of health data with respect to big data. Besides this, it also classifies research efforts in treatment planning on the basis of the informatics domain being used, which include medical informatics, imaging informatics and translational bioinformatics.

Method: This is a survey paper that reviews existing literature on the use of big data technologies for treatment planning in the healthcare ecosystem. Therefore, a qualitative research methodology was adopted for this work.

Result: Review of existing literature has been analyzed to identify potential gaps in research, identifying and providing insights into high prospect areas for potential future research.

Conclusion: Use of big data for treatment planning is rapidly evolving and findings of this research can head start and streamline specific research pathways in the field.

Keywords: big data, treatment planning, medical informatics, medical imaging, translational bioinformatics, smart healthcare

1. INTRODUCTION

Big data technology is recognized to have far-reaching implications on diverse domains, including healthcare. Although, the use of assorted technologies to transform healthcare is rather new, the challenges associated with healthcare data have existed and been investigated for long. The concept of big data finds inherent applications in the healthcare sector in view of the large and complex datasets that this field generates.

*Address correspondence to this author at the School of Mathematics and Computer Science, University of Wolverhampton, United Kingdom; E-mail: samiya.khan@wlv.ac.uk

Typically, a big health dataset is complex, rapidly generated, huge volumes of data that cannot be stored, processed and managed using existing tools available with healthcare providers. Applications centered on big health data can play a significant role in facilitating sustainable and efficient health services to individuals of disparate age groups with focus on early and precise diagnosis, optimal management and treatment personalization. Recent times have witnessed path-breaking advancements in technology, as a result of which a lot of complex and high velocity data can be captured from clinical trials.

In entirety, evolution of such tools and technologies has made collection and processing of health data possible, enhancing the potential of analytical systems that work within and over clinical activities. With respect to the use of big data in

1573-4056/19 $58.00+.00

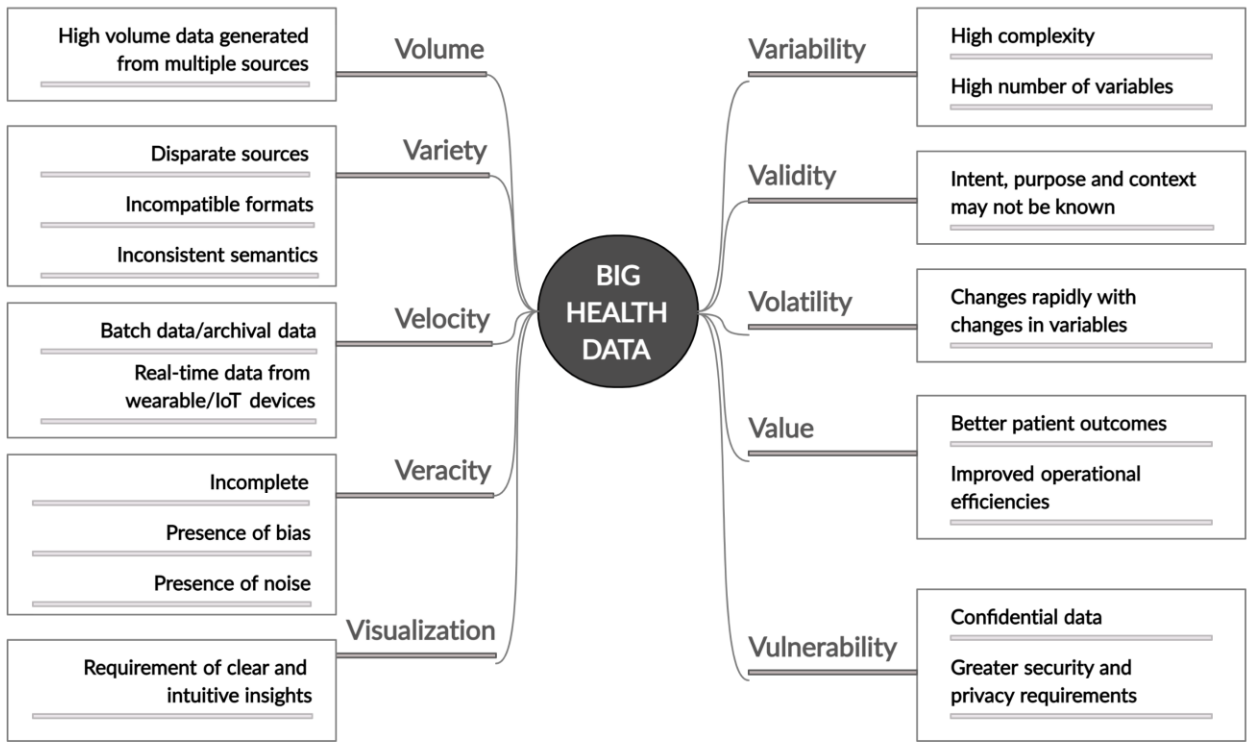

Figure 1. Characteristics of Big Health Data

healthcare, three critical components have been identified. Firstly, data needs to be captured, cleaned and stored in an appropriate and reusable manner. Secondly, this data needs to be analyzed and processed in such a way that a single environment can support large amounts of multi-type data. Finally, it should be possible to facilitate decision-making and make precise clinical interpretations on the basis of presented analysis.

The use of big data in healthcare is primarily based on the notion that use of large volumes of data coming from multiple, heterogeneous sources shall allow identification of clinical signals that traditional approaches fail to support. In congruence with this, big data technologies can support a wide range of healthcare applications ranging across descriptive, diagnostic, predictive and prescriptive analytics for disparate domains such as personalized healthcare, improved efficacy of signal detection and identification, drug design and discovery, health risks identification and implementation of preventive approaches for effective healthcare strategy. These applications individually and collectively support treatment planning by allowing development of decision support systems to help healthcare providers in streamlining the process of tailoring applications and available resources to the needs of and treatment goals for a patient. Big data-enabled systems are capable of performing multi-dimensional assessments, taking personalized treatment to an altogether different level of individual-specific adaptability.

Big data and its applications in healthcare is a widely surveyed and researched field with domain-specific survey papers for generic healthcare applications [1] as well as specific areas such as precision health [2], drug discovery [3], healthcare environments [4][5], and genomics [6], in addition to many others. However, none of the available survey papers dwell into the synergistic use of big data for treatment planning. This work is an attempt to bridge this gap in existing literature. This research paper broadly surveys literature related to the use of big data in treatment planning to comprehend existing state of the art systems and explore future possibilities in this field of research.

The rest of the paper is organized in the following manner: Section 2 provides a background of big data characterization model and uses the same to describe health data from the big data perspective. Besides this, it also gives a brief account of the different segments identified for big data application in healthcare to isolate informatics domains, which directly or indirectly relate to treatment planning. Section 3 investigates the different dimensions, components, perspectives and perceptions related to the field of treatment planning. Moreover, it also gives an account of existing literature that mentions the use of Artificial Intelligence (AI) driven medical, imaging and translational informatics for treatment planning. Section 4 provides an analysis of trends and outlines future scope. Finally, Section 5 synopsizes the contributions of the paper.

2. HEALTH DATA: FROM BIG DATA PERSPECTIVE

One of the most complex challenges faced by healthcare big data analytics is extraction of actionable insights. Big data can unquestionably become one of the core technologies for driving domains like clinical decision support, predictive analytics and population health management. The complexity of big data for healthcare can be broken down using the Multi-V model for big data characterization [7]. The standard multi-V model [8] includes volume, variety and velocity. However, this model has evolved with time to include many other characteristics like value, veracity, variability, vulnerability, validity and volatility, to name a few. This section explores these dimensions of big health data to comprehend how big data can drive clinical, operational and financial initiatives of healthcare organizations. The 5 characteristics for big health data are illustrated in Figure 1.

According to EMC statistics, the total amount of data existing in the world was estimated to be 4.4 zettabytes in 2013 with an average rate of increase expected to be more than double per year [9]. Although, most of this data was composed of streamed videos and audio files, it was projected that around 35\% of this data will be useful and could be used for generating insights. Health data, unlike social media data, falls in the useful category and is typically generated from gene sequences, clinical notes, lab results, insurance claim results, imaging data and medical device data. The digitization initiatives and advancements in transformative technologies has made it possible for systems to extract, identify, store, manage and process health data in variable forms, making 'volume' a significant characteristic of big health data.

Volume-wise healthcare institutions managed 8.41 Petabytes of data in 2018, which increased by a multiplication factor of 9 from 2016, while genomic data is expected to be 25 Petabytes by 2030 [10]. Healthcare data forms a large chunk of total data created in the world everyday with multiple sources of data generation. One of the fundamental characteristics of big data is variety or data complexity spawned when two or more datasets are fused together to gain meaningful insights. Health data is a classical case of complex data with data residing in disparate and heterogeneous locations, kept in formats that are not necessarily compatible and demonstrating inconsistent semantics. With initiatives such as FHIR [11], the support for higher variety data for analysis is improving. Besides this, data generated at medical facilities from ICU and emergency systems is generated and needs to be analyzed in real-time, making response time a critical characteristic for these systems.

Veracity or certainty is a crucial characteristic from the healthcare perspective because the accuracy of the insights is directly proportional to the quality of base dataset, which is extremely grave for life-critical situations in healthcare. However, bias, incompleteness and noise are inherent to healthcare data. Standardization in this domain are dependent on the robustness of information governance. On similar lines, data validity is also a serious concern for clinicians and healthcare researchers. There may be scenarios that the dataset is complete, but its intent, purpose and context may not be clear. This stems from the fact that data may or may not be up to date or have been collected using standard procedures. The extensive use of metadata for data curation is encouraged in healthcare for these reasons.

Healthcare data, owing to its complexity, manifests high variability. Therefore, one of the crucial tasks is to determine which of the data variables affect, correlate or associate to the desired results. This shall have a direct impact on how informative the generated analyses will be for the users. An example of this assertion is the use of social media analysis to determine correlations between parameters such as twitter mentions of air pollution and increased asthma cases in emergency departments. Although, these analyses may provide some interesting results, their value and dependability remains questionable.

Data generated by healthcare systems is heavily volatile as it changes rapidly with change in variables such as seasons, time during the day and other external factors. For instance, an emergency department may witness more accident cases around holiday season than otherwise. Therefore, it makes it imperative for researchers to carefully investigate which of the historical metrics need to be considered for an analysis and how much of archival data would be required to support a statistically valid analysis. These decisions convolute as the size of data increases. Moreover, data storage haunts healthcare providers particularly with the HIPAA regulation that requires them to store patient data of a certain level for a minimum of six years [12]. Therefore, decisions of data that qualifies for archival and deletion are based on their long-term reusability for analyses. For instance, genomic results remain valid for a individual, but pathology test results vary and may become unusable after a length of time.

Data confidentiality has always been a matter of prime concern for healthcare sector, making vulnerability one of the fundamental characteristics of big health data. HIPAA regulations make it imperative for organizations and individuals to pay heed to compliance of security and privacy guidelines [13]. Finally, visualization is also a significant data characteristic for healthcare application in view of the fact that timely action is of prime importance in most scenarios. Therefore, if the visualization systems are not simple and intuitive, their adoption and usability will be questionable.

3. TREATMENT PLANNING THROUGH AI-DRIVEN INFORMATICS

Clinically, comprehensive assessment of an individual's condition generates data, which is further inferred and then a treatment plan is negotiated with the individual. The objective of data analysis is to clearly identify areas that require attention. These areas must be specifically and explicitly managed in the treatment plan. As a rule, every identified area of concern should have corresponding information with it in the form of nature of problem, treatment goals (long and short term) and the management strategy [14]. Moreover, if there are multiple areas that require clinician's attention, then they must be prioritized in the order of their importance. Alternatively, if a problem is identified, but a management plan for the same is not laid out, a justification for the same must be provided in the treatment plan.

Therefore, treatment planning is a collaborative process that requires inputs from the clinician as well as the individual who needs to agree to a plan before it can be executed. With that said, clinician takes control and needs to justify the plan to reduce resistance, if any, from the individual's end. Different conditions require distinctive treatment planning approaches prioritizing patient outcomes over other factors. Different elements [15] of the treatment planning process are illustrated in Figure 2.

For instance, treatment planning for chronic conditions or cancer prioritizes patient health and survivability [15, 16] while treatment planning for Rheumatologic and Connective Tissue Disorders [17] is directed by the physical disabilities in the patient. Other scenarios such as individuals dealing with substance abuse or psychological conditions are centered on rehabilitation or restoring the individual's everyday life to normalcy as much as possible [18].

Treatment planning, in general, can be facilitated by big data in a multi-faceted manner. On the basis of the different big data technologies and their indirect or direct application to treatment planning, three main technological pathways namely medical informatics, imaging informatics and translational bioinformatics, are identified as relevant to this research and illustrate in Figure 3. These aspects have been elaborated upon in this section.

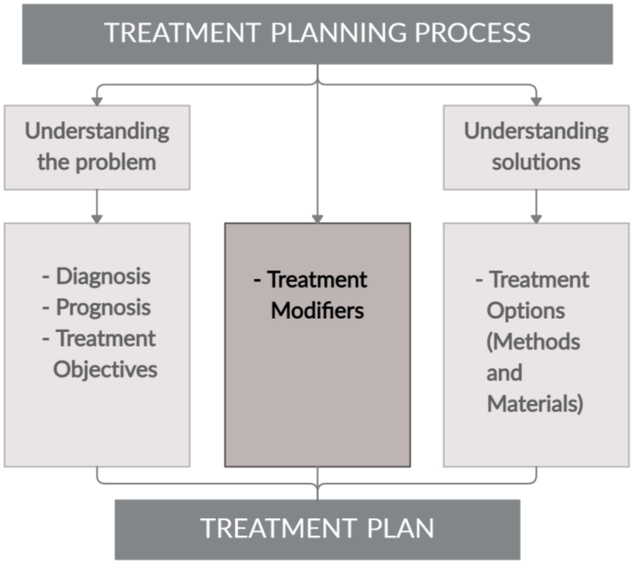

Figure 2. Treatment Planning: Elements and Planning Process

3.1. Medical Informatics

Electronic health records (EHRs) continue to remain one of the primary sources of health data in the digital world. However, this reserve is majorly underutilized when it comes to information mining and knowledge discovery. Health data centres capture humongous amounts of data related to every patient, which typically includes laboratory test reports, diagnostics, ancillary clinical data and medication or treatment plans. In medical informatics, medical images with various characteristics have an important role and used for several purposes such as automated abnormality detection and classification [19-22]. Clinical notes are written in natural language, natural language processing (NLP) plays an instrumental role in comprehending this information using semantic and systematic analysis. Such analyses are used in clinical research for usages such as discovery of phenotype information [23].

In addition, EHR data mining is proven to have significantly important applications in pharmacovigilance [24], disease management [25-26], survival analytics [27, 28], health risk assessment [29, 30], comorbidities evaluation and building systems or models for recommending therapeutics or generally support clinical trials [27, 31, 32]. Most of the considered literature and proposed frameworks required long-term collection of multi-dimensional health data. However, as a limitation of healthcare frameworks, it is usually not possible to collect time series data with such dimensionality and time length. With the advent of technology and increased use of mobile phones (for instance in dermatology to classify lesions [33]) and alternate devices for integrated healthcare applications, several lateral sources of health data have come up. These sources are collectively being used to fill the void of information deficit in healthcare systems and develop comprehensive clinical decision support systems.

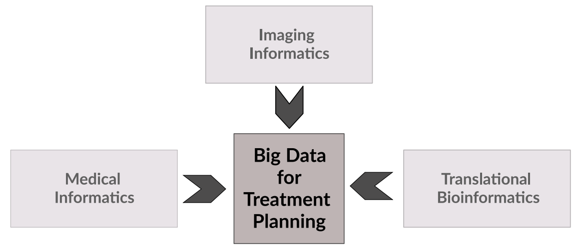

Figure 3. Different Informatics Domains in Treatment Planning

Such systems can be substantially useful in identifying and predicting chronic conditions that start as an acute event, but it is generally not possible to predict their existence solely on the basis of analysis of the event. Therefore, in such cases analysis of hospital and clinical data may be insufficient. In order to create a patient-specific models that can make predictions of that level require integration of clinical data with data coming from non-clinical sources to make a context-aware prediction. For instance, thoracic aortic dissection is a rare condition that starts out as a tear in aorta's intimal tear [base]. From this point onwards, this condition may evolve into a type A or type B dissection. The treatment plans for these two conditions are distinct in the sense type A patients require immediate surgery whereas type B patients require continuous and long term monitoring of blood pressure.

The clinical inferences of whether a patient will develop aortic aneurysms or not, is based on the correlation between aortic dissection and Marfan syndrome (MFS). People suffering from MFS are considered susceptible to aortic aneurysms. Consequently, clinical inferences related to whether aortic aneurysms would evolve into a type A or type B aortic dissection is based on an analysis of flow patterns near the location of the tear. Imaging technologies can be effectively used for providing insights into such scenarios. However, disease progression studies require time series analysis of hemodynamic variations, blood pressure levels and the influence of external factors such as lifestyle on the occurrence or progression of the condition.

3.2. Imaging Informatics

In order to understand biological processes, it is imperative to identify structures and the way they are related to specific functions. This assertion is true not just for organs and tissues, but it also applies to cells and proteins. Taking the example of cardiac abnormalities, modeling scenarios require haemodynamic data, myocardial perfusion data and contractile analyses. Several integrative heart function models have been developed [34-35] that scale up tissue-level analysis to describe organ behavior with some analyses exploring the sub-cellular levels [36] as well.

Modeling of brain structures and neural networks has been of immense research interest recently with extensive studies carried out to explore the function and structure of the human brain. Study of each type of cell and synapse involves examination of many markers [37], resulting in high data volumes. Sources of data for these studies comprise of functional MRI (fMRI) that provides data as microstructural, macrostructural and dense connectivity matrices with improved versions providing time series data associated with multiple volumes [38]. Besides this, non-invasive methods such as functional near-infrared spectroscopy and wearable electroencephalography allows gathering of fine-grained functional data associated with the brain.

In order to make a definite inference, data from multiple modalities such as laboratory results, demographics and medical records needs to be analyzed. Moreover, in order to characterize the function and structure to study progression of diseases, linking metadata with extracted features is the key. This shall involve generation, extraction and segmentation of features and objects [46]. An important consideration is this regard is the development of querying systems that can demonstrate high efficacy [39]. Besides this, as research progresses towards enhancing spatiotemporal constraints of imaging techniques, the computational complexity of these tasks is also correspondingly increasing, challenging capabilities of existing infrastructures.

Table 1. Summary of Challenges

Informatics Domain	Data Sources	Challenges	Possible Future Research
Medical Informatics	Medical data, clinical as well as non-clinical	In order to assess individual level risk, detect condition and initiate strategy for treatment, the number and variability of datasets used is increasing by the day [58].	Development of innovative machine/deep learning techniques for supporting applications/services such as therapeutic decision making and outcome prediction
Imaging Informatics	High resolution images for different modalities complemented with HER/EMR data	(1) Recurring challenge of sharing, storage and analysis without causing security or privacy concerns [59-60]	(1) Use of transfer learning approaches for feature extraction and fine-tuning to widen the scope of the existing deep learning knowledge base used for this purpose.Investigate the use of explainable artificial intelligence for 3D reconstruction and advanced reasoning
		(2) Analysis, indexing and querying of pathological data because of its multi-dimensional and large-scale aspects [61]	(2) Traditional machine learning techniques offer high accuracy, but methods need to be devised for dealing with lack of generalization and smaller datasets particularly in multi-population context for enhanced clinical translation
		(3) Deep learning based computer vision techniques are developed for training on large datasets, which are generally difficult to acquire [62].	(3) Development of analytical approaches for heterogeneous data and its integrated use with imaging data for studying disease progression and improving treatment outcomes for evolving Radiogenomics paradigms
Translational Bioinformatics	Genomics data along with environmental, clinical and socio-economic data	(1) Inconsistency in file formats, vocabulary and file structure from multiple sources of data are used [63].	(1) Efforts are being made for standardization. However, it is still in progress and needs research attention.
		(2) Data analysis may require statistical analysis and/or biological analysis, and there is no standardization [64].	(2) New statistical methods must be developed to compensate for the lack of standard analytical methods for multi-faceted analysis required for translational bioinformatics
			(3) Development of custom network or learning based regressive models for drug response prediction

3.3. Translational Bioinformatics

Transformational bioinformatics is a term that is used to describe the intersection of the fields, clinical informatics, statistical genetics, molecular biology and biostatistics. Research in this field accelerated after mapping of human genome. Amongst the different sub-fields of transformational bioinformatics, the sub-field that is of specific importance to the subject of this work is pharmacogenomics, which deals with the study of variations in how different individuals respond to drugs and its relationships with their genetic composition. Therefore, this field applies to the concept of risk stratification and improved scope of personalization in medicine and treatment plans. These studies have become feasible in the last decade owing to the reducing cost of sequencing per genome [40].

From the data point of view, a single human genome sequenced using next generation sequencing takes 3GB of disk memory, which can go up to 200 GB depending upon the coverage and depth [41]. Therefore, this data is evidently voluminous. However, an important point to consider is that variation in genome data at the individual level occurs in just 0.1\% of the genome, leading to around 3 million variations. Theoretically, it would not be wrong to say that compressed genotyping is feasible. However, its adoption in pragmatic scenarios is limited. This field of study is extensively used in cancer research owing to the fact that drugs have heterogeneous responses even when tested for cancers of the same type [42-43]. Individual-level pharmacogenomics allow demystification of person-specific signaling patterns, which can then allow development of drugs that effectively target these patterns.

NIH's cancer genome atlas project [44] conducted a study on 10,000 individual genomic profiles with 20 different cancer types and found that several sub-types of cancers exist depending upon individual profile characteristics [45]. Moreover, the differences in drug responses are a result of genomic variations [47]. Thus, these studies can be significantly helpful in predicting drug combinations [48, 49] and drug repositioning [50, 51], making them an inevitable inclusion in drug development and allowing healthcare systems to take significant steps towards provisioning precision medicine. Research in pharmacogenomics has evolved to include complex features such as gene expression to investigate factors related to drug targets. However, research on gene expression profiling is still underway. In addition to drug discovery, pharmacogenomics can also be used to study vascular diseases such as acute coronary syndrome (ACS), in which different patients demonstrate different responses to drugs [52].

The association of health informatics applications with statistical sampling issues makes interpretations questionable and adoption debatable in consideration of the criticality of healthcare applications. With that said, the fact remains that study of causes and risks associated with chronic health conditions such as cancer require large-scale multimodal studies, which makes technological intervention useful and inevitable.

4. DISCUSSION

In order to improve the reliability of data-driven studies, data quality becomes a weighty trait. Although, training of machine and deep learning models require datasets of considerable size to achieve desirable accuracy, data quality should never be compromised to ensure that the pragmatic accuracy of these models match their theoretical accuracy. In addition, a noteworthy characteristic of health data is the variations that it exhibits. Therefore, models need to change and be updated at the same pace as the data variations to maintain system reliability.

Big data analytics for healthcare typically requires real-time storage and processing of large volumes of high-dimensional data. This data may be generated from sensor fusion [53] and collection of health attributes or biomarkers [54]. Pervasive health monitoring applications require processing of continuous streams of data, which adds the requirement of high capacity to support low-latency, iterative computations. Several big data frameworks are capable of supporting these requirements with the help of in-memory computation and data caching.

However, the biggest challenge in working with health data is posed by high data dimensionality. Although, feature selection and dimensionality reduction methods have been successfully applied to deal with this issue, the fact that a machine or deep learning model needs to learn a set of parameters, which are determined using optimization, increases the complexity of the process substantially when working with big data. Scalable learning [55, 56] and iterative parameter learning [57] are considered as potential solutions to this problem.

In addition to the above-mentioned, health data requires consideration of several other aspects such as security, privacy, governance and ownership. Considering the fact that the whole process of data collection is based on personal data records aggregation, which is stored at central locations that are virtually spread across multiple data centres and servers. Data security and privacy, particularly for personal data, is governed by the domestic legal framework of most countries. Smart healthcare systems require robust security and privacy frameworks that perform data security at multiple levels, starting with the sensor or data source.

CONCLUSION

Treatment planning is a collaborative process that involves agreement between the healthcare provider and the individual on the short and long term care goals of treatment. Therefore, the efficacy of a treatment plan has a direct impact on patient outcomes, which makes it a critical aspect of the healthcare ecosystem. Transformative healthcare technologies have been seeping into this sector to automate processes and streamline functions, treatment planning being no exception to this rule. However, technical and non-technical challenges exist in development and adoption of technology in healthcare. The data sources, research challenges and possibilities for future research are summarized in Table 1. This paper explores the state of the art in big data-driven treatment planning classifying research efforts on the basis of domain of informatics performed, ranging from medical informatics, imaging informatics and translational bioinformatics. Finally, the paper identifies research challenges to provide insights on the scope of future research in this field.

ETHICAL STATEMENT

No experiments involving humans or animals have been performed for this research work.

CONSENT FOR PUBLICATION

Not applicable

FUNDING

None.

CONFLICT OF INTEREST

The authors declare that there is no conflict of interest.

ACKNOWLEDGEMENTS

Declared none.

REFERENCES

[1] Agrawal R, Prabakaran S. Big data in digital healthcare: lessons learnt and recommendations for general practice. Heredity. 2020 Apr;124(4):525-34.

[2] Bakken S, Koleck TA, Dreisbach C, Hickey KT. Enabling Precision Health Approaches for Symptom Science Through Big Data and Data Science. InGenomics of Pain and Co- Morbid Symptoms 2020 (pp. 239-255). Springer, Cham.

[3] Zhu H. Big data and artificial intelligence modeling for drug discovery. Annual review of pharmacology and toxicology. 2020 Jan 6;60:573-89.

[4] Goceri, E. Future Healthcare: Will Digital Data Lead to Better Care?. New Trends and Issues Proceedings on Advances in Pure and Applied Sciences, 2017, 8:07–11.

[5] Goceri E, Songul C. Mobile health technologies for patients with mental illness. InInternational conference on advanced technologies, Antalya, Turkey 2018 (Vol. 146).

[6] Celesti F, Celesti A, Carnevale L, Galletta A, Campo S, Romano A, Bramanti P, Villari M. Big data analytics in ge- nomics: The point on Deep Learning solutions. In2017 IEEE Symposium on Computers and Communications (ISCC) 2017 Jul 3 (pp. 306-309). IEEE.

[7] Wu C, Buyya R, Ramamohanarao K. Big data analytics= machine learning+ cloud computing. arXiv preprint arXiv:1601.03115. 2016 Jan 13.

[8] Khan S, Liu X, Shakil KA, Alam M. A survey on scholarly data: From big data perspective. Information Processing & Management. 2017 Jul 1;53(4):923-44.

[9] EMC Digital Universe with Research & Analysis by IDC, The Digital Universe of Opportunities: Rich Data and the Increasing Value of the Internet of Things, EMC [Internet]. 2014. Avail- able from: https://www.emc.com/leadership/digital-universe/2014iview/index.htm

[10] Banks MA. Sizing up big data. Nature medicine. 2020 Jan 1;26(1):5-6.

[11] HL7.org, FHIR, HL7 [Internet] 2019. Available from: https://www.hl7.org/fhir/

[12] HHS, Summary of the HIPAA Security Rule, HHS [Internet]. n.d. Available from: https://www.hhs.gov/hipaa/for-professionals/security/laws-regulations/index.html, (n.d.)

[13] camH, Treatment planning, Knowledge ex- change archive [Internet]. n.d. Available from: https://www.porticonetwork.ca/web/knowledgex-archive/amh-specialists/overview-mmt/assessment-treatment-planning-counsellor/treatment-planning

[14] Mondello, S., Treatment planning Demystified, Quintessence Publishing Blog [Internet]. n.d. Available from: https://www.quintpub.net/news/2019/12/treatment-planning-demystified/,

[15] [13] Araujo LH, Horn L, Merritt RE, Shilo K, Xu-Welliver M, Carbone DP. Cancer of the lung: non–small cell lung cancer and small cell lung cancer. InAbeloff's Clinical Oncology 2020 Jan 1 (pp. 1108-1158). Content Repository Only!.

[16] Gomez DR, Li H, Zhang X, Chang JY, Liao Z, Lin SH, Lung cancer, Editor(s): Steven J. Frank X. Ronald Zhu, Pro- ton Therapy, Elsevier, 2021, (205-214).

[17] Little JW, Falace DA, Miller CS, Rhodus NL, Rheumatologic and Connective Tissue Disorders, Editor(s): Little JW, Falace DA, Miller CS, Rhodus NL, Little and Falace's Dental Management of the Medically Compromised Patient (Eighth Edition), Mosby, 2013 (pp. 320-346).

[18] Melchert TP, Treatment Planning, Editor(s): Melchert TP, Foundations of Professional Psychology, Elsevier, 2011. (pp. 147-166).

[19] Goceri E, Songul C. Biomedical information technology: image based computer aided diagnosis systems. InInternational Conference on Advanced Technologies, Antalaya, Turkey 2018 Apr.

[20] Goceri E. CapsNet topology to classify tumours from brain images and comparative evaluation. IET Image Processing. 2020 Apr;14(5):882-9.

[21] Dura E, Domingo J, Göçeri E, Martí-Bonmatí L. A method for liver segmentation in perfusion MR images using probabilistic atlases and viscous reconstruction. Pattern Analysis and Applications. 2018 Nov;21(4):1083-95.

[22] Göçeri E. A comparative evaluation for liver segmentation from spir images and a novel level set method using signed pressure force function (Doctoral dissertation, İzmir Institute of Technology, İzmir).

[23] Friedman C, Shagina L, Lussier Y, Hripcsak G. Auto-mated encoding of clinical documents based on natural lan- guage processing. Journal of the American Medical Infor- matics Association. 2004 Sep 1;11(5):392-402.

[24] Eriksson R, Werge T, Jensen LJ, Brunak S. Dose-specific adverse drug reaction identification in electronic patient records: temporal data mining in an inpatient psychiatric population. Drug safety. 2014 Apr;37(4):237-47.

[25] Sun J, McNaughton CD, Zhang P, Perer A, Gkoulalas-Divanis A, Denny JC, Kirby J, Lasko T, Saip A, Malin BA. Predicting changes in hypertension control using electronic health records from a chronic disease management program. Journal of the American Medical Informatics Association. 2014 Mar 1;21(2):337-44.

[26] Forrest GN, Van Schooneveld TC, Kullar R, Schulz LT, Duong P, Postelnick M. Use of electronic health records and clinical decision support systems for antimi- crobial

stewardship. Clinical Infectious Diseases. 2014 Oct 15;59(suppl_3):S122-33.

[27] Xu H, Aldrich MC, Chen Q, Liu H, Peterson NB, Dai Q, Levy M, Shah A, Han X, Ruan X, Jiang M. Validating drug repurposing signals using electronic health records: a case study of metformin associated with reduced cancer mortality. Journal of the American Medical Informatics Association. 2015 Jan 1;22(1):179-91.

[28] Hagar Y, Albers D, Pivovarov R, Chase H, Dukic V, Elhadad N. Survival analysis with electronic health record data: Experiments with chronic kidney disease. Statistical Analy- sis and Data Mining: The ASA Data Science Journal. 2014 Oct;7(5):385-403.

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

[30] Boland MR, Hripcsak G, Albers DJ, Wei Y, Wilcox AB, Wei J, Li J, Lin S, Breene M, Myers R, Zimmerman J. Dis- covering medical conditions associated with periodontitis using linked electronic health records. Journal of clinical peri-odontology. 2013 May;40(5):474-82.

[31] Cars T, Wettermark B, Malmström RE, Ekeving G, Vikström B, Bergman U, Neovius M, Ringertz B, Gustafs- son LL. Extraction of electronic health record data in a hospital setting: comparison of automatic and semi-automatic methods using anti-TNF therapy as model. Basic & clinical pharmacology & toxicology. 2013 Jun;112(6):392-400.

[32] Marcos M, Maldonado J A, Martínez-Salvador B, Boscá D, Robles M. Interoperability of clinical decision-support systems and electronic health records using archetypes: a case study in clinical trial eligibility. Journal of biomedical informatics. 2013 Aug 1;46(4):676-89.

[33] Goceri E. Diagnosis of skin diseases in the era of deep learning and mobile technology. Computers in Biology and Medicine. 2021 Jul 1;134:104458.

[34] Le Grice IJ, Hunter PJ, Smaill BH. Laminar structure of the heart: a mathematical model. American Journal of Physiology-Heart and Circulatory Physiology. 1997 May 1;272(5):H2466-76.

[35] LeGrice I, Hunter P, Young A, Smaill B. The architecture of the heart: a data–based model. Philosophical Transactions of the Royal Society of London. Series A: Mathematical, Physical and Engineering Sciences. 2001 Jun 15;359(1783):1217-32.

[36] Luo CH, Rudy Y. A dynamic model of the cardiac ventricular action potential. I. Simulations of ionic currents and concentration changes. Circulation research. 1994 Jun;74(6):1071-96.

[37] Kaya B, Goceri E, Becker A, Elder B, Puduvalli V, Winter J, Gurcan M, Otero JJ. Automated fluorescent miscroscopic image analysis of PTBP1 expression in glioma. Plos one. 2017 Mar 10;12(3):e0170991.

[38] Feinberg DA, Moeller S, Smith SM, Auerbach E, Ramanna S, Glasser MF, Miller KL, Ugurbil K, Yacoub E. Multiplexed echo planar imaging for sub-second whole brain FMRI and fast diffusion imaging. PloS one. 2010 Dec 20;5(12):e15710.

[39] Aji A, Wang F, Saltz JH. Towards building a high performance spatial query system for large scale medical imaging data. In Proceedings of the 20th international conference on advances in geographic information systems 2012 Nov 6 (pp. 309-318).

[40] National Institutes of Health, National Human Genome Research Institute (NHGRI) [Internet], 2017, Avail- able from: https://www.nih.gov/about-nih/what-we-do/nih-almanac/national-human-genome-research-institute-nhgri

[41] Andreu-Perez J, Poon CC, Merrifield RD, Wong ST, Yang GZ. Big data for health. IEEE journal of biomedical and health informatics. 2015 Jul 10;19(4):1193-208.

[42] Zlotta AR. Words of wisdom: Re: Genome sequencing identifies a basis for everolimus sensitivity. European urol- ogy. 2013 Sep;64(3):516.

[43] Iyer G, Hanrahan A J, Milowsky MI, Al-Ahmadie H, Scott SN, Janakiraman M, Pirun M, Sander C, Socci ND, Ostrovnaya I, Viale A. Genome sequencing identifies a basis for everolimus sensitivity. Science. 2012 Oct 12;338(6104).

[44] TCGA-web, National Cancer Atlas Program [Inter- net], 2017. Available from: https://tcga-data.nci.nih.gov/tcga/tcgaHome2.jsp

[45] Cancer Genome Atlas Network. Comprehensive molecular portraits of human breast tumours. Nature. 2012 Oct;490(7418):61.

[46] Banday SA, Mir AH. Statistical textural feature and deformable model based brain tumor segmentation and volume estimation. Multimedia Tools and Applications. 2017 Feb 1;76(3):3809-28.

[47] Weinstein JN, Collisson EA, Mills GB, Shaw KR, Ozenberger BA, Ellrott K, Shmulevich I, Sander C, Stuart JM. The cancer genome atlas pan-cancer analysis project. Nature genetics. 2013 Oct;45(10):1113-20.

[48] Huang L, Li F, Sheng J, Xia X, Ma J, Zhan M, Wong ST. DrugComboRanker: drug combination discovery based on target network analysis. Bioinformatics. 2014 Jun 15;30(12):i228-36.

[49] Lee JH, Kim DG, Bae TJ, Rho K, Kim JT, Lee JJ, Jang Y, Kim BC, Park KM, Kim S. CDA: combinatorial drug discov- ery using transcriptional response modules. PloS one. 2012 Aug 8;7(8):e42573.

[50] Dudley JT, Sirota M, Shenoy M, Pai RK, Roedder S, Chiang AP, Morgan AA, Sarwal MM, Pasricha PJ, Butte AJ. Computational repositioning of the anticonvulsant

topira- mate for inflammatory bowel disease. Science translational medicine. 2011 Aug 17;3(96).

[51] Sirota M, Dudley JT, Kim J, Chiang AP, Morgan AA, Sweet-Cordero A, Sage J, Butte AJ. Discovery and preclini- cal validation of drug indications using compendia of public gene expression data. Science translational medicine. 2011 Aug 17;3(96).

[52] Kim KA, Park PW, Hong SJ, Park JY. The effect of CYP2C19 polymorphism on the pharmacokinetics and phar- macodynamics of clopidogrel: a possible mechanism for clopidogrel resistance. Clinical Pharmacology & Therapeu- tics. 2008 Aug;84(2):236-42.

[53] Yang GZ, Andreu-Perez J, Hu X, Thiemjarus S. Multi-sensor fusion. InBody sensor networks 2014 (pp. 301-354). Springer, London.

[54] Hilario M, Kalousis A. Approaches to dimensionality re-duction in proteomic biomarker studies. Briefings in bioin- formatics. 2008 Mar 1;9(2):102-18.

[55] Aggarwal CC, editor. Data streams: models and algorithms. Springer Science & Business Media; 2007 Apr 3.

[56] Andreu J, Angelov P. Real-time human activity recognition from wireless sensors using evolving fuzzy systems. In- International Conference on Fuzzy Systems 2010 Jul 18 (pp. 1-8). IEEE.

[57] Bottou L, Bousquet O. The tradeoffs of large scale learning Advances in Neural Information Processing Systems 20, 2008.

[58] Rose SM, Contrepois K, Moneghetti KJ, Zhou W, Mishra T, Mataraso S, Dagan-Rosenfeld O, Ganz AB, Dunn J, Hornburg D, Rego S. A longitudinal big data approach for preci- sion health. Nature medicine. 2019 May;25(5):792-804.

[59] Tresp V, Overhage JM, Bundschus M, Rabizadeh S, Fasching PA, Yu S. Going digital: a survey on digitalization and large-scale data analytics in healthcare. Proceedings of the IEEE. 2016 Oct 19;104(11):2180-206.

[60] Abouelmehdi K, Beni-Hessane A, Khaloufi H. Big healthcare data: preserving security and privacy. Journal of Big Data. 2018 Dec;5(1):1-8.

[61] Zhang X, Xing F, Su H, Yang L, Zhang S. High-throughput histopathological image analysis via robust cell segmentation and hashing. Medical image analysis. 2015 Dec 1;26(1):306-15.

[62] Panayides AS, Amini A, Filipovic ND, Sharma A, Tsaftaris SA, Young A, Foran D et al. AI in medical imaging informatics: Current challenges and future directions, IEEE Journal of Biomedical and Health Informatics, 2020;24(7): 1837-1857.

[63] Caudle KE, Dunnenberger HM, Freimuth RR, Peterson JF, Burlison JD, Whirl-Carrillo M, Scott SA, Rehm HL, Williams MS, Klein TE, Relling MV. Standardizing terms for clinical pharmacogenetic test results: consensus terms from the Clinical Pharmacogenetics Implementation Consortium (CPIC). Genetics in Medicine. 2017 Feb;19(2):215-23.

[64] Barrot CC, Woillard JB, Picard N. Big data in pharma-cogenomics: current applications, perspectives and pitfalls. Pharmacogenomics. 2019 Jun;20(8):609-20.