Eng. Sci., 2022, **19**, 42–58

Engineered Science

DOI: https://dx.doi.org/10.30919/es8d580

Demystifying the Advancements of Big Data Analytics in Medical Diagnosis: An Overview

Nithesh Naik,[1, 2] Yuvraj Rallapalli,[3] Manamohana Krishna,[4] Anoushka Suresh Vellara,[5] Dasharathraj K Shetty,[6,*] Vathsala Patil,[7] BM Zeeshan Hameed,[2, 8] Rahul Paul,[9] Nirmal Prabhu,[10] Bhavan Prasad Rai,[2, 11] Piotr Chłosta[12] and Bhaskar K Somani[2, 13]

Abstract

The healthcare industry generates a large amount of data, driven by record keeping, patient care, compliance, and regulatory requirements. The digitization of the information is called “Big Data”, which is capable of supporting a wide range of medical and healthcare functions. Big data analytics (BDA) in healthcare is evolving into a promising field for providing insight from very large data sets and has the potential to improve the quality of healthcare delivery with a reduced cost. BDA has a significant impact on healthcare delivery and holds favorable support in a wide range of medical and healthcare applications that includes clinical decision support, disease surveillance, and population health management. BDA has brought in transformation in healthcare and enabled researchers and practitioners with tools to utilize data generated by healthcare systems globally. BDA also aids in preventing adverse events via early detection and diagnosis, leading to safer cost-effective procedures. In the interest of comprehending and analyzing the complex biomedical data now accessible, BDA has become essential for modeling, validating, and interpreting medical diagnosis through the broad spectrum of bioinformatics, medical imaging techniques, and precision medicine.

Keywords: Big data Analytics, Bioinformatics, Healthcare Diagnosis, Medical Imaging Techniques, Precision Medicine.
Received: 27 September 2021; Revised: 21 November 2021; Accepted: 23 November 2021.
Article type: Review article.

1. Introduction

Big data has existed in various forms since the advent of statistical research many centuries ago, however, its utility has evolved in recent years. Characteristics of big data in the 21st century can be defined in terms of the 6V’s: Velocity - the accelerated rate at which data accumulates requires quick real-time processing in clinical decision support, Volume - the sheer magnitude of data that sources such as medical imaging modalities produce, variety - the wide range of data types and sources such as unstructured data seen in medical imaging, variability - the stability that medical data offers over time, veracity - the accuracy, reliability, and quality of medical data from different sources and value-real-time value from clinically relevant data as shown in Fig. 1.[1]

The International Data Corporation (IDC) approximates a projected growth in digital technologies of 175 zettabytes (ZB) in 2025. This number has far exceeded the earlier found estimate of 33 ZB in 2018, capturing a 4-fold growth rate.[2] The growth rate is observed primarily due to the incorporation of artificial intelligence (AI) and deep learning (DL) techniques to analyze this data. However, studies have shown that less than 3% of the potentially useful data is being utilized.[3] BDA involves various techniques such as forecasting trends, optimizing ongoing practices, simulation of events, and predictive analysis, which assists in early detection, diagnosis, and decision-making.[4,5]

The popularity of BDA in healthcare has seen rapid growth in recent times. The underlying repugnant challenges have existed in the healthcare domain which includes constraints, such as privacy concerns about patient medical records, data quality, and medical regulatory policies.[6] Modern medical

[1] *Department of Mechanical and Manufacturing Engineering, Manipal Institute of Technology, Manipal Academy of Higher Education, Manipal, Karnataka 576014, India.*
[2] *iTRUE (International Training and Research in Uro-oncology and Endourology) Group, Manipal, Karnataka 576104, India.*
[3] *Department of Biotechnology, Manipal Institute of Technology, Manipal Academy of Higher Education, Manipal, Karnataka 576014, India.*
[4] *Department of Computer Science Engineering, Manipal Institute of Technology, Manipal Academy of Higher Education, Manipal, Karnataka 576104, India.*
[5] *Department of Biomedical Engineering, Manipal Institute of Technology, Manipal Academy of Higher Education, Manipal, Karnataka 576104, India.*

technology has aided in realizing the possibility of capturing large amounts of patient data over a stipulated time frame, with tools such as Next generation sequencing (NGS) that is about to revolutionize our knowledge of medicine by enabling the possibility of high accuracy low-cost sequencing of whole genomes, proteomes, transcriptomes, which for simplicity can be characterized as Omic data.[7] By 2025, out of the 175 ZB of data postulated, nearly 40 Exabytes are expected to be from Genomic data alone, overshadowing the data expected from most other sources.[8] In addition to NGS, the widespread availability of consumer devices has led to significant developments in clinical data collection, with real-time tracking and analysis of patient vitals. These developments have occurred in tandem with data collection, storage, and analysis becoming much cheaper, leading to considerable strides in disease diagnosis and treatments.[9]

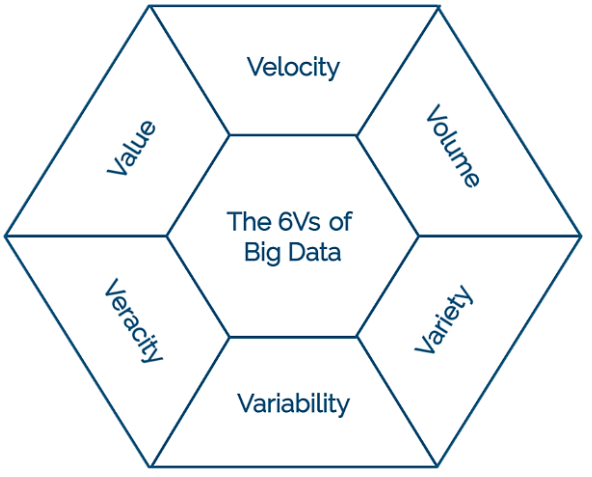


Fig. 1 Big data defined through the 6V's.

Today, medical diagnoses are significantly aided by BDA, with insightful medical data sets propelling continuous and comprehensive research to provide healthcare organizations with the right set of tools to interpret this data.[1] With diseases and medical treatments often being influenced by various factors such as genetic conditions, environmental variables, gender, exercise regimen, and diet, a practitioner can only make his decisions based on a limited subset of these. BDA would offer the opportunity to utilize the entirety of the information available to increase the accuracy of diagnoses and aid in finding additional variables. BDA can also analyze the continuous flow of real-time data at a macro level to search for patterns to predict disease outbreaks,[10] track vitals via consumer wearable devices to predict adverse events[11] before they occur, and enable diagnoses including rare and unforeseen conditions.

This review discusses the advancements in medical diagnoses such as medical imaging analysis, omic data, precision medicine, and the effect of the information gathered from these sources on healthcare and disease diagnosis. Furthermore, the review focuses to elaborate on the effects of big data on the affordability and cost-effectiveness of healthcare, effectively summarised in Fig. 2.

2. Materials and methods

A non-systematic review of all BDA in Medical Diagnoses English language literature published in the last two decades (2000-2020) was conducted using MEDLINE, Scopus, EMBASE, and Google Scholar. Our search strategy involved creating a search string based on a combination of keywords. They are: 'Big Data', 'Big Data Analytics', 'Medical Diagnostics', 'Artificial intelligence in medicine, 'EHR', 'Machine learning in medicine', 'NLP in medicine', 'Precision Medicine', 'Omics Data', 'electronic health records', 'Genomic Data Science', 'EMR,' 'bioinformatics', 'genome', 'Cancer Data analysis', 'Medical imaging', 'Medical image analytics', 'Adverse events in medicine', 'Wearables in healthcare', 'Triage', 'Affordability of healthcare', 'Bias of AI' and 'AI for cancer'. The review includes only the original research work published in English.

Inclusion criteria:
1. Articles on BDA, Medical Imaging, and Precision medicine
2. Full-text original articles on all aspects of healthcare diagnoses, affordability, and accessibility.

Exclusion criteria:
1. Commentaries, articles with no full-text context
2. Animal, laboratory, or cadaveric studies

3. Need for BDA in healthcare

Due to several constraints, clinicians find it difficult to focus on the entire bodily functions of a patient, but rather restrict themselves to providing the patient with restricted and specific treatment. Thus, a seemingly marginal error in medicine can start a cascade of events within a patient which has caused clinicians to have a narrow view of the actual diagnosis.

[6] *Department of Humanities and Management, Manipal Institute of Technology, Manipal Academy of Higher Education, Manipal, Karnataka 576104, India.*
[7] *Department of Oral Medicine and Radiology, Manipal College of Dental Sciences, Manipal, Manipal Academy of Higher Education, Manipal, Karnataka 576104, India.*
[8] *Department of Urology, Father Muller Medical College, Mangalore, Karnataka 575002, India.*
[9] *Department of Radiation Oncology, Massachusetts General Hospital, Harvard Medical School, Boston, MA 02115, USA.*
[10] *School of Computer Science and Engineering, Vellore institute of Technology, Chennai, Tamil Nadu 600127, India.*
[11] *Department of Urology, Freeman Hospital, Newcastle upon Tyne NE7 7DN, UK.*
[12] *Department of Urology, Jagiellonian University in Krakow, Kraków 31-007, Poland.*
[13] *Department of Urology, University Hospital Southampton NHS Trust, Southampton SO16 6YD, UK.*
**Email:* raja.shetty@manipal.edu (D. K. Shetty)

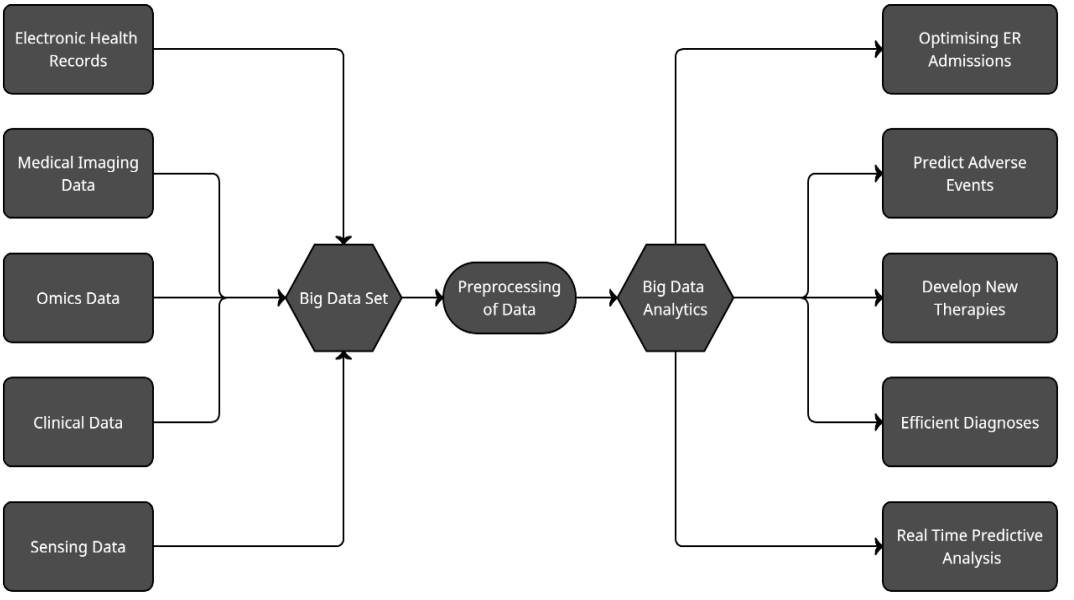


Fig. 2 Summary of the big data analysis Workflow.

The Swiss Cheese Model of Accident Causation proposed by James Reason in 1990[12] illustrated in Fig. 3 shows that each slice of the block represents a single medical process within a patient, but when multiple such processes align suitably, a desirable result can be achieved. Unlike the traditional approaches to medical diagnosis, BDA allows clinicians to analyze a medical predicament rather than parts of the puzzle.

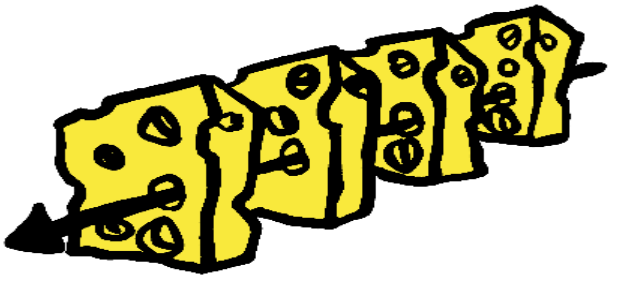


Fig. 3 The Swiss cheese model of accident causation.

For example, early disease detection by using electronic sensors is employed to monitor and target biomarkers using real-time analysis, as data is collected from a patient and processed by the HIPAA (The Health Insurance Portability and Accountability Act) compliant analysis system. Such analytics alert healthcare providers about potential adverse events in individuals suffering from allergic reactions, side effects to medication, and impeding the development of infection.

BDA plays a pivotal role in successfully linking the data with systems across different organizations in healthcare. The need for BDA in healthcare can be simplified into five pathways,[13] which are:

a. Right care

Right care is required to ensure that the patient receives suitable treatment with relevant medical data that can help clinicians oversee the treatment and provide similar objectives to avoid any redundancy in the effort of administering care.[14] BDA can help spot discrepancies in the care given at an early stage, with an element of hindsight, to ensure any data obtained from real-time sensors and wearable devices matches the expected outcomes.

b. Right provider

If equipped with tools like big data, healthcare providers can approach patients holistically by understanding the overall view of a patient from data sources such as their socioeconomic data and health statistics. Thus, ensuring targeted investigations and guaranteeing appropriate treatment.[15] BDA would also promote oversight to ensure that the medical institute is following the required protocols set by their respective government to improve the quality of the treatments offered.

c. Right innovation

Right innovation recognizes that with the exponential rise in new diseases and treatments, medical innovation will continue to evolve as well. At the moment, there have been numerous cases of certain drugs only working well for a small set of patients, with the development of Precision Medicine, BDA can be used to identify which drugs will have a higher probability of being effective on a personal basis.[16] In the case of cancer, it has been found that over 40% of cases require readjustment of dosage or modification of drug throughout a patient's therapy.[17]

d. Right value - Another critical factor for ensuring the quality of healthcare services, medical providers must pay equal attention to a patient's economic status and obtain results

identified by the patient's social insurance system.[18] With the advent of predictive models, diseases can be caught early, leading to lower medical costs.

e. Right living

An overall holistic improvement of a patient's quality of life using tools such as wearable devices allows to connect a patient with the proper diet, exercise, and preventive care by utilizing information mining to enable these users to make better choices for their wellbeing.[19] Patients taking part in clinical trials must submit regular reports and undergo check-ups for their vital parameters, and with the implementation of BDA and technology, a mobile phone might be enough to collect the required data, without any hassle.

4. Medical imaging

The novel discovery of the X-ray by Wilhelm Conrad Roentgen in 1895 revolutionized medicine by offering a visual approach to medical diagnosis.[20] Since then, medical imaging has seen an accelerated advancement, with many imaging techniques still being discovered. The necessary information is found primarily in medical images, which are produced by the interaction of tissues with different forms of radiation[21] and various other principles. Modalities such as magnetic resonance imaging (MRI), positron emission tomography (PET-Scan), nuclear medicine, magnetic resonance spectroscopic imaging (MRSI), and ultrasound offer a window of possibilities for accurate diagnosis, therapy assessment, and planning, which are well-established in a clinical setting. To leverage this increasing patient health data in a more visual and explorative manner, medical imaging plays a vital role for a clinical practitioner for a designed medical task.[22]

Medical imaging data in healthcare is primarily characterized as unstructured data.[23] Hence, streamlining this data through computer-aided algorithms becomes essential. Computational intelligence is what will support and improve diagnostic accuracy in clinical settings.[24] Fig. 4 enables us to understand the prime processes involved in handling the medical diagnostic images in BDA: capture, curate, analyze, visualize, and make decisions.[25] These steps must be aided with the appropriate protocols to meet the decision sciences involved in BDA.

4.1 Types of medical imaging data sets

The types of data obtained vary based on the imaging technique used. We focus on the various prominent medical imaging modalities that assist clinicians today, such as X- Ray radiography, computed tomography (CT), MRI, functional magnetic resonance imaging (fMRI), Ultrasound, and Nuclear Medicine illustrated in Fig. 5.

a. X-ray radiography

Wilhelm Conrad discovered that these X-ray radiations could travel through different materials and could be captured on a photographic plate to form an X-ray image. Not long after this revolutionizing discovery did X-rays find a use for medical purposes. However, X-rays depend on several parameters for the resolution of the image, including the dimension of the body part, the size of the focal spot, and properties used in the fluorescent screen.[26]

b. Computed tomography (CT)

Advancements in radiography have become viable with the development of modern computing. The basic principle of CT captures an object in an X-ray from multiple orientations and then measure the decrease in intensity along with several of the linear paths.[27] As the image is reconstructed from the measured data, the CT images are then visualized using different two-dimensional (2D) and three-dimensional (3D) processing tools. The two standard modes for 3D visualization of CT data are iso-surfacing and volume rendering. Iso-surfacing involves defining a 3D counter to the internal structures, which helps distinguish the image boundaries.[28] Volume rendering involves revealing the internal structure in the image by mapping the CT value to color and opacity.

c. Magnetic resonance imaging (MRI)

To overcome the limitations of X-Ray and CT, which failed to capture joints, ligaments, and tendons, the MRI, a powerful magnet used to generate these medical images, was developed.[29] A radio signal is produced when the electromagnetic field caused by the magnetic resonance imaging (MRI) machine reorients itself to align with the oxygen atoms in the blood, which is then measured by an internal scanner and then produces the image. Different tissues within the body generate a difference in proton signals, which allows images to be distinguished. The MRI truly revolutionized medical imaging as the data is captured at high rates but suffers from low spatial resolution. In addition, movement artifacts cause substantial noise and hinder signal capturing.[30] To overcome this, CHEFNN (Competitive Edge Finding Neural Networks), a 2-layered hop field neural network, was designed to detect the edges in MRI images.[31]

Fig. 4 Medical imaging workflow.

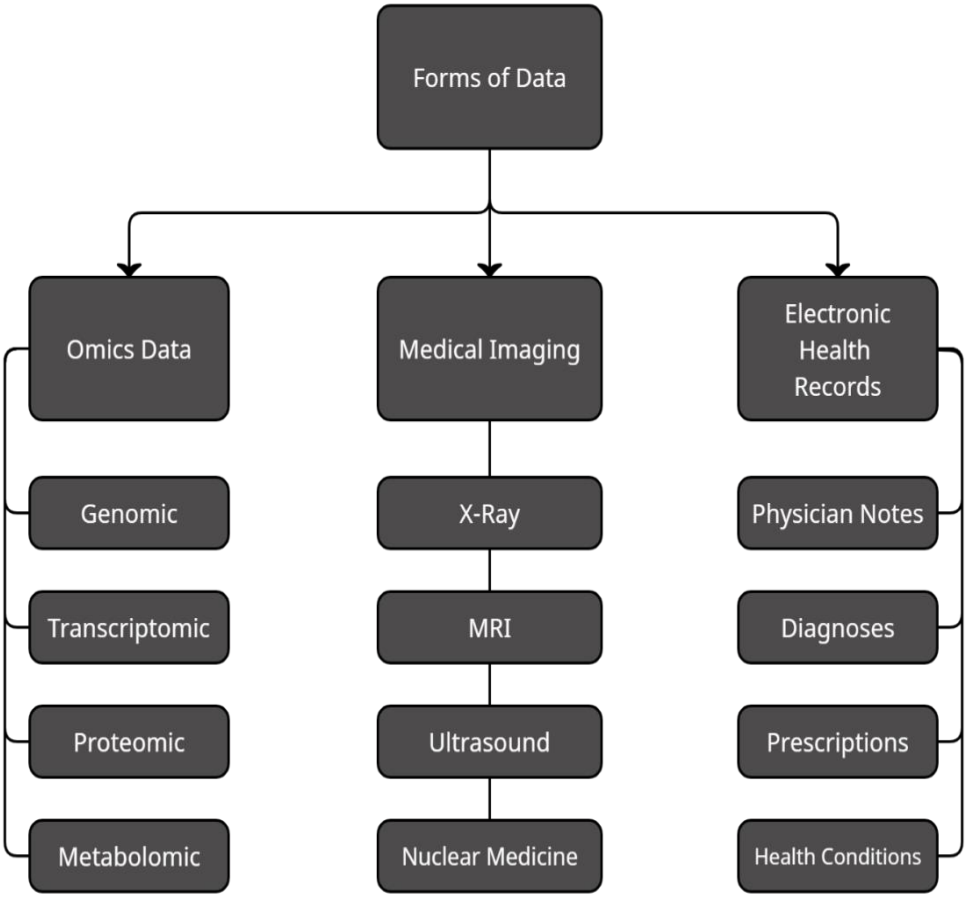


Fig. 5 The various forms of data related to medical imaging and precision medicine.

Functional magnetic resonance imaging (fMRI), a non-invasive imaging modality that measures and locates the increasing fluctuations of live brain activity is captured by the fMRI,[32] which also provides a map of the functional connectivity within the brain. However, noise and signal behavior presented in fMRI poses a limit to the spatial and temporal resolution of the imaging data that could factorize in deriving conclusive findings. Therefore, depending on the experimental requirements, either arterial spin labeling (ASL) or blood oxygen level-dependent (BOLD) signals are used to achieve the required balance between the spatial and temporal resolutions.[33]

d. Ultrasound

A powerful and ubiquitous no-radiation diagnostic and screening tool used by radiologists and clinicians globally, it is one of the most widely used imaging modalities.[34] The sonographic principle uses high-frequency sound waves transmitted to the body, and the transducer obtains the reflected signal to create an image. In addition, 3D images from acquired data from multiple scans are a popular form of examination using parallel scanners or by rotation.[35]

e. Nuclear medicine

Imaging in nuclear medicine involves administering a patient with a radioactive tracer that could be injected or orally given and then observing radiation from different parts of the body. Care is taken so that the radiation exposure of the patient is kept as low as possible. The two most widely used imaging modalities of nuclear medicine are PET scans and single photon emission computed tomography (SPECT). The main difference between both techniques is the use of the radioactive tracer involved in the process. In comparison, PET scans are far superior to SPECT in producing better contrast, resolution, and the data captured through PET scans can be either dynamic or static. The dynamic acquisition of the data by using the radioactive tracker allows clinicians to better understand the long-term behavior of the tissues concerned. It is also important to note that the cost of a PET scan is significantly higher than that of SPECT.[36]

5. Analytical methods

Medical image data sets consist primarily of unstructured data, and handling these data sets becomes a challenge. The need for automatic retrieval of medical images for accelerating predictions and diagnosis of diseases could be addressed by efficient computation. Challenges like storage and qualitative and quantitative data extraction also necessitate changes in traditional image processing methods used for segmentation, extraction, and deionizing medical images. Advancements in

cloud-based computation for lowering computation complexity[37] and methods such as the Hybrid Digital Optical Correlator (HDOC) for processing images at high speeds and computing the correlation by the use of volume holographic memory have also been developed.[38] Another challenge that hinders the coherent interpretation of medical images is the lack of professional healthcare providers globally. To compensate for this hindrance, methodical systems like picture archiving and communication systems (PACS) to access and store medical data conveniently have been developed.[39] PACS, however, focuses on using only structured data for the retrieval of medical images; Hence a lot of vital information about the patient's health status is lost. For this, image analytics in healthcare enables the use of machine learning (ML) and pattern recognition to draw insight from additional sources of unstructured data to enhance clinical decisions. For efficient computation, cloud computing, parallel programming, and various tools and frameworks like Hadoop, mapreduce, hive, yet another resource negotiator (YARN), apache-spark are used to carry out a vast amount of medical analytics at a time and have been discussed in this section.

5.1 HADOOP

The most popular framework used in medical imaging today is hadoop; it processes and stores vast amounts of different kinds of data efficiently by enabling a cluster of modalities, such as hardware.[40] The Hadoop framework consists of three specific components designed for big data.[41]

a. HADOOP distributed file system (HDFS)

acts as a storage unit that ensures that data is capitalized and stored in blocks. This allows vast amounts of data to be copied across multiple systems and ensures fault tolerance.[42]

b. Mapreduce

Hadoop primarily uses the programming tool MapReduce for implementing the data process. MapReduce follows a simple architecture of two subtasks, namely, mapping the task and reducing the task at hand. This technology has also been useful in speeding up the optimal support vector machine (SVM) parameters found in the case of the texture of the images, in analyzing the lungs, and in indexing medical images based on content and texture using wavelet analysis. mapreduce facilitates the process of data separately in Hadoop and aggregates the result to give the final output.[43] To warrant fault tolerance, the initial execution of the mapping task creates three copies of each block of the input. These blocks are then split, after which the mapper phase generates intermediate values to shuffle and sort the output locally. The final output obtained in the reduced phase is triplicated to ensure fault tolerance and is then stored in the HDFS. Although studies have shown techniques using mapreduce to be substandard as compared to the traditional approaches like parallel database management system (DBMS),[44] the MapReduce architecture in Hadoop achieves great scalability with the trade-off being significantly low efficiency that parallel DBMS has the edge over.

c. Yet another resource negotiator (YARN)

capacitates the need for independence between the resource management and programming models used for the analysis of big data for organizations. YARN as a resource negotiator also allows efficient allocation of resources required to process and manage clusters of data across Hadoop like Apache spark, dryad,[45] *etc*. The main components of YARN architecture consist of a resource manager (RM), node manager (NM), application manager (AM), and Container.

5.2 Apache SPARK

To overcome these challenges faced in the Hadoop framework, like its dependency on hardware, another popular open source was formulated with the development of apache spark in Berkeley's AMP Lab at the University of California, Berkeley. The unique feature of spark is its ability to support several programming languages like Python, Java, R, and Scala, by which the integration of AI and data has now become a possibility.[46] Hadoop cannot support real-time analysis, but the distributed cluster computing tools spark permits significantly faster memory processing as compared to Hadoop's disk-based storage, thus enabling more efficient computation of heterogeneous data like that of medical images in CT scans and the X-ray modality used in mammography for the detection of breast cancer.[47] The architecture of spark follows the master-slave architecture, and hence the data is stored as resilient distributed datasets (RDDs). The driver is responsible for all communication with the master/cluster manager, which in turn communicates with the slave/workers for executing the running processes or tasks which are distributed. spark, in turn, allows batch, and graph processing and reads the data in real-time from the clusters. The spark architecture is highly advantageous as it permits the acquisition of large-sized medical images from different sources, enables data analysis in real time, and implements the extraction of image features by the filtering method.

6. Precision medicine

Precision medicine is a model for healthcare, which heavily depends on BDA and beyond, with the national research council defining it as "The tailoring of medical treatment to the individual characteristics of each patient to classify individuals into subpopulations that differ in their susceptibility to a particular disease or their response to a specific treatment. Preventative or therapeutic interventions can then be concentrated on those who will benefit, sparing expense and side effects for those who will not".[48] The strengths of precision medicine lie in the ability to guide healthcare decisions toward the most effective and efficient treatment for a given patient and thus reduce the burden of excess diagnostic testing and therapies.[49] Precision medicine

requires a wide range of data, ranging from collection and management (Storage of data, privacy, *etc*.) to analytics (data mining, visualization, and integration).[50] The most significant factors in precision medicine are health records and omics data due to the high amount of data available and the ability to detect biomarkers.

6.1 Types of data sets in precision medicine

There are two major forms of data for the implementation of Precision Medicine, electronic health records and Omics data. Most of the data obtained from a medical source and be categorized as either of the following datasets.

a. Electronic health records (EHR) and their vitality

Over the past few years, medical records have migrated from paper-based to digital, electronic health records. As shown in Fig. 5, EHR typically consists of physician notes, problem lists, diagnoses, medication prescribed, other health conditions, and clinical decision support.[51] The data in concept appears to be a perfect foundation for big data analyses. Still, practically EHR data suffers from a lack of quality control - inconsistencies in format, variability in recording, and missing data. Data such as height/weight, gender, and clinical tests are typically universally standard and structured well, but they do not cover the clinical diagnosis. Other unstructured data cannot be entered into discretely defined fields but rather require specialized fields like radiology reports, case-specific physician comments, and other reports, which cannot be restricted to common criteria. This unstructured data is a major hurdle for big data, and there have been attempts to use natural language processing (NLP) for understanding and categorizing this data, but it has not been up to the level of a human.[52]

In addition, the data generated continuously from physiological signal monitoring medical devices have failed to store the patient data for prolonged periods, hence the inadequacy to capture what this extensive data could entail.[53] To overcome this redundancy, the dynamic data must be integrated with the static data from the EHR. The bridging of the two systems becomes a requisite to provide a more robust path to analyze the contextual and situational awareness for an analytical engine to identify.[54] The primary advantage of EHR is the immediate availability of a patient's medical history to a practitioner. This would, in general, increase the efficiency and accuracy of diagnoses on its own and has led to reduced allergic reactions to drugs and proper dosage prescriptions.[3] Secondarily, this data combined with Omic data will help precision models be made for highly accurate diagnoses. However, a major limitation of EHR is its restricted availability to developed countries, leading to a significant bias in the predicted results. In recent years, several countries have taken steps to incorporate EHR into their healthcare system.[55]

b. Omics data

Omic data consists of sequencing data from various molecular profiles such as genomic, transcriptomic, and proteomic data. This data is produced by various biochemical assays that measure and sequence all the molecules of a certain type from a biological sample. Genomic data consists of the various DNA sequences present and is usually invariant over time. The only possible changes are single nucleotide polymorphisms (SNPs), copy number variations, and frameshift mutations.[56,57] Transcriptomic data consists of the RNA sequences, which are primarily the exonic regions from the genome and help analyze gene expression and study splicing.[58,59] Proteomic data contains the protein sequences translated from the transcriptome and gives valuable data on protein expression, post-translational modifications, and interaction between proteins.[60] While these are the most common forms of omic data, epigenomic data that documents protein-DNA interactions, and DNA modification patterns;[61] and metabolomic data that gives information about the expression of metabolites[62] are also vital constituents of it. In recent years, the costs of sequencing have drastically fallen, and thereby enormous amounts of data are being produced every day. This vast database provides the basis for many treatments, with precision medicine being the most popular.[63,64]

6.2 BDA for precision medicine

Omics data and EHR are multi-dimensional, which requires high computational times and may even lead to low accuracy in results. Thus, a prerequisite step is the reduction of the dimensionality of the data and this is achieved by identifying a subset of latent factors that preserve as much of the original information as possible.[64] There are two strategies that are typically used, feature selection to identify an optimal subset of the existing data and feature extraction, which transforms the given data into a compact set of dimensions.[65]

With Omics data being produced by high throughput assays like NGS and MS, different pre-processing methods are required before the data can be used. The most common step for data from NGS is to align the sequences obtained with a reference genome using different tools while ensuring the quality of the sequence obtained is up to a suitable standard.[66] Genomic data is further processed by tools such as GATK,[57] Samtools[67] for the detection of genomic variants using per base differences between the reads and reference genome. Similarly, transcriptomic data is further processed with a focus on alternative splicing detection using de novo transcriptome assembly,[68,69] expression profiling which associates the mapped reads from the previous step with genes and isoforms,[70] and fusion gene detection.[58,59] Epigenomic data is used to identify patterns of protein-DNA binding, DNA methylation, and histone modifications, this information is built into a profile representing the density of reads, and background noises and determines statistically significant peaks.[61,66] For proteomic and metabolomic studies, MS is used with its pre-processing steps being alignment, baseline correction, and peak detection.[71] In the case of EHR data, due

to its unstructured nature, it requires some pre-processing where missing data is interpolated,[72] noise has to be filtered, and data quality must be increased by using complex waveforms to fill in gaps and correct any artifacts that are present, along with sensor fusion techniques to obtain a more reliable measure.[73,74]

Biomarker identification is a significant part of precision medicine, with samples generally taken from multiple biological conditions (healthy vs. disease) or different time frames. (Before and after treatments). Several tools such as SNP assoc,[75] edgeR, DBCh IP,[76] and Detect TLC[77] are used to identify statistically significant variations. Such as differences in gene expression,[78] alternative splicing,[79] protein-DNA binding, histone modifications[80], and DNA methylation.[81] The variants with the highest significance are selected as biomarkers.[64]

Data mining is a crucial tool for extracting data from the pre-processed EHR, with two primary strategies being followed, the first of which is static endpoint prediction, where the relationship between clinical features such as patient condition and targeted clinical endpoints are modeled. This primarily involves building statistical models using tools such as regression analysis and association rule learning, the latter of which discovers reliable associations between clinical variables which occur at a high frequency.[82] This strategy's ML techniques are more suited to large datasets. Temporal data mining can be used to model the temporal relationship between diagnosis, treatment, and outcome, chronologically from the EHR data using techniques such as the hidden Markov Model[83] and the conditional random field.[84] The temporal association can be simplified as an antecedent followed by an outcome with a given time difference. The major constraint of this strategy is the requirement of predefined clinical variables and outcomes, which are difficult to generalize for a given treatment.[64] The applications of the analyzed data from biomarker identification and data mining of EHR will be discussed in the following sections.

7. Challenges of BDA

BDA has the ability to radicalize global health for positive impact but faces legitimate hurdles such as data standardization, data security, data structure, data governance as a managerial skill, and storage and transfer of medical data.[85-87] Integration of databases and standardization of protocols followed in laboratories and medical organizations still pose a considerable challenge to BDA.[85]

EHRs are also influenced by misunderstanding or incorrect interpretation of data due to clinician or clerical error. Studies have also shown that the uploading of EHR data at the end of each day by clinicians leads to burnout. As EHR fields increase in complexity, the effort required to input data also increases, leading to a significant amount of time being spent entering the data rather than treating patients.[86] From a physician's perspective, it is a burden as they would be entering a huge amount of data every day, with little or no useful information that will be of immediate requirement or clinical benefit.[87] Despite the continuous development of new techniques to store and analyze big data, there are several challenges before BDA can be implemented in the healthcare sector globally. These range from issues with the fragmentation of data due to varying formats used for EHR in practice and its storage, to inherent bias being developed due to having biased data. Although the integration of BDA in healthcare has not been as rampant as seen in industries like marketing and finance, it has however picked up pace in recent years, such as with the evolution of wearable healthcare devices. In this section, we will address the challenges prevalent in BDA like fragmentation, retrieval, bias, regulatory compliance, and privacy and security issues seen with medical data.

7.1 Regulatory compliance and privacy

Regulatory compliance issues regarding the collection of healthcare data have also impeded the overall growth of BDA in medical diagnoses. The issues related to regulatory compliance have risen due to the legalities of the collection and management of processing high volumes of sensitive data. In healthcare organizations, the information systems handling such data must comply with all protocols and norms specific to healthcare.[88] Federal laws and regulations in the United States of America (USA) are much more stringent with its legal framework for healthcare information as compared to countries like India. The federal laws practiced in the USA include the Health Information Portability and Accountability Act of 1996 (HIPAA) which aims to govern the protected health information (PHI) of a patient's care, cost of care, and overall health condition.[89] Authorization to such patients' health data is restricted only to covered entities that fall within the legal framework, such as healthcare providers and patient insurance companies. However, in countries like India, they do not have dedicated data protection laws regarding healthcare but provision under the Information Technology Act 2000 and Information Technology rules 2011 dictate that healthcare data is included within the framework of personal and sensitive data. Failure to comply with these laws in India can attract heavy fines of up to 500,000 Rupees as well as criminal imprisonment of up to 3 years,[90] which poses a significant challenge to BDA for medical diagnosis.

Ideally, the big dataset used for analytics will involve not only medical data as discussed but also data about the patient's day-to-day life such as social data, social media information, and data from cell phones. But for this to become a reality, it would require the adoption of robust standards and require new ways to preserve privacy and prevent misuse.[91,92] Several studies have shown that apprehensions about the safety of data online have led to the slowdown of the adoption of EHR.[93] While the security of data present in the databases has improved, it is yet to strike complete confidence, but with the incorporation of encryption techniques and other measures, the challenge is being overcome.[94]

7.2 Fragmentation of Data

EHRs have started gaining mainstream use in recent years, with most hospitals primarily designing them to assist with their workflow and billing systems. Hence, they have not been optimized to accurately document the reasoning behind clinical decisions as they usually involve building an evolving narrative that cannot be reduced into a series of options on a form.[95] As a consequence, it cannot provide the well-structured data required for facilitating clinical research. With each organization following its format of documenting EHR, there is little uniformity in the systems used today, leading to fragmentation of data.[96] This leads to several issues when analyzing the data as the limited availability of data on major diseases like cancer, where critical data may not be consistently recorded across all organizations. There have been instances where an individual would get diagnosed with different ailments at different hospitals, which would lead to two different EHR databases having varying information about the same individual. This mismatch of data available can lead to confounding and selection bias.[97]

7.3 Bias

Bias can be defined as a tendency to give results favoring or avoiding one group of the population over its entirety. This can arise due to the unavailability of data from a major chunk of the population, with a majority of the countries around the world yet to fully adopt EHRs or incomplete records where complete information may not have been given either due to human error or lack of documentation. This bias may lead to misdiagnosed health issues using predictive models or differences between BDA outcomes.[98,99] The quality and applicability of the model are only as good as the data available. What holds for the majority of the population of the USA might not be true for the much higher population of India, which leads to a divide in the research being conducted, as any outcome from the BDA would not apply to all of humanity but only a part of it. In addition to this, events such as undocumented medical conditions, varying requirements in different countries, and hesitancy to document an uncertain diagnosis can all be sources of bias in the data.[100] The disproportionate documentation of medical encounters over the control group can also lead to information bias.[101] The research community also adds to the prevalent bias in further developing BDA for medical diagnoses. Leading journals tend to focus only on statistically significant data leading to publication bias, in doing so the leading publications limit research that may be present in lesser-known journals which might not be indexed in the Cumulative Index of Nursing and Allied Health Literature (CINAHL) or PubMed.

7.4 Retrieval of visual medical images

Medical content-based image retrieval (CBIR) also known as query by image content (QBIC) is a state-of-the-art technique for the retrieval of medical images used in modern healthcare. The vast size of the medical image repositories uses technology such as CBIR to exploit the visual content in medical data, making it easier to manage the large size of data now available for clinical research, medicine, and education by retrieving the most visually similar images for a given query in a medical database.[102] CBIR is applied in the retrieval of multimodality images, 2D images, and multiple dimension images specifically in medicine.[103] Retrieval of images through such techniques opens doors for applications in BDA for medical diagnoses.[102] Although the biomedical research community has actively engaged in advancing CBIR technology, several challenges have prevented it from being incorporated into solving practical medical problems. These challenges include:

7.4.1 Advancements in medical imaging

Medical imaging is rapidly evolving, giving rise to new modalities such as clinicians studying the radiology of patients through Augmented Reality (AR) and Virtual Reality (VR) technology and the improvements of CT and MRI scanners for clearer 3D images with high resolution. Such advancements in imaging modalities need the development of new algorithms for retrieval of the medical images, and these too must be adapted for a particular image modality setting. It also highlights that the databases within healthcare organizations cannot remain static but must integrate with the current setting, which is rarely seen.

7.4.2 Data Acquisition for high-quality images

A predominant limitation in medical CBIR is that while retrieving images, much of the original information is left out. The availability of ground truth in medical data reveals subtle connections for diagnosis, therefore this data must be acquired.

7.4.3 Multimodal data integration

In a clinical setting, a doctor would not require one image but rather an entire case profile to carefully assess the similarity between medical cases. Hence, all information including medical images, free text, and structured data must be accounted for while combining all information to find similar cases. However, CBIR limits the tackling of this problem, which has a great impact in a clinical setting.

8. Applications of BDA

The possibilities of the methods and techniques of BDA in healthcare applications are endless. However, in this review few prominent applications of patient diagnosis, treatment and care are discussed.

8.1 Triage

Triage is the first step once a patient is brought to a hospital; in essence, it is the determination of degrees of the urgency of wounds and illnesses. BDA can be used for triage by analyzing the level of danger of the patient based on previous data from EHRs. This would aid in quick reaction and identification of possible problems before they are apparent to the hospital staff.

Machine Learning tools have been used to predict the occurrence of sepsis and thereby saving lives in the ER.[104] In rural areas, where specialists such as ophthalmologists are not available, AI methods have been used to accurately assess the severity of the issue and identify the health condition.[105] While most works focused on chronic and serious illnesses, even in the case of outpatient departments of hospitals, BDA was found to improve efficiency and reduce the number of working staff required for the diagnosis of minor illnesses.[106]

8.2 Healthcare affordability

By exploiting the vast influx of information now available in healthcare, organizations globally can radically shift the face of healthcare delivery. Clinicians can now target medical complications more effectively by using data and rigor rather than the traditional approach of intuition. It has been suggested that BDA allows predictive modeling in medicine to ensure that costs on data-intensive tasks can be reduced by elements of computing such as processing and storage of data.[107,108] Cost reduction and savings will be experienced across the medical spectrum through effective treatments, such as with the development of big data computing technologies like telecardiology that monitors the patient's cardiac responses through telecommunication. Such advancements also curb the need for re-hospitalization by identifying emergency cases in real-time, permitting interoperability among hospitals, and allowing accessibility of data.[11]

8.3 Cancer diagnosis and treatment

The information we get from precision medicine not only helps with identifying the probability of tumor growth but also in identifying the most optimum treatment. The identification of proto-oncogenes and prediction of new carcinogenic genes help predict and identify tumors before they reach a major stage.[109] Biomarker detection, where BDA is used to detect proteins that are inherent only to tumors, is a valuable asset, but these do require clinical validation.[110] As of March 2021, there have been 189 biomarkers linked to anti-cancer drugs with pharmacogenomic information present in the labeling.[111] Data from the genomic analysis can then be utilized for selecting the most suitable drug for the patient. Precision medicine techniques have also been used for identifying suitable targets for the delivery of noncoding.[112]

In the case of prostate cancer, precision medicine tools such as Oncotype DX by Genomic Health Inc. can predict the chances of recurrence and mortality of patients after surgery.[113,114] However, these are yet to be incorporated into treatment planning for cancer. The IBM Watson for oncology support system has been shown to be considered reliable and accurate in algorithm-based decision-making for treatment recommendations for lung cancer,[115], breast cancer,[116] and gastric cancer.[117]

8.4 Adverse Events

Another major application of BDA is the prediction of patients that are at risk of having adverse medical events.[118] These events usually lead to treatments that are expensive and can cause significant morbidity and mortality, which can easily be prevented.[119] These events might be relatively low frequency but could have a significant clinical impact, several Neural network-based models have been developed for the diagnosis of such chronic disease diagnoses and prediction of patients' future disease, based on the pre-existing EHR data.[120,121]

Renal failure is particularly expensive and has a high risk of mortality.[122] But consequently, renal function is fairly simple to measure, detection of changes before major decompensation occurs can be detected using BDA, with models predicting renal failure up to 48 hours before occurrence.[123] The combination of datasets from EHR such as measures of kidney function, blood pressure, exposure to a specific medication, and urine output aid in the predictive analyses. Adverse drug events also pose a significant cost as they can frequently occur unless controlled. These can be events such as side effects of a drug, or human error in dosage.[124,125]

Chronic conditions such as arthritis, scleroderma, and lupus are autoimmune diseases that may span across more than one organ system, and it is difficult to project the trajectory of the disease.[92] This can lead to multiple organ dysfunction syndromes (MODS), which is an extraordinarily complex, heterogeneous syndrome with multiple combinations of ailments. This has led to the limited development of diagnostic models for it.[126] There have been recent strides taken to develop novel models for MODS, with the promise being shown for the future.[127]

8.5 Accessibility of high-quality and structured data

BDA has also been instrumental in converting raw and unstructured data into meaningful information[128] that can benefit a clinical setting through various open-source technology like Hadoop or apache spark as previously discussed and various other frameworks like that of mobile cloud computing environments. Advanced frameworks such as the Healthcare Information System (HIS) provide a high level of access ability and integration of medical data, which can be in the form of EHRs or different healthcare organizations where the information is stored in cloud storage areas and can be accessed by the internet which also provides useful insight to medical practitioners at crucial times for critical decision making.[129] BDA also ensures that effective medical data can be reused to form new valuable knowledge, thus initiating a process of recycling medical data to keep up with current medical trends. Data quality through BDA can also be maintained by sophisticated analytics that can streamline useful medical data and eliminate unnecessary information.[130,131]

9. Conclusion

BDA for medical diagnosis has taken large strides in recent years, with novel therapies and models being developed with

the aid of medical imaging, EHR, and Omics data. Computing frameworks like Hadoop and apache spark have also been discussed to analyze the vast volumes of unstructured medical data more efficiently. For handling medical imaging in BDA, challenges such as scalability need to be addressed with more research on automated analytical models to help better predict treatment plans and detect diseases. The measures being taken to reduce physician burden by incorporating AI techniques in data entry and reducing the time and effort required to upload EHR data is a significant morale boost. This would lead to better all-round acceptance of EHR, and with more and more countries working to implement their forms of it. BDA tools like precision medicine are bound to improve the healthcare professional's diagnostic ability.

10. Future perspectives

The continuous improvement of BDA in healthcare is indicative of what we may expect in the future. The COVID-19 pandemic put the onus on the need for more efficient healthcare systems and the need for affordable healthcare. A majority of the models developed so far are yet to be incorporated into regular clinical practice, after which pre-existing models can be improved and valuable data can be collected about the accuracy and reliability for better predictive analysis in the future. With fields such as NLP, ML and NGS constantly improving, the collection and processing of the data from EHR will be made more reliable, along with the accuracy of the sequencing and other biological data collection leading to better predictions for a wider range of ailments and medical conditions using BDA assisted tools for helping patient diagnosis.

Conflict of interest

There are no conflicts to declare.

Supporting information

Not applicable.

References

[1] J. Andreu-Perez, C. C. Y. Poon, R. D. Merrifield, S. T. C. Wong, G.-Z. Yang, *IEEE Journal of Biomedical and Health Informatics*, 2015, **19**, 1193-1208, doi: 10.1109/JBHI.2015.2450362.

[2] D. Reinsel, J. Rydning, J. F. Gantz, Worldwide global datasphere forecast, 2021–2025: The world keeps creating more data—now, what do we do with it all, *IDC Corporate USA*, 2021.

[3] S. Dash, S. K. Shakyawar, M. Sharma, S. Kaushik, *Journal of Big Data*, 2019, **6**, 54, doi: 10.1186/s40537-019-0217-0.

[4] M. Doumpos, C. Zopounidis, *Omega*, 2016, **59**, 1-3, doi: 10.1016/j.omega.2015.06.006.

[5] Y. Duan, J. S. Edwards, Y. K. Dwivedi, *International Journal of Information Management*, 2019, **48**, 63-71, doi: 10.1016/j.ijinfomgt.2019.01.021.

[6] M. Adibuzzaman, P. DeLaurentis, J. Hill, B. D. Benneyworth, *AMIA Annual Symposium Proceedings*, 2018, **2017**, 384-392.

[7] J. Shendure, H. Ji, *Nature Biotechnology*, 2008, **26**, 1135-1145, doi: 10.1038/nbt1486.

[8] Z. D. Stephens, S. Y. Lee, F. Faghri, R. H. Campbell, C. Zhai, M. J. Efron, R. Iyer, M. C. Schatz, S. Sinha, G. E. Robinson, *PLoS Biology*, 2015, **13**, e1002195, doi: 10.1371/journal.pbio.1002195.

[9] R. Agrawal, S. Prabakaran, *Heredity*, 2020, **124**, 525-534, doi: 10.1038/s41437-020-0303-2.

[10] K. Priyanka, N. Kulennavar, *International Journal of Computer Science and Information Technology*, 2014, **5**, 5865-5868.

[11] J. C. Hsieh, A. H. Li, C. C. Yang, *International Journal of Environmental Research and Public Health*, 2013, **10**, 6131-6153, doi: 10.3390/ijerph10116131.

[12] J. Reason, *Western Journal of Medicine*, 2000, **172**, 393-396. doi: 10.1136/ewjm.172.6.393.

[13] S. Kumar, M. Singh, *Big Data Mining and Analytics*, 2019, **2**, 48-57, doi: 10.26599/bdma.2018.9020031.

[14] Ishwarappa, J. Anuradha, *Procedia Computer Science*, 2015, **48**, 319-324, doi: 10.1016/j.procs.2015.04.188.

[15] J. Sun, C. K. Reddy, *Proceedings of the 19th ACM SIGKDD international conference on Knowledge discovery and data mining*, Chicago, Illinois, USA. New York: ACM, 2013, **1525**, doi: 10.1145/2487575.2506178.

[16] R. W. Peck, *Annual Review of Pharmacology and Toxicology*, 2018, **58**, 105-122, doi: 10.1146/annurev-pharmtox-010617-052446.

[17] R. J. Mody, Y. M. Wu, R. J. Lonigro, X. Cao, S. Roychowdhury, P. Vats, K. M. Frank, J. R. Prensner, I. Asangani, N. Palanisamy, J. R. Dillman, R. M. Rabah, L. P. Kunju, J. Everett, V. M. Raymond, Y. Ning, F. Su, R. Wang, E. M. Stoffel, J. W. Innis, J. S. Roberts, P. L. Robertson, G. Yanik, A. Chamdin, J. A. Connelly, S. Choi, A. C. Harris, C. Kitko, R. J. Rao, J. E. Levine, V. P. Castle, R. J. Hutchinson, M. Talpaz, D. R. Robinson, A. M. Chinnaiyan, *Journal of the American Medical Association*, 2015, **314**, 913-925, doi: 10.1001/jama.2015.10080.

[18] M. Viceconti, P. Hunter, R. Hose, *IEEE Journal of Biomedical and Health Informatics*, 2015, **19**, 1209-1215, doi: 10.1109/jbhi.2015.2406883.

[19] S. Hussain, B. H. Kang, S. Lee, *Springer International Publishing*, 2014, 236-242, doi: 10.1007/978-3-319-13102-3_39.

[20] G. W. C. Kaye, *Nature*, 1934, **133**, 511-513, doi: 10.1038/133511a0.

[21] A. Meyer-Baese, V. Schmid, Pattern recognition and signal analysis in medical imaging, Elsevier, 2014.

[22] F. Ritter, T. Boskamp, A. Homeyer, H. Laue, M. Schwier, F. Link, H. O. Peitgen, *IEEE Pulse*, 2011, **2**, 60-70, doi: 10.1109/mpul.2011.942929.

[23] R. Raja, I. Mukherjee, B. K. Sarkar, *Scientific Programming*, 2020, 1-15, doi: 10.1155/2020/5471849.

[24] A. Belle, R. Thiagarajan, S. M. R. Soroushmehr, F. Navidi, D. A. Beard, K. Najarian, *BioMed Research International*, 2015, **2015**, 370194, doi: 10.1155/2015/370194.

[25] H. Wang, Z. Xu, H. Fujita, S. Liu, *Information Sciences*, 2016, **367-368**, 747-765, doi: 10.1016/j.ins.2016.07.007.

[26] P. Suetens, Fundamentals of Medical Imaging, *Cambridge*

University Press, 2009, 159-189.
[27] A. Momose, T. Takeda, Y. Itai, K. Hirano, *Nature Medicine*, 1996, **2**, 473-475, doi: 10.1038/nm0496-473.
[28] C. I. Lee, A. H. Haims, E. P. Monico, J. A. Brink, H. P. Forman, *Radiology*, 2004, **231**, 393-398, doi: 10.1148/radiol.2312030767.
[29] S. Ogawa, T. M. Lee, A. R. Kay, D. W. Tank, *Proceedings of the National Academy of Sciences*, 1990, **87**, 9868-9872, doi: 10.1073%2Fpnas.87.24.9868.
[30] P. Jezzard, S. Clare, *Human Brain Mapping*, 1999, **8**, 80-85, doi: 10.1002/(sici)1097-0193(1999)8
[31] C. Y. Chang, P. C. Chung, *Optical Engineering*, 2000, **39**, 695-703, doi: 10.1109/TSMCC.2009.2013817.
[32] A. Ehtemami, Statistical data analysis of resting state fMRI: a study of nicotine addiction treatment, *The Florida State University ProQuest Dissertations Publishing*, 2016, 102423333.
[33] S. B. Stewart, J. M. Koller, M. C. Campbell, K. J. Black, *Peer J*, 2014, **2**, e687, doi: 10.7717/peerj.687.
[34] S. Liu, Y. Wang, X. Yang, B. Lei, L. Liu, S. X. Li, D. Ni, T. Wang, *Engineering*, 2019, **5**, 261-275, doi: 10.1016/j.eng.2018.11.020.
[35] T. R. Nelson, T. T. Elvins, *IEEE Computer Graphics and Applications*, 1993, **13**, 50-57, doi: 10.1109/38.252557.
[36] S.Misciagna, *Intech Open*, 2013, doi: 10.5772/57537.
[37] A. Tsymbal, E. Meissner, M. Kelm, M. Kramer, *IEEE-EMBS International Conference on Biomedical and Health Informatics (BHI)*, 2014, 593-596, doi: 10.1109/BHI.2014.6864434.
[38] T. Zheng, L. Cao, Q. He, G. Jin, *Optical Engineering*, 2014, **53**, 011003, doi: 10.1117/1.oe.53.1.011003.
[39] N. H. Strickland, *Archives of Disease in Childhood*, 2000, **83**, 82-86, doi: 10.1136/adc.83.1.82.
[40] A. Holmes, Hadoop in Practice, 2012, *Manning Publications*, ISBN: 9781617290237.
[41] H. S. Bhosale, D. P. Gadekar, *International Journal of Scientific and Research Publications*, 2014, **4**, 1-7.
[42] D. Borthakur, *Hadoop Apache Project*, 2008, **53**, 1-13.
[43] J. Dean, S. Ghemawat, MapReduce, *Communications of the ACM*, 2008, **51**, 107-113, doi: 10.1145/1327452.1327492.
[44] M. Stonebrakeretal, *Communications of the ACM*, 2010, **53**, 64-71, doi:10.1145/1629175.1629197.
[45] M. Isard, M. Budiu, Y. Yu, A. Birrell, D. Fetterly, Dryad: distributed data-parallel programs from sequential building blocks, *Proceedings of the 2nd ACM SIGOPS/EuroSys European Conference on Computer Systems*, 2007, **41**, 59-72, doi: 10.1145/1272998.1273005.
[46] M. Zaharia, R. S. Xin, P. Wendell, T. Das, M. Armbrust, A. Dave, X. Meng, J. Rosen, S. Venkataraman, M. J. Franklin, A. Ghodsi, J. Gonzalez, S. Shenker, I. Stoica, Apache spark, *Communications of the ACM*, 2016, **59**, 56-65, doi: 10.1145/2934664.
[47] S. Sarraf, M. Ostadhashem, Big data application in functional magnetic resonance imaging using apache spark, *2016 Future Technologies Conference (FTC)*, 2016, 281-284.
[48] N. R. Council, others. Toward precision medicine: building a knowledge network for biomedical research and a new taxonomy of disease, *National Academies Press*, Published online 2011.
[49] G. S. Ginsburg, K. A. Phillips, Precision medicine: from science to value, *Health Affairs*, 2018, **37**, 694-701, doi: 10.1377/hlthaff.2017.1624.
[50] R. Mirnezami, J. Nicholson, A. Darzi, Preparing for precision medicine, *New England Journal of Medicine*, 2012, **366**, 489-491, doi: 10.1056/nejmp1114866.
[51] S. G. Peters, M. A. Khan, *Journal of Comparative Effectiveness Research*, 2014, **3**, 515-522, doi: 10.2217/cer.15.55.
[52] L. S. Weiss, X. Zhou, A. M. Walker, A. N. Ananthakrishnan, R. Shen, R. E. Sobel, A. Bate, R. F. Reynolds, *Pharmaceutical Medicine*, 2018, **32**, 31-37, doi: 10.1007/s40290-017-0216-4.
[53] C. F. Mackenzie, P. Hu, A. Sen, R. Dutton, S. Seebode, D. Floccare, T. Scalea, Automatic pre-hospital vital signs waveform and trend data capture fills quality management, triage and outcome prediction gaps, *AMIA Annual Symposium Proceedings*, 2008, **2018**, 318-322.
[54] D. Apiletti, E. Baralis, G. Bruno, T. Cerquitelli, *IEEE Transactions on Information Technology in Biomedicine*, 2009, **13**, 313-321, doi: 10.1109/titb.2008.2010702.
[55] A. Kumar, D. Sundar, D. Agarwal, *Indian Journal of Ophthalmology*, 2020, **68**, 432, doi: 10.4103/ijo.ijo_1474_19.
[56] C. Xie, M. T. Tammi, *BMC Bioinformatics*, 2009, **10**, 80, doi: 10.1186/1471-2105-10-80.
[57] M. A. DePristo, E. Banks, R. Poplin, K. V. Garimella, J. R. Maguire, C. Hartl, A. A. Philippakis, G. del Angel, M. A. Rivas, M. Hanna, A. McKenna, T. J. Fennell, A. M. Kernytsky, A. Y. Sivachenko, K. Cibulskis, S. B. Gabriel, D. Altshuler, M. J. Daly, *Nature Genetics*, 2011, **43**, 491-498, doi: 10.1038/ng.806.
[58] C. Trapnell, A. Roberts, L. Goff, G. Pertea, D. Kim, D. R. Kelley, H. Pimentel, S. L. Salzberg, J. L. Rinn, L. Pachter, *Nature Protocols*, 2012, **7**, 562, doi: 10.1038/nprot.2012.016.
[59] F. Ozsolak, P. M. Milos, *Nature Reviews Genetics*, 2011, **12**, 87-98, doi: 10.1038/nrg2934.
[60] R. Aebersold, M. Mann, *Nature*, 2003, **422**, 198-207, doi: 10.1038/nature01511.
[61] M. Hirst, M. A. Marra, *Briefings in Functional Genomics*, 2010, **9**, 455-465, doi: 10.1093/bfgp/elq035.
[62] A. Buchholz, J. Hurlebaus, C. Wandrey, R. Takors, *Biomolecular Engineering*, 2002, **19**, 5-15, doi: 10.1016/s1389-0344(02)00003-5.
[63] F. S. Collins, H. Varmus, *New England Journal of Medicine*, 2015, **372**, 793-795, doi: 10.1056/nejmp1500523.
[64] P. Y. Wu, C. W. Cheng, C. D. Kaddi, J. Venugopalan, R. Hoffman, M. D. Wang, *IEEE Transactions on Biomedical Engineering*, 2016, **64**, 263-273, doi: 10.1109/tbme.2016.2573285.
[65] M. Cord, P. Cunningham, Machine learning techniques for multimedia: case studies on organization and retrieval, *Springer Science & Business Media*, Berlin, Heidelberg, 2008, doi: 10.1007/978-3-540-75171-7
[66] S. Pepke, B. Wold, A. Mortazavi, *Nature Methods*, 2009, **6**, S22-S32, doi: 10.1038/nmeth.1371.
[67] H. Li, *Bioinformatics*, 2011, **27**, 2987-2993, doi:

10.1093/bioinformatics/btr509.
[68] G. Robertson, J. Schein, R. Chiu, R. Corbett, M. Field, S. D. Jackman, K. Mungall, S. Lee, H. M. Okada, J. Q. Qian, M. Griffith, A. Raymond, N. Thiessen, T. Cezard, Y. S. Butterfield, R. Newsome, S. K. Chan, R. She, R. Varhol, B. Kamoh, A.-L. Prabhu, A. Tam, Y. Zhao, R. A. Moore, M. Hirst, M. A. Marra, S. J. M. Jones, P. A. Hoodless, I. Birol, *Nature Methods*, 2010, **7**, 909-912, doi: 10.1038/nmeth.1517.
[69] M. G. Grabherr, B. J. Haas, M. Yassour, J. Z. Levin, D. A. Thompson, I. Amit, X. Adiconis, L. Fan, R. Raychowdhury, Q. Zeng, Z. Chen, E. Mauceli, N. Hacohen, A. Gnirke, N. Rhind, F. di Palma, B. W. Birren, C. Nusbaum, K. Lindblad-Toh, N. Friedman, A. Regev, *Nature Biotechnology*, 2011, **29**, 644-652, doi: 10.1038/nbt.1883.
[70] A. R. Quinlan, I. M. Hall, *Bioinformatics*, 2010, **26**, 841-842, doi: 10.1093/bioinformatics/btq033.
[71] L. N. Mueller, M.-Y. Brusniak, D. R. Mani, R. Aebersold, *Journal of Proteome Research*, 2008, **7**, 51-61, doi: 10.1021/pr700758r.
[72] J. L. Schafer, *Statistical Methods in Medical Research*, 1999, **8**, 3-15, doi: 10.1177/096228029900800102.
[73] G. D. Clifford, W. J. Long, G. B. Moody, P. Szolovits, Philosophical Transactions of the Royal Society A: Mathematical, *Physical and Engineering Sciences*, 2009, **367**, 411-429, doi: 10.1098/rsta.2008.0157.
[74] Q. Li, R. G. Mark, G. D. Clifford, *Physiological Measurement*, 2008, **29**, 15-32, doi: 10.1088/0967-3334/29/1/002.
[75] J. R. Gonzalez, L. Armengol, X. Sole, E. Guino, J. M. Mercader, X. Estivill, V. Moreno, *Bioinformatics*, 2007, **23**, 654-655, doi: 10.1093/bioinformatics/btm025.
[76] Liang K, Keleş S, *Bioinformatics*, 2012, **28**, 121-122. doi: 10.1093/bioinformatics/btr605.
[77] C. D. Kaddi, R. V. Bennett, M. R. L. Paine, M. D. Banks, A. L. Weber, F. M. Fernández, M. D. Wang, *Journal of the American Society for Mass Spectrometry*, 2016, **27**, 359-365, doi: 10.1007/s13361-015-1293-9.
[78] M. D. Robinson, D. J. McCarthy, G. K. Smyth, *Bioinformatics*, 2010, **26**, 139-140, doi: 10.1093/bioinformatics/btp616.
[79] S. Shen, J. W. Park, J. Huang, K. A. Dittmar, Z.-X. Lu, Q. Zhou, R. P. Carstens, Y. Xing, *Nucleic Acids Research*, 2012, **40**, e61, doi: 10.1093/nar/gkr1291
[80] H. Xu, C.-L. Wei, F. Lin, W.-K. Sung, *Bioinformatics*, 2008, **24**, 2344-2349, doi: 10.1093/bioinformatics/btn402.
[81] Y. Zhang, H. Liu, J. Lv, X. Xiao, J. Zhu, X. Liu, J. Su, X. Li, Q. Wu, F. Wang, Y. Cui, *Nucleic Acids Research*, 2011, **39**, e58, doi: 10.1093/nar/gkr053.
[82] C.-W. Cheng, N. Chanani, J. Venugopalan, K. Maher, M. D. Wang, *IEEE Journal of Translational Engineering in Health and Medicine*, 2013, **1**, 4400110, doi: 10.1109/jtehm.2013.2290113.
[83] R. V. Andreao, B. Dorizzi, J. Boudy, *IEEE Transactions on Biomedical Engineering*, 2006, **53**, 1541-1549, doi: 10.1109/tbme.2006.877103.
[84] G. de Lannoy, D. Francois, J. Delbeke, M. Verleysen, *IEEE Transactions on Biomedical Engineering*, 2012, **59**, 241-247, doi: 10.1109/tbme.2011.2171037
[85] J. Luo, M. Wu, D. Gopukumar, Y. Zhao, *Biomedical Informatics Insights*, 2016, **8**, BII.S31559, doi: 10.4137/bii.s31559.
[86] J. Adler-Milstein, W. Zhao, R. Willard-Grace, M. Knox, K. Grumbach, *Journal of the American Medical Informatics Association*, 2020, **27**, 531-538, doi: 10.1093/jamia/ocz220.
[87] J. Hecht, *Nature*, 2019, **573**, S114-S116, doi: 10.1038/d41586-019-02876-y.
[88] R. L. Schilsky, D. L. Michels, A. H. Kearbey, P. P. Yu, C. A. Hudis, *Journal of Clinical Oncology*, 2014, **32**, 2373-2379, doi: 10.1200/jco.2014.56.2124.
[89] W. Moore, S. Frye, *Journal of Nuclear Medicine Technology*, 2019, **47**, 269-272, doi: 10.2967/jnmt.119.227819.
[90] S. Kaur, R. Kaur, R. Aggarwal, *Paradigm*, 2019, **23**, 164-174, doi: 10.1177/0971890719859943.
[91] J. D. Tenenbaum, S.-A. Sansone, M. Haendel, *Journal of the American Medical Informatics Association*, 2014, **21**, 200-203, doi: 10.1136/amiajnl-2013-002066.
[92] D. W. Bates, S. Saria, L. Ohno-Machado, A. Shah, G. Escobar, *Health Affairs*, 2014, **33**, 1123-1131, doi: 10.1377/hlthaff.2014.0041.
[93] G. Bansal, F. ". Zahedi, D. Gefen, *Decision Support Systems*, 2010, **49**, 138-150, doi: 10.1016/j.dss.2010.01.010
[94] I. Keshta, A. Odeh, *Egyptian Informatics Journal*, 2021, **22**, 177-183, doi: 10.1016/j.eij.2020.07.003.
[95] C.R. Manzetal., *JAMA Oncology*, 2020, **6**, e204759-e204759, doi: 10.1001/jamaoncol.2020.4759.
[96] J. R. Vest, L. D. Gamm, *Journal of the American Medical Informatics Association*, 2010, **17**, 288-294, doi: 10.1136/jamia.2010.003673.
[97] S. Haneuse, M. Daniels, *EGEMS (Wash DC)*, 2016, **4**, doi: 10.13063/2327-9214.1203.
[98] D. Reeves, S. M. Campbell, J. Adams, P. G. Shekelle, E. Kontopantelis, M. O. Roland, *Medical Care*, 2007, **45**, 489-496, doi: 10.1097/mlr.0b013e31803bb479.
[99] F. de Vries, C. de Vries, C. Cooper, B. Leufkens, T.-P. van Staa, *International Journal of Epidemiology*, 2006, **35**, 1301-1308, doi: 10.1093/ije/dyl147.
[100] R. A. Verheij, V. Curcin, B. C. Delaney, M. M. McGilchrist, *Journal of Medical Internet Research*, 2018, **20**, e185, doi: 10.2196/jmir.9134.
[101] J. A. Rassen, W. Murk, S. Schneeweiss, *Obesity and Metabolism*, 2021, **23**, 1453-1462, doi: 10.1111/dom.14338.
[102] H. Muller, X. Zhou, A. Depeursinge, M. Pitkanen, J. Iavindrasana, A. Geissbuhler, *IEEE International Conference on Multimedia and Expo*, 2007, 683-686.
[103] S. Antani, L. R. Long, G. R. Thoma, *21st IEEE International Symposium on Computer-Based Medical Systems*, 2008, 4-6.
[104] J Kim, H Chang, D Kim, DH Jang, I Park, K Kim, *Journal of Critical Care*, 2020, **55**, 163-170, doi: 10.1016/j.jcrc.2019.09.024.
[105] D. S. Ting, D. V. Gunasekeran, L. Wickham, T. Y. Wong, *British Journal of Ophthalmology*, 2020, **104**, 299-300, doi:

10.1136/bjophthalmol-2019-315066.
[106] Y. Hu, K. Duan, Y. Zhang, M. S. Hossain, S. M. Mizanur Rahman, A. Alelaiwi, *Multimedia Tools and Applications*, 2018, **77**, 3729-3743, doi: 10.1007/s11042-016-3719-1.
[107] A. McAfee, E. Brynjolfsson, T. H. Davenport, D. Patil, D. Barton, *Harvard Business Review*, 2012, **90**, 60-68.
[108] W. Raghupathi, V. Raghupathi, *Health Information Science and Systems*, 2014, **2,** doi: 10.1186/2047-2501-2-3.
[109] N. Coudray, P. S. Ocampo, T. Sakellaropoulos, N. Narula, M. Snuderl, D. Fenyö, A. L. Moreira, N. Razavian, A. Tsirigos, *Nature Medicine*, 2018, **24**, 1559-1567, doi: 10.1038/s41591-018-0177.
[110] A. A. Friedman, A. Letai, D. E. Fisher, K. T. Flaherty, *Nature Reviews Cancer*, 2015, **15**, 747-756, doi: 10.1038/nrc4015.
[111] Evaluation C for D, Research. Table of Pharmacogenomic Biomarkers, *US Food Drug Administration, Published online* March 2021. https://www.fda.gov/drugs/science-and-research-drugs/table-pharmacogenomic-biomarkers-drug-labeling.
[112] Y. Wang, S. Sun, Z. Zhang, D. Shi, *Advanced Materials*, 2018, **30**, 1705660, doi: 10.1002/adma.201705660.
[113] E. A. Klein, M. R. Cooperberg, C. Magi-Galluzzi, J. P. Simko, S. M. Falzarano, T. Maddala, J. M. Chan, J. Li, J. E. Cowan, A. C. Tsiatis, D. B. Cherbavaz, R. J. Pelham, I. Tenggara-Hunter, F. L. Baehner, D. Knezevic, P. G. Febbo, S. Shak, M. W. Kattan, M. Lee, P. R. Carroll, *European Urology*, 2014, **66**, 550-560, doi: 10.1016/j.eururo.2014.05.004.
[114] J. Cullen, I. L. Rosner, T. C. Brand, N. Zhang, A. C. Tsiatis, J. Moncur, A. Ali, Y. Chen, D. Knezevic, T. Maddala, H. J. Lawrence, P. G. Febbo, S. Srivastava, I. A. Sesterhenn, D. G. McLeod, *European Urology*, 2015, **68**, 123-131, doi: 10.1016/j.eururo.2014.11.030.
[115] H.-S. You, C.-X. Gao, H.-B. Wang, S.-S. Luo, S.-Y. Chen, Y.-L. Dong, J. Lyu, T. Tian, *Cancer Management and Research*, 2020, **12**, 1947-1958, doi: 10.2147/cmar.s244932.
[116] S. P. Somashekhar, M.-J. Sepúlveda, S. Puglielli, A. D. Norden, E. H. Shortliffe, C. Rohit Kumar, A. Rauthan, N. Arun Kumar, P. Patil, K. Rhee, Y. Ramya, *Annals of Oncology*, 2018, **29**, 418-423, doi: 10.1093/annonc/mdx781.
[117] Y. Tian, X. Liu, Z. Wang, S. Cao, Z. Liu, Q. Ji, Z. Li, Y. Sun, X. Zhou, D. Wang, Y. Zhou, *Journal of Medical Internet Research*, 2020, **22**, e14122, doi: 10.2196/14122.
[118] T. P. Miller, Y. Li, K. D. Getz, J. Dudley, E. Burrows, J. Pennington, A. Ibrahimova, B. T. Fisher, R. Bagatell, A. E. Seif, R. Grundmeier, R. Aplenc, *British Journal of Haematology*, 2017, **177**, 283-286, doi: 10.1111/bjh.14538.
[119] P. M. Amisha, M. Pathania, V. K. Rathaur, *Journal of Family Medicine and Primary Care*, 2019, **8**, 2328. doi: 10.4103/jfmpc.jfmpc_440_19.
[120] F. M. Sanchez, V. A. Pulido, G. L. Campos, N. Peek, L. Sacchi, *Yearbook of Medical Informatics*, 2017, **26**, 28-35, doi: 10.15265/IY-2017-008.
[121] D. Al-Jumeily, A. Hussain, C. Mallucci, C. Oliver, *Applied computing in medicine and health*, 2015, doi: 10.1016/c2014-0-02198-3.
[122] D. W. Bates, L. Su, D. T. Yu, G. M. Chertow, D. L. Seger, D. R. J. Gomes, E. J. Dasbach, R. Platt, *Clinical Infectious Diseases*, 2001, **32**, 686-693, doi: 10.1086/319211.
[123] N. Tomašev, X. Glorot, J. W. Rae, M. Zielinski, H. Askham, A. Saraiva, A. Mottram, C. Meyer, S. Ravuri, I. Protsyuk, A. Connell, C. O. Hughes, A. Karthikesalingam, J. Cornebise, H. Montgomery, G. Rees, C. Laing, C. R. Baker, K. Peterson, R. Reeves, D. Hassabis, D. King, M. Suleyman, T. Back, C. Nielson, J. R. Ledsam, S. Mohamed, *Nature*, 2019, **572**, 116-119, doi: 10.1038/s41586-019-1390-1.
[124] A. P Tafti, J. Badger, E. LaRose, E. Shirzadi, A. Mahnke, J. Mayer, Z. Ye, D. Page, P. Peissig, *JMIR Medical Informatics*, 2017, **5**, e51, doi: 10.2196/medinform.9170.
[125] N. Mehta, A. Pandit, *International Journal of Medical Informatics*, 2018, **114**, 57-65, doi: 10.1016/j.ijmedinf.2018.03.013.
[126] C. W. Seymour, H. Gomez, C.-C H. Chang, G. Clermont, J. A. Kellum, J. Kennedy, S. Yende, D. C. Angus, *Critical Care*, 2017, **21**, 257, doi: 10.1186/s13054-017-1836-5.
[127] J. Ye, L. N. Sanchez-Pinto, *AMIA Annual Symposium Proceedings*, 2020. 2020, 1345.
[128] L. M. Fernandes, M. O'Connor, V. Weaver, *Journal of AHIMA*, 2012, **83**, 38-43.
[129] A. E. Youssef, International Journal of Ambient Systems and Applications, 2014, 2, 1-11, doi: 10.5121/ijasa.2014.2201.
[130] T. Naqishbandi, C. I. Sheriff, S. Qazi, *International Journal of Engineering Research & Technology*, 2015, **4**, 1-6, doi: 10.1109/iccci.2015.7218108.
[131] E. Baro, S. Degoul, R. Beuscart, E. Chazard, *BioMed Research International*, 2015, doi: 10.1155/2015/639021.

Authors information

Nithesh Naik *is a faculty in the Department of Mechanical and Manufacturing Engineering, Manipal Institute of Technology. He has four years of industry experience in the field of planning and design of HVAC systems. He is a master graduate in Design Engineering from prestigious university Manipal Academy of Higher Education (Institute of Eminence). His research interest includes the development of FE analysis of dental sciences, artificial intelligence, composite materials and design and product development techniques. He has applied 4 patents at Indian Patent Office. He has published several research publications in international journals of repute and has keen interests in medical innovations. He received the award of South India's most exciting young teacher. His areas of expertise include FE analysis of dental and medical sciences, composite materials, design and product development, medical devices and innovations, artificial intelligence, medical web and app*

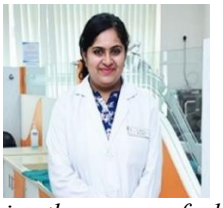
development.

Yuvraj Rallapalli *is an undergraduate student in Department of Biotechnology at Manipal Institute of Technology, Manipal, India and an aspiring researcher in cancer biology. He has worked on various facets of biological research connected to cancer along with studying the clinical benefit for patients. His research interests include nanotechnology, molecular modelling, and immunotherapy for cancer. He is former national level tennis player and fitness enthusiast. The research carried will help the cancer patients across the world by reducing the side effects and increasing quality of life.*

Manamohana Krishna *is a faculty member of the Department of Computer Science Engineering, Manipal Institute of Technology (MIT), Manipal Academy of Higher Education (MAHE), Manipal. His research interests include Computer networking, Database Management systems, Algorithms, Data mining, soft computing and Data and Network Security.*

Anoushka Suresh Vellara *is a biomedical engineering undergraduate student currently at Manipal Institute of Technology, India. Her research interests are in areas such as BioMEMS and microfluidics with focus on therapeutic and diagnostic devices. She hopes to increase the cost effectiveness of medical devices to ensure healthcare is affordable for vulnerable populations. She currently works on miniaturising the apparatus used for electrophoresis through Lab-on-Chip Technology and is looking forward to enhancing her skillset with a graduate degree in biomedical engineering.*

Dasharathraj K Shetty *is a faculty member of the Department of Humanities and Management, Manipal Institute of Technology (MIT), Manipal Academy of Higher Education (MAHE), Manipal. Formerly he was the faculty at the Department of Computer Science and Engineering. He is an Author, Columnist, Engineer and Social Entrepreneur. Dasharathraj is a B.E. (Computer Science and Engineering) and has three Post-graduation Degrees - MBA (Finance), MPhil (Management) and M.Tech (Computer Science and Engineering), PhD (Computer Vision). He is also a Certified Microsoft Certified Technology Specialist, Dale Carnegie High Impact Teaching Skills, AIMA Certified Management Trainer and RBNQA Examiner.*



Vathsala Patil *is a Reader in the Department of Oral medicine and Radiology at Manipal College of Dental Sciences, Manipal, India. She has authored various scientific publications in the area of dental imaging. She also specializes in management of oral premalignant and malignant lesions. Her research Interest includes early diagnosis of potentially malignant lesions, artificial neural network and its use and applications in dentistry, Cone beam computed tomography and its applications, Finite element analysis and its applications in dentistry. Her area of expertise is Oral mucosal lesions diagnosis and management, artificial neural network in age and gender estimation. She is actively involved in research area in Laser Spectroscopy, artificial neural network, machine learning. Professional affiliations, Life member of Indian Academy of Oral Medicine and Radiology, Member of Asian Academy of Oral and Maxillo-Facial Radiology.*

Zeeshan Hameed BM *is a Professor in the Department of Urology, Father Muller Medical College, Mangalore, India. His research interests include studies in relation to urological malignancies, artificial intelligence, and design and product development techniques. He has applied 3 patents at Indian Patent Office. He has published several research publications in international journals of repute and has keen interests in medical innovations. His areas of expertise include medical devices and innovations, artificial intelligence, medical web and app development. He has developed web application for urological stent and symptom tracking by name "UROSTENTZ.*

Rahul Paul *currently works at the Department of Radiation Oncology, Massachusetts General Hospital, Harvard Medical School, USA as a Postdoctoral Research Fellow. etts General Hospital, Harvard Medical School, as a Postdoctoral Research Fellow. His current research is focused on early diagnosis, prediction, and prognosis of cancer using different imaging modality, and clinical data. He completed his doctoral degree in Computer Science & Engineering from University of South Florida,*

Tampa, USA. During his doctoral study, he worked on lung nodule malignancy analysis, neonatal pain monitoring and COVID detection. He has published over 12 journal and over 10 conference publications and a member of IEEE and SPIE. Dr. Paul has served as a reviewer for a number of international journals (IEEE Transactions on AI, IEEE Access, IEEE/ACM TCBB, Medical Physics, EBioMedicine, Nature Scientific Reports, and others) and conferences (IEEE SMC, IEEE ICDM, and others), and he is currently an editorial board member for the International Journal of Imaging System and Technology (Wiley). Dr. Paul was selected as a PostDoc-AI-Net Fellow (DAAD) in 2021.

Nirmal Prabhu *is currently pursuing Bachelor of Technology in Computer Science from Vellore Institute of Technology, Chennai. He completed his 12th grade from Little Rock Indian School, Udupi. His current research interests include Big Data Applications and Distributed Deep Learning. His Bachelor's curriculum includes courses like Distributed Systems, Machine Learning and Social Networks to aid his research. He has multiple certifications in the field of Data Science from Kaggle, Udemy and Coursera. Nirmal is currently pursuing projects in the field of Data Science: A Deep Learning based regressor to predict and manage data in calamities, COVID-19 detection system using deep neural networks and a few other big data cloud implementations to compliment his research understanding.*

Bhavan Rai *is as a Consultant Urological Surgeon with a specialist interest in robotic surgery, minimally invasive surgery, and urological oncology at the Freeman Hospital, Newcastle United Kingdom. He is one of very few urologists world-wide to have undertaken double high-volume fellowship training in robotic surgery. Firstly in the world famous University of Leipzig Urology Clinic under the mentorship of Professor Ule Stolzenberg, who pioneered minimally-invasive radical prostatectomy. Subsequently as part of a competitively selected prestigious nationally accredited Royal College of Surgeons of England and British Association of Urological surgeons Fellowship at the Lister Hospital, Stevenage. He is chairman of the uro-oncology multidisciplinary meeting at the Freeman Hospital, UK. He is a teaching faculty member in number of international robotic and minimal invasive courses. He has published extensively on urological and surgical research and has over 100 peer-reviewed articles, national and international research presentations and authored textbook chapters. He has a MSc Research degree, evaluating the role of urinary biomarker MCM-2 as a diagnostic tool for bladder cancer. His current research interests include outcomes research, systematic reviews in uro-oncology and surgical education. He has also been sub-investigator for several regional and national trials.*

Piotr Chłosta *is an eminent specialist in the field of urology, a Fellow of the European Board of Urology. His main areas of interest include urological oncology, endourology, and laparoscopic urology. He was the first doctor in Poland and one of the first physicians in the world to perform a laparoscopic retroperitoneal lymph node dissection in metastatic testicular cancer patients, the first in Poland to create an orthotopic ileal neobladder completely laparoscopically, and the first in the world to laparoscopically implant an artificial sphincter around prostatic urethra. He has immense experience in uro-oncologic laparoscopic surgery, including radical laparoscopic prostatectomy and radical laparoscopic urinary bladder dissection with urinary tract reconstruction. Prof. Chłosta has introduced a number of innovative surgical methods to Polish urology, along with modifications of several widely used treatment methods. Continuing the work of his teacher, Professor Andrzej Borówka, he developed a minimally invasive method of biopsy, which can be applied to treat advanced urinary bladder cancer. He also co-authored and implemented the method of percutaneous cystolithotrypsy (PCCL) to remove bladder stones. Prof. Chłosta is a member of a number of scientific associations, including the European Association of Urology (EAU), the Association of Academic European Urologists (AAEU), the Laparoscopy Working Group of European Section Uro-Technology (ESUT), the American Urological Association (AUA), and, from 2014 on, the Royal College of Surgeons. He received a number of awards from Polish and international research associations, international medical journals, and the Polish Ministry of Health. Prof. Piotr Chłosta, head of the JU MC Clinic and Chair in Urology, has received the European Board of Urology Golden Pin. The distinction was conferred upon him for his significant accomplishments in urological education on the European scale.*

Bhaskar Somani *is a Professor of Urology and a Consultant Endourologist at University Hospital Southampton. He has been involved in clinically innovative patient-centred treatments. His research includes minimally invasive surgical techniques (MIST) in management of kidney stone disease and BPH, urinary tract infections, role of mobile phone apps and artificial intelligence (AI) in urology. He is the Clinical Director of 'South Coast Lithotripter Services'. Over the last 10 years his clinical and research work has been covered by BBC, ITV, Daily Mail, The Telegraph and other newspapers and media articles on a number of occasions. He has been a member of BAUS Academic and Endourology sub-sections and is the Wessex Clinical Research Network and Simulation Lead for Urology. He is also the founding member and President of PETRA (Progress in Endourology, Technology and Research Association) Urogroup and i-TRUE (International training and research in uro-oncology and endourology) group, an active member of European School of Urology (ESU) Training and Research group and EAU section of uro-technology (ESUT) endourology group, besides being in the EAU Live surgery and undergraduate training committees. For his work he got awarded the fellowship of Royal college of Edinburgh (Fellow of faculty of Surgical Trainers) in 2017, honorary fellowship of Royal College of Physicians and Surgeons of Glasgow in 2020, Endourology society 'Arthur Smith' award in 2020 and BAUS 'Golden Telescope' award and Zenith Global awards in 2021. With dedication, commitment and passion for research and teaching, he collaborates nationally and internationally sharing his research and teaching successfully across the world. He has published excellent clinical outcomes and his outcome and research translate into patient benefit.*

