

Complex predictive analysis for health care: a comprehensive review

Dolley Srivastava¹, Himanshu Pandey², Ambuj Kumar Agarwal³

¹Department of Computer Science, Maharishi University of Information Technology, Lucknow, India

²Department of Computer Science and Engineering, Faculty of Engineering and Technology, University of Lucknow, Lucknow, India

³Department of Computer Science and Engineering, School of Engineering and Technology, Sharda University, Greater Noida, India

Article Info

Article history:

Received Jul 6, 2022

Revised Aug 24, 2022

Accepted Oct 14, 2022

Keywords:

Data mining

Decision making

Machine learning

Predictive analytics

Smart decision support system

ABSTRACT

Healthcare organizations accept information technology in a management system. A huge volume of data is gathered by healthcare system. Analytics offers tools and approaches for mining information from this complicated and huge data. The extracted information is converted into data which assist decision-making in healthcare. The use of big data analytics helps achievement of improved service quality and reduces cost. Both data mining and big data analytics are applied to pharma co-vigilance and methodological perspectives. Using effective load balancing and as little resources as possible, obtained data is accessible to improve analysis. Data prediction analysis is performed throughout the patient data extraction procedure to achieve prospective outcomes. Data aggregation from huge datasets is used for patient information prediction. Most current studies attempt to improve the accuracy of patient risk prediction by using a commercial model facilitated by big data analytics. Privacy concerns, security risks, limited resources, and the difficulty of dealing with massive amounts of data have all slowed the adoption of big data analytics in the healthcare industry. This paper reviews the various effective predictive analytics methods for diverse diseases like heart disease, blood pressure, and diabetes.

This is an open access article under the [CC BY-SA](#) license.



Corresponding Author:

Dolley Srivastava

Department of Computer Science, Maharishi University of Information Technology

IIM Road, Lucknow, Uttar Pradesh, India

Email: dolley89@gmail.com

1. INTRODUCTION

Information mining in the healthcare company can significantly reduce expenses by boosting efficiencies, improving the quality of life of the patient, and potentially even more substantially, helping to save many more patients' lives [1], [2]. The word data processing will imply different things to completely distinct individuals like analytics and business intelligence. Data mining's most fundamental concept is the analysis of big information sets to find patterns and use those patterns to predict or predict the probability of future occurrences.

- Descriptive analytics—description of what occurred
- Predictive analytics—what will occur?
- Prescriptive analytics—determination of what to do about it

Data processing is applied in the core category of predictive analysis. Data mining includes uncovering patterns from large data stores and building predictive models using that information. Many sectors are using information mining effectively. It enables the client reaction to the retail sector model. It enables banks to predict the profitability of their customers. It provides medium, manufacturing, automotive,

higher education, life sciences, and more comparable use cases. However, information processing stays attentive these days, with only several pragmatic success tales for the primary instructional practice. Academics are exploiting data mining methods to publish studies such as decision trees, clusters, neural networks, and time series [3], [4].

People with physical and mental disabilities may be diagnosed, treated, and prevented by healthcare professionals in a variety of ways. In most nations, the healthcare business is advancing at a breakneck rate. Medical records, administrative reports, and other benchmarking results generated by the healthcare business are considered to be rich in data because of their volume. However, these medical records are being underutilised. Large amounts of data may be mined to find new and important information. Data mining in healthcare is mostly used to aid clinicians in making clinical choices by forecasting illnesses and providing diagnostic assistance [5]. The following is a breakdown of the numerous approaches used in the healthcare business.

Anomaly detection is a technique for spotting the most important changes to a data collection. In order to test the accuracy of anomaly detection on the uncertain dataset [6] anomaly detection methodologies such as normal support vector data description, density induced support vector data description, and Gaussian combination are utilised. Clustering is a frequent evocative activity that involves identifying a limited collection of categories or clusters to characterise the data. In the clustering method, the technology of vector quantization was applied to predict readmissions in advanced medicine [7]. Classification is the discovery of a prediction characteristic of learning that classifies a data item into one of several specified classes [8]:

- Statistical

Taguchi and Jugulum [9] used the mahalanobis Taguchi system (MTS) approach for multivariable statistical assessment. The mahalanobis distance (MD) is used to compare the abnormality of two clusters, whereas the mahalanobis room (MS) is used to compare the abnormality of the known reference group data. Pressure ulcers were predicted using the MTS. Class imbalance is common in healthcare datasets. When dealing with skewed or unbalanced data sets, skewed distribution commonly impacts data mining processes. This difficulty typically results in very precise classification accuracy for the majority class but low precision for the minority class. This method is excellent for analysing data since it can indicate the degree of identified abnormality. This strategy is also used for scaling the MD. This method's performance is based on the wide disparity between normal and atypical instances. During the test phase, the MTS shows better sensitivity and g-means scaling. The MTS has improved sensitivity efficiency.

- Discriminant analysis

Linear discriminant analysis (LDA) is often used in discriminating assessments to estimate the class from a series of data based on fresh, unlabeled observations. The LDA method was utilised by Duda *et al.* [10] to estimate the severity of Parkinson's disease patient's non-motor symptoms.

- Decision tree

Several studies have examined the use of decision trees to analyse clinical data [11], [12]. Using the tree and its principles to forecast on a dataset is the heart of this method. The information set employed in this investigation has a very even distribution of information. For predictive purposes, the decision tree cannot be employed since it repeatedly separates data into branches in order to form a tree. Some of the best minds in the field of sickness treatment are analysing a decision tree approach that was just presented to them. In a choice tree, each branch shows an attribute value, each inner node suggests a test on an attribute utilised for, and a leaf node reflects the courses or class distributions anticipated by the tree structure. Predictive value of each attribute is used to determine where in the tree to start classifying. There are four steps in the method: data partitioning and categorization, choosing a decision tree category, and requesting a reduction in fault trimming. Testing with or without a vote is part of information partitioning. The gini index, the improvement of information, and the gain ratio are three types of decision trees. Finally, a reduction in the amount of mistake trimming is useful in order to give more closed choice recommendations.

- Swarm intelligence

The particle swarm optimization (PSO) algorithm is able to locate in big search spaces the ideal or near-optimal alternatives effectively. If fewer characteristics are used, the classification method will be quicker and more precise. As PSO is used to select appropriate parameters in the classifiers concerned, the PSO-based strategy demonstrates to enhance the general classification outcomes [13].

- K-nearest neighbour

A classification approach based on an instance is the k-nearest neighbour (k-NN). Since the n-dimensional space points in the technique are assumed to be connected to all instances, this algorithm's parameter units are samples. Because no data is ever lost in the process of training, the method is very valuable [14].

- Logistic regression

It is possible to use logistic regression (LR) in a variety of ways, including continuous, discrete, or a combination of both. Logistic regression is then applied to the resultant linear combination of inputs. These methods are popular because they are easy to implement and provide effective outcomes [15].

- Bayesian classifier

The Bayesian classifier is highly effective in computational efficiency and capable of handling naturally and effectively missing information. The Bayesian classifier also demonstrates by getting the models applied that the model is appropriate as the average strategy has resulted in enhanced forecast precision and enables writers to extract more characteristics from the information without being overfitted. Bayesian classifiers predict the likelihood of class affiliation so that the likelihood of a specified sample belongs statistically to a specific class. In view of the observation, the Bayes theorem used to determine the probability of a suggested diagnosis being right. The model of Naive Bayes defines the physical characteristics and characteristics of a disease patient. It provides the option of an attribute for the expected state for each input [16].

- Support vector

Because of their success in a wide range of pattern categorization tasks, support vector machines (SVM) have recently attracted a lot of interest. Supervised learning using SVM is a technique used in this field. The SVM method predicts disease incidence by arranging disease predicting features on a multidimensional hyperplane and calculating the margin between two data clusters. Kernels, which are nonlinear characteristics, are used to attain high accuracy in this technique. SVM, a generalization-efficient approach, has been shown to be helpful in handling classification responsibilities. By limiting structural hazard, a strategy seeks to reduce generalisation error's upper bounds [17].

- Neural network

Neural networks are well known for generating extremely precise outcomes in practical applications. By using the neural network model feed-forward, variable learning speed and momentum learning algorithm back-propagation, the neural network trained with the database. The model's structure is as follows: it begins with clinical information input and moves towards the development of an artificial neural networks (ANN) algorithm [18].

- Hybrid

Prediction of disease is one of the main problems facing the healthcare sector. Motivated by the growing death rate of patients globally, scientists use various information mining methods in disease diagnosis. Every method possesses its own merits and demerits. Each algorithm used by each method includes certain features that are useful for diagnosing the disease [19]. Here the output mixture is regarded as "hybridization."

2. TECHNOLOGY AND METHODS FOR DATA ANALYSIS IN HEALTHCARE

It is difficult to analyse dynamic and complicated data such as magnetic resonance imaging (MRI) images, X-rays and biological signals electrocardiography (ECG), electroencephalograph (EEG), and electromyography (EMG) because of their dynamic and complex character. They are multi-dimensional. There aren't many methods for analysing data of this kind and making decisions based on it [20]. Some of the analytical methodologies that may be used in healthcare and medicine are mentioned in the literature (Table 1).

Table 1. Medical big data analytical methods

Technique	Healthcare application	Studies
Cluster analysis	Identification of at-risk populations via locating obesity clusters; Finding population groups with unique health variables for chronic illness treatment	Clark <i>et al.</i> [21] Swain [22], Schatz [23]
Data mining	Biosignal monitoring for health issues; Detecting outbreaks; Healthcare data analysis based on inductive and exploratory reasoning	Forkan <i>et al.</i> [24] Ghani <i>et al.</i> [25] Roski <i>et al.</i> [26]
Graph analytics	Examination of hospital performance on a variety of different qualitative criteria	Downing <i>et al.</i> [27]
Machine learning	Diagnosis and treatment planning;	Chen <i>et al.</i> [28], Khullar <i>et al.</i> [29]
Natural language processing (NLP)	Review of the hospital's effectiveness; Epidemiological surveillance, enhanced care delivery, and cost containment	Downing <i>et al.</i> [27], Lo and Jamrozy [30]
	Assistance in identifying high-risk variables, as well as training, consulting and therapy	Khalifaand and Meystre [31] Martin-Sanchez <i>et al.</i> [32]
	Extracting data from medical records; lowering the risk of disease and death	Ford <i>et al.</i> [33]
Neural networks	Chronic illness diagnosis and treatment Prediction of a patient's sickness in the future	Al-Jumeily <i>et al.</i> [34] Martin-Sanchez <i>et al.</i> [32]
Pattern recognition	Improvements in public health monitoring	Luxton [35]
Spatial analysis	Using visual, geographical, and sophisticated analytics to get insights on the population at large	Amirian <i>et al.</i> [36]

3. CLASSIFICATION MODEL TO PREDICT THE DIABETICS MELLITUS AND RETINOPATHY

Faruque and Sarker [37] had been worked on the diabetes mellitus information mining concept. The investigator utilizes the information classification technique of Naive Bayes, k-NN, C4.5, and SVM algorithm. Furthermore, in this document, the investigator cites the fine points about the knowledge discovery in database (KDD), and explores the risk factors for prediction of disease. The article discusses how to manage the data's null values and noise. It provides a vivid overview of the meaning of classification. It utilizes classification WEKA software. The research compares two well-known algorithms, the algorithm of Naive Bayes, and the algorithm of J48. The precision approached by the algorithm is 81%.

Nurrahman *et al.* [38] clarified that type 2 diabetes mellitus is a metabolic disease that affects an individual with high blood glucose levels in easy terms called diabetes. Diabetes is a metabolic disorder triggered by the body's inability to make insulin or use insulin correctly. This situation occurs when there is not enough insulin generated by the body or because the cells do not react to the insulin generated and leads to obesity. A key technique for diagnosing diabetes is the blood glucose test. Numerous computerized techniques for diagnosing diabetes have also been suggested. All of these techniques have some input values that would result from various trials to be performed in hospitals. This article recommends a method to relieve patients experiencing countless medical exams, which are considered by most as a tedious and time-consuming job. The parameters known for polygenic disease detection are intended so that if the patient is impacted by the polygenic disease itself, the user will predict. Diagnosis is based on the backpropagation algorithm.

Singh and Singh [39] used UCI machine learning repository's pima indians diabetes database which is a very rich dataset and 768 tuples and about 8 constant characteristics. The sigmoid, linear SVM kernel model and radial basis function is used. The ensemble approach is used by C4.5, sequential minimal optimization (SMO) and NB. The stacking approach of SMO generated the highest accuracy of 79%. Rashid and Abdullah [40] suggested the diabetes diagnosis hybrid artificial bee colony (ABC). In this work, the writers used the genetic algorithm and back propagation neural network (BPNN). This improved ABC's diversity without compromising the quality of the solution. The altered ABC was used to produce an ideal fuzzy classifier as an evolutionary algorithm. The research is highly accurate and reliable in tuning optimal guidelines that trained BPNN mutation with hybrid ABC method. This classifier is more optimal. This has supplied the diabetes information with a good diagnostic instrument. The author overcome the over fitting problem. However, in order to attain precision and efficiency, the method requires more training samples.

Sejdinovic *et al.* [41] used the ANN for classification. The aim was to concentrate on type 2 diabetes and prediabetes. 2 layered, feed-forward neural networks were implemented by the scientists. In the concealed layer, the number of neurons needed was 15. The neural network is with 2 key parameters of fasting glucose rate and HbA1c testing. The training and test information was distinct from the experiment. The precision for prediabetes was nearly 94.1% and for type 2 diabetes's about 93.3%.

Choubey *et al.* [42] implemented the two approaches. The first approach is classification via regression, k-NN and radial basis network function classification on the Adaboost. The second approach is feature reduction with the same classification models. The classification models with and without the principal component analysis (PCA) and linear discriminant analysis is compared. The result shows that both are useful to remove insignificant features. The classification accuracy, receiver operating characteristic (ROC) curve is measured for validation.

4. CLASSIFICATION TECHNIQUE TO PREDICT HEART DISEASE

In order to analyze the performance of many information handling methods, Zunaidi *et al.* [43] present classification of carcinoma data may be used to search for outcomes of some disease or to explore the prevalent nature of heart disease. The many information handling methods are used to heart disease analysis; this new strategy usually concludes the comparative performance of decision tree classifiers such as SMO, random forest, k-NN, J48, and multi-layer perceptron (MLP). The outcome demonstrates that MLP output offers intelligent results in terms of precision, low error rate and efficiency compared to separate classifiers.

Latha and Jeeva [44] implemented ensemble bagging and boosting classification technique for predicting heart disease by measuring accuracy. Latha and Jeeva [44] implemented a faithfully greater performance comparison analysis on BPNN and logistic regression for heart disease prediction. BPNN is non parametric and LR is parametric analysis. Olatunde *et al.* [45] used Adaboost and Bagging for naive Bayes, neural network, and random forest. This ensemble approach is supportive for weak classifiers but performance is low for rigorous dataset.

Shah *et al.* [46] conducted a heart disease forecast survey with the assistance of the regulations of the association. They used a straightforward algorithm for decision tree. This algorithm treats attributes continuously as numerical or categorical. This is used to transform to a transaction format medical records for application. The decision tree is used for mining information as it splits numerical values automatically.

The split point selected by the tree of decision is of little use. Clustering is used to gain comprehensive information knowledge.

Dora *et al.* [47] scheduled the Gauss-Newton novel carcinoma classification-based algorithm (GNRBA). It utilizes the distributed illustration with the selection of function and evaluates the sparseness in an extremely economical approach to computation. Then this method scheduled fresh, mostly classifier-based gauss, Newton, to search for optimum weights for classification coaching specimens. These methods are studied from the UCI machine learning repository using Wisconsin carcinoma database and Wisconsin diagnosing carcinoma database. The outcome demonstrates that this method offers greater precision, sensitivity, specificity, matrices of confusion compared to old methods.

Reis *et al.* [48] provides an assessment of the categorization and automatic classification of carcinoma by the victimization of multi-scale fundamental picture characteristics basic image features (BIF) and local binary patterns (LBP) in conjunction with the random decision trees classifier used in carcinoma classification. These methods show the text-based classification of invasive breast carcinoma (IBC) images based on haematoxylin and eosin (H&E). The outcome demonstrates the good precision is provided by the multi-scale strategy.

5. CLASSIFICATION TECHNIQUE TO PREDICT VARIOUS STAGES OF CANCER

Xiao *et al.* [49] addressed five different classification models with ensemble approach to cancer prediction. This deep learning approach analyzed the gene expression on different cancer tissues such as lung, breast and stomach. The generation rate is reduced. Rahimi *et al.* [50] have introduced a cancer forecast classifier, the most prevalent cancers for males and females alike. The author also examined the usefulness of the fresh rule pruning method and showed the suggested procedure that has a beneficial impact on both multi objective PSO and PSO precision. This research also likened their strategy to popular information mining algorithms on the classification of breast cancer and pulmonary cancer and showed that combined with the suggested rule pruning, multi objective PSO is efficient in predicting prevalent cancer kinds. An experiment has shown that the fresh pruning technique has improved the precision of classification, and this approach has been discovered to be efficient in cancer forecast. The effective classification techniques were presented by Mavaddat *et al.* [51]. They used polygenic risk factors for analyzing breast cancer and its subtypes. In the prediction of the result, the family history is jointly correlated with the polygenic risk factors. However, further improvement is needed for the classification of trained dataset. Dutta *et al.* [52] evaluated the breast cancer. The fuzzy logic with inference system is used to distinguish breast lumps with the tumour. The classification model used are naive based, decision tree and logic boost.

6. BIG DATA IN HEALTHCARE

Business and clinical models may be transformed by big data analytics in order to offer care more intelligently and efficiently [53]. De-identified health data may be integrated to allow for secondary usage of the data. It may also help with self-decision making by spotting patterns and interpreting connections. Analytics of big data in healthcare practise may assist to discover diseases earlier, to accurately forecast their course, to identify deviations from a healthy condition, to influence the course of diseases and to detect fraud. As a result of this information, healthcare organisations are able to tailor their forecasts and treatment plans to the individual, so reducing waste and increasing efficiency; and individual patients are encouraged to keep a healthy lifestyle by receiving practical advice. It's possible to discover low-frequency occurrences that have a major clinical effect thanks to the power of big data. Several usability studies have been summarised in Table 2 based on their intended use.

Table 2. Healthcare big data use cases

Application area	Studies by:
Genomics	[52], [54]
Pharmacology and clinical pharmacology	[55], [56]
Patient-centered medical care	[57]–[59]
Precision medicine	[60]–[62]
Elderly care	[63], [64]
Mental health	[65], [66]
Cardiovascular disease	[67], [68]
Diabetes	[69]–[72]
Gynecology	[73]–[75]
Nephrology	[76]–[78]
Oncology	[79], [80]
Ophthalmology	[81], [82]
Urology	[83], [84]

To keep up with the tremendous growth in biomedical and healthcare papers, we've undertaken a comprehensive assessment of healthcare analytics across five subfields as illustrated in Table 3. Big data analytics is being used in the field of bioinformatics because to its complicated and enormous datasets, according to the results of the study. Many bioinformatics tools, methodologies, and platforms are available for analysing biological, genomic, protein, and gene sequencing data.

Table 3. Comparison of the literature

Healthcare discipline	Analytical methods for large data sets	Studies
Imaging informatics and medical image processing	Image visualization	[85], [86]
	Image classification	[18], [29], [49]
	Image retrieval	[87], [88]
	Data and workflow sharing	[89]
Bioinformatics	Feature selection	[90], [91]
	Classification	[11], [12], [14], [41], [42], [44], [47]
	Clustering	[92]
	Microarray data analysis	[93]
	Protein-protein interaction	[94]
	Protein sequencing	[95]
Clinical informatics	Storage of EHR	[93]
	Retrieval of EHR	[87]
	Data sharing via the use of interactive data retrieval	[88]
	Recommendation for treatment	[46]
	Disease prediction, progress and diagnosis	[24], [28], [37]–[39]
Public health informatics	Surveillance of infectious diseases	[95]
	Management of population health	[53]
	Management of mental illness	[15], [35], [64]

7. THEORETICAL FOUNDATION FOR BIG DATA ANALYTICS

A standard healthcare informatics or analytics project's conceptual framework is quite similar to that of a big data analytics project. The main difference is in the way processing is carried out. If you are working on a standard health analytics project, you may use an in-house business intelligence product installed on a desktop or laptop to conduct your analysis. Processing is distributed over numerous nodes due to the sheer size of big data. Distributed processing has been around for a long time. Although its usage in huge data sets is relatively new, healthcare practitioners are beginning to leverage their massive data repositories to get insights for making better health-related choices. As a result of the availability of cloud-based open source technologies like Hadoop/MapReduce, healthcare organisations are increasingly relying on big data analytics. Traditional health analytics solutions have grown more user-friendly and clearer, despite the fact that their algorithms and models are quite similar to those of big data. It is difficult to learn and use big data analytics technologies because of their complexity and the need to use a wide range of skill sets. Open-source tools and platforms that have arisen in an ad hoc way are lacking in features that vendor-driven proprietary products provide, such support and usability. Complexity starts with data, as seen in Figure 1.

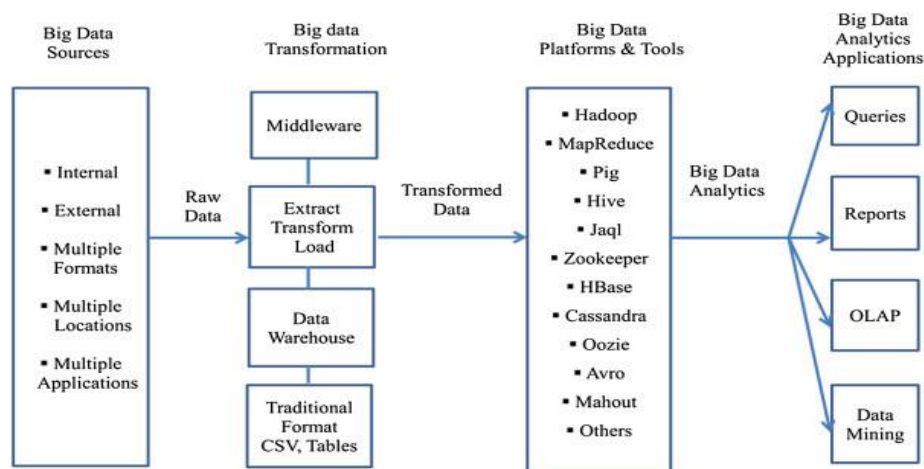


Figure 1. Applied big data analytics conceptual architecture

Big data analytics platforms in healthcare need to be able to perform the essential activities of data processing. There are a variety of factors to consider when evaluating a platform, including the following: accessibility, consistency, usability, usability at various levels of granularity, privacy and security, and the capacity to ensure quality. Open source platforms have their pros and disadvantages, and this is no different for those platforms. Big data analytics in healthcare must be packaged in a way that is menu-driven, user-friendly, and transparent in order to thrive. In healthcare, real-time big data analytics is essential. Is there a time lag between data gathering and analysis? Another need for widespread acceptance is the ability to access various analytics techniques, models, and methodologies through a drop-down menu. The management concerns of ownership, governance, and standards must be taken into account. The concerns of continual data collection and data purification are also intertwined. There is a lack of standardisation in health care data since information is typically fragmented, or created in outdated IT systems that are incompatible. This is a huge problem that has to be addressed.

8. ADVANTAGES OF HEALTHCARE ANALYTICS DATA

The use of prospective healthcare analytics by Big Data alters the healthcare industry. As a consequence of using big data analytics, doctors are able to make rapid treatment choices. Analytics is helping the healthcare industry make money. Analytical tools, data collecting, data sharing via electronic health records (EHR), electronic medical record (EMR), and interchange of medical information may be used to analyse healthcare data. Disease may be detected and predicted earlier and treated more quickly with the improvements in healthcare standards. There must be a process in place for new enrollees to be assessed for health risk. IBM utilises unstructured information management architecture (UIMA) to anticipate and analyse cardiac failures using big data from EHR data at the earliest stages of the illness. These open data sources may also be used to detect the transmission of infectious illnesses and analyse them to anticipate how they will spread in the future. When it comes to big data analytics, the challenge is to figure out the probabilities from databases of different sizes. The benefits of healthcare analytics programmes are evident, even if certain challenges are left unaddressed. By combining, digitising, and making effective use of big data, hospital networks may reap enormous advantages for their healthcare operations. Benefits include detecting illnesses at an early stage, analysing them, and then treating them efficiently. Applied analytics will be used to manage the associated dataset since each record is unique and its entry was made in. Big data analytics can answer a slew of questions.

9. CHALLENGES IN HEALTHCARE

From new disease outbreaks to ensuring a high level of operational efficiency, the healthcare industry faces several problems. In the development of healthcare applications, data mining and data analytics hold great potential for addressing these difficulties, but success is based on the availability of high-quality data, and there is no magic recipe for efficiently applying data analytics approaches to every circumstance. As a result, the creation of data analytics-based applications is dependent on the storage, preparation, and mining of data. As a result of the sheer volume and complexity of the data involved, chemical analytics may be difficult to implement. Data complexity, regulatory compliance, access to data, information security, inter-operability, efficient analytics methodologies, security, manageability, development, maturity, re-usability are some of the issues that need to be addressed to overcome these obstacles.

9.1. Multiple source information management

Despite the massive expansion in EHR adoption, making this data meaningful, readable, and relevant to clinicians and patients remains a challenge. The healthcare industry must find out how to keep, manage, and distribute all this data. Interoperability may help us tackle this problem. Lack of EHR compatibility complicates healthcare big data analytics. The establishment of a new infrastructure that enables all data providers to communicate and share data is required.

9.2. Security, privacy, and trust

Everyone involved in the healthcare business has a stake in protecting the privacy and security of patient data. Each of us bears some of the weight of the burden. A well-functioning healthcare system is dependent on patient privacy and information security to achieve better health results, healthier individuals, and more cost-effective expenditure. If a patient has any doubts about the confidentiality of their health information, he or she may refuse to divulge it or ask their physician not to record it. As a result, the patient is placed at danger, as well as the organisation, in terms of clinical results and operational efficiency analyses, as a result of this mentality. If patients are to benefit, providers and people must have confidence in the privacy and security of their health information.

9.3. Advanced analysis methods

Wearable technologies, patient-centered care, and other technological breakthroughs are reshaping the health care business. Even while EHRs have made data collecting easier, they lack the capacity to combine, convert or do analytics on the information collected. An inadequate amount of data may be gleaned from the little information provided by retroactive reporting. Analyzing large amounts of complicated data may be done using any number of methods, strategies, or instruments at one's disposal.

Based on a sample of the complete dataset, traditional machine learning uses statistical analysis. Traditional machine learning approaches are inefficient and computationally infeasible when applied to this data set. Analysts are able to concentrate on complicated data analysis approaches that can be scaled to suit the volume, velocity, and diversity in the healthcare data.

9.4. Data quality

Most hospitals utilize real-time data monitors (particularly ICUs), however real-time data analytics are not in use. With real-time data collecting, hospitals will be able to identify infections before they spread, monitor treatment progress, and choose better treatments in order to minimize morbidity and mortality. This will be possible in the near future thanks to real-time data analytics in the healthcare business. Data standards and device interoperability are essential if we are to accomplish real-time processing of data.

10. CONCLUSION

Most hospitals utilise real-time data monitors (particularly ICUs), however real-time data analytics are not in use. With real-time data collecting, hospitals will be able to identify infections before they spread, monitor treatment progress, and choose better treatments in order to minimise morbidity and mortality. This will be possible in the near future thanks to real-time data analytics in the healthcare business. Data standards and device interoperability are essential if we are to accomplish real-time processing of data.

REFERENCES

- [1] W. B. van Dijk *et al.*, "Text-mining in electronic healthcare records can be used as efficient tool for screening and data collection in cardiovascular trials: A multicenter validation study," *Journal of Clinical Epidemiology*, vol. 132, pp. 97–105, Apr. 2021, doi: 10.1016/j.jclinepi.2020.11.014.
- [2] A. K. Agarwal, R. G. Tiwari, R. K. Kaushal, and N. Kumar, "A systematic analysis of applications of blockchain in healthcare," in *2021 6th International Conference on Signal Processing, Computing and Control (ISPCC)*, Oct. 2021, pp. 413–417, doi: 10.1109/ISPCC53510.2021.9609339.
- [3] R. G. Tiwari, A. K. Agarwal, R. K. Kaushal, and N. Kumar, "Prophetic analysis of bitcoin price using machine learning approaches," in *2021 6th International Conference on Signal Processing, Computing and Control (ISPCC)*, Oct. 2021, pp. 428–432, doi: 10.1109/ISPCC53510.2021.9609419.
- [4] R. G. Tiwari, M. Husain, B. Gupta, and A. Agrawal, "Amalgamating contextual information into recommender system," in *2010 3rd International Conference on Emerging Trends in Engineering and Technology*, Nov. 2010, pp. 15–20, doi: 10.1109/ICETET.2010.110.
- [5] M. L. Kolling *et al.*, "Data mining in healthcare: Applying strategic intelligence techniques to depict 25 years of research development," *International Journal of Environmental Research and Public Health*, vol. 18, no. 6, pp. 1–21, Mar. 2021, doi: 10.3390/ijerph18063099.
- [6] X. Liu, Y. Ding, H. Tang, and F. Xiao, "A data mining-based framework for the identification of daily electricity usage patterns and anomaly detection in building electricity consumption data," *Energy and Buildings*, vol. 231, pp. 1–22, Jan. 2021, doi: 10.1016/j.enbuild.2020.110601.
- [7] S. Saxena, P. Kumar, and R. Gaurang, "Two-step technique for prediction analysis using k-means clustering algorithm," *International Journal of Computer Applications*, vol. 166, no. 9, pp. 9–12, May 2017, doi: 10.5120/ijca2017914110.
- [8] V. Khullar, S. Ahuja, R. G. Tiwari, and A. K. Agarwal, "Investigating efficacy of deep trained soil classification system with augmented data," in *2021 9th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions) (ICRITO)*, Sep. 2021, pp. 1–5, doi: 10.1109/ICRITO51393.2021.9596515.
- [9] G. Taguchi and R. Jugulum, *The mahalanobis-taguchi strategy: A pattern technology System*. New York: John Wiley & Sons, 2002.
- [10] R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern classification*, 2nd ed. New York: John Wiley & Sons, 2000.
- [11] Y. Tang, X. Li, Y. Chen, Y. Zhong, A. Jiang, and C. Wang, "High-accuracy classification of attention deficit hyperactivity disorder with l 2, 1-norm linear discriminant analysis and binary hypothesis testing," *IEEE Access*, vol. 8, pp. 56228–56237, 2020, doi: 10.1109/ACCESS.2020.2982401.
- [12] J. P. Li, A. U. Haq, S. U. Din, J. Khan, A. Khan, and A. Saboor, "Heart disease identification method using machine learning classification in e-healthcare," *IEEE Access*, vol. 8, pp. 107562–107582, 2020, doi: 10.1109/ACCESS.2020.3001149.
- [13] L. M. Abualigah, A. T. Khader, and E. S. Hanandeh, "A new feature selection method to improve the document clustering using particle swarm optimization algorithm," *Journal of Computational Science*, vol. 25, pp. 456–466, Mar. 2018, doi: 10.1016/j.jocs.2017.07.018.
- [14] Y. Su *et al.*, "Colon cancer diagnosis and staging classification based on machine learning and bioinformatics analysis," *Computers in Biology and Medicine*, vol. 145, p. 105409, Jun. 2022, doi: 10.1016/j.compbiomed.2022.105409.
- [15] A. Y. Naser *et al.*, "Mental health status of the general population, healthcare professionals, and university students during 2019 coronavirus disease outbreak in Jordan: A cross-sectional study," *Brain and Behavior*, vol. 10, no. 8, pp. 1–13, Aug. 2020, doi: 10.1002/brb3.1730.
- [16] S. McLachlan, K. Dube, G. A. Hitman, N. E. Fenton, and E. Kyrimi, "Bayesian networks in healthcare: Distribution by medical




- condition,” *Artificial Intelligence in Medicine*, vol. 107, pp. 1–7, Jul. 2020, doi: 10.1016/j.artmed.2020.101912.
- [17] J. Jin, W. Sun, F. Al-Turjman, M. B. Khan, and X. Yang, “Activity pattern mining for healthcare,” *IEEE Access*, vol. 8, pp. 56730–56738, 2020, doi: 10.1109/ACCESS.2020.2981670.
- [18] M. Z. Uddin, M. M. Hassan, A. Alsanad, and C. Savaglio, “A body sensor data fusion and deep recurrent neural network-based behavior recognition approach for robust healthcare,” *Information Fusion*, vol. 55, pp. 105–115, Mar. 2020, doi: 10.1016/j.inffus.2019.08.004.
- [19] D. V. Dimitrov, “Medical internet of things and big data in healthcare,” *Healthcare Informatics Research*, vol. 22, no. 3, pp. 156–163, 2016, doi: 10.4258/hir.2016.22.3.156.
- [20] S. Akter, R. Bandara, U. Hani, S. F. Wamba, C. Foropon, and T. Papadopoulos, “Analytics-based decision-making for service systems: A qualitative study and agenda for future research,” *International Journal of Information Management*, vol. 48, pp. 85–95, Oct. 2019, doi: 10.1016/j.ijinfomgt.2019.01.020.
- [21] A. Clark, J. Q. Ng, N. Morlet, and J. B. Semmens, “Big data and ophthalmic research,” *Survey of Ophthalmology*, vol. 61, no. 4, pp. 443–465, Jul. 2016, doi: 10.1016/j.survophthal.2016.01.003.
- [22] A. K. Swain, “Mining big data to support decision making in healthcare,” *Journal of Information Technology Case and Application Research*, vol. 18, no. 3, pp. 141–154, Sep. 2016, doi: 10.1080/15228053.2016.1245522.
- [23] B. R. Schatz, “National surveys of population health: Big data analytics for mobile health monitors,” *Big Data*, vol. 3, no. 4, pp. 219–229, Dec. 2015, doi: 10.1089/big.2015.0021.
- [24] A. R. M. Forkan, I. Khalil, and M. Atiquzzaman, “ViSiBiD: A learning model for early discovery and real-time prediction of severe clinical events using vital signs as big data,” *Computer Networks*, vol. 113, pp. 244–257, Feb. 2017, doi: 10.1016/j.comnet.2016.12.019.
- [25] K. R. Ghani, K. Zheng, J. T. Wei, and C. P. Friedman, “Harnessing big data for health care and research: Are urologists ready?,” *European Urology*, vol. 66, no. 6, pp. 975–977, Dec. 2014, doi: 10.1016/j.eururo.2014.07.032.
- [26] J. Roski, G. W. Bo-Linn, and T. A. Andrews, “Creating value in health care through big data: Opportunities and policy implications,” *Health Affairs*, vol. 33, no. 7, pp. 1115–1122, Jul. 2014, doi: 10.1377/hlthaff.2014.0147.
- [27] N. S. Downing *et al.*, “Describing the performance of U.S. hospitals by applying big data analytics,” *PLOS ONE*, vol. 12, no. 6, pp. 1–14, Jun. 2017, doi: 10.1371/journal.pone.0179603.
- [28] M. Chen, Y. Hao, K. Hwang, L. Wang, and L. Wang, “Disease prediction by machine learning over big data from healthcare communities,” *IEEE Access*, vol. 5, pp. 8869–8879, 2017, doi: 10.1109/ACCESS.2017.2694446.
- [29] V. Khullar, R. G. Tiwari, A. K. Agarwal, and S. Dutta, “Physiological signals based anxiety detection using ensemble machine learning,” in *Cyber Intelligence and Information Retrieval*, 2022, vol. 291, pp. 597–608, doi: 10.1007/978-981-16-4284-5_53.
- [30] S. W. Lo and D. Jamroz, “Genomics and epidemiological surveillance,” *Nature Reviews Microbiology*, vol. 18, no. 9, pp. 478–478, Sep. 2020, doi: 10.1038/s41579-020-0421-0.
- [31] A. Khalifa and S. Meystre, “Adapting existing natural language processing resources for cardiovascular risk factors identification in clinical notes,” *Journal of Biomedical Informatics*, vol. 58, pp. 128–132, Dec. 2015, doi: 10.1016/j.jbi.2015.08.002.
- [32] F. J. Martin-Sanchez, V. Aguiar-Pulido, G. H. Lopez-Campos, N. Peek, and L. Sacchi, “Secondary use and analysis of big data collected for patient care,” *Yearbook of Medical Informatics*, vol. 26, no. 01, pp. 28–37, Sep. 2017, doi: 10.15265/IY-2017-008.
- [33] E. Ford, J. A. Carroll, H. E. Smith, D. Scott, and J. A. Cassell, “Extracting information from the text of electronic medical records to improve case detection: A systematic review,” *Journal of the American Medical Informatics Association*, vol. 23, no. 5, pp. 1007–1015, Sep. 2016, doi: 10.1093/jamia/ocv180.
- [34] D. Al-Jumeily, A. Hussain, C. Mallucci, and C. Oliver, *Applied computing in medicine and health*. Amsterdam: Elsevier, 2016, doi: 10.1016/C2014-0-02198-3.
- [35] D. D. Luxton, “An introduction to artificial intelligence in behavioral and mental health care,” in *Artificial Intelligence in Behavioral and Mental Health Care*, Elsevier, 2016, pp. 1–26, doi: 10.1016/B978-0-12-420248-1.00001-5.
- [36] P. Amirian *et al.*, “Using big data analytics to extract disease surveillance information from point of care diagnostic machines,” *Pervasive and Mobile Computing*, vol. 42, pp. 470–486, Dec. 2017, doi: 10.1016/j.pmcj.2017.06.013.
- [37] M. F. Faruque, Asaduzzaman, and I. H. Sarker, “Performance analysis of machine learning techniques to predict diabetes mellitus,” in *2019 International Conference on Electrical, Computer and Communication Engineering (ECCE)*, Feb. 2019, pp. 1–4, doi: 10.1109/ECCE.2019.8679365.
- [38] Widiyawati, A. Nurrahman, B. Darmawan, A. I. Jaya, and R. Ratianingsih, “Prediction of obesity, type 2 diabetes mellitus, metabolic syndrome and coronary heart disease using Backpropagation algorithm,” *Journal of Physics: Conference Series*, vol. 1218, no. 1, pp. 1–5, May 2019, doi: 10.1088/1742-6596/1218/1/012039.
- [39] N. Singh and P. Singh, “A stacked generalization approach for diagnosis and prediction of type 2 diabetes mellitus,” in *Advances in Intelligent Systems and Computing*, 2020, pp. 559–570, doi: 10.1007/978-981-13-8676-3_47.
- [40] T. Rashid and S. Abdullah, “A hybrid of artificial bee colony, genetic algorithm, and neural network for diabetic mellitus diagnosing,” *ARO-The Scientific Journal of Koya University*, vol. 6, no. 1, pp. 55–64, Jun. 2018, doi: 10.14500/aro.10368.
- [41] D. Sejdinović *et al.*, “Classification of prediabetes and type 2 diabetes using artificial neural network,” in *CMBEIH 2017*, vol. 62, 2017, pp. 685–689, doi: 10.1007/978-981-10-4166-2_103.
- [42] D. K. Choubey, M. Kumar, V. Shukla, S. Tripathi, and V. K. Dhandhanian, “Comparative analysis of classification methods with PCA and LDA for Diabetes,” *Current Diabetes Reviews*, vol. 16, no. 8, pp. 833–850, Sep. 2020, doi: 10.2174/1573399816666200123124008.
- [43] W. H. A. W. Zunaidi, R. R. Saedudin, Z. A. Shah, S. Kasim, C. S. Seah, and M. Abdurrohman, “Performances analysis of heart disease dataset using different data mining classifications,” *International Journal on Advanced Science, Engineering and Information Technology*, vol. 8, no. 6, pp. 2677–2682, Dec. 2018, doi: 10.18517/ijaseit.8.6.5042.
- [44] C. B. C. Latha and S. C. Jeeva, “Improving the accuracy of prediction of heart disease risk based on ensemble classification techniques,” *Informatics in Medicine Unlocked*, vol. 16, pp. 1–9, 2019, doi: 10.1016/j.imu.2019.100203.
- [45] Y. Olatunde, L. Omotosho, and C. Akanbi, “Comparison of adaboost and bagging ensemble method for prediction of heart disease,” *Annals. Computer Science Series*, vol. 17, no. 1, pp. 268–279, 2019.
- [46] M. Shah, A. Modi, J. Jethwa, and H. Bhor, “Heart disease prediction and recommendation,” *International Research Journal of Engineering and Technology (IRJET)*, vol. 6, no. 4, pp. 2334–2337, 2019.
- [47] L. Dora, S. Agrawal, R. Panda, and A. Abraham, “Optimal breast cancer classification using Gauss–Newton representation based algorithm,” *Expert Systems with Applications*, vol. 85, pp. 134–145, Nov. 2017, doi: 10.1016/j.eswa.2017.05.035.
- [48] S. Reis *et al.*, “Automated classification of breast cancer stroma maturity from histological images,” *IEEE Transactions on Biomedical Engineering*, vol. 64, no. 10, pp. 2344–2352, Oct. 2017, doi: 10.1109/TBME.2017.2665602.
- [49] Y. Xiao, J. Wu, Z. Lin, and X. Zhao, “A deep learning-based multi-model ensemble method for cancer prediction,” *Computer Methods and Programs in Biomedicine*, vol. 153, pp. 1–9, Jan. 2018, doi: 10.1016/j.cmpb.2017.09.005.

- [50] S. Rahimi, A. Abdollahpouri, and P. Moradi, "A multi-objective particle swarm optimization algorithm for community detection in complex networks," *Swarm and Evolutionary Computation*, vol. 39, pp. 297–309, Apr. 2018, doi: 10.1016/j.swevo.2017.10.009.
- [51] N. Mavaddat *et al.*, "Polygenic risk scores for prediction of breast cancer and breast cancer subtypes," *The American Journal of Human Genetics*, vol. 104, no. 1, pp. 21–34, Jan. 2019, doi: 10.1016/j.ajhg.2018.11.002.
- [52] S. Dutta, S. Ghatak, A. Sarkar, R. Pal, R. Pal, and R. Roy, "Cancer prediction based on fuzzy inference system," in *Smart Innovations in Communication and Computational Sciences*, 2019, pp. 127–136, doi: 10.1007/978-981-13-2414-7_13.
- [53] S. Dash, S. K. Shakyawar, M. Sharma, and S. Kaushik, "Big data in healthcare: Management, analysis and future prospects," *Journal of Big Data*, vol. 6, no. 1, pp. 1–25, Dec. 2019, doi: 10.1186/s40537-019-0217-0.
- [54] K. Y. Ngiam and I. W. Khor, "Big data and machine learning algorithms for health-care delivery," *The Lancet Oncology*, vol. 20, no. 5, pp. 262–273, May 2019, doi: 10.1016/S1470-2045(19)30149-4.
- [55] G. Singh, D. Schulthess, N. Hughes, B. Vannieuwenhuyse, and D. Kalra, "Real world big data for clinical research and drug development," *Drug Discovery Today*, vol. 23, no. 3, pp. 652–660, Mar. 2018, doi: 10.1016/j.drudis.2017.12.002.
- [56] M. Viceconti, P. Hunter, and R. Hose, "Big data, big knowledge: Big data for personalized healthcare," *IEEE Journal of Biomedical and Health Informatics*, vol. 19, no. 4, pp. 1209–1215, Jul. 2015, doi: 10.1109/JBHI.2015.2406883.
- [57] V. Jagadeeswari, V. Subramaniaswamy, R. Logesh, and V. Vijayakumar, "A study on medical internet of things and big data in personalized healthcare system," *Health Information Science and Systems*, vol. 6, no. 1, pp. 1–20, Dec. 2018, doi: 10.1007/s13755-018-0049-x.
- [58] E. M. Cahan, T. Hernandez-Boussard, S. Thadaneys-Israni, and D. L. Rubin, "Putting the data before the algorithm in big data addressing personalized healthcare," *NPJ Digital Medicine*, vol. 2, no. 1, pp. 1–6, Dec. 2019, doi: 10.1038/s41746-019-0157-2.
- [59] T. Hulsen *et al.*, "From big data to precision medicine," *Frontiers in Medicine*, vol. 6, pp. 1–14, Mar. 2019, doi: 10.3389/fmed.2019.00034.
- [60] T. Qian, S. Zhu, and Y. Hoshida, "Use of big data in drug development for precision medicine: An update," *Expert Review of Precision Medicine and Drug Development*, vol. 4, no. 3, pp. 189–200, May 2019, doi: 10.1080/23808993.2019.1617632.
- [61] O. Anya and H. Tawfik, "Leveraging big data analytics for personalized elderly care: Opportunities and challenges," in *Applied Computing in Medicine and Health*. Elsevier, 2016, pp. 99–124, doi: 10.1016/B978-0-12-803468-2.00005-9.
- [62] R. G. Tiwari, M. Husain, V. Srivastava, and K. Singh, "A hypercube novelty model for comparing e-commerce and M-commerce," in *Proceedings of the 2011 International Conference on Communication, Computing & Security - ICCCS '11*, 2011, pp. 616–619, doi: 10.1145/1947940.1948068.
- [63] M. Conway and D. O'Connor, "Social media, big data, and mental health: Current advances and ethical implications," *Current Opinion in Psychology*, vol. 9, pp. 77–82, Jun. 2016, doi: 10.1016/j.copsyc.2016.01.004.
- [64] R. Stewart and K. Davis, "'Big data' in mental health research: Current status and emerging possibilities," *Social Psychiatry and Psychiatric Epidemiology*, vol. 51, no. 8, pp. 1055–1072, Aug. 2016, doi: 10.1007/s00127-016-1266-8.
- [65] Y. Liang, X. Zheng, and D. D. Zeng, "A survey on big data-driven digital phenotyping of mental health," *Information Fusion*, vol. 52, pp. 290–307, Dec. 2019, doi: 10.1016/j.inffus.2019.04.001.
- [66] J. A. Leopold, B. A. Maron, and J. Loscalzo, "The application of big data to cardiovascular disease: paths to precision medicine," *Journal of Clinical Investigation*, vol. 130, no. 1, pp. 29–38, Jan. 2020, doi: 10.1172/JCI129203.
- [67] J. Kim, "Big data, health informatics, and the future of cardiovascular medicine," *Journal of the American College of Cardiology*, vol. 69, no. 7, pp. 899–902, Feb. 2017, doi: 10.1016/j.jacc.2017.01.006.
- [68] J. M. M. Rumbold, M. O'Kane, N. Philip, and B. K. Pierscionek, "Big data and diabetes: The applications of big data for diabetes care now and in the future," *Diabetic Medicine*, vol. 37, no. 2, pp. 187–193, 2020, doi: 10.1111/dme.14044.
- [69] M. Chen, J. Yang, J. Zhou, Y. Hao, J. Zhang, and C. H. Youn, "5G-smart diabetes: Toward personalized diabetes diagnosis with healthcare big data clo," *IEEE Communications Magazine*, vol. 56, no. 4, pp. 16–23, 2018, doi: 10.1109/MCOM.2018.1700788.
- [70] B. G. M. Bai, B. M. Nalini, and J. Majumdar, "Analysis and detection of diabetes using data mining techniques—a big data application in health care," in *Emerging Research in Computing, Information, Communication and Applications*, 2019, pp. 443–455, doi: 10.1007/978-981-13-5953-8_37.
- [71] T. Nibareke and J. Laassiri, "Using big data-machine learning models for diabetes prediction and flight delays analytics," *Journal of Big Data*, vol. 7, no. 1, pp. 1–18, Dec. 2020, doi: 10.1186/s40537-020-00355-0.
- [72] R. Khamisy-Farah, L. B. Furstenuau, J. D. Kong, J. Wu, and N. L. Bragazzi, "Gynecology meets big data in the disruptive innovation medical era: State-of-art and future prospects," *International Journal of Environmental Research and Public Health*, vol. 18, no. 10, pp. 1–13, May 2021, doi: 10.3390/ijerph18105058.
- [73] E. A. Ereksion and C. B. Iglesia, "Improving patient outcomes in gynecology: The role of large data registries and big data analytics," *Journal of Minimally Invasive Gynecology*, vol. 22, no. 7, pp. 1124–1129, 2015, doi: 10.1016/j.jmig.2015.07.003.
- [74] A. Goodin *et al.*, "The power and pitfalls of big data research in obstetrics and gynecology: A consumer's guide," *Obstetrical and Gynecological Survey*, vol. 72, no. 11, pp. 669–682, 2017, doi: 10.1097/OGX.0000000000000504.
- [75] J. Saez-Rodriguez, M. M. Rinschen, J. Floege, and R. Kramann, "Big science and big data in nephrology," *Kidney International*, vol. 95, no. 6, pp. 1326–1337, 2019, doi: 10.1016/j.kint.2018.11.048.
- [76] N. Kaur, S. Bhattacharya, and A. J. Butte, "Big data in nephrology," *Nature Reviews Nephrology*, vol. 17, no. 10, pp. 676–687, Oct. 2021, doi: 10.1038/s41581-021-00439-x.
- [77] C. Thongprayoon *et al.*, "Promises of big data and artificial intelligence in nephrology and transplantation," *Journal of Clinical Medicine*, vol. 9, no. 4, pp. 1–14, Apr. 2020, doi: 10.3390/jcm9041107.
- [78] J. M. Purswani, A. P. Dicker, C. E. Champ, M. Cantor, and N. Ohri, "Big data from small devices: The future of smartphones in oncology," *Seminars in Radiation Oncology*, vol. 29, no. 4, pp. 338–347, 2019.
- [79] A. Major, S. M. Cox, and S. L. Volchenboum, "Using big data in pediatric oncology: Current applications and future directions," *Seminars in Oncology*, vol. 47, no. 1, pp. 56–64, 2020, doi: 10.1053/j.seminoncol.2020.02.006.
- [80] C.-Y. Cheng *et al.*, "Big data in ophthalmology," *Asia-Pacific Journal of Ophthalmology*, vol. 9, no. 4, pp. 291–298, Jul. 2020, doi: 10.1097/APO.0000000000000304.
- [81] C. Kern *et al.*, "Big data simulations for capacity improvement in a general ophthalmology clinic," *Graefe's Archive for Clinical and Experimental Ophthalmology*, vol. 259, no. 5, pp. 1289–1296, May 2021, doi: 10.1007/s00417-020-05040-9.
- [82] B. M. Z. Hameed *et al.*, "Big data analytics in urology: The story so far and the road ahead," *Therapeutic Advances in Urology*, vol. 13, pp. 1–13, Jan. 2021, doi: 10.1177/1756287221998134.
- [83] W.-J. Kim, P. Jin, W. H. Kim, and J. Kim, "Utilizing machine learning to discern hidden clinical values from big data in urology," *Investigative and Clinical Urology*, vol. 61, no. 3, pp. 239–241, 2020, doi: 10.4111/icu.2020.61.3.239.
- [84] P. A. Glemser, K. Engel, D. Simons, J. Steffens, H.-P. Schlemmer, and B. Orakcioglu, "A new approach for photorealistic




- visualization of rendered computed tomography images,” *World Neurosurgery*, vol. 114, pp. 283–292, Jun. 2018, doi: 10.1016/j.wneu.2018.02.174.
- [85] C. Karmonik, T. B. Boone, and R. Khavari, “Workflow for visualization of neuroimaging data with an augmented reality device,” *Journal of Digital Imaging*, vol. 31, no. 1, pp. 26–31, Feb. 2018, doi: 10.1007/s10278-017-9991-4.
- [86] A. Khatami, M. Babaie, H. R. Tizhoosh, A. Khosravi, T. Nguyen, and S. Nahavandi, “A sequential search-space shrinking using CNN transfer learning and a Radon projection pool for medical image retrieval,” *Expert Systems with Applications*, vol. 100, pp. 224–233, Jun. 2018, doi: 10.1016/j.eswa.2018.01.056.
- [87] K. Karthik and S. S. Kamath, “A deep neural network model for content-based medical image retrieval with multi-view classification,” *The Visual Computer*, vol. 37, no. 7, pp. 1837–1850, Jul. 2021, doi: 10.1007/s00371-020-01941-2.
- [88] P. Ross and H. Pohjonen, “Images crossing borders: Image and workflow sharing on multiple levels,” *Insights into Imaging*, vol. 2, no. 2, pp. 141–148, 2011, doi: 10.1007/s13244-010-0059-y.
- [89] Z. Chen *et al.*, “Feature selection may improve deep neural networks for the bioinformatics problems,” *Bioinformatics*, vol. 36, no. 5, pp. 1542–1552, Oct. 2019, doi: 10.1093/bioinformatics/btz763.
- [90] R. J. Urbanowicz, R. S. Olson, P. Schmitt, M. Meeker, and J. H. Moore, “Benchmarking relief-based feature selection methods for bioinformatics data mining,” *Journal of Biomedical Informatics*, vol. 85, pp. 168–188, 2018, doi: 10.1016/j.jbi.2018.07.015.
- [91] D. T. Patel, “Big data analytics in bioinformatics,” *Biotechnology: Concepts, Methodologies, Tools, and Applications*, pp. 1967–1984, 2019, doi: 10.4018/978-1-5225-8903-7.ch080.
- [92] J. Xu *et al.*, “Microarray expression profiling and bioinformatics analysis of circular rna expression in lung squamous cell carcinoma,” *American Journal of Translational Research*, vol. 10, no. 3, pp. 771–783, 2018.
- [93] L.-P. Li, Y.-B. Wang, Z.-H. You, Y. Li, and J.-Y. An, “PCLPred: A bioinformatics method for predicting protein–protein interactions by combining relevance vector machine model with low-rank matrix approximation,” *International Journal of Molecular Sciences*, vol. 19, no. 4, pp. 1–13, Mar. 2018, doi: 10.3390/ijms19041029.
- [94] M. Tu, S. Cheng, W. Lu, and M. Du, “Advancement and prospects of bioinformatics analysis for studying bioactive peptides from food-derived protein: Sequence, structure, and functions,” *TrAC Trends in Analytical Chemistry*, vol. 105, pp. 7–17, Aug. 2018, doi: 10.1016/j.trac.2018.04.005.
- [95] G. Sun, N. V. Trung, L. T. Hoi, P. T. Hiep, K. Ishibashi, and T. Matsui, “Visualisation of epidemiological map using an internet of things infectious disease surveillance platform,” *Critical Care*, vol. 24, no. 1, pp. 1–2, 2020, doi: 10.1186/s13054-020-03132-w.

BIOGRAPHIES OF AUTHORS






Dolley Srivastava    is an Assistant Professor at the Department of Information Technology, Babu Banarasi Das Institute of Technology and Management, Lucknow, where she has been as faculty member since 2012. She received the B.Tech engineering degree from Gautam Budh University, Noida in Information technology branch. She has done her M.Tech in Department of Computer Science from babu Banarasi Das University from Lucknow in 2016. Now she is doing Ph.D. from Maharishi University of Information Technology, Lucknow. She has 10 years of experience in teaching. Her research areas are web semantics, artificial intelligence, machine learning, big data, and deep learning. She has published two national patents and research papers in Scopus conferences. She can be contacted at email: dolley89@gmail.com.



Himanshu Pandey    has a Ph.D. in Computer Science and Engineering. Currently, he is working as an Assistant Professor in the Department of Computer Science and Engineering and has also worked as an Assistant Director in the Central Placement Cell and Information, Publication & Public Relations (IPPR) at Lucknow University, Lucknow. He has more than 10 years of extensive experience in the IT industry, academics, and administration. His research areas are software engineering, DSA, machine learning, and deep learning. He is guiding six Ph.D students from different Universities. He has organised more than 150 webinars, seminars, and workshops. He has published more than 45 papers in national and international conferences and journals. He has published 12 papers in Scopus journals. He can be contacted at email: hpandey010@gmail.com.



Ambuj Kumar Agarwal    has a Ph.D. in Computer Science. He has completed his M. Tech from MDU, Rohtak and he has done his M.Phil (CS) and MCA from Aligappa University and UPTU, Lucknow. Currently, he is working as a Professor in the CSE department at Sharda University, Greater Noida, India. He has more than 18 years of extensive experience in the IT industry and Academics in India. His research areas are software engineering, software testing, operating system, cloud, machine learning, and deep learning. He is guiding four Ph.D. Students of different universities. He is highly disciplined and organized. He has organized more than 20 conferences/seminars and workshops. He has chaired more than 10 conference special sessions. He is serving in an advisory capacity in international conferences as a technical programme committee/organizing committee. He has published more the 50 papers in national and international conferences and journals. He is also published 14 Indian patents (both utility and design) and 2 Australian patents are granted. He has published 10 papers in SCI/Scopus journals. He can be contacted at email: ambuj4u@gmail.com.