UTILIZING MACHINE LEARNING IN AREAS OF HEALTH CARE

Abstract

The current state of the health care industry is pushing businesses in the direction of implementing new health IT infrastructure strategies. As a result of the daily increase in data quantities, businesses need to make investments in new information technology in order to satisfy the requirements of hospitals and patients. This is necessary because hospitals require access in real time to key diagnostic information that has the potential to improve the quality of care. The Centers for Medicare and Medicaid Services (CMS) have implemented substantial changes that affect the ways in which healthcare providers create digital data, store that data, and evaluate that data. The World Wide Web currently makes a massive quantity of data, the majority of which is unstructured and of varying types, very simple to access. Techniques from the field of machine learning (ML) can be utilized to automatically gather, categorize, or cluster observations from huge amounts of data. The analytics of Big Data (BD) has resulted in numerous recent projects in both theory and practice and has spurred interest within the community of machine learning researchers. As hospitals begin to implement sophisticated data analytics capabilities, the next big thing in the healthcare business will be developing prediction models for BD problems using ML as a service. This is all poised to alter Health care Analytics (HcA) towards a better future. Because BD is connected with a high variety, variability, and velocity of data from the internet of things (IoT), powerful machine learning techniques are necessary to develop knowledge that can be utilized to enhance the results of the process of providing patient care.

Keywords : HealthCare Analytics, Machine Learning, Big Data, Internet of Things, Artificial Intelligence, Deep Learning;

Authors

Venkata Pavan Kumar Savala
Associate Professor & Head,
Dept. of CSE-DS,
Siddhartha Institute of Engineering & Technology,
Ibrahimpatnam, Hyderabad.
venkatapavankumarsavala@gmail.com

G N R Prasad
Associate Professor,
Dept. of MCA,
Chaitanya Bharathi Institute of Technology,
Hyderabad.
gnrp@cbit.ac.in

Ponnaboyina Ranganath
Research Scholar,
Acharya Nagarjuna University,
Guntur.
ranganathponnaboyina@gmail.com

P. V. Ravi Kumar
Associate Professor,
Dept. of CSE,
Krishna Chaitanya Institute of Technology & Sciences,
Markapur.
putta.msc@gmail.com

P M Yohan
Professor & Principal,
CSI Wesley Institute of Technology and Sciences,
Secunderabad.
pmyohan@rediff.com

SK Althaf Hussain Basha
Professor and R&D Coordinator,
Krishna Chaitanya Institute of Technology & Sciences,
Markapur.
althafbashacse@gmail.com

I. INTRODUCTION

John McCarthy introduced the concept of artificial intelligence (AI) for the first time in 1956. AI refers to computers that have the ability to do activities that are unique to humanoid intelligence as well as the potential to learn without being specifically programmed. Processing, identifying, comprehending, planning, problem-solving, and learning are all components of AI. Machine learning techniques are methods in which computers may learn on their own. It is possible to accomplish this without resorting to obvious automation. Machine learning is a function of artificial intelligence that provides a framework to automatically learn and advance analysis based on prior information. The input and output of a code may be assimilated, which can then be used to construct ML algorithms. Expert systems that use AI are distinct from those that use ML, which stands for machine intelligence.

II. THE GOAL OF THIS CHAPTER'S LEARNING

Despite the fact that the healthcare business has been somewhat sluggish to absorb ML for a variety of very reasonable reasons, ML will soon completely revolutionize the healthcare industry. This is made feasible by systematically applying ML to utilize unstructured data that is accessible in the form of blood tests, biopsy findings, vital signs, medical histories, medication histories, photographs, genetic profiles, epidemiological data, physician notes, and even medical research publications. This data may be obtained from a variety of sources. To reorganize the universe, a combination of AI, Cognitive Learning (CL), and Deep Learning (DL) is being developed as a set of tools that can make use of both organized and unstructured data to establish relationships, rules, and patterns. This set of technologies is known as the "AI-CL-DL stack." Businesses in a variety of industries, including retail, financial services, entertainment, and transportation, are already making use of ML. The implementation of machine learning may be experienced by users on a daily basis in the form of tailored internet adverts and other examples. The use of ML in HcA allows for diagnosis, the prediction of outcomes, the provision of follow-up, and the customization of therapies; it also has the potential to make healthcare more intelligent, efficient, safe, and cost-effective for everyone. With the data that is now accessible, there is a big and enormous opportunity for the use of AI and ML in the field of healthcare. The primary purpose of this chapter is to lead to the identification of the issues that arise when attempting to use AI and ML approaches from the perspective of developers and researchers in order to construct and offer the technological infrastructure, processing, and storage capacity that will enable them to produce ML solutions.

III. THE CONTEXT AND THE HIGHLIGHTS

Big data refers to the practice of carrying out a variety of operations on huge datasets, including organizing, processing, and the collection of insights from the data. In spite of the fact that dealing with data is still a challenge, value has been placed on computers. The task of big data systems is to deal with massive amounts of data. As a result, it includes a variety of techniques that provide it the ability to work on such data even when it is stored and formatted in a variety of ways. Problems with heterogeneity, timeliness, scalability, privacy, and complexity in Big Data slow development in all phases that might derive value from data by using machine learning. The challenges begin at the very beginning of the data collecting

stage, when it is necessary to make judgments in an ad hoc way based on the massive information that was acquired. Understanding the data, maintaining, rejecting, or saving the data in a reliable manner, as well as producing the appropriate metadata, is the true research problem posed by the data collection. One further significant obstacle is that the data that will be accessible over the web will not be presented in an organized fashion. Blogs and messages are two examples of non-adequately organized text. Videos and photos, on the other hand, are structured for display and storage, but not for semantic content or search. It will be a big and demanding undertaking to convert this knowledge into a format that is appropriately organized for future research or reference. Methods of data linking are implemented in order to combine data obtained from a variety of sources into a single, more comprehensive dataset. This helps to enhance the value of the information. At this time, data may be found in a digital version on the World Wide Web. Because of this, enabling and making a connection for the data presents a great number of options but also presents a great deal of difficulty. In healthcare, the majority of linking strategies may integrate records from separate records into a single record if those data pertain to the same object. (An entity might be a patient, an organization, a geographic location, etc.). Due to the complexity of the data and the lack of scalability of the algorithms that need to be evaluated, there are only a few additional major obstacles. These issues include evaluating the data, obtaining the data, organizing the data, and modeling the data.

Wei Fan and Albert Bifet have proposed three variables that should be considered while managing big data:

1. **Volume:** The amount of data on the web is much bigger than it was in the past, and the size of the data is expanding at an alarming rate; nevertheless, the tools that are now accessible can only analyze a small portion of the data that is already available.

2. **Variety:** Because of the Internet of Things (IoT), users now have access to a wide variety of data in a variety of formats, including text, video, audio, graphs, GPS coordinates, sensor data, and many more.

3. **Velocity:** Although a sizable amount of data is accessible in real time in the form of data streams, the study that is being conducted right now is primarily concerned with gathering relevant and valuable information.

Other characteristics also started exist in the big data are:

1. **Variability**: The end user desires to use the data in a variety of contexts and formats.

   It is now feasible to develop judgments based on the answers to questions that could not be quantified in the past. This presents a valuable opportunity. This confers value on the company for use by an organization that possesses compelling benefits.

   The development of machine learning techniques can provide Web Analytics (WA) with assistance in locating the tendencies that are promising, yet there are still opportunities for improvement in areas such as:

- **Data from a variety of sources:** Comprehending the many datasets that may be found on the internet is required for this step.

  Format of the Data because the data may have originated from a variety of sources, its formats and languages may be distinct from one another.

- **Preparation of the Data:** In order to conduct analyses, the data that has been obtained must first be prepared.

  Data security requires that all of the data that has been gathered and prepared be protected, as well as maintaining their confidentiality, integrity, and availability.

  On the market, there are a number of tools for Big Data, Data Mining (DM), and Warehouse, however these tools aren't always the best choice for some jobs, including ones that involve things like

  Association refers to the process of discovering links between various data sets.

  Classification refers to the process of determining and organizing data into a set of categories based on a few similar characteristics of the data.

  Clustering is the process of recognizing previously ignored and unrecognized facts and displaying them in a visually appealing manner.

  The process of identifying data trends that can lead to accurate predictions is known as forecasting.

  The goal of this line of study is to address the traditional issues that have been plaguing WA and devise an innovative approach to discovering interesting patterns by using the streams that are available on the real-time Web.

IV. PERSPECTIVE TAKEN AT THIS TIME

The Internet is now the most widely utilized, constantly expanding, and most extensive data storage platform available. The Internet has emerged as the most important source of knowledge due to the breadth and depth of the material it contains. Because the data format and the content of the web are so diverse, the accuracy of the data is dependent on the algorithms used in Web Mining (WM). As an example of current Internet usage, Google processes one billion search inquiries every single day. Every every day, Twitter processes 250 million new messages. There are 800 million and 4 billion views received each day on Facebook and YouTube, respectively. More than ninety percent of all online data was generated in the most recent couple of years, and the annual growth rate of data is predicted to be forty percent, with units of measurement in zettabytes. The widespread availability of the Internet as well as the fact that every individual as well as every business creates a significant amount of data on a daily basis are the two primary reasons for this phenomenon. The Internet has undergone significant development over the past decade and now plays a significant role in many day-to-day activities, such as shopping, reading the newspaper, seeking information and service provided by governments and enterprises helped web

marketing to overgrow in many developed countries, investment to web advertisements have grown big in comparison to television and newspaper. The significance of the Internet will only become more apparent as a result of this expansion, which will continue unabated.

In 2005, it was believed that there were 11.5 billion pages on the Internet; as of now, it is expected that there are around 100 billion pages, and the growth rate is faster than Moore's law. The number of pages is increasing by a factor of two every eight months, which results in the generation of a significant amount of data with an ever-increasing value.

The modern healthcare system is a highly linked network that is comprised of a wide variety of clinical and administrative health care personnel who generate complicated data from a wide variety of systems and pieces of medical equipment. The continued growth of the Internet contributed to the conception of a Personal Health Record (PHR) that is based on the internet. This makes people's health information available throughout their whole lives. Researchers and doctors are able to gain access to this health information by utilizing a variety of methods. This information can reveal vitally important facts such as the patient's current health state, the results of various tests performed on patients, human resources, and the coordination of medical units, among other things. As a result of the challenges that are now being faced in health care and the medical sector, such as rising medical costs, an aging population, and an increase in the number of chronic patients, it is imperative that prerequisites for delivering outstanding medical services be met. A lack of medical resources and health care is caused by an aging population, which is a problem in many industrialized nations. This deficit contributes to the gap that exists between people's demand for healthcare and the available medical resources. In order to meet the requirements of this demand, a brand-new medical service system has to be established, and information technologies should be made to play a significant part in both the development and improvement of a central service system. The medical community started making the required preparations to accept and advance the use of information technology in medical treatment. The National Health Information Infrastructure (NHII) has presented a strategy to develop a national health care infrastructure in three different facets: public health, personal health, and medical institution health. A significant quantity of information should be shared amongst patients as part of an effort to enhance the quality of the health care system. This procedure will be beneficial to patients and will also increase the moderators' medical knowledge. The health care system is supposed to provide quality, improved access, and continuous treatment to patients at cheap investment and operating costs in infrastructures and technology. This is a requirement that has been placed upon the system.

Over the course of the past several years, there have been substantial advancements made in the use of ML in a variety of fields, including research. In the sphere of medicine and health care, there have been many different applications of ml methods, but categorization has been the most common one. In the realm of medicine, classification methods have often been used as an integral component in order to analyze patient data and produce a prediction model or set of association rules. Researchers have made use of a hybrid intelligent system to develop a novel approach to the categorization of medical data. This approach is founded on the hybrid combinatorial structure of the Fuzzy max-min neural network, and the classification of the data is carried out with the help of the Random forest. This approach was applied to a range of datasets, including those pertaining to PIDD and Breast Cancer, and it performed quite well in comparison to other methods that were already in use.

1. In their study, Pradeep K. R. et al. presented a more accurate method of diabetes diagnosis. by making an estimate of the glucose level in the blood. The approach that was suggested demonstrates that the accuracy is improved by employing a Support Vector Machine, and it illustrates that the pre-processing of the data set has an influence on both the performance and the accuracy of the prediction of disease. [3,4] Research has shown that an early diagnosis of diabetes at a lower cost is always better, and it also demonstrated that the J48 algorithm is renowned for the accuracy it provides.

2. The naive Bayes weighted approach (NBwa) technique was utilized to develop a decision assistance function that was suggested by P. for the heart disease prediction system. Priyanga and others, etc. [5]. The physicians will be able to benefit from this system by providing better judgments, and it may also be utilized to give nice perspectives. These systems made use of a method known as the weighted approach in order to improve the accuracy of their predictions.

3. As detailed by Zahra Beheshti et al., machine learning techniques have seen widespread use in a variety of fields, including but not limited to education, pharmacy, manufacturing industries, commerce, and engineering. [6], Machine learning may now be utilized in the field of medical epidemic analysis to pick an appropriate architecture and learning algorithms that are capable of providing higher levels of efficiency. ML is a challenging endeavor, but there are already accessible learning algorithms that are efficient and effective, and these algorithms have the potential to play a significant part in improving both accuracy and performance.

   It is conceivable to provide high quality and efficiency gains in the delivery of enhanced health care by utilizing techniques from artificial intelligence and machine learning. According to Victoria Espinel, president and chief executive officer of the BSA, artificial intelligence has the potential to enhance health outcomes by between 30 and 40 percent by the year 2020. It was also mentioned that AI and ML would result in enhanced outcomes for better decision-making by healthcare professionals as a result of providing them with fresh insights into enormous amounts of data that are readily available. Espinel's testimony indicated that artificial intelligence technologies are enabling machine-assisted diagnostics, and surgical applications are being employed to improve treatment options and results. "Algorithms for image analysis are assisting pathologists in accurately interpreting patient data, which in turn is assisting physicians in providing patients with more effective therapy.

   The capacity of machine learning algorithms to evaluate enormous amounts of data and discover intriguing patterns originating from a variety of sources is also driving significant advancements in the fields of biological and epidemiological research. The field of medical imaging is one that makes major use of AI and ML applications, and neuroradiologists are already employing these methods to locate genetic markers in MRI images and forecast the presence of brain tumors. Researchers at Stanford have created algorithms that are capable of diagnosing 14 distinct types of medical diseases, including the ability to identify pneumonia from photographs of medical disorders. The Memorial Sloan Kettering Cancer Center is now focusing on developing pictures that will increase the accuracy of diagnosing prostate cancer. Radiologists can benefit from the use of machine learning algorithms to gather more information in a more efficient manner and so better meet the needs of their patients.

V. RECENT PROGRESS IN ML

In order to frame the models and acquire techniques for analysis, a comprehensive collection of machine learning algorithms has been developed. These algorithms may be broken down into the following categories according to the type of learning they employ: Instance-based Algorithms, abbreviated as "I-bA,"

This is a challenge of decision-making among the instances of training data that the model considers to be dynamic. These are the kinds of systems that routinely acquire data through training, then use that data to assess test data through evaluation, and finally construct a prediction model known as a lazy learner. This prediction model essentially stores the training data and waits till it can process the test data before making any predictions. Therefore, a slow learner agrees to a condensed amount of time for training, followed by extended amounts of time for predicting. Locally Weighted Learning (LWL), Self-Organizing Map (SOM), Learning Vector Quantization (LVQ), and k-Nearest Neighbor (kNN) are some of the most prominent I-bA algorithms.

VI. ALGORITHMS FOR CLUSTERING DATA

Clustering is the process of organizing a set of data into a collection of subsets that share similarities. Each subset can have a regularly restricted regular feature that is associated to a limited specified distance matrix. Methods such as Hierarchical Clustering, k-Means, Expectation-Maximization (EM), and k-Medians are examples of general clustering approaches.

1. **Algorithms for Regression Analysis:** The processing done by these algorithms begins with the generation of an association between the variables, which then moves forward by continuously improving its error rate in predictions. The task of forecasting the value of a variable that is constantly changing requires the use of some of the most used methods, such as the Locally Estimated Scatterplot Smoothing (LOESS) methodology, the Logistic Regression Multivariate Adaptive Regression Splines (MARS) algorithm, and the Stepwise Regression, Ordinary LeastSquares Regression (OLSR) program.

2. **Algorithms Based on Decision Trees:** These methods efficiently segment the dataset in order to deliver classification approaches and regression procedures. In this case, the unconditional goal values that are reinforced by the decision trees are quick and precise, and there is a vast range of options linked with ML. The most common decision tree methods are the Chi-squared Automatic Interaction Detection (CHAID) algorithm, the ID3 algorithm coupled with the C4.5 and C5.0 algorithms, and the Classification and Regression Tree (CART) algorithm.

3. **Artificial neural network algorithms:** The ANN model makes use of supervised learning approaches, and it was developed based on the structure of Genetic Neural Networks (GNN), which have AN weighted interrelationships in the intermediate units. These are also known as parallel distributed processing networks and contain methods such as Back-Propagation, Hopfield Network, and Radial Basis Function Network (RBFN).

4. **Deep Learning:** Deep learning techniques are a relatively recent addition to ANN; they are an innovation that follows in the footsteps of beneficial design and creates intricate neural networks. The Deep Belief Network (DBN), the Convolutional Neural Network (CNN), Stacked Auto encoders, and the Restricted Boltzmann Machine (RBM) are all examples of deep learning algorithms.

5. **Algorithms for Ensembles:** Unsupervised learning, which divides the training data into subsets of the data base and serves as the foundation for the creation of self-governing learning models, is required for ensemble algorithms to function properly. The most common ensemble algorithms include AdaBoost, Gradient Boosting Machines (GBM), Random Forest, Bootstrapped Aggregation, Stacked Generalization (blending), Boosting, Bagging, and Gradient Boosted Regression Trees (GBRT). Other popular algorithms are Bootstrapped Aggregation, Stacked Generalization (blending), and Bootstrapped Aggregation.

6. **Into the Future: Trends:** Data analysis is a big difficulty in application, and presenting the results by domain experts who do not have technical expertise is a particularly tough task to do. This is because presenting the results requires technical understanding. Specifically in the sphere of healthcare, addressing these difficulties calls for more advanced research in Big Data that provides a comprehensive picture together with a future projection. The fact that the following study is still needed despite other considerable work being done in Big Data and being presented at a number of conferences or published in publications specifically devoted to "DM and KDD" or "ML" suggests that this work needs to be done.

7. **Scalable Big Data Structure:** The experience of analyzing the Twitter data [7], provides clear facts about the infrastructures of Big Data that offer the experience of conducting the analysis of the Twitter data. According to the author, this is not an easy topic, and it is difficult to undertake analytics owing to the lack of DM tools that are now accessible in the market. It has also been shown that the majority of time is used in the process of planning the application of DM approaches and transforming pilot models into outcomes that are effective.

8. **The article "Mining Heterogeneous Information Networks (MHIN):** A Organizational Analysis Method" [8] demonstrates that mining heterogeneous information networks (HIN) is an area of Big Data research that shows promise. HIN that is organized, semi-structured, and unstructured are all taken into consideration in this work. These semi-structured HIN models have an impact on the unexpected interpretation of collected linkages and nodes in a grid, which results in the disclosure of fascinating information.

9. **Big Graph Mining (BGM):** Procedures and findings [9], provide mining huge graphs, employment of the Pegasus tool, and the study demonstrates relatively intriguing results in Twitter and Web Graph. The results of this study effort provide research directions for work that will be done in the future about massive graph mining.

VII. USE CASES AND OBSTACLES RELATING TO MACHINE LEARNING IN THE HEALTHCARE INDUSTRY

The use of machine learning in the field of healthcare faces a number of problems, including developing a comprehensive and open-minded computational architecture for machine learning that is suitable for the healthcare industry.

Capability to determine, construct, and differentiate the uniqueness of data prior to the administration of machine learning algorithms and tools. It involves the capacity to reduce the complexity of the input and to learn on several levels simultaneously via combination learning.Through stubborn implication, there is the potential to achieve numerical base development using Structured Prediction.

Significant machine learning strategies, like association rule learning, ensemble learning, large-scale recommender systems, and Natural Language Processing (NLP), are still struggling with scaling issues. The main issues that have been brought up to enable ML systems that are unsuited for establishing the categorization of Big Data are as follows:

Training on a particular dataset based on ML approaches is not appropriate for other datasets; in this case, classification accuracy may no longer be relevant over other types of datasets.ML approaches are not ideal for use with the present Big Data since there are multiple learning tasks to be undertaken. In most cases, ML approaches are used after training on a predetermined number of different class types.

Traditional machine learning techniques often do not scale well when used to BD. The essential trickery is in the limitations of one's recollection. Despite the fact that the algorithms normally assume that training data trials survive only within primary memory where BD doesn’t succeed into it. The data distribution method is the conventional way that is taken while gaining information on a massive dataset. Here, different computations are carried out in order to advance to the discrete subgroups, and this is done by altering the batch size on the exclusive training dataset. When this is taken into account, training the various prediction models results in a loss in precision. The alternative method is done using online learning, where the amount of memory used does not rely on the size of the dataset. Distributed learning and online learning, on the other hand, are insufficient for learning the BD division for the same reasons, which are detailed below:

The degree of unpredictability associated with the data amount is significantly higher, i.e., adaptability to online or dispersed learning. Learning BD in a sequential fashion online requires a significant amount of time spent practicing on each unique computer. On the other hand, distributed learning for a large number of machines slows down the efficiency that may be achieved by each unit and alters the overall performance.

There is no longer any thought being given to integrating immediate training and prediction into the BD. In the situation where BD that has been operated on once is being retained in the distributed storage, the learning process validates the attempt that has been supplied by the batch approach as a result of this.

The development of medical imaging machine learning algorithms that are capable of producing a technological infrastructure that is able to handle the vast amounts of unstructured data that are now accessible will be the next challenge for academics. In order to solve the primary challenge presented by these methods, a significant amount of processing power and storage performance is required. The phenomenon known as "GPU starvation," which occurs when a processor fails because it can't access the data that it needs, is something that machine learning applications need to avoid. The amount of processing power that machine learning algorithms demand is growing at an alarming rate, while the processing power that can be given by GPUs has risen by a factor of ten. The development of machine learning applications that may significantly improve both the quality and the effectiveness of healthcare delivery presents the primary obstacle for any company.

To provide a brief summary, the following are some examples of the breadth, services, and benefits that may result from the use of ML in the healthcare industry.

- The availability of internet at medical facilities will be of benefit to patients, and it can also facilitate improved communication.
- Requests for consultation made by rural health care providers can be met with a prompt response from medical professionals working at metropolitan reference institutions.
- Plan and diagnose in the same manner as a remote health care worker is possible for physicians.
- It is not necessary for physicians from the same or a different nation to physically visit remote regions since they are able to assist health care personnel effectively.
- Both primary care physicians and medical experts can benefit from the usage of live interactive video.
- In conjunction with patient information, the use of store and forward communication of diagnostic photos, highly essential indicators, and video clips is a method that may be carried out effectively.
- Monitoring of patients remotely in addition to providing medical and health information to consumers
- Patients can benefit from access to specialized health information and online conversation with teams through the use of the Internet and wireless devices, which facilitates the provision of peer-to-peer assistance.

VIII. CONCLUSION

The current state of healthcare research is beset by falling success rates; hence, Big Data and ML with helpful analytics might be a crucial component. These technologies are gaining acceptance not just in the pharmaceutical industry but also in other industries. It is anticipated that managing Big Data techniques would lead to effective decision making, which might create up to one hundred billion dollars in the structure of healthcare, and machine learning is playing a vital part in the field of medical diagnostics. The purpose of the trials is to advance innovation, enhance the effectiveness of clinical trials and research, and provide unique tools for physicians, consumers, and other interested parties.

Data growths are created from numerous sources in the healthcare business, and the proper use of this data will assist companies in identifying better and developing a predictive model of being successful that is safe and effective. Many machine learning diagnostic applications are classified as chatbots, which are forms of artificial intelligence equipped with voice recognition capabilities. Chatbots are able to recognize patterns in patient complaints, which allows for improved diagnosis and the provision of appropriate therapy. As a result of the fact that patient data can take the form of structured data, unstructured data, high-definition 3D medical photos, and videos, businesses are faced with the problem of developing value-based healthcare that can improve clinical, financial, and technological operations. Biometric readings, safety medical equipment, and wearable activity trackers can all contribute to the generation of patient data, and these readings should be encrypted before being sent to an organization. In addition to this, they need to simplify the design of the storage system so that it is flexible enough to manage Big Data while maintaining adequate backup speed. Deep learning algorithms are being used by researchers in order to detect and construct models that are equivalent to physicians. The ability of doctors to detect uncommon diseases can be improved by machine vision and other machine learning technology. Better infrastructure that can support performance, advanced data analytics, usage of private, hybrid, or public clouds, and evolving clinical processes has to be developed by organizations. This infrastructure needs to be adaptable.

REFERENCES

[1] William Carroll; G. Edward Miller, “Disease among Elderly Americans: Estimates for the US civilian non institutionalized population, 2010,” *Med.Expend. Panel Surv.*, no. June, pp. 1–8,2013.
[2] V. Kirubha and S. M. Priya, “Survey on Data Mining Algorithms in Disease Prediction,” vol. 38, no. 3, pp. 124–128,2016.
[3] M. A. Jabbar, P. Chandra, and B. L. Deekshatulu, “Prediction of risk score for heart disease using associative classification and hybrid feature subset selection,” *Int. Conf. Intell. Syst. Des. Appl. ISDA*, pp. 628–634,2012.
[4] Michael W.Berryet.al,”Lecture notes in data mining”,*WorldScientific*(2006)
[5] S. Shilaskar and A. Ghatol, “Feature selection for medical diagnosis : Evaluation for cardiovascular diseases,” Expert Syst. Appl., vol. 40, no. 10, pp. 4146–4153, Aug.2013’
[6] C.-L. Chang and C.-H. Chen, “Applying decision tree and neural network to increase quality of dermatologic diagnosis,” Expert Syst. Appl., vol. 36, no. 2, Part 2, pp. 4035– 4041, Mar. 2009.
[7] S. Kumra, R. Saxena, and S. Mehta, “An Extensive Review on Swarm Robotics,” pp. 140–145, 2009.
[8] T. M. Lakshmi, A. Martin, R. M. Begum, and V. P. Venkatesan, “An Analysis on Performance of Decision Tree Algorithms using Student‟s Qualitative Data,” *Int. J. Mod. Educ. Comput. Sci.*,vol. 5, no. 5, pp. 18–27,2013.
[9] P. Sharma and A. P. R. Bhartiya, “Implementation of Decision Tree Algorithm to Analysis the Performance,” *Int. J. Adv. Res. Comput. Commun. Eng.*, vol. 1, no. 10, pp. 861–864, 2012.
[10] A. L. Bui, T. B. Horwich, and G. C. Fonarow, “Epidemiology and risk profile of heart failure,” Nature Reviews Cardiology, vol. 8, no. 1, pp. 30–41, 2011.
[11] Dr. Sunil Butada and Dr. Subhani Shaik, “IPL Match Prediction using Machine Learning”, IJAST, Vol.29, Issue 5, April-2020.
[12] KP Surya Teja, Vigneswar Reddy and Dr. Subhani Shaik,” Flight Delay Prediction Using Machine Learning Algorithm XGBoost”, Jour of Adv Research in Dynamical & Control Systems, Vol. 11, No. 5, 2019.
[13] Shiva Keertan J and Dr. Subhani Shaik,” Machine Learning Algorithms for Oil Price Prediction”, International Journal of Innovative Technology and Exploring Engineering, Volume-8 Issue-8, 2019.
[14] P. Santosh and Dr. Subhani Shaik,” Heart disease prediction with PCA and SRP”, International Journal of Engineering and Advanced Technology,” Volume-8, Issue-4, 2019.

