# IJARETY

# International Journal of Advanced Research in Education and TechnologY (IJARETY)


**Volume 11, Issue 2, March 2024**

**Impact Factor: 7.394**

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

INNO SPACE
SJIF Scientific Journal Impact Factor

doi crossref

निस्केयर NISCAIR

International Journal of Advanced Research in Education and TechnologY(IJARETY)

| ISSN: 2394-2975 | www.ijarety.in| | Impact Factor: 7.394| A Bi-Monthly, Double-Blind Peer Reviewed & Referred Journal |

|| Volume 11, Issue 2, March 2024 ||

DOI:10.15680/IJARETY.2024.1102023

# Improving the Accuracy of Medical Diagnosis Detection using Machine Learning

**Mrs.S.NAZEEMA, PRABHAKARAN K, NAVEENKUMAR V, VIGNESH S**

Assistant Professor, Department of CSE, Muthayammal Engineering College (Autonomous), Rasipuram, Tamil Nadu, India

Department of CSE, Muthayammal Engineering College (Autonomous), Rasipuram, Tamil Nadu, India

Department of CSE, Muthayammal Engineering College (Autonomous), Rasipuram, Tamil Nadu, India

Department of CSE, Muthayammal Engineering College (Autonomous), Rasipuram, Tamil Nadu, India

**ABSTRACT:** The use of ML and DL techniques for healthcare prediction has the potential to change the way traditional healthcare services are delivered. In the case of ML and DL applications, healthcare data is deemed the most significant component that contributes to medical care systems. This paper aims to present a comprehensive review of the most significant ML and DL techniques employed in healthcare predictive analytics. In addition, it discussed the obstacles and challenges of applying ML and DL Techniques in the healthcare domain. As a result of this survey, a total of 41 papers covering the period from 2019 to 2022 were selected and thoroughly reviewed. In addition, the methodology for each paper was discussed in detail.

**KEYWORDS**: Artificial Intelligence, Autism Spectrum Disorder, Convolutional Neural Networks, Digital in-line holographic microscopy

## I. INTRODUCTION

Machine Learning (ML), a branch of Artificial Intelligence (AI), learns from the data using various algorithms and is a self-improving process in terms of performance as making adjustments during the learning process. ML has been successfully applied to practically every domain such as robotics, education, travel to health care. In the healthcare domain, the ML approaches are mainly used for the purpose of disease diagnosis.

The machine learning approaches came into the health sector domain in the 1970s and an international AI journal Artificial Intelligence in Medicine was established in 1980. In the next two decades, disease diagnosis domain adopted the classical ML approaches such as Support Vector Machine, Naïve Bayes, and some artificial neural networks. The introduction of AlexNet in 2012 initiated the current wave of deep learning in this field as neural networks demonstrated superior performance. Also, in this past decade, the investment in AI in healthcare applications has increased significantly. The studies in show that the use of AI and ML technologies in healthcare is leading to the development of software, platforms, automated systems and devices to check as well as improve the health condition of people.

The analysis of the clinical data can lead to the timely diagnosis of the disease which will help to start cure for the patient in time as well. Traditional approach of diagnosing disease is generally costly and time-consuming. As well, the potential of time and cost-proficient machine learning-based disease diagnosis approaches are proven by the researchers. ML techniques have not only been able to diagnose the common diseases but are also equally capable of diagnosing the rare diseases. Authors in demonstrate the significance and robustness of AI and ML techniques to solve health care problems.

In general, a dataset table used to build an ML model for diagnosing a disease has columns for different attributes and a column variable for the class variable. Here, class variable indicates whether the instance in the table indicated is positively diagnosed with the disease under consideration. Usually, class values of 1 means positively diagnosed and 0 means negatively diagnosed. Supervised and unsupervised ML approaches have been in practice for analyzing the health care data. In general, disease diagnosis problems are based on supervised learning. We will present a detailed analysis of the used dataset and ML algorithms in .

| ISSN: 2394-2975 | www.ijarety.in| | Impact Factor: 7.394| A Bi-Monthly, Double-Blind Peer Reviewed & Referred Journal |

|| Volume 11, Issue 2, March 2024 ||

DOI:10.15680/IJARETY.2024.1102023

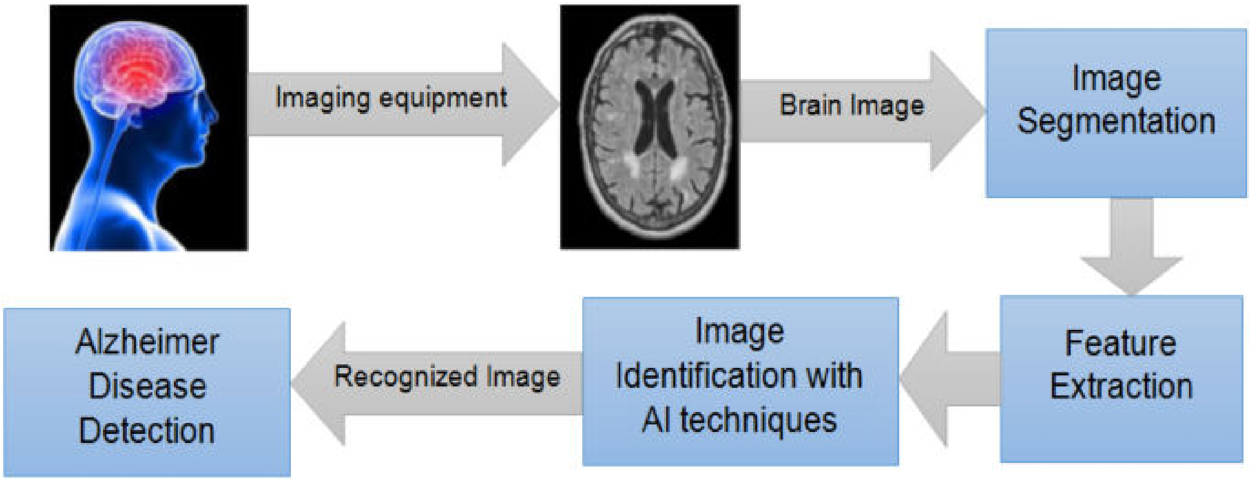


**Fig 1: Artificial intelligence in disease diagnosis**

Although ML offers systematic and sophisticated algorithms of multi-dimensional clinical data, the accuracy of the ML in diagnosing the diseases is still a concern [16]. As well, the improvement in the performance of ML to diagnose disease is a hot topic in this domain. As different ML approaches perform differently for different healthcare dataset, we are also in need to find the way to apply many state-of-the-art algorithms to same dataset in reasonable time with minimal lines of codes, so that the search of best ML method can be pursued efficiently to diagnose a particular disease.

## II. RELATED WORK

Healthcare is the collective effort of society to ensure, provide, finance, and promote health. In the twentieth century, there was a significant shift toward the ideal of wellness and the prevention of sickness and incapacity. The delivery of healthcare services entails organized public or private efforts to aid persons in regaining health and preventing disease and impairment [6]. Health care can be described as standardized rules that help evaluate actions or situations that affect decision-making [7]. Healthcare is a multi-dimensional system. The basic goal of health care is to diagnose and treat illnesses or disabilities. A healthcare system's key components are health experts (physicians or nurses), health facilities (clinics and hospitals that provide medications and other diagnostic services), and a funding institution to support the first two [8].

With the introduction of systems based on computers, the digitalization of all medical records and the evaluation of clinical data in healthcare systems have become widespread routine practices. The phrase "electronic health records" was chosen by the Institute of Medicine, a division of the National Academies of Sciences, Engineering, and Medicine, in 2003 to define the records that continued to enhance the healthcare sector for the benefit of both patients and physicians. Electronic Health Records (EHR) are "computerized medical records for patients that include all information in an individual's past, present, or future that occurs in an electronic system used to capture, store, retrieve, and link data primarily to offer healthcare and health-related services," according to Murphy, Hanken, and Waters [8].

Daily, healthcare services produce an enormous amount of data, making it increasingly complicated to analyze and handle it in "conventional ways." Using machine learning and deep learning, this data may be properly analyzed to generate actionable insights. In addition, genomics, medical data, social media data, environmental data, and other data sources can be used to supplement healthcare data. Figure 1 provides a visual picture of these data sources. The four key healthcare applications that can benefit from machine learning are prognosis, diagnosis, therapy, and clinical workflow, as outlined in the following section

## III. METHODS

Machine learning (ML) is a subfield of AI that aims to develop predictive algorithms based on the idea that machines should have the capability to access data and learn on their own. ML utilizes algorithms, methods, and processes to detect basic correlations within data and create descriptive and predictive tools that process those correlations. ML is usually associated with data mining, pattern recognition, and deep learning. Although there are no clear boundaries between these areas and they often overlap, it is generally accepted that deep learning is a relatively new subfield of

| ISSN: 2394-2975 | www.ijarety.in| | Impact Factor: 7.394| A Bi-Monthly, Double-Blind Peer Reviewed & Referred Journal |

|| Volume 11, Issue 2, March 2024 ||

DOI:10.15680/IJARETY.2024.1102023

ML that uses extensive computational algorithms and large amounts of data to define complex relationships within data. As shown in ML algorithms can be divided into three categories: supervised learning, unsupervised learning, and reinforcement learning

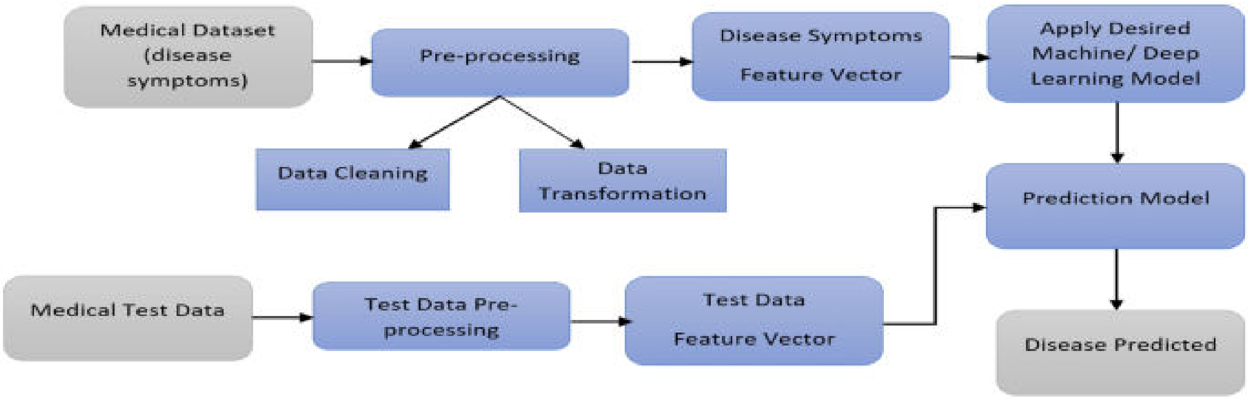


**Fig 2: Artificial intelligence in disease diagnosis**

Logistic regression, also known as the logistic model, investigates the correlation between many independent variables and a categorical dependent variable and calculates the probability of an event by fitting the data to a logistic curve. Discrete mean values must be binary, i.e., have only two outcomes: true or false, 0 or 1, yes or no, or either superscript or subscript. In logistic regression, categorical variables need to be predicted and classification problems should be solved. Logistic regression has many benefits; for example, it shows the linear relationship between dependent and independent variables with the best results. It is also simple to understand. On the other hand, it can only predict numerical output, is not relevant to nonlinear data, and is sensitive to outliers.

Random forest (RF) is a basic technique that produces correct results most of the time. It may be utilized for classification and regression. The program produces an ensemble of DTs and blends them. In the RF classifier, the higher the number of trees in the forest, the more accurate the results. So, the RF has generated a collection of DTs called the forest and combined them to achieve more accurate prediction results. In RF, each DT is built only on a part of the given dataset and trained on approximations. The RF brings together several DTs to reach the optimal decision.

## IV. RESULT ANALYSIS

SVM is a statistics-based learning method that follows the principle of structural risk minimization and aims to locate decision bounds, also known as hyperplanes, that can optimally separate classes by finding a hyperplane in a usable N-dimensional space that explicitly classifies data points . SVM indicates the decision boundary between two classes by defining the value of each data point, in particular the support vector points placed on the boundary between the respective classes. SVM has several advantages; for example, it works perfectly with both semi-structured and unstructured data. The kernel trick is a strong point of SVM. Moreover, it can handle any complex problem with the right functionality and can also handle high-dimensional data. Furthermore, SVM generalization has less allocation risk. On the other hand, SVM has many downsides. The model training time is increased on a large dataset. Choosing the right kernel function is also a difficult process.

| ISSN: 2394-2975 | www.ijarety.in| | Impact Factor: 7.394| A Bi-Monthly, Double-Blind Peer Reviewed & Referred Journal |

IJARETY

|| Volume 11, Issue 2, March 2024 ||

DOI:10.15680/IJARETY.2024.1102023

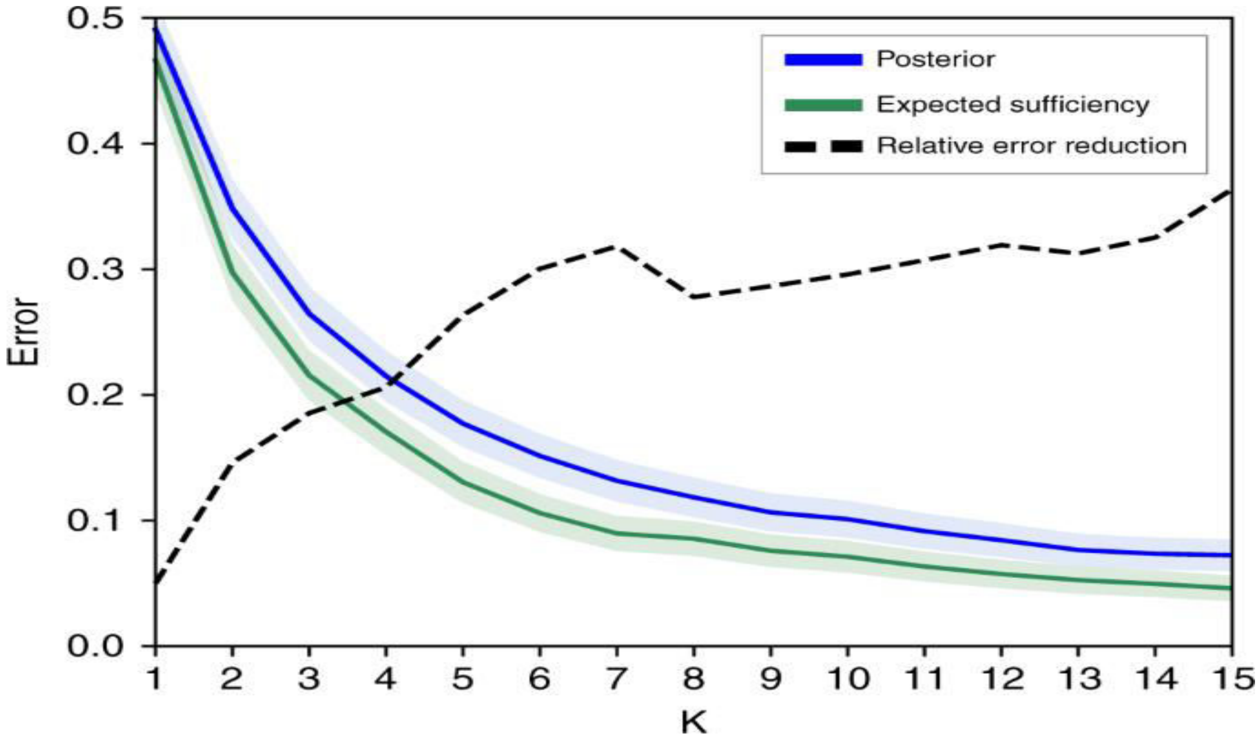


**Fig 3: Improving the accuracy of medical diagnosis**

Progressively and unevenly. The tree policy is utilized to get the critical node of the current tree for each iteration of the method. The tree strategy seeks to strike a balance between exploration and exploitation concerns. Then, from the specified node, simulation 2 is run, and the search tree is then updated according to the obtained results. This comprises adding a child node that matches the specified node's activity and updating its ancestor's statistics. During this simulation, movements are performed based on some default policy, which in its simplest case is to make uniform random movements.

## V. CONCLUSION

Machine Learning (ML) algorithms have been successfully applied in the healthcare domain to diagnosing diseases. In our work we show that, the use of libraries such as AutoGluon can help to compare the performances of different ML approaches in diagnosing a disease for a given dataset with optimal lines of code. This helps in finding the best performing ML algorithm for a particular dataset or a particular type of disease as well. Furthermore, it decreases the probability of inaccurate diagnosis, which is a significantly important consideration while dealing with the health of the people. In this study we have tested the performance of 20 ML approaches in diagnosing diabetes based on the Pima Indian Diabetes Dataset. For the dataset considered in this study, the Naïve Bayes algorithm performed better among the other algorithms. This shows that using complex and computationally costly algorithms does not necessarily improve the accuracy of diagnosing a disease.

The possibility of the improvement in the performance of ML models in the future can be started by finding the correlation among each attribute and dropping the highly correlated attributes, because the highly correlated attributes can confuse a model in the learning phase. The evidence of applying multiple ML algorithms with optimal lines of codes in this study strongly suggests that such investigations are to be pursued.

## REFERENCES

1. Mitchell, T.M. *Machine Learning*; McGraw-Hill: New York, NY, USA, 1997. [**Google Scholar**]
2. Fatima, M.; Pasha, M. Survey of machine learning algorithms for disease diagnostic. *J. Intell. Learn. Syst. Appl.* **2017**, *9*, 73781. [**Google Scholar**] [**CrossRef**]
3. Machine Learning Use Cases|Neural Designer. Available online: **https://www.neuraldesigner.com/solutions** (accessed on 13 January 2022).

| ISSN: 2394-2975 | www.ijarety.in| | Impact Factor: 7.394| A Bi-Monthly, Double-Blind Peer Reviewed & Referred Journal |

|| Volume 11, Issue 2, March 2024 ||

DOI:10.15680/IJARETY.2024.1102023

4. Demystifying AI in Healthcare: Historical Perspectives and Current Considerations. Available online: **https://www.physicianleaders.org/news/demystifying-ai-in-healthcare-historical-perspectives-and-current-considerations** (accessed on 13 January 2022).
5. Kononenko, I. Machine learning for medical diagnosis: History, state of the art and perspective. *Artif. Intell. Med.* **2001**, *23*, 89–109. [**Google Scholar**] [**CrossRef**] [**PubMed**]
6. Krizhevsky, A.; Sutskever, I.; Hinton, G.E. ImageNet Classification with Deep Convolutional Neural Networks. *Adv. Neural Inf. Process. Syst.* **2012**, 1097–1105. [**Google Scholar**] [**CrossRef**]
7. Massaro, A.; Ricci, G.; Selicato, S.; Raminelli, S.; Galiano, A. Decisional Support System with Artificial Intelligence oriented on Health Prediction using a Wearable Device and Big Data. In Proceedings of the 2020 IEEE International Workshop on Metrology for Industry 4.0 & IoT, Rome, Italy, 3–5 June 2020; pp. 718–723. [**Google Scholar**] [**CrossRef**]
8. Habib, M.; Faris, M.; Qaddoura, R.; Alomari, M.; Alomari, A.; Faris, H. Toward an automatic quality assessment of voice-based telemedicine consultations: A deep learning approach. *Sensors* **2021**, *21*, 3279. [**Google Scholar**] [**CrossRef**] [**PubMed**]
9. Massaro, A.; Galiano, A.; Scarafile, D.; Vacca, A.; Frassanito, A.; Melaccio, A.; Solimando, A.; Ria, R.; Calamita, G.; Bonomo, M.; et al. Telemedicine DSS-AI Multi Level Platform for Monoclonal Gammopathy Assistance. In Proceedings of the 2020 IEEE International Symposium on Medical Measurements and Applications (MeMeA), Bari, Italy, 1 June–1 July 2020. [**Google Scholar**] [**CrossRef**]
10. Niculescu, M.S.; Florescu, A.; Pasca, S. LabConcept—A new mobile healthcare platform for standardizing patient results in telemedicine. *Appl. Sci.* **2021**, *11*, 1935. [**Google Scholar**] [**CrossRef**]
11. Massaro, A.; Maritati, V.; Savino, N.; Galiano, A. Neural Networks for Automated Smart Health Platforms oriented on Heart Predictive Diagnostic Big Data Systems. In Proceedings of the 2018 AEIT International Annual Conference, Bari, Italy, 3–5 October 2018. [**Google Scholar**] [**CrossRef**]

IJARETY

International Journal of Advanced Research in Education and Technology

ISSN: 2394-2975 Impact Factor: 7.394

www.ijarety.in editor.ijarety@gmail.com