International Journal of Environmental Sciences
ISSN: 2229-7359
Vol. 11 No. 2s, 2025
https://theaspd.com/index.php

Machine Learning-Based Predictive Modelling For Healthcare

Lukeshwari Sahu[1], Dr. Nidhi Mishra[2], Praveen Kumar Joshi[3]
[1]Assistant Professor, Department of Pharmacy, Kalinga University, Raipur, India.
ku.lukeshwarisahu@kalingauniversity.ac.in, 0009-0002-3045-6538
[2]Assistant Professor, Department of CS & IT, Kalinga University, Raipur, India. ku.nidhimishra@kalingauniversity.ac.in, 0009-0001-9755-7950
[3]Assistant Professor, New Delhi Institute of Management, New Delhi, India., E-mail: praveen.joshi@ndimdelhi.org, https://orcid.org/0009-0007-7777-8178

Abstract
machine learning and deep learning approaches can be used to efficiently analyse this data and produce insightful findings. By adding data from social media, genetics, medical records, environmental data, and other sources to healthcare data, a comprehensive understanding of various data streams can also be attained. Additionally, the suggested framework analyzing health trends in relation to the symptoms experienced by certain populations and individual treatment decisions. This method can be used to calculate the proportion of the population that sought medical attention within a given time frame. In recent decades, the 'ecology of medical care' theory has gained widespread acceptance in academic circles. Because the system is dynamic and scalable, medical practitioners often face new difficulties, changing duties, and continuous disruptions. Because of this heterogeneity, healthcare providers frequently make identifying diseases a secondary concern. In addition to highlighting the inherent challenges of implementing machine learning and deep learning techniques in the healthcare sector, this paper attempts to provide a thorough overview of the methods currently employed in healthcare prediction.
Keywords: *Modelling for Healthcare, incorrect prediction, transforming*

1. INTRODUCTION

Human life continues to advance daily, yet each generation's health may either become better or get worse. Uncertainties abound in life, and people frequently experience serious health problems as a result of delayed disease discovery [1]. Early diagnosis, however, can stop the disease's progression. By using machine learning to forecast diseases, it is possible to identify common disorders early on. Presently, health is often regarded as a secondary priority, resulting in numerous complications. Many patients are unable to pay for medical consultations, and others have very hectic schedules that cause them to ignore chronic symptoms for extended periods of time, which can have detrimental effects on their health [2]. Because diseases are a global problem, academics and medical professionals are working hard to reduce disease-related mortality [9]. Predictive analytics has grown in significance in the medical sector as a result of the recent surge in the volume of healthcare data from disparate and incompatible sources. However, the enormous volumes of historical data and the continuous flow of data produced by healthcare services present an unparalleled challenge to typical database systems in terms of processing, storing, and analyzing. In the actual world, a medical diagnosis is a crucial concern and a means of problem-solving [3]. Interpreting observational data into distinct disease classifications is a step in the diagnosis process. It is believed that diseases are conceptual medical constructs that identify abnormalities in observed data, which includes information gleaned from biological samples and patient evaluations. The term "healthcare" refers to the collective efforts of society to guarantee, provide, finance, and promote health. The twentieth century saw a dramatic change toward wellness and the prevention of illness and incapacity [13]. In order to help people regain their health and prevent illness and disability, healthcare services are provided through coordinated public or private programs [4]. Healthcare can be characterized as a set of standardized guidelines that facilitate the evaluation of actions or circumstances influencing decision-making [10]. It is a complex, multi-faceted system.

2. REVIEW OF LITERATURE

Clinical data analysis and medical record digitization have become commonplace in healthcare systems with the introduction of computer-based systems. The Institute of Medicine, a member of the National

Academies of Sciences, Engineering, and Medicine, coined the term "electronic health records" in 2003 to describe medical records that continue to improve healthcare for patients and medical professionals [6]. It is defined by Murphy, Hanken, and Waters as "Computerized medical records for patients that encompass all information regarding an individual's past, present, or future, stored in an electronic system designed to capture, store, retrieve, and link data primarily for the provision of healthcare and health-related services" (EHR). Healthcare services generate massive amounts of data every day, which makes traditional methods of management and analysis increasingly challenging [5]. Nevertheless, The following section outlines the four primary healthcare applications that machine learning can help with: diagnosis, treatment, clinical workflow, and prognosis. The design and execution of curative treatments, which constitute a component of health care, are highly valued in the notion of medical care [11]. The concept of the ecosystem of medical treatment was initially put forth by White in 1961.

3. MATERIALS AND METHODS

The suggested supervised classification-based machine learning algorithms or approaches are the best method for the issue at hand. Stochastic gradient descent, logistic regression, multilayer perceptrons, and the ensemble models Random Forest and XG boost are the machine learning models that have been suggested. The suggested machine learning models are under the category of supervised machine learning. The XG boost method and Random Forest are the ensemble algorithms. While the boosting strategy is used with the XG boost algorithm, the bagging approach is used with Random Forest. When it comes to prediction models, both of these machine learning techniques are more effective [7]. The multilayer perceptron method uses a different technique to prediction and has a high prediction capacity. The SGD technique is intended to be more efficient, which is basically linear in the training data set, easy to implement, and offers a wide range of parameter adjustment choices. The SGD algorithm process, which is a slope plunge method enhanced by the union pace, deals with the random probability for i in the range (m). The components are taken into consideration separately, which is different from the conventional slope approaches. SGD uses a single piece of information to approximate the inclination. Accordingly, evaluating angle saves a great deal of time in comparison to inputting all the information. It is appropriate for SGD, which enhances learning capacity, and it is computationally less costly. Finding this derivative of faults essentially involves looking at each epoch, as we are only looking at one data point at a time. Stochastic gradient descent, or SGD, is the process of collecting a sample, determining the Y, and then updating the weight in the back propagation one point at a time. Assuming that K is always smaller than N, the total number of data points (samples), we examined k data points as key data points (samples) from the dataset in the case of mini-batch stochastic gradient descent. We suggested using the short batch SGD technique to classify COPD data points.

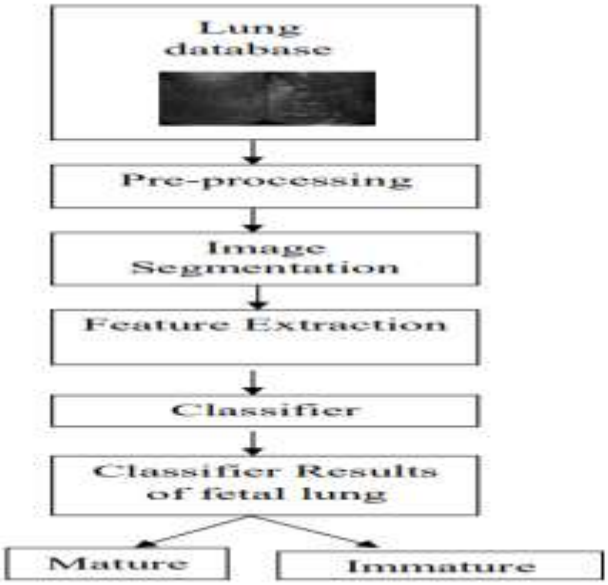

Figure 1: proposed flow

Because the logistic function, also referred to as the sigmoid function, always yields values between 0 and 1, we always use it when performing logistic regression for classification problems. This is because the classification problems are either true or false, meaning that the values are either one or zero [7]. It refers to binary classification issues; when there are two target classes out of the potential classes, the logistic function is used.

4. RESULT

Another way to describe it is that it is a function to estimate probabilities, which is what we obtain by applying logistic regression. After obtaining the value of the active function, we apply a specific threshold, which instructs us to call the value 0 if it is less than 0.5 and 1 if it is larger than 0.5 [12]. One training instance serves as the logistic regression's cost function, which is parameterized using the logarithmic function to make it convex [8]. In order to reach a global minimum, the convex function must be implemented. The minimum value is determined by theta values. Thus, logarithmic functions are used to parameterize the cost functions [15]. The tagged dataset that was used to build this model is comparable to how a teacher teaches a student the fundamentals of mathematics [14].

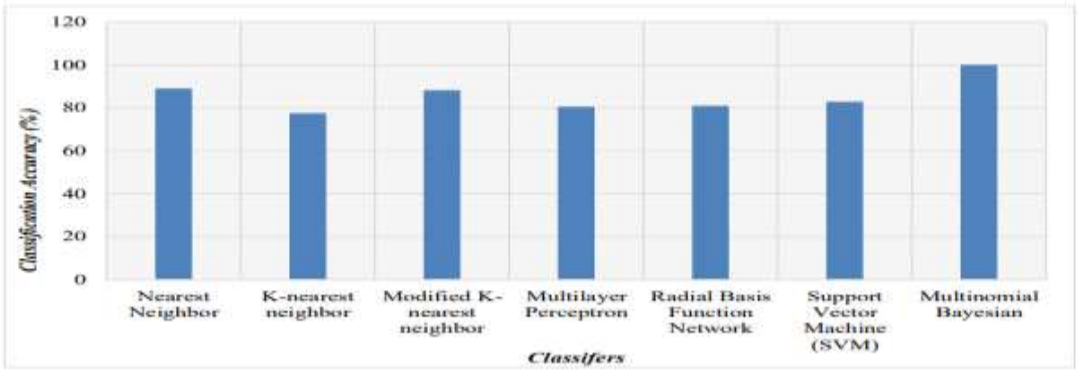

Figure 2: Classification Accuracy

Labelled data with expected right answers produced from algorithmic outputs are required for this learning strategy. Famous supervised learning techniques include logistic regression, support vector machines, decision trees, random forests, K-nearest neighbours, naive Bayes, and linear regression. The logistic model, also known as logistic regression, fits data to a logistic curve to estimate the chance of an event by analysing the connection between several independent factors and a categorical dependent variable.

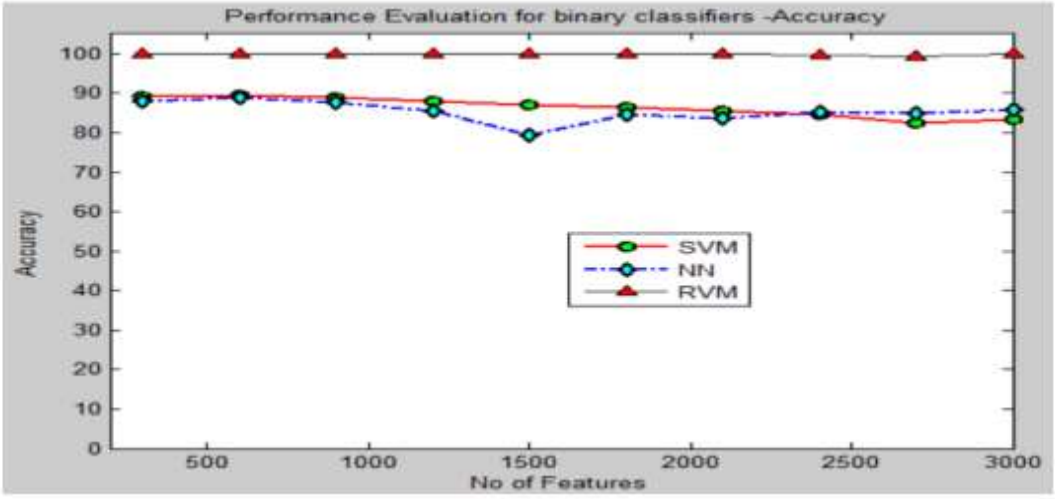

Figure 3: Comparison of accuracy for the existing model

For the discrete mean values, there can only be two possible outcomes: superscript or subscript, true or false, 0 or 1, yes or no. In the context of logistic regression, categorical variable prediction and categorization problems must be addressed. R, Python, Java, and MATLAB are just a few of the programming tools that can be used for logistic regression.

5. CONCLUSION

The application of machine learning (ML) and deep learning (DL) techniques in healthcare prediction has the potential to completely transform the provision of conventional healthcare services. In the context of ML and DL applications, healthcare data is recognized as the most significant component influencing medical care systems. Presenting a thorough examination of the primary ML and DL techniques used in healthcare predictive analytics is the aim of this paper. It also talks about the challenges and obstacles associated with using ML and DL techniques in the healthcare sector. This inquiry led to the selection and careful analysis of 41 academic papers published between 2019 and 2022. Furthermore, a thorough analysis of each study's methodology was conducted. According to the literature review, artificial intelligence methods ML and DL in particular help predict and analyse medical data and make accurate diagnoses by connecting multiple clinical records and using that information to reconstruct a patient's medical history. Through the application of ML and DL approaches, this study contributes to the body of existing literature and future research in the field of healthcare predictive analytics by providing researchers and other academics with a valuable resource.

REFERENCES

1. Badawy, Mohammed, Nagy Ramadan, and Hesham Ahmed Hefny. "Healthcare predictive analytics using machine learning and deep learning techniques: a survey." Journal of Electrical Systems and Information Technology 10, no. 1 (2023): 40.
2. Suganthi, K., & Raja, R. (2018). Optimal Relay Node Placement Algorithm for Energy Balancing in Green Cellular Network. International Journal of Advances in Engineering and Emerging Technology, 9(3), 31–43.
3. Ogunpola, Adedayo, Faisal Saeed, Shadi Basurra, Abdullah M. Albarrak, and Sultan Noman Qasem. "Machine learning-based predictive models for detection of cardiovascular diseases." Diagnostics 14, no. 2 (2024): 144.
4. Dutta, R., & Alam, T. (2024). Policy Harmonization for SDG Implementation in the ASEAN Region: Challenges and Opportunities. International Journal of SDG's Prospects and Breakthroughs, 2(3), 4-6.
5. Mahoto, Naeem Ahmed, Asadullah Shaikh, Adel Sulaiman, Mana Saleh Al Reshan, Adel Rajab, and Khairan Rajab. "A machine learning based data modeling for medical diagnosis." Biomedical Signal Processing and Control 81 (2023): 104481.
6. Kapoor, S., & Sharma, V. (2024). A Comprehensive Framework for Measuring Brand Success and Key Metrics. In Brand Management Metrics (pp. 16-30). Periodic Series in Multidisciplinary Studies.
7. Tran, Nguyen Duy Thong, Carson K. Leung, Evan WR Madill, and Phan Thai Binh. "A deep learning based predictive model for healthcare analytics." In 2022 IEEE 10th International Conference on Healthcare Informatics (ICHI), pp. 547-549. IEEE, 2022.
8. Alam, R., & Subramanian, V. (2024). Fertility Decline in High-income Countries: Socioeconomic Drivers and Cultural Resistance. Progression Journal of Human Demography and Anthropology, 2(2), 1-4.
9. Butt, Umair Muneer, Sukumar Letchmunan, Mubashir Ali, Fadratul Hafinaz Hassan, Anees Baqir, and Hafiz Husnain Raza Sherazi. "Machine learning based diabetes classification and prediction for healthcare applications." Journal of healthcare engineering 2021, no. 1 (2021): 9930985.
10. Khan, I., & Siddiqui, S. (2024). Machine Design a Systematic Approach to Designing Mechanical Systems. Association Journal of Interdisciplinary Technics in Engineering Mechanics, 2(3), 6-11.
11. Aitzaouiat, Charafeddine E., Adnane Latif, Abderrahim Benslimane, and Hui-Hsin Chin. "Machine learning based prediction and modeling in healthcare secured internet of things." Mobile Networks and Applications 27, no. 1 (2022): 84-95.
12. Hameed, A. Z., Balamurugan, R., Rizwan, A., & Shahzad, M. A. (2025). Analyzing And Prioritizing Healthcare Service Performance in Hospitals Using Serqual Model. Archives for Technical Sciences, 1(32), 165–175. https://doi.org/10.70102/afts.2025.1732.165
13. Qureshi, Kashif Naseer, Sadia Din, Gwanggil Jeon, and Francesco Piccialli. "An accurate and dynamic predictive model for a smart M-Health system using machine learning." Information Sciences 538 (2020): 486-502.
14. Kim, H. (2020). 5G core network security issues and attack classification from network protocol perspective. Journal of Internet Services and Information Security, 10(2), 1-15.
15. Stiglic, Gregor, Primoz Kocbek, Nino Fijacko, Marinka Zitnik, Katrien Verbert, and Leona Cilar. "Interpretability of machine learning-based prediction models in healthcare." Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery 10, no. 5 (2020): e1379.