Review Article

ACCENTS Transactions on Information Security, Vol 8(30)
ISSN (Online): 2455-7196
http://dx.doi.org/10.19101/ TIS.2023.829002

Machine learning in disease detection: a review of advancements, challenges, and implications for healthcare

Tanushree Pandya* **and Md Zuber**
School of Computer Science and Engineering, Madhyanchal Professional University, Bhopal, India

Received: 15-03-2023; Revised: 14-04-2023; Accepted: 17-04-2023

Abstract
The integration of machine learning into healthcare, particularly in disease detection, has the potential to transform patient outcomes and alleviate healthcare burdens. This specialized review explores the vast landscape of literature surrounding the application of machine learning in disease detection. Highlighted advantages include early detection, accuracy, personalized medicine, reduction in healthcare costs, and significant public health impact. However, challenges such as data quality, interpretability, data privacy, and integration into clinical practice persist. As machine learning techniques evolve to address these challenges, their role in disease detection is poised to become more integral to modern healthcare, promising better healthcare delivery.

Keywords
Machine learning. Disease detection. Early diagnosis. Personalized medicine. Healthcare challenges.

1.Introduction
In our rapidly advancing world, where the boundaries of technology continue to expand, the role of machine learning in healthcare is becoming increasingly prominent [1-4]. One of the most significant areas where machine learning has shown immense promise is disease detection and prediction [1-4]. Chronic diseases, including heart diseases, kidney diseases, and neurocircuitry disorders, are a growing global concern. Early diagnosis is often the key to effective treatment and improved patient outcomes. As traditional diagnostic methods may have limitations in detecting diseases at their nascent stages, the integration of machine learning techniques into healthcare has become a promising solution. This specialized review aims to explore into the vast landscape of literature surrounding the application of machine learning in disease detection.

Disease detection has always been a pivotal component of the healthcare system [5, 6]. Traditionally, healthcare providers have relied on clinical assessments, medical tests, and patient history to diagnose various ailments.

While these methods have proven effective, their limitations become evident when dealing with diseases that exhibit subtle or no symptoms in their early stages [7, 8]. In many cases, patients remain unaware of their conditions until they reach advanced and often irreparable stages of the disease. This situation indicating the pressing need for early detection methods that can significantly enhance the chances of successful treatment and recovery [9, 10].

This literature encompasses a diverse range of diseases, including but not limited to cardiovascular diseases, chronic kidney disease, Alzheimer's disease, and various neurocircuitry disorders. Machine learning models like logistic regression, random forest, support vector machine, and decision trees have proven instrumental in this context, demonstrating exceptional performance in classifying diseases and predicting their progression [11–15]. In the realm of heart disease, machine learning has been employed to predict coronary diseases, making early intervention and treatment a reality [16, 17]. These predictive models offer clinicians a powerful tool to assess the risk of heart disease and prescribe preventative measures or treatments in a timely fashion. Additionally, the application of machine learning in chronic kidney disease detection has gained momentum [18, 19]. Researchers have harnessed algorithms like k-nearest neighbor, support vector machine, and decision trees to create

*Author for correspondence

predictive models that can identify kidney disease in its early stages. These models not only improve patient outcomes but also significantly reduce the burden on healthcare systems. Neurocircuitry disorders, such as Alzheimer's and Parkinson's disease, have posed significant challenges in terms of early detection [20, 21]. Machine learning techniques have shown promise in analyzing neuroimaging data and clinical data to develop models that can detect and predict these diseases in their incipient stages. The potential to revolutionize early diagnosis and intervention in these disorders has made this area of research particularly appealing.

The motivation behind exploring the role of machine learning in disease detection is fueled by the urgency to address global health challenges. Machine learning presents a transformative approach to mitigate this issue by offering early and accurate disease detection. Through the integration of machine learning models, healthcare providers can make timely interventions, tailor treatment plans, and ultimately save lives. The motivation stems from the realization that this innovative approach has the potential to reshape healthcare and significantly enhance patient care.

The primary objective of this specialized review is to synthesize the existing literature on machine learning in disease detection and prediction. By critically evaluating the methodologies, findings, and limitations of various studies, we aim to provide a holistic overview of the state-of-the-art in this field. *Figure 1* shows the applications of machine learning in disease detection.

This paper is structured as follows: Section 2 presents the literature review. Section 3 explores the discussion and analysis. Finally, Section 4 provides the concluding remarks.

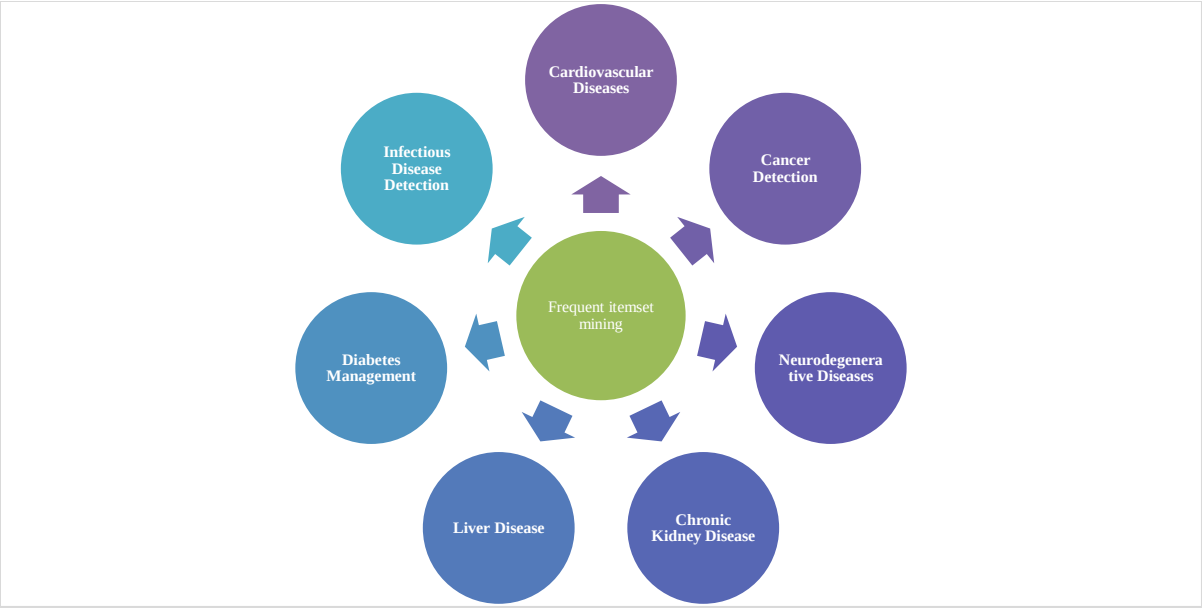

Figure 1 Applications of machine learning in disease detection

2.Literature review

In 2021, a study was conducted by Swamy and Divya [22] to address the challenge of diagnosing skin diseases due to complexities in skin texture, hair presence, and color variations. Machine learning techniques were applied for more accurate diagnosis. The process involved three stages: feature extraction, training, and testing, using color and texture features to classify skin diseases. Entropy, variance, and HSV features were employed in Decision Tree and Support Vector Machine algorithms to enhance accuracy. The algorithm's performance was evaluated using accuracy metrics.

In 2021, Ramanathan et al. [23] addressed global visual impairment causes, developing a system using Logistic Regression, Random Forest, Gradient Boosting, and Support Vector Machine algorithms for early detection of cataract, glaucoma, and retinal diseases. Their work aimed to reduce blindness rates, with Gradient Boosting achieving 90% accuracy for

cataract, and logistic regression and random forest at 89% and 86%, respectively.

In 2021, Likitha et al. [24] addressed the increasing prevalence of heart disease by developing a machine learning-based system to enable early detection. The study utilized patient characteristics such as age, chest pain, blood pressure, gender, cholesterol, and heart rate to distinguish high-risk individuals. Various algorithms, including Logistic Regression, K-Nearest Neighbor, Decision Tree, Naive Bayes, Random Forest, and Support Vector Machine, were employed to enhance heart disease diagnosis effectively.

In 2022, Sahoo et al. [25] developed a personalized at-home system for early heart disease risk assessment. Machine learning models, including Logistic Regression, K-Nearest Neighbor, Support Vector Machine, Naive Bayes, Decision Tree, Random Forest, and XG Boost, were applied. The study emphasized the importance of timely detection and evaluated the model's performance using the Cleveland Heart Disease dataset, where Random Forest achieved the highest accuracy at 90.16%. This approach can reduce hospital burdens and prioritize critical patients for more efficient healthcare delivery.

In 2022, Dixit et al. [26] explored the use of deep learning and machine learning in the study of large-scale multimodal neuroimaging data to detect and prevent neurocircuitry disorders. Their work focused on early detection and classification of Alzheimer's disease, Parkinson's detection, generalized anxiety disorder, and stress prediction, aiming to facilitate early treatment. The study also reviewed relevant literature on neurological diseases, emphasizing the potential of deep learning for early disease diagnosis.

In 2022, Prabhakar and Bhargavi [27] addressed the challenges of diagnosing Alzheimer's disease in its early stages. They highlighted the gradual onset of Alzheimer's disease, making early intervention critical. The study focused on non-amyloid blood plasma proteins as potential biomarkers for early Alzheimer's disease detection. A machine learning model was developed to identify these proteins, aiding in timely diagnosis and intervention.

In 2022, Gulati et al. [28] focused on addressing the global issue of heart diseases, which account for 32% of all deaths. Coronary heart diseases are a major contributor. The study emphasized the importance of early recognition and timely treatment to reduce fatality rates. Supervised Machine Learning was employed for detection and classification, achieving high accuracy. Six algorithms were tested on the Cleveland Heart Disease dataset using the Weka tool, with Naive Bayes demonstrating superior accuracy in predicting coronary diseases.

In 2022, Rasheed and Glob [29] recognized the importance of leveraging machine learning algorithms, especially in the medical field, for disease detection and prediction. They conducted a review of recent studies using machine learning techniques for early disease detection. Naive Bayes outperformed other algorithms with an 84% accuracy rate in detecting various diseases, highlighting the significance of these techniques in healthcare.

In 2023, Kaur and Kaur [30] addressed the public health concern of chronic kidney disease (CKD). They explored the use of machine learning algorithms for CKD prediction, employing Naive Bayes, Support Vector Machine, k-Nearest Neighbor, and linear regression models. After preprocessing a CKD dataset from Kaggle, the SVM model stood out with a remarkable 99.04% accuracy. This study highlights the potential of ML in enhancing CKD prediction, offering opportunities for improved early diagnosis and treatment, ultimately benefitting patients and alleviating the healthcare system's burden.

In 2023, Umbare et al. [31] recognized the importance of early liver disease detection due to the limited symptoms in its early stages. They employed machine learning techniques to facilitate early diagnosis, identify factors leading to severe liver impairment, and categorize the disease's stage and type. By addressing the challenge of diagnosing liver disease in its early, asymptomatic stages, this initiative aims to improve patient outcomes.

In 2023, Shukla et al. [32] addressed the critical issue of chronic kidney disease (CKD) and the challenges of early detection. CKD poses a significant health risk, as it often remains asymptomatic until the late stages, making early detection crucial for saving lives. Machine learning algorithms, including KNN, Decision Tree, and ANN, were employed to predict CKD. Decision Tree exhibited the highest accuracy at 98.60%. The study emphasized the importance of specific attributes in early CKD detection and reviewed the previous research in this domain.

The reviewed papers collectively demonstrate the transformative impact of machine learning in disease detection across various medical domains. Researchers have harnessed machine learning techniques to address the critical challenge of early disease detection, saving lives and reducing healthcare burdens. Notable applications include predicting and preventing cardiovascular diseases, detecting cancer in its early stages, and aiding in the diagnosis of neurodegenerative disorders. The use of machine learning also extends to addressing kidney and liver diseases, improving diabetes management, and predicting the spread of infectious diseases. These studies highlight the potential of machine learning to reshape modern healthcare by offering early, accurate diagnoses and personalized treatment plans.

3.Discussion and analysis

The papers reviewed here collectively indicate the transformative power of machine learning in disease detection and early intervention, which has significant implications for improving patient outcomes and alleviating the burden on healthcare systems. The major advantages are as under:

Early Detection: One of the primary advantages across these studies is the ability of machine learning to detect diseases in their early stages, often before clinical symptoms become apparent. This early detection is critical for timely intervention and treatment, which can significantly improve patient prognosis.

Accuracy: Machine learning algorithms can analyze vast amounts of patient data and medical images with high accuracy, often outperforming traditional diagnostic methods. This enhanced accuracy not only aids in disease detection but also reduces the likelihood of misdiagnosis.

Personalized Medicine: Machine learning models can tailor treatment plans to individual patients based on their unique characteristics. This personalized approach is especially relevant in heart disease and cancer management, where treatment efficacy varies from person to person.

Reduction in Healthcare Costs: Early disease detection and accurate diagnosis can lead to cost savings in the long run. Patients with early-stage diseases often require less invasive and less expensive treatments compared to those with advanced-stage conditions.

Public Health Impact: The ability of machine learning to predict the spread of infectious diseases, , contributes to public health efforts by enabling better containment strategies and resource allocation.

The major challenges are as under:

Data Quality: The success of machine learning models relies on the quality and quantity of data. In some cases, it may be challenging to obtain large and diverse datasets for training, especially for rare diseases or when dealing with sensitive medical information.

Interpretability: Machine learning models are often seen as "black boxes," making it difficult for healthcare professionals to understand how the model arrived at a specific diagnosis. This can hinder trust and acceptance of these models in clinical practice.

Data Privacy and Security: Healthcare data is sensitive and subject to strict privacy regulations.

Integration with Clinical Practice: To be effective, machine learning models need to be seamlessly integrated into clinical workflows. Healthcare professionals must trust and understand how to use these tools effectively.

Overfitting: Machine learning models can be prone to overfitting, where they perform exceptionally well on training data but poorly on unseen data. This can result in false positives or negatives in clinical settings.

The application of machine learning in disease detection is rapidly changing the landscape of healthcare. These studies have demonstrated the potential to detect diseases early, improve accuracy, and personalize treatment plans, ultimately leading to better patient outcomes. Despite the many advantages, challenges remain, including data quality, interpretability, data privacy, and the need for seamless integration into clinical practice. As machine learning techniques continue to evolve and address these challenges, their role in disease detection is expected to become even more integral to modern healthcare practices. The potential for early disease detection, personalized treatment, and improved patient outcomes is promising, making these advancements a significant stride towards better healthcare delivery.

4.Conclusion and future work

The reviewed literature collectively demonstrates the transformative power of machine learning in disease detection and early intervention. This approach offers numerous advantages, including early disease detection, improved accuracy, personalized medicine, and a reduction in healthcare costs. It also contributes to public health efforts by predicting the spread of infectious diseases. Nevertheless, challenges like data quality, interpretability, data privacy, and integration with clinical practice persist. As machine learning

continues to evolve and overcome these challenges, its role in disease detection is expected to play a more crucial role in modern healthcare, with the potential for earlier diagnosis, personalized treatment, and better patient outcomes.

Acknowledgment
None.

Conflicts of interest
The authors have no conflicts of interest to declare.

References
[1] Nagavelli U, Samanta D, Chakraborty P. Machine learning technology-based heart disease detection models. Journal of Healthcare Engineering. 2022; 2022.
[2] Chang V, Bhavani VR, Xu AQ, Hossain MA. An artificial intelligence model for heart disease detection using machine learning algorithms. Healthcare Analytics. 2022; 2:100016.
[3] Kamal M, Pratap AR, Naved M, Zamani AS, Nancy P, Ritonga M, et al. Machine learning and image processing enabled evolutionary framework for brain MRI analysis for alzheimer's disease detection. Computational Intelligence and Neuroscience. 2022; 2022.
[4] Konstantonis G, Singh KV, Sfikakis PP, Jamthikar AD, Kitas GD, Gupta SK, et al. Cardiovascular disease detection using machine learning and carotid/femoral arterial imaging frameworks in rheumatoid arthritis patients. Rheumatology International. 2022; 42(2):215-39.
[5] Aljalal M, Aldosari SA, AlSharabi K, Abdurraqeeb AM, Alturki FA. Parkinson's disease detection from resting-state EEG signals using common spatial pattern, entropy, and machine learning techniques. Diagnostics. 2022; 12(5):1033.
[6] Ketu S, Mishra PK. Empirical analysis of machine learning algorithms on imbalance electrocardiogram based arrhythmia dataset for heart disease detection. Arabian Journal for Science and Engineering. 2022: 1-23.
[7] Dubey AK, Gupta U, Jain S. Comparative study of K-means and fuzzy C-means algorithms on the breast cancer data. International Journal on Advanced Science, Engineering and Information Technology. 2018; 8(1):18-29.
[8] Altaf I, Butt MA, Zaman M. Disease detection and prediction using the liver function test data: A review of machine learning algorithms. In international conference on innovative computing and communications: proceedings of ICICC 2021, 2022 (pp. 785-800). Springer Singapore.
[9] Dubey AK, Sinhal AK, Sharma R. Heart disease classification through crow intelligence optimization-based deep learning approach. International Journal of Information Technology. 2023:1-6.
[10] Abdualgalil B, Abraham S, Ismael WM. Efficient ensemble machine learning techniques for early prediction of diphtheria diseases based on clinical data. International Journal of Advanced Technology and Engineering Exploration. 2022; 9(90):583-603.
[11] Alqahtani A, Alsubai S, Sha M, Vilcekova L, Javed T. Cardiovascular disease detection using ensemble learning. Computational Intelligence and Neuroscience. 2022; 2022.
[12] Nemade V, Pathak S, Dubey AK. A systematic literature review of breast cancer diagnosis using machine intelligence techniques. Archives of Computational Methods in Engineering. 2022; 29(6):4401-30.
[13] Rela M, Rao SN, Patil RR. Performance analysis of liver tumor classification using machine learning algorithms. International Journal of Advanced Technology and Engineering Exploration. 2022; 9(86):143-54.
[14] Travis WD, Brambilla E, Noguchi M, Nicholson AG, Geisinger KR, Yatabe Y, et al. International association for the study of lung cancer/american thoracic society/european respiratory society international multidisciplinary classification of lung adenocarcinoma. Journal of thoracic oncology. 2011; 6(2):244-85.
[15] Djirackor L, Halldorsson S, Niehusmann P, Leske H, Capper D, Kuschel LP, et al. Intraoperative DNA methylation classification of brain tumors impacts neurosurgical strategy. Neuro-Oncology Advances. 2021; 3(1).
[16] Chahar R, Kaur D. A systematic review of the machine learning algorithms for the computational analysis in different domains. International Journal of Advanced Technology and Engineering Exploration. 2020; 7(71):147-64.
[17] Dubey AK, Choudhary K. A systematic review and analysis of the heart disease prediction methodology. International Journal of Advanced Computer Research. 2018; 8(38):240-56.
[18] Iftikhar H, Khan M, Khan Z, Khan F, Alshanbari HM, Ahmad Z. A comparative analysis of machine learning models: a case study in predicting chronic kidney disease. Sustainability. 2023; 15(3):2754.
[19] Islam MA, Majumder MZ, Hussein MA. Chronic kidney disease prediction based on machine learning algorithms. Journal of Pathology Informatics. 2023; 14:100189.
[20] Nancy Noella RS, Priyadarshini J. Machine learning algorithms for the diagnosis of Alzheimer and Parkinson disease. Journal of Medical Engineering & Technology. 2023; 47(1):35-43.
[21] Rajab MD, Jammeh E, Taketa T, Brayne C, Matthews FE, Su L, et al. Cognitive function and ageing neuropathology study group. assessment of alzheimer-related pathologies of dementia using machine learning feature selection. Alzheimer's Research & Therapy. 2023; 15(1):47.
[22] Swamy KV, Divya B. Skin disease classification using machine learning algorithms. In 2nd international

conference on communication, computing and industry 4.0 (C2I4) 2021 (pp. 1-5). IEEE.
[23] Ramanathan G, Chakrabarti D, Patil A, Rishipathak S, Kharche S. Eye disease detection using machine learning. In 2nd global conference for advancement in technology (GCAT) 2021 (pp. 1-5). IEEE.
[24] Likitha KN, Nethravathi R, Nithyashree K, Kumari R, Sridhar N, Venkateswaran K. Heart disease detection using machine learning technique. In second international conference on electronics and sustainable communication systems (ICESC) 2021 (pp. 1738-43). IEEE.
[25] Sahoo GK, Kanike K, Das SK, Singh P. Machine learning-based heart disease prediction: a study for home personalized care. In 32nd international workshop on machine learning for signal processing (MLSP) 2022 (pp. 1-06). IEEE.
[26] Dixit S, Gaikwad A, Vyas V, Shindikar M, Kamble K. United neurological study of disorders: alzheimer's disease, parkinson's disease detection, anxiety detection, and stress detection using various machine learning algorithms. In international conference on signal and information processing (IConSIP) 2022 (pp. 1-6). IEEE.
[27] Prabhakar B, Bhargavi MS. A machine learning model to identify best blood plasma proteins for early detection of alzheimer's disease. In 4th international conference on cybernetics, cognition and machine learning applications (ICCCMLA) 2022 (pp. 190-5). IEEE.
[28] Gulati S, Guleria K, Goyal N. Classification and detection of coronary heart disease using machine learning. In 2nd international conference on advance computing and innovative technologies in engineering (ICACITE) 2022 (pp. 1728-32). IEEE.
[29] Rasheed SS, Glob IH. Classifying and prediction for patient disease using machine learning algorithms. In information technology to enhance e-learning and other application (IT-ELA) 2022 (pp. 196-200). IEEE.
[30] Kaur M, Kaur G. Comparative Analysis of Machine Learning Algorithms for Early Detection of Chronic Kidney Disease. In 4th international conference on electronics and sustainable communication systems (ICESC) 2023 (pp. 1754-9). IEEE.
[31] Umbare RT, Ashtekar O, Nikhal A, Pagar B, Zare O. Prediction and Detection of Liver Diseases using Machine Learning. In 3rd international conference on technology, engineering, management for societal impact using marketing, entrepreneurship and talent (TEMSMET) 2023 (pp. 1-6). IEEE.
[32] Shukla G, Dhuriya G, Pillai SK, Saini A. Chronic kidney disease prediction using machine learning algorithms and the important attributes for the detection. In IAS global conference on emerging technologies (GlobConET) 2023 (pp. 1-4). IEEE.

Tanushree Pandya is pursuing M.Tech in Computer Science & Engineering at Madhyanchal Professional University, Bhopal (MP), India , and holds a B.Tech degree in Information Technology from Gyan Ganga Institute of Technology and Sciences Jabalpur (MP), India . His primary area of interest is Machine Learning.
Email: tanushreepandya08@gmail.com