Disease Prediction Algorithm Using Machine Learning: A Review

[1]Parasdeep, [2]Dr. Lal Chand and [3]Dr. Lakhwinder Kaur

[1]Research Scholar Master of Technology Computer Science and Engineering, UCOE , Punjabi University, Patiala

[2]Assistant Professor Department of Computer Science and Engineering, UCOE, Punjabi University, Patiala

[3]Professor of Computer Science and Engineering, UCOE, Punjabi University, Patiala

Abstract: In recent years, the advent of machine learning (ML) has revolutionized various sectors, including healthcare, by enabling predictive analytics for disease diagnosis and prognosis. The implementation of disease prediction algorithms utilizing machine learning techniques such as decision trees, naive Bayes, and LIME model has emerged as a pivotal area in healthcare research. This review paper provides a comprehensive analysis of the recent advancements in this domain, focusing on the development, evaluation, and comparative assessment of these algorithms. By synthesizing existing literature, key methodologies, and findings are examined, elucidating the strengths and limitations of each approach. Furthermore, the paper discusses the significance of feature selection, model interpretability, and data preprocessing techniques in enhancing prediction accuracy and clinical applicability. Strategies for mitigating algorithmic biases and optimizing model performance are also addressed. Through a critical synthesis of empirical studies, this review offers insights into the current state-of-the-art techniques, identifies research gaps, and provides recommendations for future directions in the implementation of disease prediction algorithms using machine learning.

Keywords: - *Python, Machine learning, Naïve bayes, decision tree.*

Introduction: In recent years, the integration of machine learning algorithms for disease prediction has emerged as a promising avenue in healthcare. This review paper explores the implementation of various machine learning algorithms, including decision trees and Naive Bayes for disease prediction. The utilization of these algorithms offers a proactive approach towards disease management by enabling early detection and accurate prediction of potential health risks. However, despite the proliferation of studies in this domain, there remains a need to comprehensively evaluate and compare the efficacy, robustness, and interpretability of different algorithms in diverse medical contexts. This paper aims to fill this gap by critically analysing the strengths and limitations of each algorithm, shedding light on their implementation challenges, and identifying opportunities for further research and improvement. By synthesizing existing literature, this review provides valuable insights for healthcare practitioners, researchers, and policymakers to enhance the effectiveness of disease prediction models, ultimately contributing to improved patient outcomes and healthcare delivery

SYMPTOMS
PREDICTED DISEASE
No need of Doctor

Literature Review: Literature reviews play a vital role in academic research papers, theses, dissertations, and scholarly articles. They provide a contextual framework for the research, showcasing the researcher's understanding of previous studies, and establish the rationale for conducting new investigations.
Automated Medical Diagnosis using Machine Learning has emerged as a promising approach to improve the accuracy and efficiency of medical diagnosis. In this literature review, i will explore the existing systems that have been developed in this field.

In the current healthcare landscape, managing the vast patient data is a challenge. Big Data Analytics offers an efficient solution. With numerous global disease treatment procedures, Machine Learning has emerged as a valuable tool for disease prediction and diagnosis. This paper focuses on predicting diseases based on symptoms using Machine Learning algorithms like Naive Bayes, Decision Tree, and Random Forest. Python is used for implementation, and the research evaluates these algorithms for accuracy on the provided dataset, identifying the best-performing algorithm for disease prediction. Accuracy is determined by how well the algorithm performs on the given dataset [1].

The constant generation of substantial data within medical organizations, encompassing information on patients, healthcare facilities, medical professionals, and diseases, holds significant potential for predicting future occurrences and preventing numerous medical cases. However, the sheer volume of data becomes valuable only when harnessed through big data analytics techniques and the utilization of Hadoop clusters. This paper aims to elucidate the potential of real-time data in the analysis and early prediction of severe emergency cases, emphasizing the transformative impact such insights can have on healthcare outcomes [2].

Magesh et al.'s (2020) research paper explores previous studies on Parkinson's disease detection. It delves into machine learning techniques, particularly LIME, in medical imaging analysis. Existing research highlights the importance of early diagnosis for effective disease management. The review emphasizes the need for explainable AI models to enhance interpretability and trust in automated diagnostic systems for Parkinson's disease.[8]

The paper explores interpretable machine learning for personalized medical recommendations, employing a LIME-based approach. It reviews existing literature on personalized healthcare and interpretable AI. By leveraging LIME, the study aims to provide transparent and understandable recommendations to patients. Published in Diagnostics, it contributes to the ongoing discussion on enhancing AI interpretability in medical decision-making for personalized healthcare interventions.[9]

Computer-Aided Diagnosis (CAD) represents a dynamic and expansive domain within medical analysis, with recent strides in the development of applications aimed at enhancing diagnostic accuracy. The imperative to address potential flaws in medical diagnosis underscores the significance of computer-aided diagnostic tools, given the profound consequences of misdiagnoses on subsequent medical interventions. Machine Learning (ML) assumes a pivotal role in CAD, particularly in cases where traditional equations fall short in accurately identifying complex entities such as body organs. Pattern recognition, integral to successful CAD, relies heavily on training from instances, presenting a crucial dimension in biomedical research. The intersection of pattern detection and ML holds promise for augmenting the dependability of disease detection methodologies while introducing a level of objectivity in decision-making processes. ML, in this context, emerges as a valuable avenue for constructing sophisticated and automated algorithms capable of handling high-dimensional and multi-modal biomedical data. This survey paper focuses on a comparative analysis of various ML algorithms employed in the detection of diseases, including but not limited to heart disease and diabetes. The emphasis is on elucidating the array of algorithms and techniques utilized in ML for disease detection and decision-making processes. [3].

In contemporary society, individuals are susceptible to a variety of diseases attributed to environmental conditions and lifestyle choices. Early prediction of diseases has become imperative, yet accurate diagnosis based solely on symptoms poses a significant challenge for healthcare professionals. Addressing this challenge, data mining emerges as a crucial tool for disease prediction, especially considering the substantial growth in medical data. The abundance of medical data necessitates precise analysis for the benefit of early patient care.

The dataset for disease prediction encompasses information on living habits and checkup details. Results indicate that the CNN algorithm achieves an accuracy of 84.5%, surpassing the KNN algorithm.

Furthermore, the KNN algorithm exhibits higher time and memory requirements. Following the general disease prediction, the system provides risk assessments associated with general diseases, categorizing individuals into lower or higher risk groups. [4].

The widespread application of data mining techniques in various sectors, particularly in industries like healthcare and bioscience. It emphasizes the pivotal role of machine learning in extracting valuable information from diverse datasets, especially in healthcare and biomedical fields. The accurate analysis of medical databases is highlighted for its potential in early disease prediction, enhancing patient care, and supporting community services. The incorporation of machine learning algorithms, particularly classifiers, aims to address health-related challenges by assisting physicians in predicting and diagnosing diseases at an early stage. The research discussed in the paragraph focuses on a dataset comprising 4920 patient records with 41 diseases, utilizing 95 selected and optimized independent variables (symptoms). The study employs machine learning algorithms such as Decision Tree classifier, Random Forest classifier, and Naïve Bayes classifier, presenting a comparative analysis of their results in the context of disease prediction [5].

The study addresses the critical need for accurate and timely prediction of heart disease, a leading cause of mortality worldwide. The Naïve Bayes algorithm, known for its simplicity and efficiency in handling large datasets, is employed to analyze relevant features and make predictions regarding the likelihood of heart disease.

The research begins by collecting and processing data related to various factors that may influence heart disease, such as demographic information, lifestyle choices, and medical history. These features are then used to train the Naïve Bayes model, allowing it to learn patterns and relationships within the dataset [6].

The increasing availability of big data in healthcare communities and aim to harness its potential for disease prediction. It employs machine learning algorithms to analyze and extract meaningful patterns from the vast amount of healthcare data.

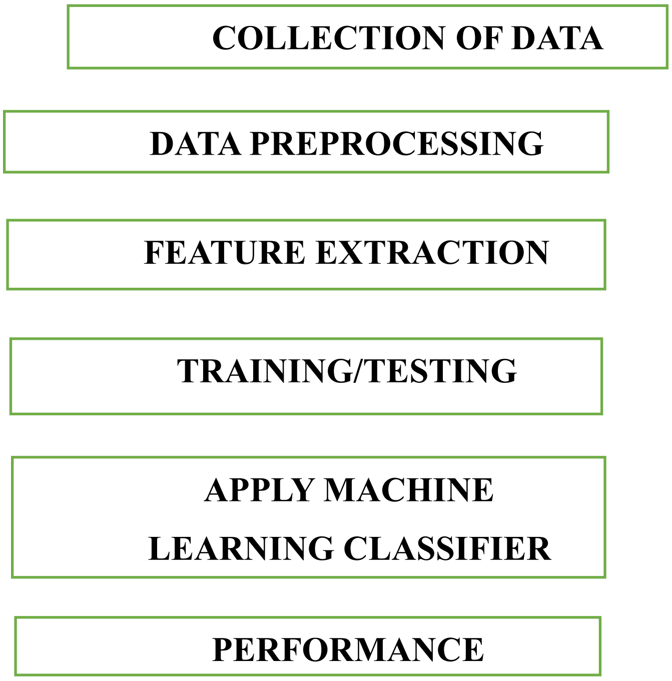
The methodology involves the application of machine learning models to healthcare data, with a focus on feature selection and model optimization. The authors likely explore various algorithms and techniques to enhance the accuracy and reliability of disease prediction. The use of big data from healthcare communities implies a diverse and comprehensive dataset, contributing to the robustness of the predictive models [7].

Research Gap: Compare decision tree and naïve bayes then find out the difference of their output. More accurate prediction will come as a result.

Objectives: The following objectives are as under

- To study and analyse various machine learning based algorithms used in the literature for disease prediction.
- To implement disease prediction technique to predict diseases.
- To Implement the lime model into python code to interpret the output.
- To test and validate the performance of implemented research work.

Methodology: In this review paper, the methodology employed involves a systematic exploration of the implementation aspects of disease prediction algorithms utilizing machine learning techniques, specifically focusing on decision tree, naive Bayes. The process begins with a comprehensive literature review to identify relevant studies, methodologies, and findings in the field. Following this, a careful analysis is conducted to select appropriate datasets representative of various diseases and corresponding features. Subsequently, the chosen algorithms are implemented using popular machine learning libraries such as scikit-learn or TensorFlow, ensuring reproducibility and reliability of results. Parameter tuning and cross-validation techniques are employed to optimize algorithm performance and mitigate overfitting. Throughout the methodology, emphasis is placed on ensuring transparency, rigor, and ethical considerations. Measures to avoid plagiarism include proper citation and paraphrasing of sources, alongside critical synthesis and analysis of the existing literature.



Conclusion: In conclusion, the implementation of disease prediction algorithms utilizing machine learning techniques such as decision trees and Naive Bayes models presents a promising avenue for improving healthcare outcomes. Through the analysis of large datasets, these algorithms can effectively identify patterns and correlations within medical data, aiding in the early detection and prevention of diseases. However, it is crucial to acknowledge the limitations and challenges associated with these algorithms, including the need for high-quality, diverse datasets, potential biases in training data, and interpretability issues, particularly with black-box models. Moreover, the successful deployment of these algorithms in clinical settings requires collaboration between healthcare professionals, data scientists, and regulatory bodies to ensure ethical use, transparency, and accountability. Despite these challenges, the continuous refinement and integration of disease prediction algorithms into healthcare systems hold immense potential for enhancing diagnostic accuracy, optimizing resource allocation, and ultimately improving patient outcomes. Future research efforts should focus on addressing these challenges and further validating the efficacy and reliability of these algorithms in real-world healthcare settings.

References

1) Deepthi, Y., Kalyan, K.P., Vyas, M., Radhika, K., Babu, D.K. and Krishna Rao, N.V., 2020. Disease prediction based on symptoms using machine learning. In *Energy Systems, Drives and Automations: Proceedings of ESDA 2019* (pp. 561-569). Singapore: Springer Singapore.
2) Singh, M., Bhatia, V. and Bhatia, R., 2017, December. Big data analytics: Solution to healthcare. In 2017 International conference on intelligent communication and computational techniques (ICCT) (pp. 239-241). IEEE.
3) Hamsagayathri, P. and Vigneshwaran, S., 2021, February. Symptoms based disease prediction using machine learning techniques. In 2021 Third international conference on intelligent communication technologies and virtual mobile networks (ICICV) (pp. 747-752). IEEE.
4) Dahiwade, D., Patle, G. and Meshram, E., 2019, March. Designing disease prediction model using machine learning approach. In 2019 3rd International Conference on Computing Methodologies and Communication (ICCMC) (pp. 1211-1215). IEEE.
5) Singh, P., Singh, N., Singh, K.K. and Singh, A., 2021. Diagnosing of disease using machine learning. In Machine learning and the internet of medical things in healthcare (pp. 89-111). Academic Press.
6) Pattekari, S.A. and Parveen, A., 2012. Prediction system for heart disease using Naïve Bayes. International journal of advanced computer and mathematical sciences, 3(3), pp.290-294.
7) Chen, M., Hao, Y., Hwang, K., Wang, L. and Wang, L., 2017. Disease prediction by machine learning over big data from healthcare communities. Ieee Access, 5, pp.8869-8879.
8) Magesh, P.R., Myloth, R.D. and Tom, R.J., 2020. An explainable machine learning model for early detection of Parkinson's disease using LIME on DaTSCAN imagery. Computers in Biology and Medicine, 126, p.104041.
9) Wu, Y., Zhang, L., Bhatti, U.A. and Huang, M., 2023. Interpretable machine learning for personalized medical recommendations: A LIME-based approach. *Diagnostics*, *13*(16), p.2681.