STM JOURNALS

Research & Reviews: A Journal of
Immunology

https://journals.stmjournals.com/rrjoi

ISSN: 2277-6206 (Online)
ISSN: 2349-1280 (Print)
Volume 14, Issue 3, 2024
September–December
DOI (Journal): 10.37591/RRJoI

Review RRJoI

Multiple Disease Prediction Using Machine Learning Algorithms

Ankit Sharma[1,*], A. N. Kshirsagar[2], Anish Kannawar[3], Abhishek Mishra[4]

Abstract

The incorporation of machine learning algorithms into healthcare has transformed disease prediction and diagnosis. This research introduces a method for predicting various diseases using machine learning techniques. A comprehensive dataset, consisting of patient records, medical histories, and key disease-related features, was utilized to build predictive models. Data preprocessing methods, including feature selection and normalization, were implemented to clean and prepare the dataset. Several machine learning algorithms, such as Decision Trees, Random Forest, Support Vector Machines (SVM), and k-Nearest Neighbors (k-NN), were applied to train and assess the performance of the models. The primary goal of this initiative is to significantly improve healthcare delivery by offering timely and precise predictions for a range of chronic conditions, including diabetes, cardiovascular diseases, cancer, and respiratory disorders. These predictive models will rely on advanced algorithms and data analysis techniques, which will process patient information to generate real-time insights. These insights will then be integrated into a user-friendly, digital platform designed specifically for healthcare professionals. This platform aims to streamline diagnosis and treatment planning, enabling more personalized, proactive care that can lead to better patient outcomes and more efficient healthcare management.

Keywords: Machine learning, disease prediction, predictive modeling, decision trees, random forest, support vector machines, k-nearest neighbors

INTRODUCTION

Healthcare is one of the most concerning businesses when it comes to the collection and processing of data. A significant amount of multidimensional data about patients is produced because of the digital age and technical improvements, including clinical factors, hospital resources, disease diagnostic data, patient records, and medical equipment. To extract knowledge for efficient decision-making, the vast, dense, and complicated data must be processed and assessed. Finding hidden patterns in medical data sets is very possible with the help of medical data mining. The application of various data mining tools and machine learning techniques has transformed healthcare organizations by uncovering important patterns and revealing correlations and relationships among numerous variables in large databases. Medical professionals and doctors need to be well-versed in all relevant diagnostic criteria, patient history, and a combination of drug therapy for effective treatment. However, since they make decisions instinctively, mistakes are possible. Their cognitive abilities are constrained due to several issues.

Data mining and machine learning techniques play a crucial role in intelligently converting

***Author for Correspondence**
Ankit Sharma
E-mail: ankitsharma.fresher@gmail.com

[1,3,4]Student, Department of E&TC, SKNCOE, SPPU, Pune, Maharashtra, India.
[2]Assistant Professor, Department of E&TC, SKNCOE, SPPU, Pune, Maharashtra, India.

Received Date: July 17, 2024
Accepted Date: October 15, 2024
Published Date: November 14, 2024

Citation: Ankit Sharma, A. N. Kshirsagar, Anish Kannawar, Abhishek Mishra. Multiple Disease Prediction Using Machine Learning Algorithms. Research & Reviews: A Journal of Immunology. 2024; 14(3): 35–39p.

available data into meaningful insights to enhance the efficiency of the diagnostic process. Numerous studies have been conducted to explore the use of machine learning for improving diagnostic accuracy and capabilities. It was found that machine learning algorithms could detect with 91.1% accuracy, outperforming the highest-skilled doctor's diagnosis accuracy of 79.97% [1]. To extract characteristics for the best possible sickness diagnosis, prediction, prevention, and therapy, machine learning techniques are specifically applied to datasets of illnesses.

LITERATURE SURVEY

Bayesian classifiers use a structured model along with a set of conditional probabilities, operating under the assumption that each factor impacts the outcome independently of the others. Initially, prior probabilities for each class are computed, and then these are used to assess the likelihood of variable values in an unknown situation. The Bayes network classifier is based on a Bayesian network, which represents a joint probability distribution over a set of categorical features. Kidney disease was predicted using the SVM and Nave Bayes techniques [2]. The authors tried to classify different kidney disease stages using the specified ANFIS algorithm. The goal of the project was to develop an efficient classification algorithm employing a variety of evaluation criteria, including execution time and accuracy. The Nave Bayes algorithm performed better since it gave results faster, despite the SVM algorithm having higher classification accuracy. The findings demonstrate that SVM performs better at predicting renal disease than the Nave Bayes Approach. To predict heart disease, a fuzzy approach utilizing a membership function was applied. The authors aimed to reduce ambiguity and uncertainty in the data by using the Fuzzy KNN Classifier. From a dataset of 550 records, 25 classes were created, each containing 22 items. The dataset was evenly split into training and testing sets. After applying preprocessing techniques, the fuzzy KNN method was implemented. Evaluation metrics such as accuracy, precision, and recall were used to assess its performance. The results showed that the fuzzy KNN classifier outperformed the standard KNN classifier in terms of accuracy [3].

Additionally, support vector machines (SVM) with adaptive features were employed to diagnose breast cancer and diabetes. The goal was to develop an automatic, fast, and flexible diagnostic method using adaptive SVM, where the bias value in traditional SVM was adjusted for improved results. The output of this proposed classifier was “if-then” rules, achieving 100% accuracy in diagnosing both breast cancer and diabetes. Future research should focus on developing more effective methods for adjusting the bias value in conventional SVM [4].

For type 2 diabetes prediction, a hybrid model combining clustering and classification was proposed. The model utilized K-means clustering followed by the C4.5 classification algorithm, with k-fold cross-validation for accurate prediction [5]. With a classification accuracy of 88.38 percent utilizing the hybrid technique, the model produced positive results that could be very helpful for physicians in making wise therapeutic decisions regarding diabetes.

Authors in [6] reported an automated method for resolving challenging questions regarding the prognosis of cardiac illness. This intelligent system was developed using the Naive Bayes methodology to deliver faster, better, and more accurate results. It might help medical professionals make clinical decisions concerning heart attacks. A pacemaker may be added to the order, SMS capabilities added, and mobile applications created for Android and iOS.

A novel method built on the ANN algorithm was developed for the prediction of heart illness [7]. The researchers created an interactive prediction method based on categorization, utilizing an artificial neural network algorithm and considering the thirteen most important clinical factors. The recommended strategy demonstrated 80% accuracy in predicting cardiac problems and can be highly helpful for medical professionals.

MOTIVATION OF SYSTEM

The motivation behind the “Multiple Disease Prediction Using Machine Learning Algorithms” project is to revolutionize healthcare by leveraging advanced technology to predict the likelihood of

various diseases in individuals. By harnessing the power of machine learning algorithms, the system aims to provide accurate and timely predictions, enabling early intervention and personalized healthcare strategies. This initiative strives to enhance preventive care, reduce healthcare costs, and ultimately improve the overall well-being and longevity of individuals [8]

IMPLEMENTATION DETAILS OF MODULE

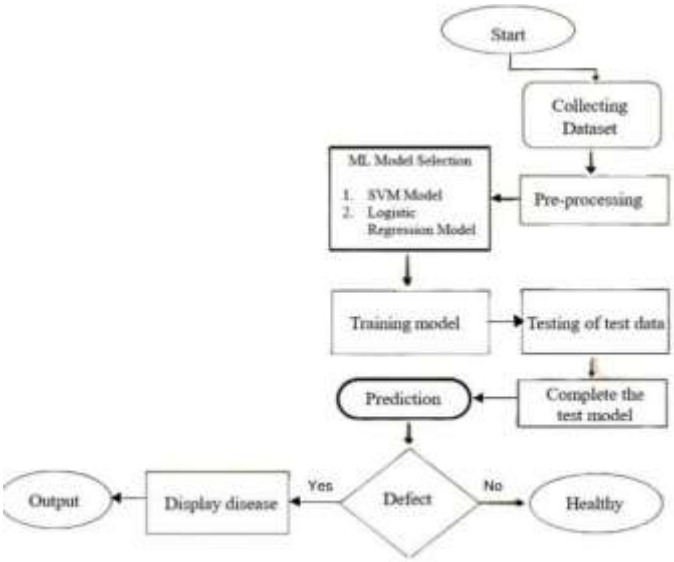

Figure 1. Disease detection process flowchart using machine learning models.

In the initial stages, relevant health data such as patient demographics, medical history, and diagnostic test results are collected and curated. Preprocessing techniques are applied to clean and structure the data, ensuring it is suitable for machine learning analysis. Feature selection is vital for identifying the most relevant variables that enhance the model's predictive accuracy. This process is crucial for optimizing performance and minimizing computational complexity. Machine learning algorithms, including logistic regression, support vector machines, and neural networks, are used to train the predictive model. These algorithms learn from patterns in the training data to develop predictive capabilities. The implementation of various disease prediction models through machine learning has the potential to transform preventive healthcare. By delivering accurate and timely predictions, these models facilitate early intervention and personalized treatment strategies, ultimately leading to better patient outcomes and more efficient resource allocation within the healthcare system [9, 10].

Support Vector Machine

- Import the dataset.
- Explore the data to figure out what they look like.
- Pre-process the data.
- Split the data into attributes and labels.
- Divide the data into training and testing sets.
- Train the SVM algorithm.
- Make some predictions.
- Evaluate the results of the algorithm.

Logistic Regression

Step 1: Data Collection: Gather dataset.
Step 2: Preprocessing: Handle missing values, scale features.
Step 3: Split Data: Divide into train and test sets.

Step 4: Initialize Parameters: Set initial weights and bias.
Step 5: Define Sigmoid Function: Maps feature to probabilities.
Step 6: Compute Cost Function: Measure model error.
Step 7: Gradient Descent: Update parameters to minimize error.
Step 8: Iterate: Repeat steps 5–7 until convergence.
Step 9: Evaluate Model: Assess performance on test set.
Step 10: Prediction: Use trained model for new data.

OUTPUT

CONCLUSION

The objective of the Disease Prediction Using Machine Learning Algorithm project is to create an intelligent and interactive system that can forecast potential diseases based on users' health information. The project intends to utilize machine learning techniques to analyze user input and offer valuable insights into their health risks. This initiative can be beneficial in various ways. It provides users with a convenient tool to assess their health risks and obtain preliminary predictions for potential diseases. This can help users become more proactive in monitoring their health, seeking appropriate medical guidance, and taking necessary precautions.

It can also serve as a valuable resource for individuals who may lack easy access to healthcare services or expert medical guidance. Overall, the project aims to contribute to promoting individual health awareness and preventive healthcare practices. Further one can extend the scope of image-processing diseases like malaria and pneumonia. For the existing diseases, prediction accuracy can be improved by fine-tuning models by having a huge input dataset. Machine learning can leverage the information obtained from big data analysis to produce valuable insights. A chatbot feature in UI could be added as well for ease of communication between the user and system and to provide the user with a subtle UI UX experience.

REFERENCES

1. Kazemi Y, Mirroshandel SA. A novel method for predicting kidney stone type using ensemble learning. Artif Intell Med. 2018; 84:117–26.
2. Barakat H, Andrew P, Bradley H, Mohammed Nabil Barakat H. Intelligible support vector machines for diagnosis of diabetes mellitus. IEEE Trans Inf Technol Biomed. 2009;14(4):1–7.
3. Patil RT, Sherekar SS. Performance analysis of Naive Bayes and J48 classification algorithm for data classification. Int J Comput Sci Appl. 2013;6(2):256–61.
4. Princy RJP, Parthasarathy S, Hency Jose PS, Lakshminarayanan AR, Jeganathan S. Prediction of cardiac disease using supervised machine learning algorithms. In: 2020 4th International Conference on Intelligent Computing and Control Systems (ICICCS); 2020. p. 570–5. doi: 10.1109/ICICCS48265.2020.9121169.

5. Deepika P, Sasikala S. Enhanced model for prediction and classification of cardiovascular disease using decision tree with particle swarm optimization. In: 2020 4th International Conference on Electronics, Communication and Aerospace Technology (ICECA); 2020. p. 1068–72. doi: 10.1109/ICECA49313.2020.9297398.
6. Grampurohit S, Sagarnal C. Disease prediction using machine learning algorithms. In: 2020 Int Conf Emerg Technol (INCET); 2020. p. 1–7. doi: 10.1109/INCET49848.2020.9154130.
7. Ratnakar S, Rajeswari K, Jacob R. Prediction of heart disease using genetic algorithm for selection of optimal reduced set of attributes. Int J Adv Comput Eng Netw. 2013;1(2):51–5.
8. Liang H, Tsui BY, Ni H, et al. Evaluation and accurate diagnoses of pediatric diseases using artificial intelligence. Nat Med. 2019;25(3):433–8.
9. Al-Mallah MH, Aljizeeri A, Ahmed AM, et al. Prediction of diabetes mellitus type-II using machine learning techniques. Int J Med Inform. 2014;83(8):596–604.
10. Dey S, Raza SA, Saha S. Machine learning techniques for diagnosis of diabetes and cardiovascular diseases. Int J Comput Appl. 2016;148(5):7–13.

