MULTI DISEASE PREDICTION SYSTEM USING MACHINE LEARNING

Article · June 2024

DOI: 10.56726/IRJMETS49550

1 author:

Swaroop Sana
Lendi Institute of Engineering and Technology
6 PUBLICATIONS **4** CITATIONS

SEE PROFILE

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:06/Issue:02/February-2024 Impact Factor- 7.868 www.irjmets.com

MULTI DISEASE PREDICTION SYSTEM USING MACHINE LEARNING

Kallepalli Reshma[*1], Pasumarthi Niharika[*2], Javvadi Haneesha[*3], Kodithala Rajavardhan[*4], Sana Swaroop[*5]

[*1,2,3,4,5]Computer Science And Systems Engineering, Lendi Institute Of Engineering And Technology, Vizianagaram, Andhra Pradesh, India.

DOI : https://www.doi.org/10.56726/IRJMETS49550

ABSTRACT

Multi Disease Prediction System (MDPS) that leverages the capabilities of machine learning (ML) algorithms, specifically Logistic Regression and Support Vector Machines (SVM), are utilized for prediction of a particular disease integrated into a user-friendly interface powered by Streamlit library. The objective is to create an accessible and efficient tool for predicting the onset of multi diseases, enhancing early detection and personalized healthcare. This article examines diabetes, heart disease, and Parkinson's disease using basic parameters like blood pressure, pulse rate, cholesterol, and heart rate. It also uses a prediction model with good accuracy and precision to identify the risk factors associated with each condition. The research highlights the potential of machine learning in multi disease prediction and on public health. This training model trains itself to predict disease using sample data. Though there are a lot of algorithms and techniques to predict a disease, there is no proper system to identify multi diseases in a single system. In this paper we focus on the prediction of multi diseases using machine learning. This helps to make a better prediction of disease.

Keywords: Streamlit, Machine Learning, Diabetes, Heart Disease, Parkinson's Disease, SVM, Logistic Regression.

I. INTRODUCTION

In recent years, the field of machine learning has witnessed remarkable advancements and applications in various domains, including healthcare. Machine learning algorithms that can forecast many diseases at once have the potential to transform medical diagnostics and enhance patient outcomes.. All patient-related information is referred to as data in the healthcare sector. A large number of current models focus on a single disease per analysis. Examples of such analyses include those for diabetes, cancer, and skin conditions. There is no common system present that can analyze more than one disease at a time. As a result, our focus is on giving customers precise and quick disease predictions, coupled with information about the symptoms that correspond with the projected disease.. This research paper explores the utilization of the Support Vector Machines (SVM) model to predict the presence of three prevalent diseases: heart disease, diabetes, and Parkinson's disease. As all the three diseases are correlated to each other. The user provides the required parameters, and the application displays the prediction result based on the input. To implement multi disease analyses we are going to use machine learning algorithms and Streamlit. The Streamlit library simplifies the process of developing interactive and user-friendly web applications. The main feature will be the machine learning, in which we will be using algorithms such as Naïve Bayes Algorithm, KNearest Algorithm, Decision Tree Algorithm, Random Forest Algorithm and Support Vector Machine, which will predict accurate disease .The SVM algorithm can handle both linear and nonlinear relationships between input features and target variables, making it suitable for a wide range of medical diagnostic applications. The aim of this study was to create a multi-disease prediction framework with support vector machines (SVMs) and assess how well it predicted Parkinson's disease, diabetes, and heart disease. SVM as a valuable tool in the multi disease prediction domain. We can get closer to developing more precise, timely, and individualized healthcare interventions, which will enhance patient outcomes and create more effective healthcare systems, by utilizing machine learning.

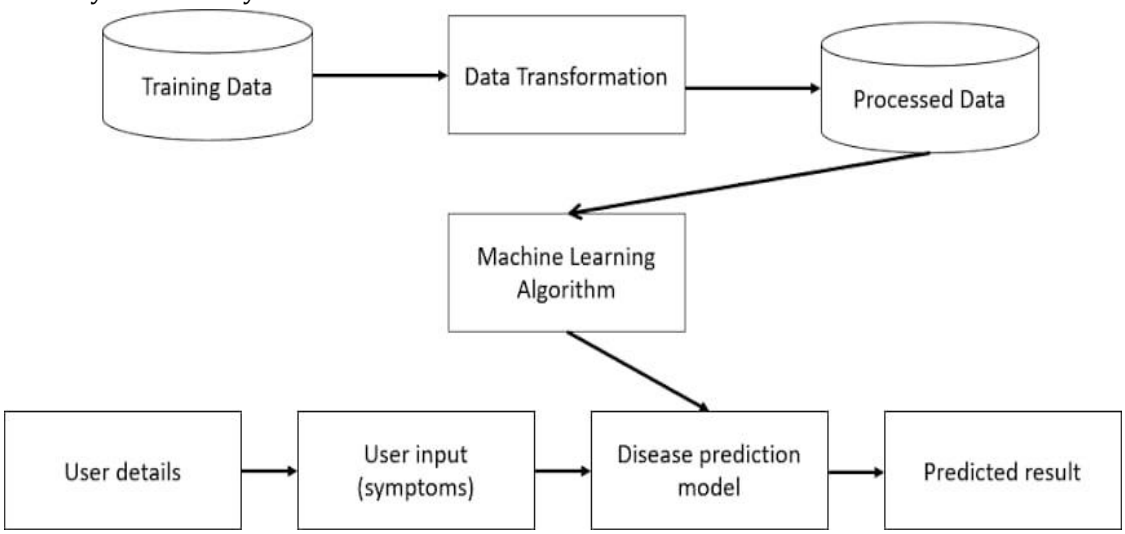
II. METHODOLOGY

Predicting multi diseases simultaneously is a complex task that requires a comprehensive methodology. The methodology for the Multi Disease Prediction project can be summarized as follows:

1. Data Collection**:** Data is collected from various sources such as electronic health records (EHRs), medical literature, public health databases. The data is obtained specifically for diabetes, heart disease and Parkinson's disease.

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:06/Issue:02/February-2024 Impact Factor- 7.868 www.irjmets.com

2. Data preprocessing: It involves two main tasks: data cleaning and transformation, which are essential for preparing the raw data into a suitable format for machine learning algorithms
3. Model Selection**:** Model selection is the initial phase of model training, where the most suitable machine learning algorithms are chosen for each disease. These selected algorithms will then undergo training using the preprocessed data and subsequently be evaluated based on performance metrics such as accuracy, precision to enhance the overall model performance.
4. Data Splitting**:** The preprocessed data will be split into training and testing sets to evaluate the performance of the machine learning models. The models will be trained on the training set, and their performance will be assessed on the testing set.
5. Deployment and Integration**:** It involve deploying the trained models accompanied by its cloud deployment functionalities to develop an interactive web application. This application provides offering with options for predicting various diseases, prompting users to input required parameters upon selection, thus enhancing accessibility and usability.



III. SYSTEM ANALYSIS

FUNCTIONAL REQUIREMENT

- The system provide an intuitive interface that allows users (patients) to input relevant information related to their symptoms, medical history, demographic data, and any other pertinent factors.
- Upon receiving user input, the system should process the data using the integrated prediction models and display the predicted diseases. The output will be presented in a clear and understandable format to the user.

NON FUNCTIONAL REQUIREMENT

- The system has to specify the range of values or confidence intervals associated with predicted disease outcomes. This information helps users understand the uncertainty associated with predictions and aids in making informed decisions about further medical consultation or interventions.
- The system should be reliable and consistent.

IV. PROBLEM STATEMENT

Predicting the onset and progression of multi disease typically focus on analyzing one disease at a time, leading to fragmentation in predictive analysis. For instance, separate models are designed for heart disease, diabetes, Parkinson's disease, and so on. Based on a patient's medical history and symptoms, the Multi Disease Prediction System using Machine Learning is an innovative healthcare application that uses machine learning algorithms to properly forecast the chance of a patient having multiple diseases. The approach involves implementing advanced machine learning techniques to build models that can accurately predict the probability of various diseases. The goal is to identify patients at an early stage, particularly those at risk of diseases such as heart disease, kidney disease, and diabetes, in order to facilitate timely treatment to the improvement of healthcare outcomes and patient care.

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:06/Issue:02/February-2024 Impact Factor- 7.868 www.irjmets.com

V. PROPOSED SYSTEM

The proposed methodology for this project involves leveraging multiple training models for disease prediction, where their performances are compared. This process involves utilizing various libraries like pandas for data handling, numpy for numerical operations, scikit-learn for model training and evaluation, and pickle for exporting the trained model for future application use. Additionally, the approach enables the simultaneous prediction of multi diseases, streamlining the prediction process for users and potentially reducing mortality rates. Compared to existing systems, this method offers faster predictions and numerous advantages, contributing to improved healthcare outcomes.

VI. EXISTING SYSTEM

The existing system focuses on predicting diabetes, heart disease, and Parkinson's disease using a variety of machine learning algorithms, including Naive Bayes, Decision Trees, Random Forest, SVM, and Logistic Regression. Streamlit library facilitate model deployment, offering a user-friendly interface. SVM achieved 76% accuracy in diabetes prediction and 71% in Parkinson's disease, highlighting its effectiveness. Other algorithms like logistic regression and Decision Trees are also employed, each leveraging distinct data characteristics for accurate predictions. Overall, the system demonstrates the potential of ML algorithms in disease prediction, with further enhancements to improve accuracy.

VII. RESULTS AND DISCUSSION

The fields of disease diagnosis and prediction hold significant potential for transformation through ML techniques. Accuracy and correctness in diagnosis are most important aspects in effective treatment procedures for managing sickness. In our system, we employ the SVM algorithm for prediction. Patients input values into the system to determine disease presence. Parameters display required value ranges, and warnings prompt correction for invalid or missing inputs. The accuracy of predictions relies on the support vector machine algorithm's ability to retrieve precise outputs, particularly effective for linear data.

SN.	Disease Name	Algorithm Name	Existing system accuracy	Proposed system accuracy
1	Diabetes	SVM Classifier	76%	79%
2	Heart disease	Logistic Regression	80%	85%
3	Parkinson’s disease	SVM Classifier	72%	89%

VIII. CONCLUSION

The main objective of this research was to create a system that could reliably forecast several illnesses. The user is spared from having to visit multiple websites. Early diagnosis of diseases can both lengthen your life and spare you from financial hardship. To achieve the highest level of accuracy, we have used a variety of machine learning algorithms, including logistic regression, svm. In this study, we investigated the use of machine learning methods for the prognosis of various illnesses, with an emphasis on heart disease, diabetes, and Parkinson's disease. Using the Support Vector Machines (SVM) model, we created a framework for multi-

disease prediction. To guarantee data quality, we gathered information from Kaggle.com and preprocessed it. We used SVM algorithm and obtained 78% accuracy for diabetes prediction. Similarly, we used SVM to predict Parkinson's disease with an 89% accuracy rate. An accuracy of 85% was obtained when heart disease was predicted using logistic regression. The results of this study show how machine learning has the power to transform illness prediction and enhance patient outcomes. The SVM model's implementation required handling and filtering. When a trained model is included into an application, it may forecast diseases in real-world situations, enabling researchers, healthcare providers, and individuals to make well-informed decisions about managing and assessing disease risk. Targeted illness management techniques, individualized treatment regimens, and early interventions may all be made easier with the help of machine learning models for accurate disease prediction. It can improve patient care, help healthcare providers make well-informed decisions, and optimize the distribution of resources within healthcare systems. It also has potential for population-level disease surveillance, which would allow for the early identification of disease outbreaks and the application of preventative measures. In summary, this study highlights the potential of SVM models in multi-disease prediction and advances the field of disease prediction using machine learning. We can get closer to developing more precise, timely, and individualized healthcare interventions by utilizing machine learning, which will ultimately enhance patient outcomes and create more effective healthcare systems.

ACKNOWLEDGEMENT

We would like to convey our heartfelt gratitude to Prof. Sana Swaroop, my mentor, for his valuable advice and assistance in completing my project. He was there to assist us every step of the way, and his motivation is what enabled us to accomplish our task effectively. We would also like to thank all of the other supporting personnel who assisted us by supplying the equipment that was essential.

We would also like to thank the Department of Computer Science and Systems Engineering, Lendi Institute of Engineering and Technology for accepting my project in my desired field of expertise.

IX. REFERENCE

[1] R. Katarya and P. Srinivas, “Predicting heart disease at early stages using machine learning: A survey,” in 2020 International Conference on Electronics and Sustainable Communication Systems (ICESC), 2020,pp. 302–305

[2] Poudel RP, Lamichhane S, Kumar A, et al.Predicting the risk of type 2 diabetes mellitus using data mining techniques. J Diabetes Res.2018;2018:1686023.

[3] S. Ismaeel, A. Miri, and D. Chourishi, “Using the extreme learning ma-chine (elm) technique for heart disease diagnosis,” in 2015 IEEE Canada International Humanitarian Technology Conference (IHTC2015), 2015,pp. 1–3.

[4] A. Gavhane, G. Kokkula, I. Pandya, and K. Devadkar, “Prediction of heart disease using machine learning,” in 2018 Second International Conference on Electronics, Communication and Aerospace Technology(ICECA), 2018, pp. 1275–1278.

[5] Deo RC. Machine learning in medicine. Circulation. 2015;132(20):1920-1930.

[6] Breiman L, Friedman JH, Olshen RA, Stone CJ.Classification and Regression Trees. Wadsworth and Brooks; 1984.

[7] Parashar A, Gupta A, Gupta A. Machine learning techniques for diabetes prediction. Int J Emerg Technol Adv Eng. 2014;4(3):672-675.

[8] Breiman L, Friedman JH, Olshen RA, Stone CJ. Classification and Regression Trees. Wadsworth and Brooks; 1984.

[9] Paniagua JA, Molina-Antonio JD, Lopez-Martinez F, et al. Heart disease prediction using random forests. J Med Syst. 2019;43(10):329.