iJRASET
International Journal For Research in Applied Science and Engineering Technology



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 11 **Issue:** IV **Month of publication:** April 2023

DOI: https://doi.org/10.22214/ijraset.2023.50573

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 11 Issue IV Apr 2023- Available at www.ijraset.com

Disease Prediction using Python

Shantanu Rajesh Mokal[1], Darshil Rajendra Chorghe[2], Neha Jitendra Madhavi[3], Pratik Vijay Kelaskar[4], Prof. S. N. Jadhav[5]
Student , Department of Computer Engineering,, G. M. Vedak Institute Of Technology, Tala, Raigad ,India

Abstract: Disease Prediction based on Symptoms with Machine Learning is a system that predicts diseases based on the user's knowledge of clinical manifestations, ensuring solid conclusions based on such facts. Given how essential the health industry is in treating prescribers' difficulties. This method can be used to learn a little bit about small illnesses if the user only needs to be aware of the illness's basics and the patient isn't in any danger. It's a system that offers clients medical guidance and strategies, as well as a tool to help them identify their illness using this forecast. The healthcare industry as well as those who don't wish to visit a hospital or clinic for their initial diagnosis. By just entering the side effects and other crucial information, the user can learn a great deal about the illness that has been revealed to him or her, and the health sector can profit from this strategy by simply asking the patient for symptoms and providing a diagnosis. We employed machine learning techniques, Python programming with the Tkinter interface, and a dataset collected from hospitals to achieve Illness Prediction based on Symptoms.
Keywords: Training data, Machine learning, Disease prediction, Python etc.

I. INTRODUCTION

The advent of the Android app ushers in the mobile technology era. The economy and the welfare of humanity depend on a functional healthcare system. There has been a significant amount of change between the world we live in today and the one we did a few decades ago. Everything has become more disorganized and ugly. In this case, medical professionals are risking their own lives in order to save as many lives as they possibly can. Board-certified physicians who prefer to practice online via phone and video consultations over in-person consultations are known as "virtual doctors," albeit this is not always feasible in an emergency. Machines are considered to be superior to humans in the absence of human error because they can do tasks more quickly while keeping a constant degree of precision. Without involving a person, a disease predictor, also referred to as a virtual doctor, can correctly forecast a patient's illness. In severe cases, like COVID-19 and EBOLA, a disease predictor can save a person's life by identifying their health without the need for physical contact. There are virtual doctors available now, but they cannot deliver the necessary level of precision. Machines are considered to be superior to humans in the absence of human error because they can do tasks more quickly while keeping a constant degree of precision. Without involving a person, a disease predictor, also referred to as a virtual doctor, can correctly forecast a patient's illness. In severe cases, like COVID-19 and EBOLA, a disease predictor can save a person's life by identifying their health without the need for physical contact. There are virtual doctors available now, but they cannot deliver the necessary level of precision. The technology compares the symptoms to the data that was previously saved. By fusing these datasets with the patient's symptoms, we can predict the patient's disease % with accuracy. Before the user selects the characteristics and enters the symptoms, the dataset and symptoms are uploaded to the system's prediction model, where the data is pre-processed for future references. Following that, the data is categorized using a range of algorithms and techniques, including Decision Tree, KNN, and Naive Bayes, to name a few.

II. PROBLEM STATEMENT

Predicting diseases is a crucial endeavor in healthcare that can aid in early diagnosis and disease prevention. Based on medical characteristics, machine learning algorithms can be used to forecast the incidence of diseases. The goal of this research is to create disease prediction models utilizing machine learning algorithms, specifically Naive Bayes, Decision Tree, and Random Forest, and to assess how well these models perform in foretelling the development of heart disease based on specific medical characteristics.

III. OBJECTIVES

1) To prepare the Heart Disease dataset for machine learning modelling by converting categorical attributes to numerical ones.
2) To put into practice the Decision Tree, Random Forest, and Naive Bayes algorithms for disease prediction based on medical characteristics.
3) To compare the accuracy, precision, recall, F1-score, and ROC curve of the Naive Bayes, Decision Tree, and Random Forest algorithms.

©IJRASET: All Rights are Reserved

4) To determine the benefits and drawbacks of each algorithm for foretelling the development of heart disease based on specific medical characteristics.
5) To shed light on the variables that influence the development of cardiac disease and their respective weight in diagnosing it.
6) Based on the results of this project, propose additional enhancements and future directions for disease prediction using machine learning techniques.

IV. LITERATURE SURVEY

Kaur, H., & Singh, A. (2019). Machine learning for Heart Disease Prediction: A Review. 3rd International Conference on Computing Methodologies and Communication Proceedings (pp. 658-664). Springer. This paper offers a thorough analysis of different machine learning techniques for predicting cardiac disease. The authors evaluate the effectiveness of these algorithms and point out their advantages and disadvantages.

Krittanawong, C., Zhang, H., Wang, Z., Aydar, M., & Kitai, T. (2020). Precision cardiovascular medicine using artificial intelligence. American College of Cardiology Journal, 75(23), 2952-2964. The use of machine learning algorithms for disease prediction in cardiovascular medicine is discussed in this article. The authors outline the most recent state-of-the-art techniques and draw attention to the difficulties and possibilities for upcoming study.

Mirza, A. M., & Ali, A. (2019). Decision trees and the naive Bayes algorithm are compared for the purpose of predicting cardiac disease. Computing and Information Sciences Journal of King Saud University, 31(2), 179–184. The effectiveness of the decision tree and naive Bayes algorithms for heart disease prediction is compared in this study. The authors assess the algorithms using a variety of performance indicators using the Cleveland heart disease dataset.

A. Limitations

1) Just 303 instances make up the dataset utilized in this project, which may reduce the precision of the machine learning models.
2) There are only 14 variables in the dataset, which may not account for all the important factors influencing the development of heart disease.
3) The dataset utilized and the hyperparameters used may have an impact on how well the machine learning models perform.

B. Research Gap

The dataset utilized in this study is somewhat dated and might not accurately reflect the state of health of the current populace. Improved disease prediction might result from updating the dataset and adding more recent data.

The goal of this effort is to forecast heart disease using certain medical characteristics. However, a variety of other elements, including dietary habits, lifestyle choices, and genetics, may also play a role in the development of heart disease. These elements could increase the prediction models' accuracy if they are taken into account.

For disease prediction, the research solely used the three machine learning algorithms Naive Bayes, Decision Tree, and Random Forest. There might be more effective machine learning algorithms for tasks involving disease prediction.

V. PROPOSED SYSTEM

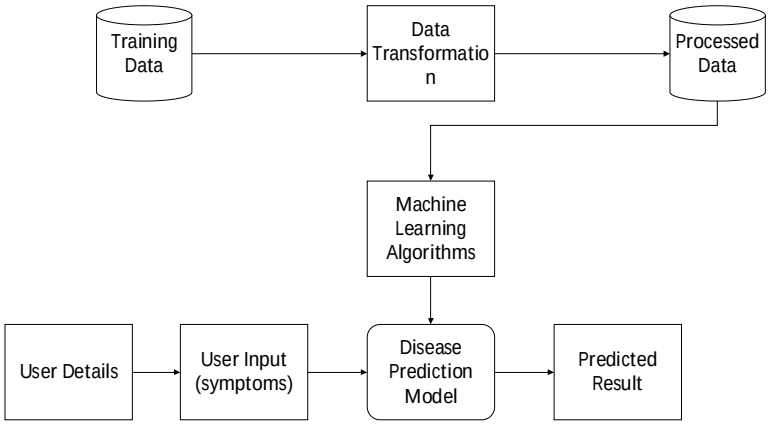


Fig.1. Proposed system

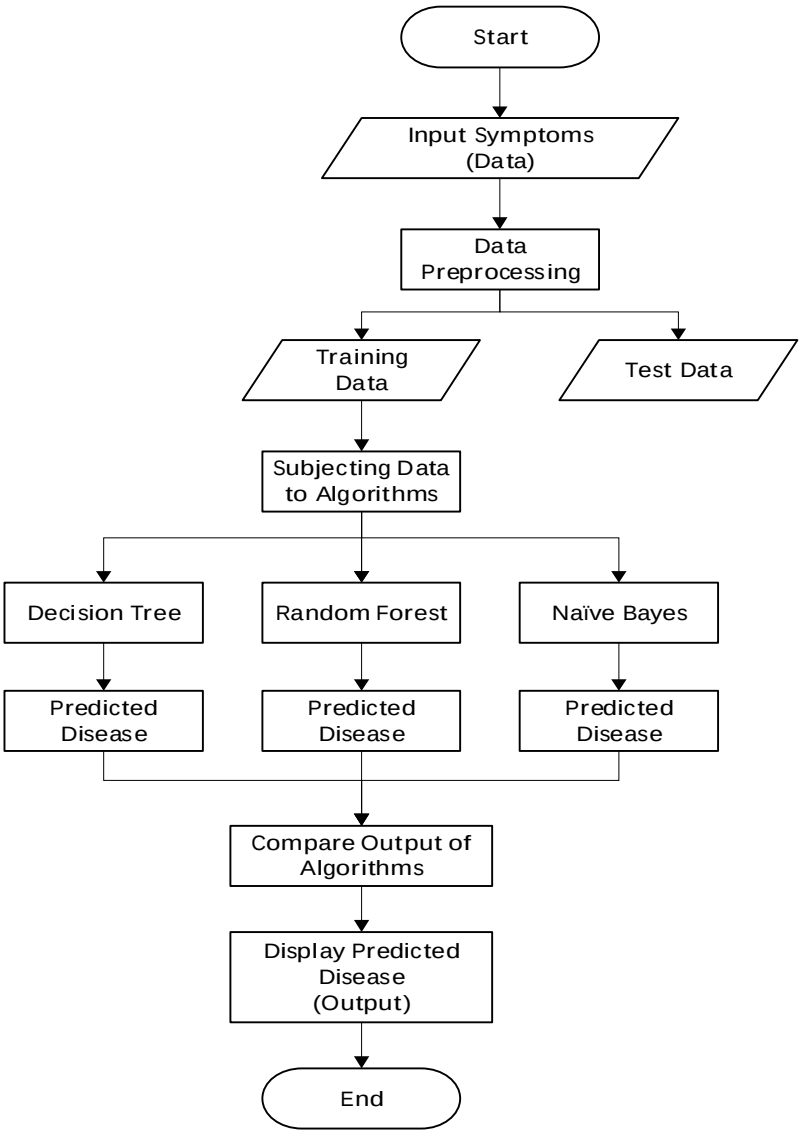


Fig.1. Methodology of Flow Chart

A. Algorithm

1) Naive Bayes Classifier: The Naive Bayes algorithm is a visual representation of the supervised machine learning classification method. By calculating the probabilities of the outcomes/outputs, it uses a probabilistic model. It is applied to analytical and prognostic issues. Noise in the input dataset is tolerated by Naive Bayes.

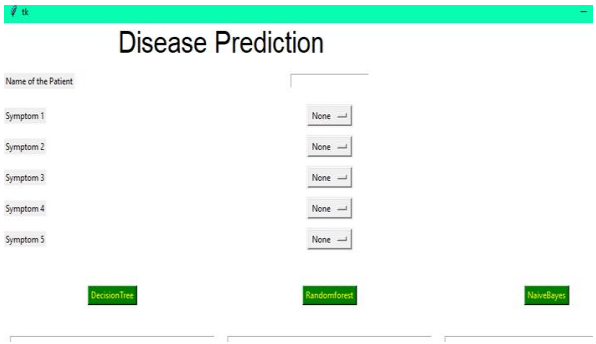
2) Decision Tree: The decision tree learning algorithm works similarly to a decision tree, mapping input about an object to the item's output. Classification trees are tree models with output divided into a finite number of classes. These tree structures have leaves that represent class labels and branches that represent relationships between system attributes and those class labels. Regression trees are decision trees with continuous output classes. A decision tree can be a decision-making input in data mining.

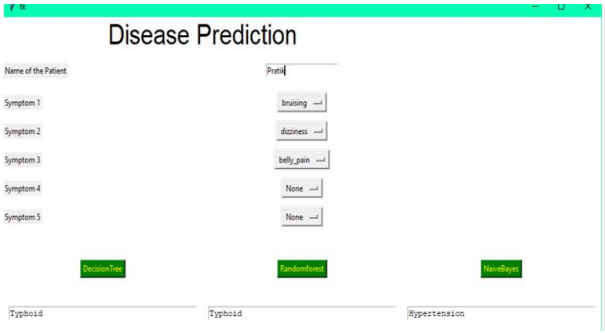
3) Random Forest Algorithm: Trees algorithm and bagging algorithm are used to mimic the Random Forest algorithm. The algorithm's creators discovered that it might increase categorization accuracy. Also, it performs well when applied to data sets with several input factors. The method begins by building a collection of trees, each of which will cast a vote for a class. In the proposed approach, we employ machine learning techniques to accurately predict the illness that the patient has been experiencing. The outcomes are more precise when historical healthcare records are used as a dataset. We employ machine learning algorithms to train the model and forecast user diseases based on the symptoms they enter.

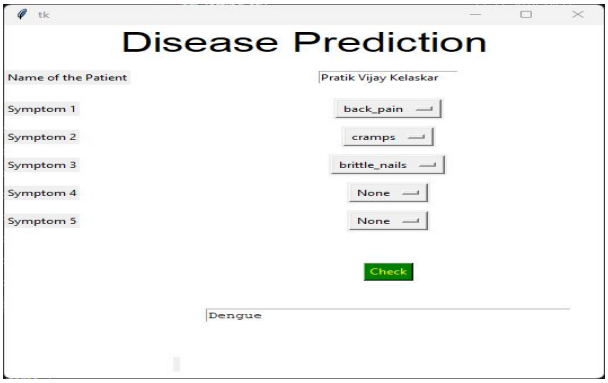
©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |

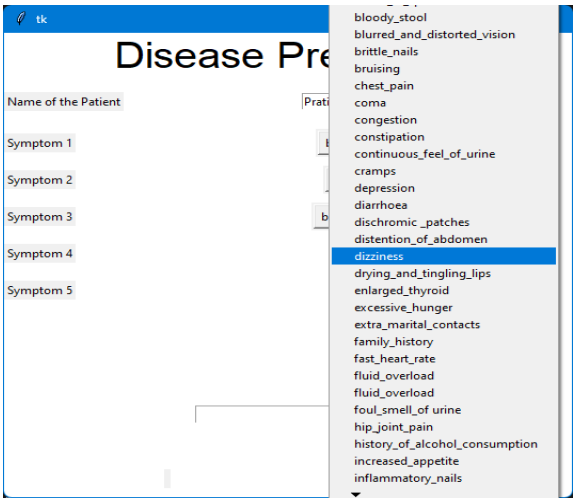
VI. RESULTS

A. *Output*









VII. CONCLUSION

Last but not least, I want to stress how crucial this project—disease prediction using machine learning—is to everyone's daily lives, but notably to those in the healthcare sector, who use these systems frequently to forecast patients' diseases based on their general characteristics and symptoms. The user can learn about the disease they are suffering from by simply entering the symptoms and any other relevant information, and the health industry can benefit from this because the health industry now plays such a large role in treating patients' diseases. This is frequently quite helpful for the health industry to inform the user, and it's also helpful for the user if he or she doesn't want to visit the hospital or other clinics. If the healthcare sector adopts this idea, doctors' workloads will be reduced and they will be better able to predict a patient's illness. A method for predicting the onset of certain common diseases that, if mistreated or ignored, can cause mortality and a host of additional issues for the patient and their family, is known as disease prediction.

VIII. FUTURE SCOPE

In order to create the best infrastructure and quickest, easiest route in the medical sectors, future work will primarily focus on providing medical assistance and appropriate medication to patients as quickly as feasible.

1) Other features could be added, like one where the model recommends a particular group of doctors to see for a given ailment.

2) If appropriate, suggest medications for diseases that pose less risk.

REFERENCES

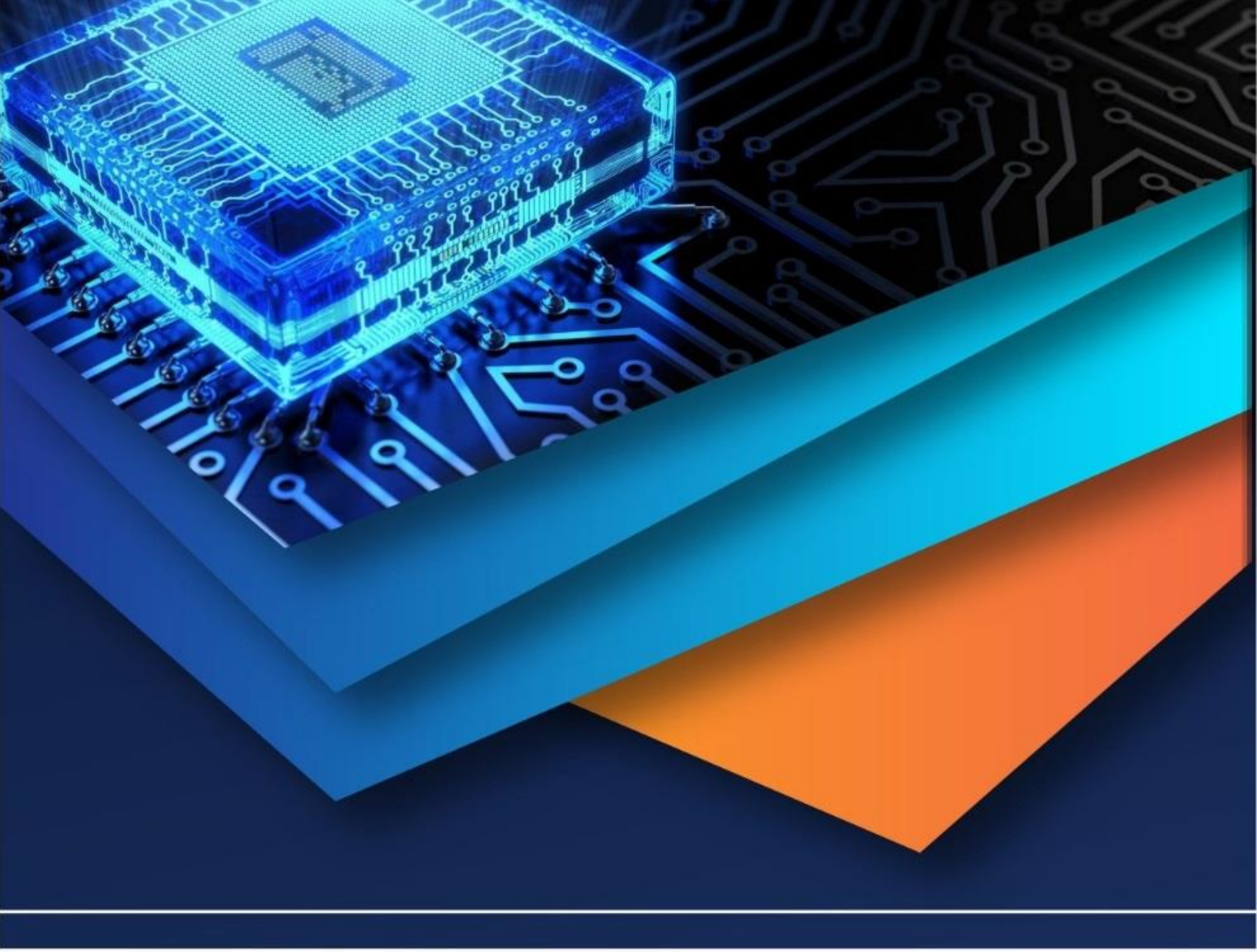
[1] 2020 International Conference for Emerging Technology(INCET) Belgaum,india

[2] M. Chen, Y. Hao, K. Hwang, L. Wang and L. Wang, "Disease Prediction by Machine Over Learning Over Big Data From Healthcare Communities," in IEEE Access, vol. 5, pp. 8869-8879, 2017, doi: 10.1109/ACCESS.2017.2694446.

[3] J. Gao, L. Tian, J. Wang, Y. Chen, B. Song and X. Hu, "Similar Disease Prediction With Heterogeneous Disease Information Networks," in IEEE Transactions on Nano Bioscience, vol. 19, no. 3, pp. 571-578, July 2020, doi: 10.1109/TNB.2020.2994983.

[4] P. S. Kohli and S. Arora, "Application of Machine Learning in Disease Prediction," 2018 4th International Conference on Computing Communication and Automation (ICCCA), Greater Noida, India, 2018, pp. 1-4, doi: 10.1109/CCAA.2018.8777449.

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)