iJRASET
International Journal For Research in Applied Science and Engineering Technology



# INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

**Volume:** 12 **Issue:** IV **Month of publication:** April 2024

**DOI:** https://doi.org/10.22214/ijraset.2024.59817

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 12 Issue IV Apr 2024- Available at www.ijraset.com*

# Machine Learning Based Multiple Disease Prediction using Streamlit

Vikas More[1], Suman Srivastav[2], Shubhangi Gaud[3], Shekhar Suman[4], PulkitSoni[5]
[1]*Professor,* [2, 3, 4, 5]*Student, Dhole Patil College Of Engineering, Wagholi, Pune 412207, India*

***Abstract:*** ***There are multiple techniques in machine learning that can in a variety of industries, do predictive analytics on large amounts of data. Predictive analytics in healthcare is a difficult endeavour, but it can eventually assist practitioners in making timely decisions regarding patients' health and treatment based on massive data. Diseases like Breast cancer, diabetes, and heart- related diseases are causing many deaths globally but most of these deaths are due to the lack of timely check-ups of the diseases. The above problem occurs due to a lack of medical infrastructure and a low ratio of doctors to the population. The statistics clearly show the same, WHO recommended, the ratio of doctors to patients is 1:1000 whereas India's doctor-to- population ratio is 1:1456, this indicates the shortage of doctors. The diseases related to heart, cancer, and diabetes can cause a potential threat to mankind, if not found early. Therefore, early recognition and diagnosis of these diseases can save a lot of lives. This work is all about predicting diseases that are harmful using machine learning classification algorithms. In this work, parkinsons, heart, and diabetes are included. To make this work seamless and usable by the mass public, our team made a medical test web application that makes predictions about various diseases using the concept of machine learning. In this work, our aim to develop a disease-predicting web app that uses the concept of machine learning-based predictions about various diseases like Parkinson, Diabetes, and Heart disease***
***Keywords: Logistic Regression, Support Vector Machine (SVM), data analysis,***

## I. INTRODUCTION

In recent years, the intersection of machine learning and healthcare has presented unprecedented opportunities for improving patient care and outcomes. Early disease detection plays a pivotal role in effective healthcare management, allowing for timely interventions and personalized treatment plans. This project endeavors to contribute to this paradigm shift by developing a machine learning-based system for predicting multiple diseases, with a focus on creating a user-friendly interface using the Streamlit framework.

The advent of machine learning in healthcare has brought forth innovative approaches to disease prediction and diagnosis. Traditionally, the identification of potential health risks has relied heavily on retrospective analysis and clinical expertise. However, with the abundance of health data and advancements in machine learning algorithms, it is now possible to harness the power of predictive analytics for early detection, leading to more proactive and tailored healthcare strategies.

The significance of this project lies in its potential to revolutionize the way diseases are predicted and managed. By combining the analytical power of machine learning with the user-friendly interface of Streamlit, we aim to bridge the gap between complex predictive models and practical, real-world applications.

This project's outcomes have far-reaching implications, from empowering individuals to take charge of their health to assisting healthcare professionals in making more informed decisions.

The motivation behind this project lies in the pressing need for accessible tools that empower both healthcare professionals and individuals to make informed decisions regarding their health. Conventional disease prediction models often lack user-friendly interfaces, hindering widespread adoption. The integration of Streamlit, a Python library designed for creating interactive web applications, addresses this gap, providing an intuitive platform for users to input their health data and receive predictions for multiple diseases

The main contributions of this project are-

*1)* Develop a robust machine learning model capable of predicting a variety of diseases.
*2)* Integrate the model into a user-friendly web application using Streamlit to enhance accessibility.
*3)* Enable users, including healthcare professionals and individuals, to input relevant health data effortlessly.
*4)* Provide accurate and interpretable predictions, fostering early disease detection and personalized healthcare.

©IJRASET: All Rights are Reserved 

## II. METHODOLOGIES

### A. Data Collection

Identify and collect diverse medical datasets encompassing features such as demographics, vital signs, medical history, and diagnostic test results.
Ensure datasets cover a broad spectrum of diseases for a comprehensive predictive model.

### B. Preprocessing

Perform data cleaning to handle missing values, outliers, and ensure data consistency.
Explore feature engineering techniques to enhance the model's predictive capabilities.

### C. Model Development

Select suitable machine learning algorithms based on the nature of the data and the prediction task (e.g., logistic regression, random forests, or neural networks),in this
Project Logistic Regression used for heart disease prediction and SVM used for Parkinsons and diabetes prediction.
Split the dataset into training and validation sets for model training and evaluation.

### D. Streamlit Application

Develop a Streamlit web application with an intuitive user interface.
Integrate the trained machine learning model into the application's backend for disease predictions.Implement user authentication and secure data handling for privacy.

### E. Model Evaluation

Assess the model's performance using metrics such as accuracy, precision, recall, F1 score, and ROC-AUC, here in this accuracy used ,got model accuracy as follows
Parkinsons- 87% using SVM, for diabetes- 77% using SVM, for heart disease-81% using Logistic Regression .

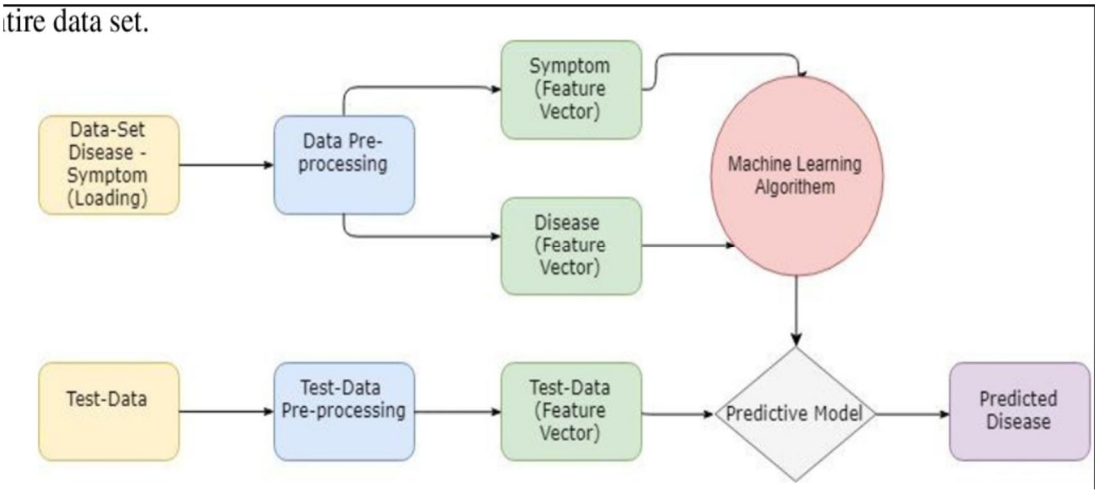


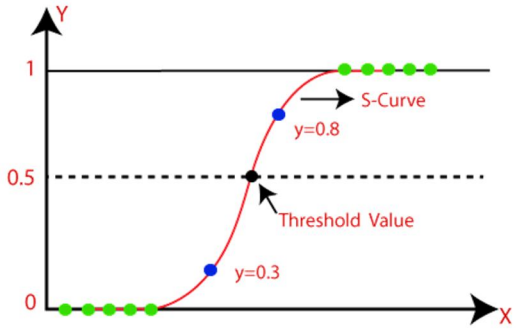

Fig. Process Flow

### F. Logistic Regression

Logistic regression is one of the most popular Machine Learning algorithms, which comes under the Supervised Learning technique. It is used for predicting the categorical dependent variable using a given set of independent variables.Logistic regression predicts the output of a categorical dependent variable. Therefore the outcome must be a categorical or discrete value. It can be either Yes or No, 0 or 1, true or False, etc. but instead of giving the exact value as 0 and 1, The sigmoid function is a mathematical function used to map the predicted values to probabilities.It maps any real value into another value within a range of 0 and 1.The value of the logistic regression must be between 0 and 1, which cannot go beyond this limit, so it forms a curve like the "S" form. The S-form curve is called the Sigmoid function or the logistic function.

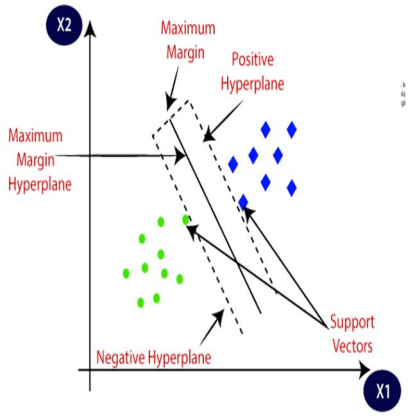
©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |

*ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538*
*Volume 12 Issue IV Apr 2024- Available at www.ijraset.com*




*G. SVM(Support Vector Machine)*

Support Vector Machine or SVM is one of the most popular Supervised Learning algorithms, which is used for Classification as well as Regression problems. However, primarily, it is used for Classification problems in Machine Learning.The goal of the SVM algorithm is to create the best line or decision boundary that can segregate n-dimensional space into classes so that we can easily put the new data point in the correct category in the future. This best decision boundary is called a hyperplane.SVM chooses the extreme points/vectors that help in creating the hyperplane. These extreme cases are called as support vectors, and hence algorithm is termed as Support Vector Machine




## III. LITERATURE SURVEY

This section describes the study of previously proposed models for predicting the diseases which are related to our proposed work. Several studies have been made for detecting various diseases. They have applied various data mining techniques for efficiently predicting a variety of diseases.[1] Akkem Yaganteeswarudu conducted comparative study on the effectiveness of Decision Tree, Random Forest and logistic regression algorithms in predicting multi Disease which resulted in logistic regression results 92% accuracy, for heart disease classification Randomforest yield 95% accuracy and for cancer detection SVM yield 96 % accuracy.

[2]Pahulpreet Singh Kohli et al, suggested disease prediction by using applications and methods of machine learning and used techniques like Logistic Regression, Decision Tree,Support Vector Machine, Random Forest and Adaptive Boosting. This paper focuses on predicting Heart disease, Breast cancer, and Diabetes. The highest accuracies are obtained using Logistic Regression that is 95.71% for Breast cancer, 84.42% for Diabetes, and 87.12%for Heart disease.

[3]G Naveen Kishore and few other authors proposed the work named Prediction Of Diabetes Using Machine Learning Classification Techniques proposed. In this work, various classification algorithms like SVM, Logistic Regression, Decision Tree, KNN, Random Forest are utilized on the 769 instances of the Pima dataset which contain features like Pregnancies, Blood pressure, body mass index, etc. They have reported the highest accuracy as 74.4 %for the classification algorithm Random Forest and the lowest accuracy in this work is attained by the KNN reported as 71.3%.

[4]The work “Understanding the lifestyle of people to identify the reasons for Diabetes using data mining” proposed by Gavin Pinto, Radhika Desai, and Sunil Jangid discussed reducing the risk of diabetes disease using data mining techniques and also discussed diabetes sub-classification. The authors used Naïve Bayes and SVM classification algorithms on the dataset collected by a survey using google forms and reported the accuracy of 64.92 for SVM and 60.44 for Naïve Bayes.

©IJRASET: All Rights are Reserved 

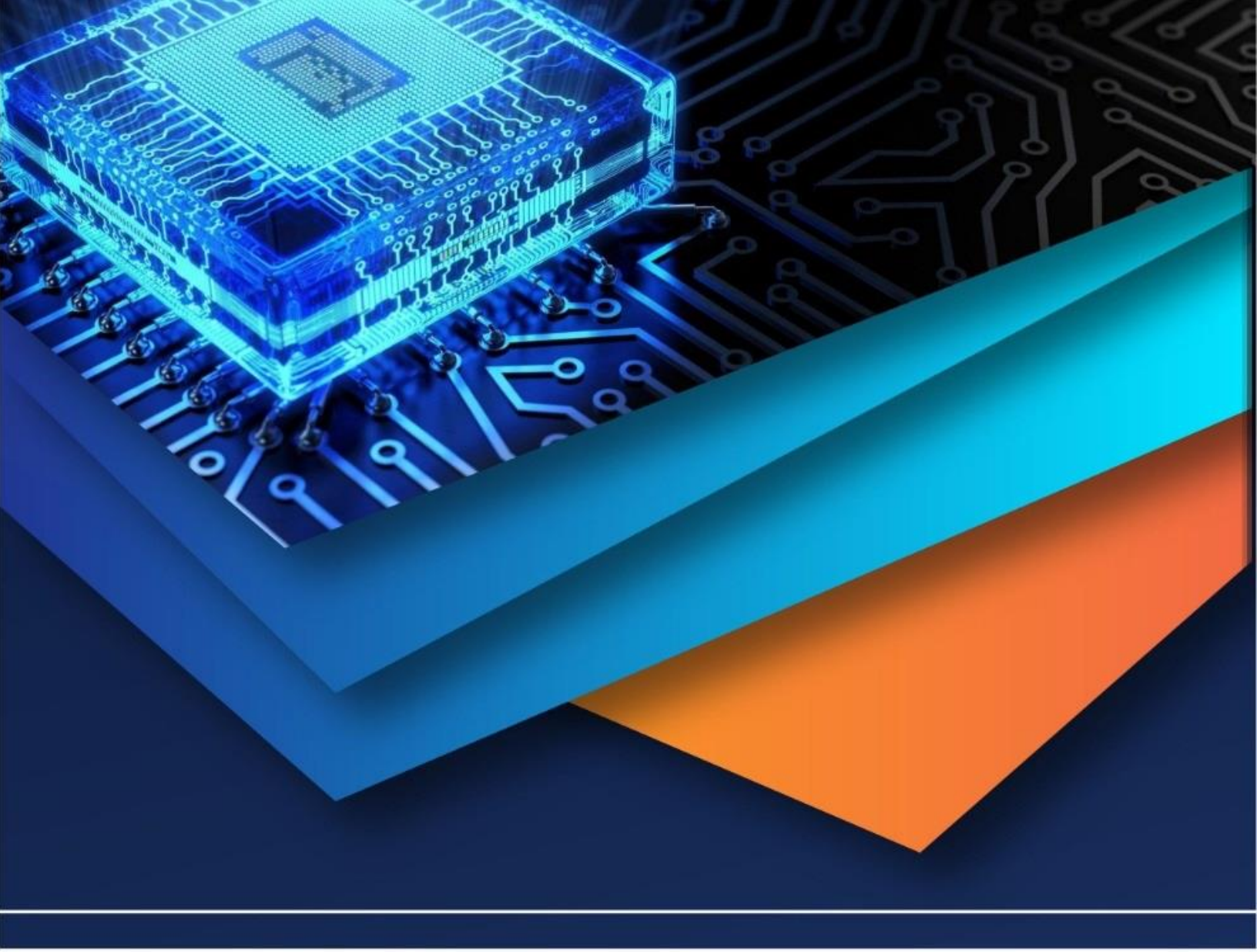
[5]In the work presented by M.Marimuthu, S.DeivaRani, Gayatri. R described the cardio diseases in a detailed manner and also applied the classification algorithms like SVM, Decision Tree, Naïve Bayes, K-Nearest Neighbors on the Framingham dataset from Kaggle. The authors compared various machine learning algorithms for the forecast of the risk of heart disease. The highest reported accuracy in this work is 83.60% for the KNN classification algorithm.
[6]Amandeep Kaur and Jyothi Arora presented a study that covered the examination of algorithms such as KNN, SVM, ANN, and Decision Tree on the heart disease dataset and plotted the accuracies graph.

## IV. CONCLUSION

This project combines machine learning and web development to create a valuable tool for disease prediction. Early detection of health issues can lead to timely interventions, potentially improving patient outcomes. The integration of Streamlit ensures a user-friendly interface, making the tool accessible to both healthcare professionals and individuals concerned about their health.

## REFERENCES

[1] Gavin Pinto, Sunil Jangid, Radhika Desai, "Understanding the Lifestyle of people to identify the reasons ofDiabetes using data mining".
[2] M.Marimuthu ,S.Deivarani ,R.Gayatri, "Analysis of Heart Disease Prediction using Machine Learning Techniques".
[3] Purushottam, Richa Sharma ,Dr. Kanak Saxena, "Efficient Heart Disease Prediction System".
[4] Adil Hussain She, Dr. Pawan Kumar Chaurasia," A Review on Heart Disease Prediction using Machine Learning Techniques".
[5] M. Chinna Rao ,K. Ramesh, G. Subbalakshmi,"Decision Support in Heart Disease Prediction System using Naïve Bayes".
[6] Amandeep Kaur , Jyothi Arora," Heart Disease Prediction using data mining Techniques :A survey".
[7] Noreen Fatima , Li Liu , Sha Hong, Haroon Ahmed ,"Prediction of Breast Cancer, Comparitive Review Of Machine Learning Algorithms and their analysis".
[8] Ch .Shravya ,K.Pravallika , Shaik Subhani, "Prediction of Cancer using supervised machine learning Algorithms".
[9] Nikita Rane, Jean Sunny, Rucha Kanade, Sulochana Devi," Breast Cancer classification and prediction usingmachine learning ".
[10] Deepti Sisodia, Dilip Singh Sisodia," Prediction of Diabetes using classification Techniques".
[11] Dr.B.Santhosh Kumar, T.Daniya, Dr. J.Ajayan," Breast Cancer Prediction using Machine Learning Algorithms".
[12] Mümine KAYA KELEŞ ,"Cancer Prediction using and Detection using Machine Learning Algorithms : AComparitive Study".
[13] Heart Disease Dataset" by UCI.
[14] Pima Indians Diabetes Dataset" by Kaggle.



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)