Presidency University Bangalore

Disease Prediction based on symptoms using machine learning algorithms

[1]T.G. Pranav Sri Vasthav, [2]Tejaswini G J, [3]Vada Swetha, [4]Vanishree R and

[5]Mr. Raghavendra T.S

[1, 2, 3, 4] Students, School of Computer Science and Engineering, Presidency University, Bangalore, India

[5] Assistant Professor, School of Computer Science and Engineering, Presidency University, Bangalore, India

Abstract---

Disease prediction based on symptoms is an essential task in healthcare that can lead to timely treatment and better patient outcomes. Machine learning algorithms have proven to be effective in predicting diseases based on symptoms. In this research paper, we explore various machine learning algorithms that can be used for disease prediction based on symptoms. We use a dataset of patients with various symptoms and their corresponding diagnosis to train and evaluate the performance of the algorithms. We also discuss the potential challenges and limitations of using machine learning for disease prediction and the future directions of this research.

I. INTRODUCTION

Disease prediction based on symptoms is a crucial task in healthcare that can help in early detection and timely treatment. The traditional approach to disease prediction involves medical experts analysing patient symptoms and making a diagnosis based on their experience and knowledge. However, this approach has limitations, such as the potential for human error and the inability to handle large amounts of data. In recent years, machine learning algorithms have emerged as a promising solution to disease prediction based on symptoms.

II. RESEARCH OBJECTIVE

We are concentrating on providing immediate and accurate disease prediction to the users about the symptoms they choose using different machine learning algorithms such as Support Vector Machine, Random Forest Classifier, Naïve Bayes and Decision Tree Classifier and then deploying the model into the website to be more user-friendly for the users. Whole work includes Registration, OTP login, uploading files, pre-processing the data, Splitting the data, choosing the symptoms then predicting the disease.

III. PROPOSED SYSTEM

The main dairy of your project is to build the user interface for prediction of disease with sheer accuracy. Building the website both for the patient and the doctor utility services. The steps we are taking to achieve this work is Test and train with different algorithm, screening out and sticking with the best algorithm, User interface. In this we have four different Machine Learning algorithms such as Gaussian Naive Bayes, Random Tree Classifier, Decision Tree Classifier, Support Vector Machine.

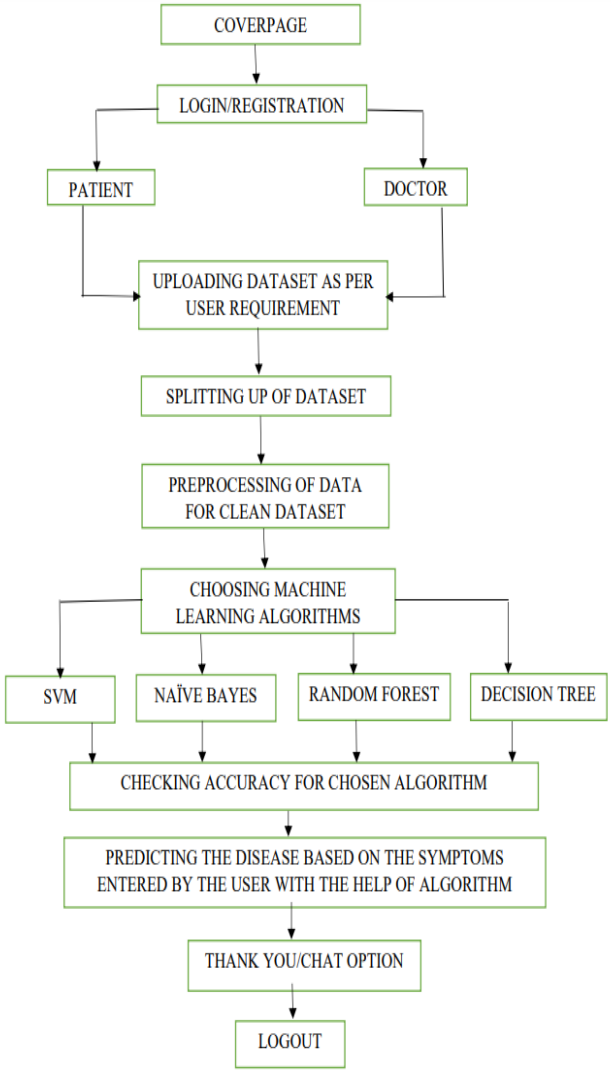

Fig 1: System Model

IV. IMPLEMENTATION

A. Division

This project is modularized as the following:

- **Patient space:** This space consists of login, registration which is connected through SQL and data will be stored in the Database.
- **Doctor Space:** This space consists of login, registration which is connected through SQL and data will be stored in the Database.
- **Load Dataset:** Here we have to upload the datasets.
- **Pre-process:** The two datasets will be joined and will have an option to split the combined dataset.
- **Model Training:** Here we will be having the option to test the accuracy of the given algorithms for the dataset uploaded

B. Training and Testing

Training and testing the data sets and combing dataset for checking greater accuracy and the images include dealing with missing values, checking data types, splitting the data for training and testing the model again, and Model training Model training is done in four algorithms namely Gaussian Naïve Bayes, Random Forest, Decision Tree, and Support Vector Machine.

C. Model Building

Many methods are used to perform data mining. Machine learning is one of the approaches. Naive Bayes Algorithm is fast and efficient, you can use it to make real-time predictions. Random Forest is a classifier that contains a number of decision trees on various subsets of the given dataset and takes the average to improve the predictive accuracy of that dataset. We selected Random Forest algorithm because the result will be displayed by considering the maximum number of outputs i.e., takes the majority vote rule. Next the Decision Tree Algorithm, the main reason for using this algorithm is because of the categorical numeric data (i.e., yes/no or 0/1). The last algorithm we included for testing and training the dataset is Support Vector Machine. We tried this algorithm since it can handle the high dimensionality data.

V. RESULTS AND CONCLUSION

Model	Accuracy
Gaussian Naive Bayes	100
Random Forest	100
Decision Tree Algorithm	99.93
Support Vector Machine	99.79

Table 1: Accuracy achieved using 4 different algorithms.

Fig 2: Home Page

Fig 3: Designation Page

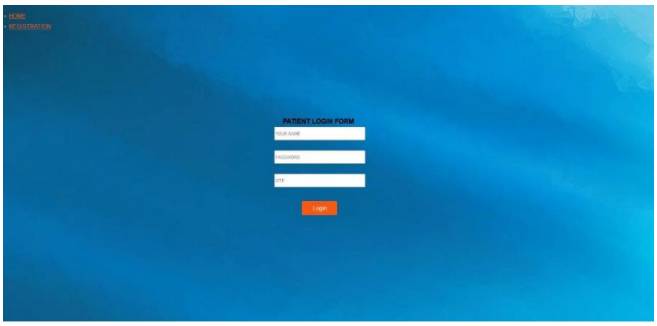

Fig 4: Login Page

Fig 5: Index Page

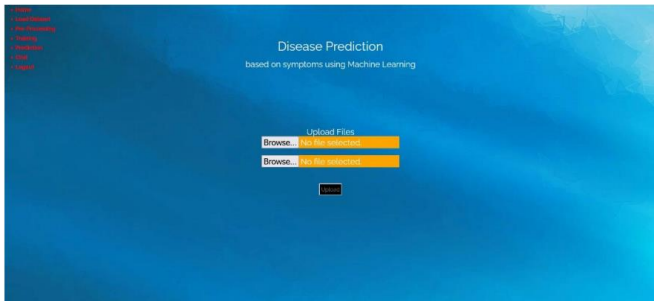

Fig 6: Dataset Upload Action Page

Fig 7: Splitting Dataset Page

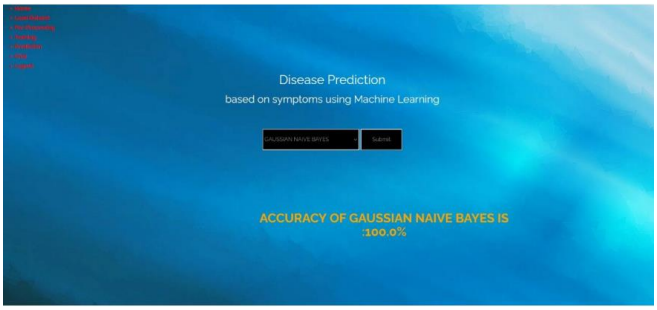

Fig 8: Choosing Algorithm Page

Fig 9: Disease Prediction Page

Fig 10: Contact Us Page

Conclusion

In this research paper, we explored various machine learning algorithms for disease prediction based on symptoms. Our results show that the Gaussian Naive Bayes and Random Forest Classifier algorithms are the most effective algorithms for this task, achieving perfect accuracy. However, there are challenges and limitations to using machine learning for disease prediction, such as the availability and quality of data and the interpretation of the results. Future research can address these challenges.

REFERENCE

[1]. Gaurav Shalimar, Shivam Pisal (2021) 'Disease Prediction Using Machine Learning'.

[2]. Manasvi Srivastava, Vikas Yadav, Swati Singh (2020) 'Implementation of Web Application for Disease Prediction Using AI'.

[3]. MIN CHEN, (Senior Member, IEEE), YIXUE HAO, KAI HWANG, (Life Fellow, IEEE), LU WANG, AND LIN WANG (2017) 'Disease Prediction by Machine Learning Over Big Data from Healthcare Communities'.

[4]. Rudra A. Godse, Smita S. Gunjal, Karan A. Jagtap, Neha S. Mahamuni, Prof. Suchita Wankhade (2019) 'Multiple Disease Prediction Using Different Machine Learning Algorithms Comparatively'.

[5]SVMalgorithmreferences https://towardsdatascience.com/support vector-machine-python-example-d67d9b63f1c8

[6]. Varinder Garg, Harish Kumar, Surinder Rana, Bikramjeet Singh Kalsi, Siddhant Mukherjee (2020) 'Machine learning based disease prediction website using symptoms of a patient'.