A Machine Learning Model for Disease Prediction System

Aditya Singh[1], Ankur Kumar Tiwari[2], Ankur Pandey[3], ChandanNishad[4], Ajay Kumar[5]

[1,2,3,4] Student, Department of Computer Science and Engineering, ITM Gida Gorakhpur, India.
[5] Assistant Professor, Department of Computer Science and Engineering, ITM Gida Gorakhpur, India.

Abstract: *The development of modern technologies such as data science and machine learning has paved the way for healthcare communities and medical institutions to detect diseases as early as possible and contribute to better patient care. The accuracy of detecting possible diseases will be reduced if we do not have complete medical data. Also, certain diseases are based on regions, which could lead to poor disease prediction. Our body shows the symptoms when something bad happens in our body, sometimes it can be a small problem but sometimes we can have a serious illness and if we don't take care of these symptoms early it can be too late for the to cure the disease. Therefore, we propose a disease prediction system that can predict possible diseases based on symptoms so that you can be cured at an early stage. It saves the time required to make the full diagnosis of the patient we can diagnose the patient for only the required diseases. In this work, we use machine learning algorithms that try to accurately predict possible diseases. The results generated by the proposed system have an accuracy of up to 87%. The system has incredible potential to anticipate possible diseases even more precisely. The primary purpose of this Study is to help non-technical staff and first-year physicians form proper opinions about disease.*
Key Word: Disease Prediction System; Machine Learning; Multilinear Regression (MLR); Support Vector Machine (SVM)
.

I. INTRODUCTION

It is a system that runs using machine learning algorithms to guess possible diseases based on the patient's symptoms [1]. The growth of technology has improved our life until now. It offers many tools that can save millions of lives, and machine learning is one of them. Machine learning is being used to develop systems that can help us predict so many diseases based on symptoms. It can suggest doctors the likelihood of possible diseases.

And diagnosis can be made based on suggestions, which could reduce costs. We live in the age of technology and nowadays people can say that almost anything is possible with the help of technology. Today we have so many tools and methods to access information from all regions of this world and information is so important in this age that we would not survive without it. We have tools that can give us or suggest relevant information within our reach, and the internet is one of those tools. Today billions of searches are made every day and sometimes the results provided are relevant and sometimes not.

In these searches, thousands of searches are related to medical advice. People often want to know if they have a serious illness based on their signs and symptoms. However, no tools are available to give them the right information. This research seeks to give them tools to provide the end user with information about possible disease predictions.

II. LITERATURE SURVEY

Numerous disease prediction studies have been conducted using various machine learning techniques and algorithms that can be used by medical institutions. This paper gives an overview of some of these studies conducted in research papers using the techniques and results they used.

MIN CHEN et al. [1] proposed in their article a disease prediction system in which they used automatic learning algorithms. In disease prediction, he used techniques such as the CNN UDRP algorithm, the CNN MDRP algorithm, Naive Bayes, K-Nearest Neighbor, and Decision Tree. This proposed system had an accuracy of 94.8%.

SayaliAmbekar et al. [2] recommended Disease Risk Prediction and used a convolution neural network to perform the task. This article uses machine learning techniques such as the CNN UDRP algorithm, Naive Bayes, and KNN algorithm. The system uses structured data to be trained and its accuracy reaches 82% and is achieved by using Naive Bayes.

Naganna Chetty et al. [3] have developed a system that provides better results for disease prediction using a fuzzy approach. And techniques like KNN classifier, fuzzy media clustering and KNN fuzzy classifier were used. This article makes a prediction of diabetes and liver disease, and the accuracy of diabetes is 97.02% and liver disease is 96.13.

Dhiraj Dahiwade et al. [4] developed a model for predicting diseases using machine learning approaches and used techniques such as ANN and CNN. This document proposes the prediction of the disease, i.e. based on the patient's symptoms. The accuracy of ANN is 95% and the accuracy of CNN is 98%.

Lambodar Jena et al. [5] focused on predicting chronic disease risk by leveraging distributed machine learning classifiers and used techniques such as Naïve Bayes and Multilayer Perceptron. This article attempts to predict chronic kidney disease and the accuracy of Naive Bayes and Multilayer Perceptron is 95% and 99.7%, respectively.

Dhomse Kanchan B et al. [6] studied the prediction of special diseases by principal component analysis using machine learning algorithms with techniques such as naive Bayes classification, decision tree and support vector machine. The accuracy of this system is 34.89% for diabetes and 53% for heart disease.

Pahulpreet Singh Kohli et al. [7] proposed disease prediction using machine learning applications and methods and techniques such as logistic regression, decision tree, support vector machine, random forest, and adaptive boost. This white paper focuses on predicting heart disease, breast cancer and diabetes. The highest accuracies are obtained using logistic regression, which is 95.71% for breast cancer, 84.42% for diabetes, and 87.12% for heart disease.

Deeraj Shetty et al. [8] investigated the use of data mining for diabetes disease prediction using naive Bayes and ANN algorithms. This system predicts diabetes and the accuracy achieved by ANN is better than Naive Bayes.

Rashmi G. Saboji et al. [9] tried to find a scalable solution that could predict heart disease using classification mining and used the random forest algorithm. This system compares to the Naive Bayes classifier, but RandomForest returns more than accurate results with 98% accuracy.

Rati Shukla et al. [10] proposed breast cancer prediction and detection using machine learning techniques such as Decision Tree, Support Vector Machine, Random Forest, Naive Bayes, Neural Network and ANN. In this system, the support vector machine gives more accurate results than any other algorithm.

III.PROPOSED METHOD

Our proposed methodology includes the following steps:

i. First I will collect the records of symptoms and their functional problems in the body.

ii. Then I will collect the information that connects the symptoms to possible diseases, so information about the disease is collected.

iii.I then receive the symptoms as input from the patient and process them using multilinear regression.

iv. Then multilinear regression predicts the diseases that are possible for those acquired symptoms.

v. The system then displays the diagnosis in the form of Highest Possible Disease and Lowest Possible Disease.

The flowchart of the methodology is shown below:

Fig1: Flowchart of proposed method

Algorithm Used:

As the name suggests, we use in our disease prediction system Support Vector Machine(SVM) for classification and multilinear regression(MLR) for predicting the result.MLR is a form of regression algorithm which involves multiple independent values, which what we tried at a value to predict is based on two or more variables.

Simple Linear Regression, where a unique Independent/Predictor(X) Variable is used to model the response variable (Y). But there can be many situations in where the response variable is influenced by multiple predictor variables for like in cases we use the MLR algorithm.

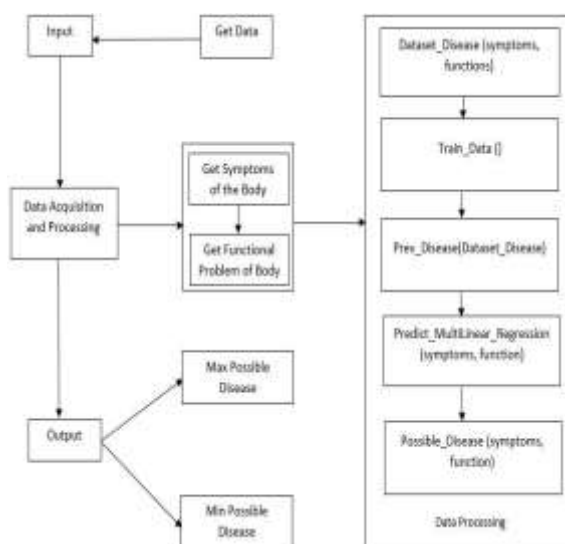**The Architecture of Disease Prediction System:**

Fig2: The architecture of Disease Prediction System

IV.CONCLUSION

In our research, we used a support vector machine and a multilinear regression algorithm to predict disease. And we also tested different algorithms like k-Nearest Neighbor, Convolution Neural Network, Decision Tree, etc. Despite testing these algorithms, I have found that the combination of support vector machine and multilinear regression gives higher accuracy than other algorithms. In this research, we found that the potential disease prediction can be as high as 87% for some diseases and as high as 68% for some diseases, but these results are obtained with the minimum number of datasets, but if we can feed the system with one huge amount of data sets, this disease prediction system can provide up to 95% accuracy.

Collecting a large number of datasets on diseases and their symptoms is time consuming and cannot be done in a year or two. It takes several years to collect these datasets and to train the system with these data searches. Doctoral students can use this system for further research work. With the help of a disease prediction system, it was possible to diagnose people based on symptoms.

The disease prediction system only provides possible results, it does not guarantee that it will correctly predict the disease. But it is far more accurate in predicting possible diseases. In our research we tested the accuracy of this system for 5 different diseases and our accuracy can be up to 87%.

References

[1]. M. Chen, Y. Hao, K. Hwang, L. Wang, and L. Wang, "Disease prediction by machine learning over big data from healthcare communities" IEEE Access, vol. 5, no. 1, pp. 8869–8879, 2017.

[2]. SayaliAmbekar, Rashmi Phalnikar, "Disease Risk Prediction by Using Convolutional Neural Network" IEEE, 978-1-5386-5257-2/18, 2018.

[3]. Naganna Chetty, Kunwar Singh Vaisla and Nagamma Patil, "An Improved Method for Disease Prediction using Fuzzy Approach" IEEE, DOI 10.1109/ICACCE.2015.67, pp. 569-572, 2015.

[4]. Dhiraj Dahiwade, Gajanan Patle and EktaaMeshram, "Designing Disease Prediction Model Using Machine Learning Approach" IEEE Xplore Part Number: CFP19K25-ART; ISBN: 978-1-5386-7808-4, pp. 1211-1215, 2019.

[5]. Lambodar Jena and Ramakrushna Swain, "Chronic Disease Risk Prediction using Distributed Machine Learning Classifiers" IEEE, 978-1-5386-2924-6/17, pp. 170-173, 2017.

[6]. Dhomse Kanchan B. and Mahale Kishor M., "Study of Machine Learning Algorithms for Special Disease Prediction using Principal of Component Analysis" IEEE, 978-1-5090-0467-6/16, pp. 5-10, 2016.

[7]. Pahulpreet Singh Kohli and Shriya Arora, "Application of Machine Learning in Disease Prediction" IEEE, 978-1-5386-6947-1/18, pp. 1-4, 2018.

[8]. Deeraj Shetty, Kishor Rit, Sohail Shaikh and Nikita Patil, " Diabetes Disease Prediction Using Data Mining" IEEE, 978-1-5090-3294-5/17, 2017.

[9]. Rashmi G Saboji and Prem Kumar Ramesh,"A Scalable Solution for Heart Disease Prediction using Classification Mining Technique" IEEE, 978-1-5386-1887-5/17, pp. 1780-1785, 2017.

[10]. Rati Shukla, Vikash Yadav, Parashu Ram Pal and Pankaj Pathak, "Machine Learning Techniques for Detecting and Predicting Breast Cancer" IJITEE, ISSN: 2278-3075, Volume-8, pp. 2658-2662, 2019.

[11]. Senthilkumar Mohan, ChandrasegarThirumalai and Gautam Srivastava, "Effective Heart Disease Prediction Using Hybrid Machine Learning Techniques" IEEE Access, DOI 10.1109/ACCESS.2019.2923707, pp. 81542-81554, 2019.