Volume 9, Issue 4, April – 2024 International Journal of Innovative Science and Research Technology
ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/IJISRT24APR1453

Multiple Disease Prediction Using Machine Learning

Vrutika Bagul[1]; Vrushali Bagul[2]; Sadichha Patil[3]; Swati Bhoir[4]
[1]Computer Engineering AMRIT, University of Mumbai, India.
[2,3,4]Computer Engineering AMRIT, University of Mumbai, India

Abstract:- Machine learning, which is a type of computer technology, has changed healthcare a lot. It helps doctors predict diseases better and faster. In healthcare, using machine learning algorithms decision tree (DT), logistic regression (LR), support vector machine (SVM) that can help predict lots of different diseases at the same time. This helps doctors find and treat illnesses early, which makes patients better and saves money on healthcare. This paper looks at how we can use computer programs that learn from data to predict many diseases. It talks about why this is good, what problems we might face, and where we might go next with it. We give a summary of the several machine learning models and information sources that are often employed in illness prediction. We also go over the significance of feature selection, model assessment, and combining several data modalities for improved illness prediction. We give a summary of the several machine learning models and information sources that are often employed in illness prediction. We also go over the significance of feature selection, model assessment, and combining several data modalities for improved illness prediction. The research shows that using machine learning algorithms to predict many diseases at once could really help public health. Again, we use a machine learning model to determine whether or not an individual is impacted by a few diseases. This training model trains itself to predict illness using sample data.

Keywords:- *Disease Prediction, Disease Data, Machine Learning, Decision Tree (DT), Logistic Regression (LR), Support Vector Machine (SVM).*

I. INTRODUCTION

In recent years, machine learning has made big progress and is being used in lots of industries, like healthcare, to do really amazing things. Using computer systems that learn from data can help doctors detect diseases more accurately and improve patient outcomes by predicting many diseases at the same time. This study used the Support Vector Machines (SVM) and logistic regression (LR) algorithms to predict the presence of five prevalent diseases: Parkinson's disease, diabetes, heart disease, lung cancer, and breast cancer. Diabetes, Parkinson's disease, lung cancer, and breast cancer are important public health concerns that have a significant influence on people's lives and healthcare systems all over the world. Reducing healthcare expenditures, optimizing treatment strategies, and improving patient prognosis are all dependent on early identification and correct diagnosis of these disorders. Early detection and precise diagnosis of these illnesses are critical for lowering healthcare costs, optimizing treatment techniques, and improving patient outcomes. Because of its ability to study huge amounts of data and identify subtle patterns, machine learning offers fascinating pathways for multi-disease prediction. Support vector machines (SVMs) are powerful supervised learning models that are commonly used in classification problems. SVMs aim to optimize the margin between unique classes in data by determining the best hyperplane to separate them. The SVM approach is suitable for a wide range of medical diagnostic applications since it can handle both linear and nonlinear relationships between input data and target variables. This study wanted to make a system that could predict several diseases using SVMs. It checked how good this system was at predicting Parkinson's disease, diabetes, and heart disease. . Using this dataset, the SVM model was trained to understand the complex correlations between the existence of the three diseases and the input features. Targeted illness management techniques, individualized treatment plans, and early interventions can all be made easier with the help of machine learning models for accurate disease prediction. It may help medical professionals make better judgments, improve patient care, and better allocate resources within healthcare systems. It also has potential for population-level disease surveillance, which would help public health officials quickly identify illness outbreaks and put preventative measures in place. The investigation and analysis of the SVM model's performance in predicting heart disease, diabetes, and Parkinson's disease revealed the utility and practicality of applying machine learning algorithms to complex medical diagnosis. The study and evaluation of the LR model's performance in predicting lung cancer and breast cancer . As a consequence, this work points out the potential of SVM and LR as effective tools in the field of multi-disease prediction. Machine learning can help us move closer to producing more precise, timely, and tailored healthcare interventions, which will improve patient outcomes and build more successful healthcare systems.

II. LITERATURE SURVEY

In this project, we studied existing research about using machine learning methods, like Support Vector Machines (SVM) and logistic regression (LR), random forest , to predict several diseases such as diabetes, heart disease, and Parkinson's disease. We looked at other studies that did similar things to understand more about how they did it and what they found. This helped us set up our own project.

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/IJISRT24APR1453

- According to the journal, diabetes is one of the world's most dangerous illnesses, affecting a variety of afflictions including blindness. In this article, scientists used machine learning algorithms to identify diabetes illness since they are straightforward and versatile in predicting whether a patient is suffering or not. The purpose of this study was to develop a system that would help people properly diagnose their diabetes. They compared the accuracy of the key algorithms (LR, Random Forest, DT, and SVM): 72%, 74%, 72.91%, and 73%.
- The main purpose of this study is to show how important the heart is to living people. As a result, detection of heart disease and forecasting must be precise and accurate since they are crucial and can result in cardiac-related passing away. As a result, machine learning may assist in the prediction of any natural disaster. In this paper, they examine the accuracy of machine learning for predicting heart disease using k-nearest neighbor, decision tree, and naïve bayes with training and testing datasets. The authors additionally analyzed the techniques and their accuracy: DT 52%, KNN 45%, and Naïve Bayes 52.33%.
- Parkinson's disease is a common condition that affects the nervous system and nerve-controlled organs in the body. It gets worse over time. The SVM model correctly identified if someone had Parkinson's disease or not in about 71% of cases.

III. METHODOLOGY

The technique for the Multiple Disease Prediction project is stated as follows:

- Data Collection : We got the data from Kaggle.com, a popular website for getting datasets. The data was collected specifically for diabetes, heart disease, lung cancer, Parkinson's disease, and breast cancer.
- Data Preprocessing :We examine and correct the data to ensure that it is of excellent quality and suitable for training our machine-learning models. This includes addressing missing information, deleting duplicates, and altering the data for the purpose of making it more convenient to utilize.
- Model Selection : We choose various machine learning approaches for each sickness prediction challenge depending on their effectiveness and suitability for that specific prediction. For example, we can apply Support Vector Machine (SVM) or Logistic Regression, depending on which is more suited for the job..
- Training and Testing: We divided the prepared data into two sets: training and testing. We utilize the training data to educate the models and then test them to see how well they perform. We assess their performance using accuracy to determine how excellent each model is.
- Model Deployment: We utilize Streamlit cloud deployment tools to build a web app that users can interact with. The program has a simple UI and allows users to anticipate five diseases: heart disease, lung cancer, diabetes, Parkinson's disease, and breast cancer. When an illness is picked, the user is prompted to enter the information required for the prediction.

IV. PROBLEM SYSTEM

Currently, many machine learning models in healthcare are focused on assessing a single ailment at once. For example, one model may concentrate on evaluating liver problems, another on cancer, and a third on lung disorders. If someone wishes to anticipate more than one ailment, they must utilize many websites or techniques. There is no standard method for analyzing numerous illnesses using one system. Some of these models are inaccurate, which might be detrimental to patients. When an organization wants to review its patients' health information, they must utilize several models, which takes time and money. Also, some systems only consider a few factors, which might lead to incorrect results.

V. EXISTING SYSTEM

Disease prediction using machine learning algorithms, existing system project aims to predict diabetes, heart disease, and Parkinson's disease using different machine learning methods. This project uses several different types of machine learning algorithms to make predictions. These include Naive Bayes, Decision Trees, Random Forest, Support Vector Machine (SVM), and Logistic Regression. The system gathers data from many sources, prepares it, and trains models using the generated information. and analyzes how well they perform. One of the algorithms employed by the system is SVM, which achieved a 73% accuracy rate in predicting diabetes. Similarly, the SVM algorithm diagnosed Parkinson's disease with 71% accuracy. This demonstrates that the SVM model accurately identified the existence or nonexistence of Parkinson's disease in 71% of instances. SVM algorithm is best algorithm to predict Parkinson's Disease in early stage. By applying the decision tree (DT) method, the system predicted heart disease with an accuracy of 52%. The system also has other machine learning methods like Naive Bayes, Decision Trees, and Random Forest. These might work better or worse depending on the illness being predicted. Overall, the existing system uses machine learning algorithms to predict diabetes, heart disease, and Parkinson's disease. We can make the models better by improving them more, which would help us predict illnesses more accurately.

VI. PROPOSED SYSTEM

In the current system, we don't use Decision Trees (DT) and Naive Bayes for implementing the models. However, in the proposed system, we plan to add two more diseases and develop models for them using Support Vector Machine (SVM) and Logistic Regression (LR). We use new methods, including data normalization to assure data consistency, label encryption to transform text data into numbers, and decreasing dimensionality to minimize features while preserving crucial information. We apply techniques that fit the dataset well and select simple models to enhance performance. The proposed system utilizes the Streamlit library, Streamlit Cloud, Python notebooks, and Python. It's a detailed project for predicting diseases, employing machine learning algorithms like Support Vector Machine (SVM), Logistic Regression, Random Forest, and K-Nearest

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/IJISRT24APR1453

Neighbors (KNN) to predict diabetes, heart disease, lung cancer, Parkinson's disease, and breast cancer. The system wants to generate reliable forecasts for diseases by taking into account input criteria and displaying them in an easy-to-use user interface designed using Streamlit and hosted on Streamlit Cloud In order to guarantee the quality of the data we preprocess and use for training the models, we get it from Kaggle. The processed data is then used to train machine learning algorithms tailored for each illness. These trained models are tested to see how accurately they can predict illnesses. the system uses the SVM algorithm to predict diabetes with a 77% accuracy rate. This means the SVM model is likely to correctly identify whether someone has diabetes or not, helping with early detection and treatment. It's impressive that the SVM algorithm is achieving such high accuracies in predicting both Parkinson's disease and heart disease. An accuracy of 84% for predicting Parkinson's disease and 81% for heart disease suggests that the SVM model is effective at distinguishing between individuals with these conditions and those without them. These results indicate the potential utility of machine learning techniques in healthcare for assisting in disease diagnosis and management. However, it's essential to remember that while high accuracy is valuable, other metrics such as sensitivity, specificity, and positive predictive value are also crucial for evaluating the performance of a predictive model, especially in medical contexts. This way of doing things helps figure out if someone might get heart disease. It means doctors can step in quickly and give the right treatment. The technique uses LR to predict lung cancer illness, with an incredible accuracy of 94%. The method also contains breast cancer prediction, which applies LR and achieves 91% accuracy. With the use of these technologies, a machine learning model may be developed that can precisely predict the existence of breast cancer, facilitating early identification and treatment. The five sickness options available on the proposed system's menu are diabetes, Parkinson's disease, breast cancer, lung cancer, and heart disease. The interface is easy to use. When an disease is chosen, the system invites the user to provide the necessary parameters for prediction. Once you've entered the details, the system makes a prediction and shows you the result. This helps you make smart choices about your health and stay on top of things.

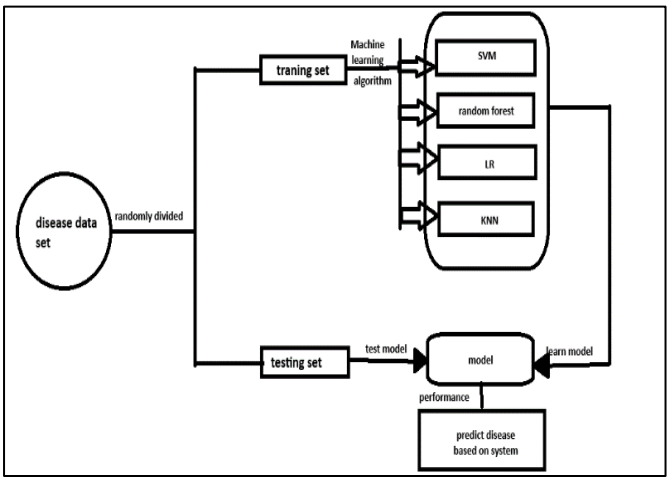

Fig 1 Proposed System Diagram

VII. RESULTS

Table 1 Results

Sr. No	Disease Name	Algorithm Name	Existing System Accuracy
1	Diabetes	SVM	73%
2	Heart disease	DT	52%
3	Parkinson's disease	SVM	71%
4	Breast cancer	LR	-
5	Lung cancer	LR	-

Table 2 Results

Sr. No	Disease Name	Algorithm Name	Proposed System Accuracy
1	Diabetes	SVM	77%
2	Heart disease	SVM	81%
3	Parkinson's disease	SVM	84%
4	Breast cancer	LR	91%
5	Lung cancer	LR	94%

VIII. CONCLUSION

Our study used machine learning methods like Support Vector Machine (SVM) and Logistic Regression to create a disease prediction model for five diseases: diabetes, heart disease, lung cancer, Parkinson's disease, and breast cancer. We got our data from Kaggle.com and made sure it was good quality. We achieved 77% accuracy in predicting diabetes using SVM and 84% accuracy for Parkinson's disease. For heart disease, we reached 81% accuracy using SVM. We used Logistic Regression for lung cancer prediction and got 94% accuracy, and for breast cancer, we got 91% accuracy.

Our system has a menu for every illness, making it simple to use. Users select a disease and enter parameters, and the system predicts the outcome. These accuracy rates show our machine learning algorithms are good at predicting these illnesses. However, accuracy changes according to the data and how the system was trained.

REFERENCES

[1]. [PDF] Prediction Of Diabetes Using Machine Learning Classification Algorithms | Semantic Scholar
[2]. https://globaljournals.org/GJCST_Volume10/6-Diagnosis-Of-Heart-Disease-Using-Datamining-Algorithm.pdf
[3]. Streamlit: Streamlit Documentation. https://docs.streamlit.io/
[4]. Kaggle: Kaggle website. https://www.kaggle.com/
[5]. https://globaljournals.org/GJCST_Volume10/6-Diagnosis-Of-Heart-Disease-Using-Datamining-Algorithm.pdf
[6]. Support Vector Machine (SVM): Corinna Cortes and Vladimir Vapnik (1995). Support-vector networks. Machine Learning, 20(3), 273-297.

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/IJISRT24APR1453

[7]. Support Vector Machine (SVM) Algorithm - GeeksforGeeks
[8]. Saeed, A., & Al-Jumaily, A. (2020). Machine learning techniques for Parkinson's disease diagnosis using handwriting: A review. Computers in Biology and Medicine, 122, 103804.
[9]. Logistic Regression: Hosmer Jr, D. W., Lemeshow, S., and Sturdivant, R. X. (2013). Applied Logistic Regression (3rd ed.). John Wiley & Sons.
[10]. scikit-learn: machine learning in Python — scikit-learn 1.4.2 documentation