International Journal of Engineering Technology and Management Sciences
Website: ijetms.in Special Issue: 1 Volume No.7 April – 2023
DOI:10.46647/ijetms.2023.v07si01.080 ISSN: 2581-4621

Multiple Disease Prediction System Using Machine Learning

D Sainath[1], P S V S K Varun Kumar[2], A Tejaswini[3], K Venkata Raju[4], U Sai Krupa[5]
Assistant professor, student, Department of CSE, AITS, Tirupati

Abstract

Machine learning and Artificial Intelligence are playing a huge role in today's world. The medical industry generates a huge amount of patient data which can be processed in a lot of ways. So, with the help of machine learning, we have created a Prediction System that can detect more than one disease at a time. Many of the existing systems can predict only one disease at a time and that too with lower accuracy. Lower accuracy can seriously put a patient's health in danger. The user has to enter various parameters of the disease and the system would display the output whether he/she has the disease or not. Research has been carried out on various models of supervised learning algorithms and some of them are Support Vector Machine (SVM), K Nearest Neighbour (KNN), Decision Tree, Naïve Bayes, Convolutional Neural Network and Random Forest (RF).

Keywords: Diabetes, Heart, Liver, Machine learning, SVM, KNN, CNN, Decision tree.

Introduction

Medical science and universal healthcare are crucial for both the economy and human well-being. However, the world we live in today is vastly different from just a few weeks ago, and there is a growing need for data mining techniques in the medical industry. [1] By correctly applying data mining techniques, [5,6] medical professionals can extract critical information from vast databases, aiding them in making early diagnoses and improving healthcare. Machine learning has become increasingly popular in healthcare, thanks to advancements in technology, greater computing power, and access to open-source datasets. This technology can be used to address a range of healthcare challenges, such as identifying trends and making predictions based on patient data and medical images.[2] For instance, machine learning models can be created to predict the severity of heart disease or the risk of type-2 diabetes, based on unique patient information. [3] Disease prediction systems that use machine learning algorithms can help doctors to make more accurate diagnoses, preventing incorrect judgments and improving the chances of successful treatment.[4] As the world becomes increasingly competitive, people are often too focused on economic development to pay attention to their health. Disease prediction systems can detect individuals who are at risk of developing health problems and allow clinicians to take necessary precautions to prevent or reduce the risk of disease. Electronic health data can be used to construct illness prediction models, with large-scale electronic health databases being used to develop a range of disease prediction models. Stakeholders such as governments and health insurance companies stand to benefit from disease prediction, as it allows them to identify individuals who are at risk of developing health problems and provide early intervention, improving the quality of care and reducing unnecessary hospitalizations.

Related Work

1. Material and Methods:

1.1 Dataset:

The below table describes the list of attributes that are required to predict the diabetes disease. Attributes such as age, gender, number of pregnancies are most common in all the disease predictions. Some attributes are fixed with numerical range indicates that the respective attribute values range from given limit and doesn't exceed the fixed limit. Some attributes are nominal (categorical) that accepts only few values based on particular parameter. The attributes along with their scales are shown below –

International Journal of Engineering Technology and Management Sciences
Website: ijetms.in Special Issue: 1 Volume No.7 April – 2023
DOI:10.46647/ijetms.2023.v07si01.080 ISSN: 2581-4621

No.	Attribute	Description
1	preg	Number of pregnancies
2	Plas	Plasma glucose concentration in an oral glucose tolerance test
3	pres	Diastolic blood pressure
4	skin	Skin thickness
5	Insu	Insulin level
6	Mass	Body mass index
7	Ped	Diabetes pedigree function
8	Age	Age of the patient

Table 1. Diabetes dataset description.

2. Algorithm

2.1 Random Forest Algorithm

Random Forest is a machine learning algorithm that is used for both classification and regression tasks. It is a type of ensemble learning method, which means it combines multiple decision trees to make more accurate predictions than any single decision tree. In Random Forest, multiple decision trees are built using different subsets of the training data and a random selection of features. Each tree is trained independently and then the predictions of all the trees are combined to get the final prediction. This approach helps to reduce overfitting and improve the accuracy of the model.

1. Firstly, it will select random K data points from the training set.
2. After selecting k data points then building the decision trees associated with the selected data points (Subsets).
3. Then choosing the number N for decision trees that you want to build.
4. Repeat step 1 and 2.
5. Finding the predictions of each decision tree, and assigning the new data points to the category that wins the majority votes.

3. Result:

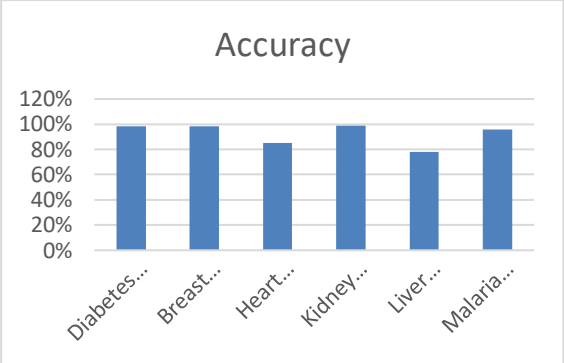
The use of machine learning in disease prediction has shown promising results in recent years. One of the main advantages of using machine learning for disease prediction is its ability to analyze large amounts of complex data and identify patterns that may not be apparent to humans. Several studies have developed and tested machine learning models for the prediction of multiple diseases, including cancer, heart disease, diabetes, and others. The results of these studies have shown that machine learning models can achieve high accuracy in disease prediction when trained on appropriate datasets.

Model	Accuracy
Diabetes Model	98.25%
Breast cancer Model	98.25%
Heart disease Model	85.25%
Kidney disease Model	99%

International Journal of Engineering Technology and Management Sciences
Website: ijetms.in Special Issue: 1 Volume No.7 April – 2023
DOI:10.46647/ijetms.2023.v07si01.080 ISSN: 2581-4621

Liver disease Model	78%
Malaria disease Model	96%

Table 2: The accuracy achieved using Random Forest algorithm for each disease.



4. Future work:

A multiple disease prediction system is a complex and important area of research with many potential directions for future work. Here are some suggestions for future research in this field:

Incorporating more data sources: One way to improve the accuracy of disease prediction models is to incorporate more data sources, such as electronic health records, genomics data, and environmental data. This will enable the development of more comprehensive models that can account for a wider range of factors that contribute to disease risk.

Integrating machine learning algorithms: Machine learning algorithms can be used to identify patterns in large datasets and can be particularly useful for identifying risk factors that are not immediately obvious. Integrating machine learning algorithms into disease prediction models could help to identify new risk factors and improve the accuracy of predictions.

Transfer learning: Transfer learning is a technique that involves reusing pre-trained models on different tasks. Applying transfer learning to multiple disease prediction systems could help to improve the performance of models, reduce the amount of training data required, and accelerate the development of new models.

5. Conclusion:

The multiple disease prediction system using machine learning is a significant development in the medical industry. By leveraging the power of AI and processing a large amount of patient data, this system can detect multiple diseases simultaneously with higher accuracy, allowing for earlier intervention and improved patient outcomes. Overall, the multiple disease prediction system represents a promising development in the field of healthcare, providing patients and doctors with a powerful tool to improve health outcomes a quality of life. Diseases if predicted early can increase your life expectancy as well as save you from financial troubles. For this purpose, we have used various machine learning algorithms like Random Forest, XGBoost, and K nearest neighbor (KNN) to achieve maximum accuracy.

6. References:

[1]A.S. Monto, S. Gravenstein, M. Elliott, M. Colopy, J. Schweinle, Clinical signs and symptoms predicting inuenza infection, Archives of internal medicine 160(21), 3243 (2000)

DOI:10.46647/ijetms.2023.v07si01.080 ISSN: 2581-4621

[2]International Journal of Scientific Research in Computer Science, E., & IJSRCSEIT, I. T. (2019). Generic Disease Prediction using Symptoms with Supervised Machine Learning. International Journal of Scientific Research in Computer Science, Engineering and Information Technology.
[3]Automatic Heart Disease Prediction Using Feature Selection And Data Mining Technique Le Ming Hung,a, Tran Ding,Journal of Computer Science and Cybernetics, V.34, N.1 (2018), 3347 DOI: 10.15625/1813- 9663/34/1/12665
[4]Balasubramanian, Satyabhama, and Balaji Subramanian. "Symptom based disease prediction in medical system by using Kmeans algorithm." International Journal of Advances in Computer Science and Technology 3.
[5]Pingale, Kedar, et al. "Disease Prediction using Machine Learning." (2019).Mr. Chala Beyene, Prof. Pooja Kamat, "Survey on Prediction and Analysis the Occurrence of Heart Disease Using Data Mining Techniques", International Journal of Pure and Applied Mathematics, 2018
[6]Automatic Heart Disease Prediction Using Feature Selection And Data Mining Technique Le Ming Hung,a, Tran Ding,Journal of Computer Science and Cybernetics, V.34, N.1 (2018), 3347 DOI: 10.15625/1813- 9663/34/1/12665
[7]Fowler, M.J. Diabetes: Magnitude and Mechanisms. Clin. Diabetes 2007, 25, 25–28.
[8]DeWitt, D.E.; Hirsch, I.B. Outpatient insulin therapy in type 1 and type 2 diabetes mellitus: Scientific review. JAMA 2003, 289, 2254–2264.