Disease Prediction Using Machine Learning Algorithms
Pranav Kumar, Subash Kumar Shah
Dept of Computer Science And Engineering
Sharda University of Engineering & Technology,
Greater Noida, Uttar Pradesh, India.
2018008769.pranav@ug.sharda.ac.in , 2018000119.subash@ug.sharda.ac.in

Abstract- Health care has become a major problem in the world. Disease cases are increasing rapidly among humans, especially among the younger generation. The healthcare sector is one of the leading research areas in the current context with the rapid development of technology and data. The latest technological advances allow for the automatic use of machine learning techniques. These machine learning methods can be used to diagnose 'or predict diseases' in advance. The healthcare industry produces a vast amount of information in its infancy in an unstructured format that a computer cannot understand. Thanks to the development of modern technology, the healthcare industry also manages information in a systematic way that can be understood through machine learning technology. In this case if we use machine learning technology to predict disease, then there is a chance of getting the disease in the early stages and informing the patient that they have received the best treatment to treat the disease. This paper uses seven controlled algorithms namely KNN, Decision Tree, Random Forest, Naive bayes to predict disease. In this paper among those seven algorithms, Neural Networks provided the best accuracy as 98.30% and this program provides results to test model accuracy in predicting diseases. The accuracy of the algorithm is determined by the performance of the data provided. The expected outcome and scope of this project is that if disease is not predicted rather than given in advance treatment that can be given to patients which can reduce health risks and save patients' lives and the cost of accessing treatment can be reduced to a minimum.

Keywords- Disease Prediction, Verification Decision, Logistic Correction, Unplanned Forest, K-Nearest Neighbor (KNN).

I. INTRODUCTION

Machine learning computer programming to improve performance using sample data or previous data. Machine learning to study computer programs that learn from data and information. The machine learning algorithm has two phases: Training, Testing. Predicting the disease using patient signals and technology to study history is difficult from decades ago. Machine learning technology provides an excellent platform in the medical field, so that health care issues can be effectively resolved. In the field of health care, Machine Learning is widely used in various fields of science such as identifying rare diseases, understanding patterns of rare disease predictions and more.

According to a study by the World Health Organization, 17.5 million people worldwide have died from heart attacks and strokes. The use of algorithms and pattern definitions can help save many lives by anticipating the disease status ahead of time. This project focuses on determining whether a patient has heart disease or not by considering the UCI database. The database originally contained seventy-six attributes that were collected in four different databases and fourteen symbols were used for our study. We use machine learning to keep complete hospital data machine learning technology that allows construction models to quickly analyse data and deliver results quickly, through machine learning technology doctors can make better decisions for patient diagnostics and treatment options, leading to improved patient health care services. Health care is the most outstanding example of how machine learning is used in the medical field. To improve accuracy from big data, existing work will be done on random data and text. Predicting existing diseases will be done on linear, KNN, decision Tree algorithm.

II. EXISTING SYSTEM

Predicting the use of a traditional model for serious illness often involves learning equipment and a supervised learning algorithm using training data with model training labels. The classification of high-risk patients was made into groups of study. But these types are very important in medical conditions and are widely studied. Sustainable health monitoring system using good systems. Studied various programs and was able to obtain good results in reducing the cost of medication and the simple case procedures for complex systems.

III. IMPLEMENTATION

The division of the Naïve Bayes: The division of the Bayesian sets the modified learning method and is a numerical method. It takes an important opportunity and allows us to maintain a model of moral uncertainty by
determining the probability of outcome. Therefore, it can solve diagnostic and predictive problems of the process.

In this Section he was named after Thomas Bayes, 1702-1761 who promoted the Bayes Theorem; the division of Bayesian provides practical learning skills, prior knowledge where I see it can be collected. Heart Disease Data: In the study of heart disease data the researcher uses a database from the Cleveland Clinic Foundation. This database contains 11 characters / characters and 303 lines / records. Every line corresponds to a single patient and every attribute is the same as observation or type of patient / object test. The description of the symbols is shown in Table 1

Table 1. Qualities and their meaning.

<table>
<thead>
<tr>
<th>No.</th>
<th>Qualities</th>
<th>Meaning</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Age of User’s</td>
<td>Age in years</td>
</tr>
<tr>
<td>2</td>
<td>Gender of User’s</td>
<td>Male Female</td>
</tr>
<tr>
<td>3</td>
<td>Type of Chest Pain</td>
<td>Typical Angina Atypical Angina Non-Angina pain Asymptomatic Pain</td>
</tr>
<tr>
<td>4</td>
<td>Blood Pressure (BP)</td>
<td>Blood pressure in mm Hg</td>
</tr>
<tr>
<td>5</td>
<td>Cholesterol</td>
<td>Cholesterol level in mg/dl</td>
</tr>
<tr>
<td>6</td>
<td>Blood Sugar</td>
<td>Is blood Sugar &gt; 120 mg/dl True False</td>
</tr>
<tr>
<td>7</td>
<td>BMI</td>
<td>0 Normal 1 Having ST-T wave abnormality Showing probable or define left ventricular hypertrophy</td>
</tr>
<tr>
<td>8</td>
<td>Diabetes</td>
<td>diabetes Achieved</td>
</tr>
</tbody>
</table>

1. Algorithms Naïve Bayes classifier:
The guided machine learning approach is represented by the Naïve Bayes algorithm. It uses a potential model by determining the probability of results / outcomes. It is used for analysis and prediction problems. Naïve Bayes has audio capabilities in the input database. Described in Figure 1.

2. Decision Tree:
The decision tree algorithm is similar to a tree that is used to map a particular object to extract an object. Types of trees with limited extraction are called decision trees (DT). In this genealogy the leaves show the class labels and the branches show the interaction between the symbols that get results from those labels in the program section. Deciduous trees with continuous extraction stages are called deciduous trees. In data mining, a decision- making tool can make decisions. An example of a decision tree is shown in Figure 2.

3. KNN Algorithm:
K-Nearest Neighbors is one of the easiest ways to learn mechanically through a targeted learning process. The KNN algorithm is similar to the similarities between new cases / data and available cases and places a new case in the same category as available categories.

The KNN algorithm stores all available data and separates the new data point accordingly. This means that where new data emerges it can be easily broken down into a workable program using the algorithm algorithm can be used for retrieval and partitioning but especially KNN problems is not a parametric algorithm, meaning logical basic data. It is also called an algorithm for lazy students because it does not learn to set up training faster than database storage and during partitioning, it works in a database.

The KNN algorithm in the training phase only saves data when it receives new information, and then separates that data into the same category. The algorithm is also compared to a real column in the database. This comparison provides the necessary precision.

4. Random Forest (RF):
Random Forest is a popular algorithm for learning machines that is a method of supervised learning. It can be used for both Classification and ML Problems in ML. It is based on the concept of learning together, which is the process of combining multiple variables to solve a complex problem and improve model performance.
IV. RESULT

Table 2. Comparison of accuracy of algorithm.

<table>
<thead>
<tr>
<th>Algorithm</th>
<th>Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Decision Tree</td>
<td>84.5%</td>
</tr>
<tr>
<td>Random Forest</td>
<td>98.95%</td>
</tr>
<tr>
<td>Naïve Bayes</td>
<td>89.4%</td>
</tr>
<tr>
<td>SVM</td>
<td>96.49%</td>
</tr>
<tr>
<td>KNN</td>
<td>71.28%</td>
</tr>
</tbody>
</table>

We found that the Support Vector Machine (SVM) algorithm is widely used (in 30 studies) followed by the Naïve Bayes algorithm (in 24 studies). However, the Random Forest algorithm showed relatively high accuracy. In the 40 studies in which it was used, RF showed the highest accuracy of 98.95%. This was followed by SVM which included 96% of the accuracy considered.

![comparison of accuracy of algorithm](image)

Fig 3. Comparison of Accuracy of Algorithm.

V. CONCLUSION

This systematic review aims to determine the performance, limitations, and future use of Software in health care. Findings may help inform future developers of Disease Predictability Software and promote personalized patient care. The program predicts Patient Diseases. Disease Prediction is done through User Symbols.

In this System Decision tree, Unplanned Forest, the Naïve Bayes Algorithm is used to predict diseases. For the data format, the system uses the Machine Learning algorithm Process Data on Database Data namely, Random Forest, Decision Tree, Naive Bayes. System accuracy reaches 98.3%. machine learning skills are designed to successfully predict outbreaks.

ACKNOWLEDGEMENT

I'd like to express my heartfelt gratitude to Dr Manish Verma, my project guide, for his relentless support, motivation, and inspiration during the project. I'd also like to say Thank you to Dr Parma Nand, Dean of Computer Science and Engineering Department, for his invaluable encouragement, ability to motivate me, and willingness to solve problems that enabled me to create a unique project and complete the mission.

REFERENCE