ISSN: 2350-0328
International Journal of AdvancedResearch in Science, Engineering and Technology
Vol. 12, Issue 5, May 2025

Using Machine Learning Algorithms to Predict Disease

Sumit Vishwanath Wakode, Prof. Minakshi Ramteke

P.G. Student, Department of Computer Science, VMIT, Nagpur, India
Assistant Professor, Department of Computer Science, VMIT, Nagpur, India,

ABSTRACT: Preventing and treating illness requires timely and accurate analysis of any health-related issue. A serious illness may not be adequately diagnosed using the conventional method. An improved diagnosis over the traditional approach can be achieved by creating a medical diagnosis system that use machine learning (ML) algorithms to predict any disease. We used a number of machine learning algorithms to create a disease prediction system. More than 230 diseases were included in the dataset that was processes. Given an individual's age, gender, and symptoms, the diagnosis system provides the output as the disease that the person may have.

I. INTRODUCTION

One of the most important aspects of the economy and human life is medicine and healthcare. The world we currently live in has changed significantly from the one that existed only a few weeks ago. Everything has become twisted and gory. In this scenario, where everything is virtual, the medical professionals are doing everything in their power to save lives, even if it means putting their own lives in danger. Additionally, some isolated communities lack access to healthcare facilities. Virtual doctors are board-certified medical professionals who prefer to practice virtually through phone and video consultations as opposed to in-person ones; however, this is not feasible in an emergency. Machines are always regarded as superior to humans because they can complete task more quickly and accurately when human error is absent. Any patient's disease can be predicted by a a disease predictor, also known as a virtual doctor, without any human error.

A. PROBLEM DEFINITION

Predicting diseases with machine learning algorithms in an effort to reduce medical costs as much as possible. The health prediction system is still lacking several features. Therefore, even though our world is technologically advanced, it is meaningless if we are unable to use it properly and efficiently. To address this, research is being done on health prediction systems. Different technologies are used in different applications. This project serves as an example of how well both technologies work together.

B. OBJECTIVES

1. Aim of the Proposed topic:

- When the quality of medical data is lacking, the analysis's accuracy decreases Moreover.
- Those previous studies, however, primarily took structured data into consideration.
- Appropriate techniques for handling seme-structured and unstructured data are lacking.
- The system that is being suggested will take into account both structured and unstructured data.
- The application of machine learning algorithms improves the accuracy of the analysis.

2.Scope of the proposed topic

- The purpose of this work is to give consumers and professionals a tool to help them identify and select diseases.
- Early disease detection will be highly beneficial to healthcare organization such as hospitals. We can add more illness to the current system in the future. To lower the death rate, we can attempt to increase prediction accuracy.

ISSN: 2350-0328
International Journal of AdvancedResearch in Science, Engineering and Technology
Vol. 12, Issue 5, May 2025

II. PROCEDURE

With the help of algorithms and other tools, we have developed a system that uses a patient's symptoms to predict their disease. Next, we compare those symptoms to the dataset that the system had access to earlier. It has been proposed that this machine learning system can predict a number of diseases. It will be possible to predict the precise percentage of the patient's disease by comparing those datasets with the patient's condition. After sending the dataset and symptoms to the system's prediction model, where they have been pre-processed for later use, the user selects the feature by entering or choosing the different symptoms.

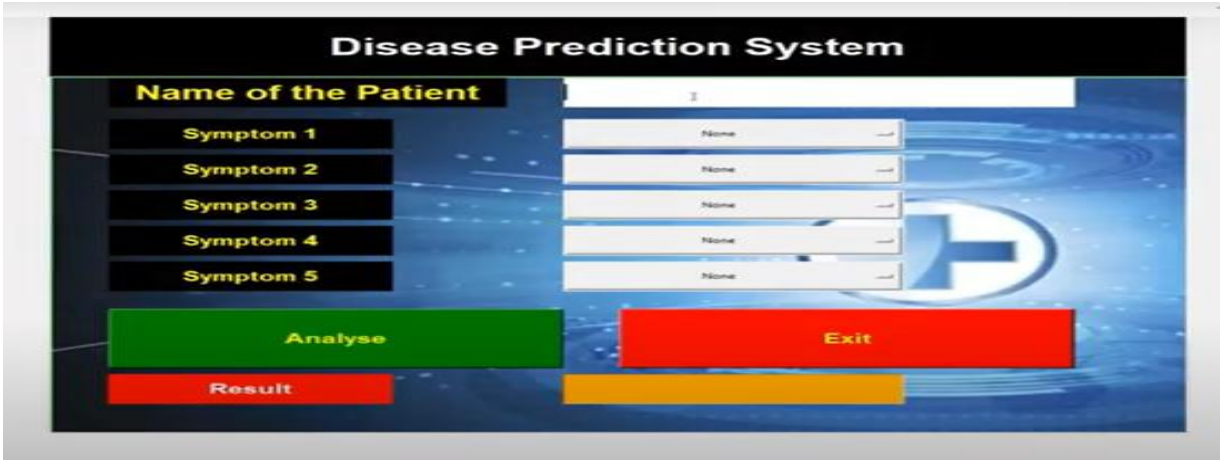

Fig. 1 Proposed system for disease prediction

Fig. 1 A disease prediction system I proposed. When needed, the doctor might not always be accessible. However, in the current situation, this prediction system can be used at any time based on necessity. The ML model can be further processed by providing the individual's symptoms, age, and gender.

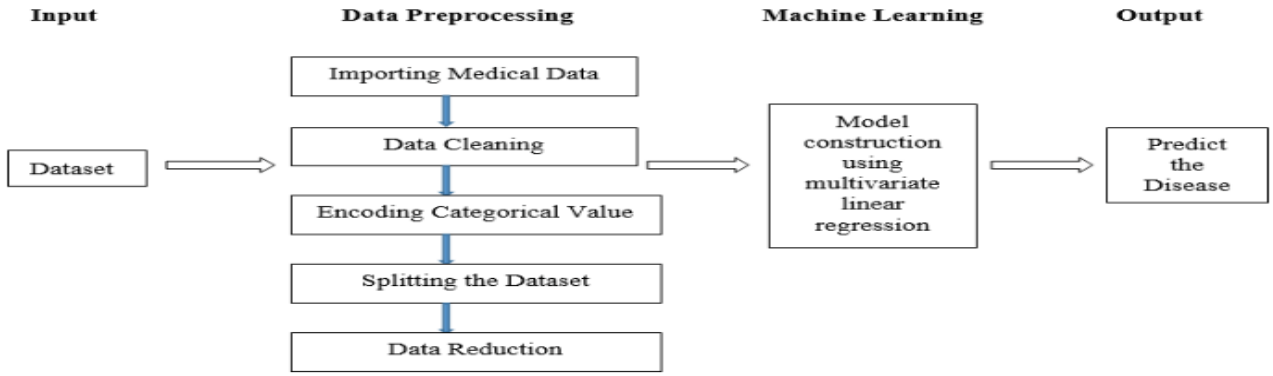

Fig. 2 Proposed system flow diagram during training.

Fig. 2 A system flow diagram that was suggested during training. The datasets, which included a person's age, gender, and symptoms, was pre-processed before being used as input by various machine learning algorithms to predict the disease. The various machine learning models that were employed included RUS Boosted trees, Fine, Medium, and Coarse Decision trees, Gaussian Naïve Bayes, Kernel Naïve Bayes, Fine, Medium, and Coarse KNN, Weighted KNN, and Subspace KNN.

ISSN: 2350-0328
International Journal of AdvancedResearch in Science, Engineering and Technology
Vol. 12, Issue 5, May 2025

III. THE WAY THE ML MODEL OPERATES

Fig. 3 How the ML models work. Based on the input factors age, gender, and symptoms the dataset was divided into input and outputs, which represented the diseases. The available data was divided into train and test sets at random.

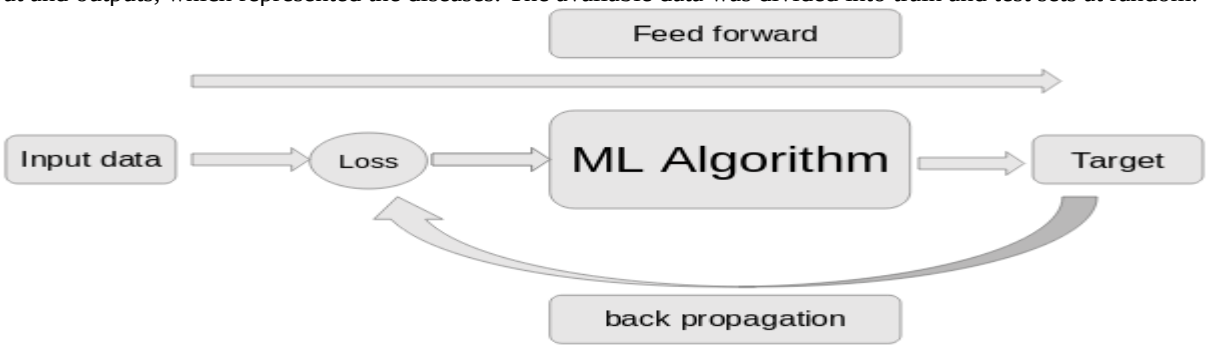

Fig. 3 Operation of Machine Learning Models

After that, these sets were encoded and subjected to additional training using various algorithms. The algorithms then evaluate that, the training set and forecast the values, determining the accuracy of various machine learning algorithms

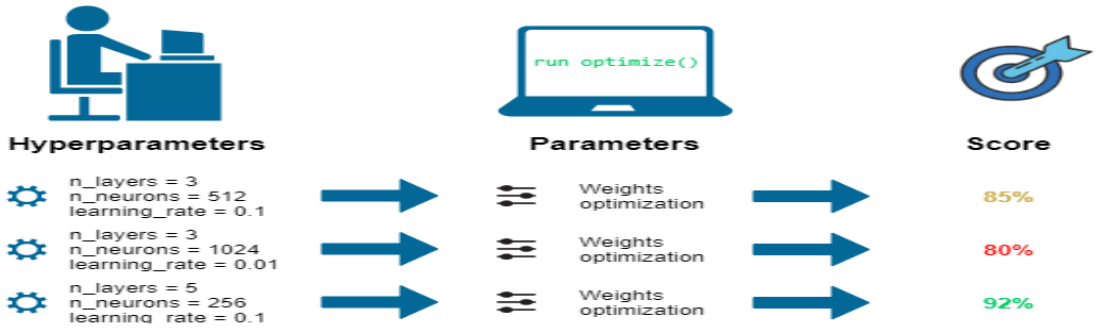

Fig .4 Accuracy values of different ML models

This supervised machine learning algorithm is one of the KNN algorithms. It merely determined how far a new data point was from every other training data point. Both Manhattan and Euclidean distances are possible. The K Closest data points are then chosen; K can be any integer. Finally, it place the data point in the class that most K data point fall into. Using the integer parameter K that we selected for KNN, we were able to determine the major predicted values.

However, the algorithm becomes much more sensitive to outlier points if K is set too low, Additionally, all of the points that are nearly identical to the K value are chosen if the K value is too high. The weight KNN addressed this problem by giving the points closest to the K value more weight and the points farther away less weight. With this model, we were able to achieve the highest accuracy. Additionally, this model produced the best results out of all the KNN models. In order to determine the distance, we must assign integer values to K. Therefore, we must assign integer values to K. Therefore, we chose a low value for K in our fine KNN model, while the medium KNN model uses about 10. The accuracy of the model varies because of the neighbours for each.

In operates in the same way as the Naïve Bayes. However, a categorical dataset is required for Naïve Bayes, and a dataset with all continuous features is required for Gaussian Naïve Bayes. This model had to be used because our dataset included continuous features of gender, age, and symptoms. This model's accuracy wasn't particularly high.

Our dataset had some numerical attributes such as age so we implied Kernel Naïve Bayes to predict the medicines. The steps followed for this algorithm are similar to the Naïve Bayes. The major benefit of using this algorithm is that it provides estimators that have a nonparametric nature. If there is no prior knowledge that the dataset used is parametric

ISSN: 2350-0328
International Journal of AdvancedResearch in Science, Engineering and Technology
Vol. 12, Issue 5, May 2025

or not this model can give more accurate results. The results given by this model were almost the same as shown by the Gaussian Naïve Bayes.

IV. ADVANTAGES AND DISADVANTAGES

ADVANTAGES

1. **Early Detection**: Machine learning can identify pattern and anomalies in large datasets, allowing for the early detection of diseases before symptoms manifest.
2. **Accuracy:** ML algorithms can analyze complex data more accurately than humans, leading to more reliable predictions.
3. **Personalized Medicine:** Machine learning can tailor treatment plans to individual patients based on their unique health data, improving outcomes.
4. **Timely Intervention:** Predictive models can prompt healthcare providers to intervene in timely manner, potentially saving lives.
5. **Cost-Efficiency:** Early detection and prevention can reduce healthcare costs by avoiding expensive treatments associated with advanced disease stages

DISADVANTAGES

1. **Data Quality:** The accuracy of predictions heavily depends on the quality of input data, which may contain errors or biases.
2. **Privacy Concerns:** Collecting and sharing sensitive health data for machine learning can raise privacy concerns if not handled properly.
3. **Overfitting:** ML models can overfit to training data, making them less generalizable to new cases.
4. **Resource Intensive:** Implementing and maintaining machine learning systems in healthcare can be resource-intensive, requiring skilled personnel and infrastructure.
5. **Ethical Issues:** Decisions bases on ML predictions may lead to ethical dilemmas, such as biased outcomes or discrimination against certain groups.

V. CONCLUSION

The manuscript presented the technique of predicting the disease based on the symptoms, age, and the gender of an individual patient. The Weighted KNN model gave the highest accuracy of 93.5% for the prediction of diseases using the above-mentioned factors. Almost all the ML models gave good accuracy values. As some models were dependent on the parameters, they couldn't predict the disease and the accuracy percentage was quite low. Once the disease is predicted, we could easily manage the medicine resources required for the treatment. This model would help in lowering the cost required in dealing with the disease and would also improve the recovery process. Machine learning has the potential to revolutionize disease prediction and improve patient outcomes. By exploring the best algorithms and themes in disease prediction research, we can make significant strides in healthcare and public health let's continue to this exciting field and make a positive impact on people's lives.

REFERENCES

[1]. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3661426
[2]. https://ijcrt.org/papers/ijcrt2105229.pdf
[3]. https://www.researchgate.net/publication/357449131_the_prediction_of_disease_using_machine_learning
[4]. https://nevonprojects.com/multiple-disease-prediction-system-using-machine-learning/
[5]. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3661426#:~:text=We%20have%20designed%20a%20disease,individual%20might%20be%20suffering%20from.