International Journal For Technological Research In Engineering
Volume 9, Issue 10, June-2022 *ISSN (Online): 2347 - 4718*

DISEASE PREDICTION USING MACHINE LEARNING

1Kritika Sharma, 2Sachin Jha, 3Vedansh Mehra, [4]Prof. Gurpreet Kaur
[1,2,3]Students, [4]Assistant Professor
Department of Computer Science Engineering
Bhagwan Mahaveer College of Engineering and Management, Sonipat, India

ABSTRACT
The Ailment Prediction System, which is based on predictive modelling, predicts the user's disease based on the symptoms that the user provides as input to the system. The algorithm analyses the symptoms provided by the user as input and generates an asian result indicating the likelihood of the disease. The implementation of the NaiveiBayes Classifier is used to predict disease. The disease probability is calculated using the Naive Bayes Classifier. With the growth of big data in the biomedical and health-care communities, accurate medical data analysis promotes early disease identification and patient care. We forecast diseases like Diabetes, Malaria, Jaundice, Dengue Fever, and Tuberculosis using linear regression and decision trees.

1. INTRODUCTION

Machine learning is the process of programming computers to maximise their performance based on examples or historical data. The study of computer systems that learn from data and experience is known as machine learning. Training and Testing are indeed the two main phases of said machine learning algorithm. Prediction of a disease based on the symptoms and medical history of the patient Machine learning has been a stumbling block for decades. Machine Learning technology provides a good platform in the medical industry for rapidly resolving healthcare challenges.

We're using machine learning to keep track of all of the hospital's data. Machine learning is a technology that allows doctors to make better decisions for patient diagnoses and treatment alternatives by allowing them to construct models swiftly evaluate data and deliver answers faster. The most prominent example of how machine learning is used in the medical profession is healthcare.

Existing work on unstructured and textual data will be done to increase the accuracy of huge data. The existing algorithm for disease prediction will be linear, KNN, and iDecision Tree. The reference order in the running text should correspond to the reference list at the end of the paper.

2. OBJECTIVE

There is a need to research and develop a system that would allow end users to predict chronic diseases without having to contact a physician or doctor for diagnosis. To diagnose various diseases by evaluating the symptoms of patients and applying various Machine Learning Models. There is no proper mechanism for handling text and structured data. Both organised and unstructured data will be considered by the proposed system. Machine Learning will improve the accuracy of predictions.

3. EXISTING SYSTEM

The system forecasts chronic diseases for a certain region and community. Disease Prediction is only performed for specific diseases.

Big Data and the CNN Algorithm are utilised in this system to forecast disease risk. The system uses K-nearest Neighbors, Decision Tree, and Nave Bayesian Machine Learning algorithms for S type data. The system's accuracy is up to 94.8 percent.

The system forecasts chronic diseases for a certain region and community. Disease Prediction is only performed for specific diseases.

Big Data and the CNN Algorithm are utilised in this system to forecast disease risk. The system uses K-nearest Neighbors, Decision Tree, and Nave Bayesian Machine Learning algorithms for S type data. The system's accuracy is up to 94.8 percent.

4. PROPOSED SYSTEM

Most chronic diseases are predicted using this system. The machine learning model accepts both structured and textual data as input. End users are the ones who use this system. On the basis of symptoms, the system will anticipate disease. Machine Learning Technology is used in this system. For disease prediction, the Nave Bayes algorithm is utilised; for clustering, the KNN algorithm is employed, and the final output is in the form of 0 or 1, for which the Logistic tree is used.

5. DATASET AND MODEL DESCRIPTION

This section describes the dataset that was used to train the machine learning model. Symptoms of various diseases will be included in the dataset.

5.1 DATASET OF HOSPITAL

The data from the hospital will be in either a textual or structural format. This project's dataset is based on real-world data. Patients' symptoms are contained in structural data, whereas unstructured data is in textual format.

The collection includes real-world hospital data as well as data from a data centre. The hospital's information includes the patients' symptoms.

6. EVALUATION METHOD

To calculate the performance evaluation in an experiment, we first define TP, TN, Fp, and Fn as true positive (number of results correctly predicted as required), true negative (number of results not required), false positive (number of results incorrectly predicted as required), and false negative (number of results incorrectly predicted as not required), respectively.

We can get four measurements: recall, precision, accuracy, and F1 measurement, which are as follows:

Accuracy = _____
Precision = _____
Recall = _____
F1-Measure = _____

7. SYSTEM ARCHITECTURE

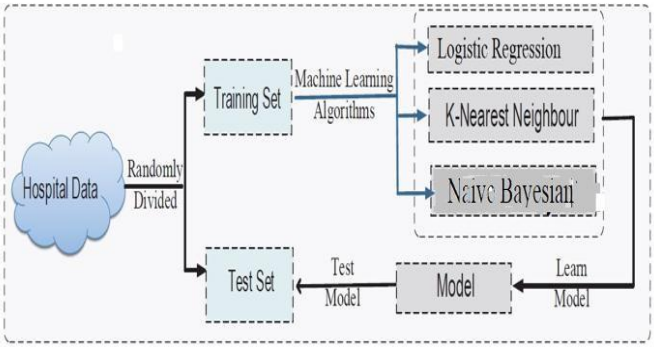

Fig -1: System Architecture

8. ALGORITHM

8.1 KNN

K Nearest Neighbor (KNN) is a basic, easy-to-understand, adaptable, and one of the most advanced machine learning algorithms. The user will be able to predict the disease in the Healthcare System. The user can forecast whether or not an illness will be detected using this approach. The proposed method divides diseases into distinct classifications, indicating which disease will occur based on symptoms. For each classification and regression problem, the KNN rule was utilised. The KNN algorithm depends on feature similarity.

8.2 NAIVE BAYES

For prognosticative modelling, Naive Bayes is a simple yet incredibly powerful rule. One of the easiest methods is to choose the most likely hypothesis based on the facts we have, which we may utilise as past information about the subject. The Bayes' Theorem explains how we can determine the likelihood of a hypothesis given our prior knowledge.

8.3 LOGISTIC REGRESSION

Logistic regression is a supervised learning classification technique used to predict the likelihood of a disease target variable. Because the nature of the target or variable is separated, there are only two possible groups.

9. CONCLUSIONS

The goal of this study is to forecast disease based on symptoms. The project is set up in such a way that the system takes the user's symptoms as input and creates an output, which is disease prediction.

Finally, the accuracy of risk prediction in diseasemrisk modelling is determined by the diversity of hospital data.

ACKNOWLEDGMENT

I wish to express my gratitude to all those who provided help and cooperation in various ways at the different stages for this research. Also, I would like to express my sincere appreciation to the director sir of Bhagwan Mahavir College Of Engineering And Management, Head of Computer Science Engineering Department Ms. Gurpreet Kaur.

REFERENCES

[1] Wang, and Lin Wang “Disease Prediction by Machine Learning over Big Data from Healthcare Communities” (2017).

[2] 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.

[3] P. Groves, B. Kayyali, D. Knott, and S. V. Kuiken, “The ’big data’ revolution in healthcare: Accelerating value and innovation,” 2016.

[4] S.-H. Wang, T.-M. Zhan, Y. Chen, Y. Zhang, M. Yang, H.-M. Lu, H.-N. Wang, B. Liu, and P. Phillips, “Multiple sclerosis detection based on biorthogonal wavelet transform, rbf kernel principal component analysis, and logistic regression,” IEEE Access, vol. 4, pp. 7567–7576, 2016.

[5] S. Patel and H. Patel, “Survey of data mining techniques used in healthcare domain,” Int. J. of Inform. Sci. and Tech., Vol. 6, pp. 53-60, March 2016.

