iJRASET
International Journal For Research in Applied Science and Engineering Technology



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 13 **Issue:** IV **Month of publication:** April 2025

DOI: https://doi.org/10.22214/ijraset.2025.69110

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 13 Issue IV Apr 2025- Available at www.ijraset.com

Special Disease Prediction System Using Machine Learning

Madhu Mitha[1], Priya Dharshini[2]
B. Sc. CS – III, Kalasalingam Academy of Research and Education, TN, India

Abstract: ***The wide variety of computer-based technologies within the healthcare industry has led to the gathering of electronic data. Due to the massive number of information, medical professionals are faced with the challenge of accurately diagnosing signs and figuring out diseases at an early stage. In medicine, misdiagnosis could be a major factor leading because of poor treatment and diagnosing the disease when it's serious. However, supervised machine learning techniques have demonstrated the potential to surpass conventional diagnostic procedures and assist medical professionals in diagnosing highrisk diseases. Mostly people feel lazy to visit a hospital, and concern a doctor for a minor complication. However, this small problem can pose significant medical risk. Since, online medical advice is readily available. The system evaluates the symptoms that person give as an input and gives the disease as an output. Naive Bayes Classifier is used in the system. Our system focuses on accuracy, the more numbers of a symptoms furnished by the person as a input the disorder prediction as a output will be better. Work can enhance the health care industry to zenith and give cure to world.***

Keywords: Supervised Machine Learning, Healthcare, Disease Prediction, Naïve Bayes classifier

I. INTRODUCTION

As we all know Machine learning algorithms use a range of developmental, mathematical, and probable development techniques to learn from data generated from past events, and use them in higher cognitive process. The enhancement of AI has enabled computer systems to acknowledge, think and use intelligently as human beings. sML algorithms are considered to be utilized in a good range of fields including network login recognition, customer acquisition behaviour detection, production process improvement, Mastercard fraud detection, and disease predictions. Many of those applications are built using the ML supervised method. During this way, records unit with labels had been delivered to hypothesis fashions that expect non-label fashions. This implies the thought that, Medical doctors can use supervised learning as a robust tool for diagnosing diseases effectively. consistent with our study general mathematical techniques, work experience and understanding of medical professionals led to unpopular selections and errors when detecting disease-related risks. With the rapid proliferation of electronic health data, medical doctors are facing the challenge of diagnosing disease accurately in advance. For this reason, advanced calculation methods such as supervised ML algorithms are introduced to detect logical patterns and hidden information from data, which can be used to make critical decisions. As a result, the burden on medical staff will be decreased, and the death rate of patients will be reduced and survival rate will be enhanced.

.

II. PROPOSED SOLUTION

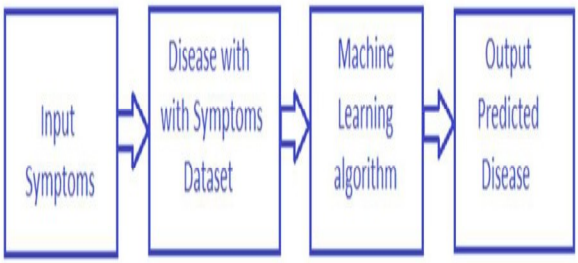


Fig.1. Block Diagram

©IJRASET: All Rights are Reserved

Classical process could be a procedure within which a patient should visit a doctor, receive various treatments and are at the end gets the results . This procedure will takes a lot of time.

To save time and money which is needed for the initial diagnostic process, this program suggests an automatic diagnostic system based on user input. The program takes input from the user and provides an inventory of possible diseases. In Fig.1 above we defined our proposed model.

The purpose of this study was to test the proposed hypothesis that ML algorithms are monitored to improve health care with accurate and rapid diagnosis. In this study, we investigate studies using more than one monitored ML model for each diagnostic problem. This way Provides more understanding and accuracy because the performance testing of a single

algorithm over different research settings creates biases that produce vague results. Analysis of ML models will be performed on a few diseases of the heart, kidneys, breast, and brain. To diagnose the disease, several methods are going to be tested like KNN, NB, DT, CNN, SVM, and LR. We've used multinomial NB as many variants mean that more signals are taken.

A. Machine Learning

The ML is a sub-field of AI. Its purpose is to enable computers to read on their own.

The ML algorithm enables it to identify patterns in targeted data, create descriptive models and predict objects without having clear predefined rules and models.

B. Different types of Machine Learning

1) Supervised Learning: -In supervised learning, we are given a group of information and already know what the proper output should appear as if, with the concept that there's a connection between the input and output.

2) unsupervised Learning: -It allows us to faced problems with little or no knowledge of what our results appears as if. We are able to experience the formation of knowledge after we don't really know the impact of the variables. We can find this structure by combining data supported relationships between variables within the data.

3) Reinforcement learning: A range of software and equipment are wont to determine the simplest behaviour or approach.

III. IMPLEMENTATION

We applied our algorithm to IDE Python 3.6.0 using the VS code (visual studio).We tested the algorithm on a system with a quad core i5 processor , RAM-8gb, 1TB hard disk.

To analyse the result statistics, we used Matplotlib and therefore the NumPy python library.

Symptoms are reported as infections. The disease is predicted using Naive bayes algorithm. According to literature searches, this algorithm leads to high accuracy in large databases. The program contains labels and other illness has its own symptoms. 70% of the database is used for training and 30% is used for testing. Training and testing is done on the system and the results obtained are also available.

1) Algorithm for Initial part

Step 1: Import all the required packages i.e., Tkinter for GUI, pandas for reading the csv files and numpy to perform numerical operations.

Step 2: Creating a list which contains all the symptoms which are according the dataset csv file

Step 3: Create another list which contains the diseases.

Step 4: Create two empty list

L1 and L2 , both have same length.

L1	S1	S2	S3	S4	S5	S6	S7	S8
L2	0	0	0	0	0	0	0	0

2) Algorithm for Dataset part

Performing the same steps for both testing and training dataset

Step 1: Using pandas module read the CSV file

Step 2: Replace with index

S1	S2	S3	S4	S5	S6	S7	S8	…	Prognosis
0	0	0	1	0	1	0	0	…	0
0	1	0	0	0	0	0	0	…	1
1	0	0	0	0	0	0	1	…	2
0	0	0	0	0	0	0	0	…	3

Step 3: Create X as symptoms and Y as disease.

a) Naïve Bayes Algorithm

This program receives information from the user and predicts the most likely diseases. This is achieved using datasets and machine learning algorithms. The algorithm we use is Naïve Bayesian that works deterministically or probabilistically. We need to import a scikit to read the library to use. In this case, we use polynomial

NB because there are many disease symptoms

Bayes theorem

- The objective of Bayes theory is to predict the disease category as a class label in our project with data in Tuple.
- Let X contain input H some assumptions, because tuple x (input (symptoms)) belongs to particular class c (output (disease)).
- In the case of separation problems, we check for the possibility that the dataset X is of type C, since we know that the attribute definition of X.

b) Dataset

The data comes from a study The study was carried out at the University of Colombia. It contains 150 diseases and each disease has an average of 810 symptoms. 70% of the data used for training is done taking into account all components of the input. The symptoms corresponding to the respective disease are marked as 1 and remain as 0.

Contains 5 drop down options where we have a list of all successful symptoms. User can select one of the symptoms and by clicking the button the correctly predicted disease will be displayed in the text box. Fig.2 contains a little detail of the dataset.



Fig: Data taken

Fig.2. Datasets

c) GUI

Tkinter GUI has been used in our project. It used for creating the GUI for desktop based applications.

Tkinter is very easy to grasp, required very little code for application, portable across all OS, and preinstalled with Python library. And it is faster way to create GUI when python with Tkinter. The developed GUI is shown in Fig.3.

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 13 Issue IV Apr 2025- Available at www.ijraset.com

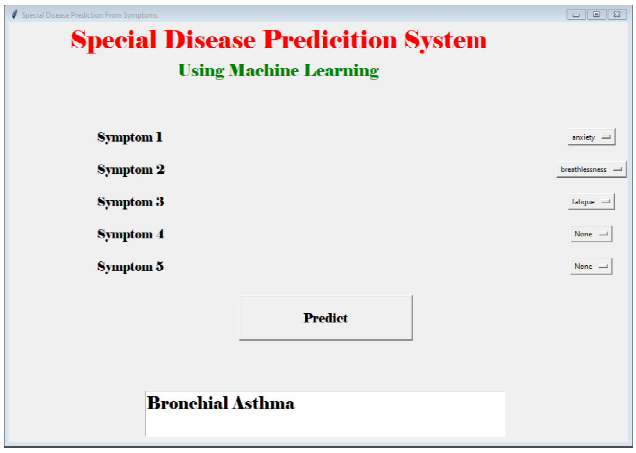


Fig.3. Developed GUI

IV. RESULT AND DISCUSSION

Classical process could be a procedure within which a patient should visit a doctor, receive various treatments and are at the end gets the results . This procedure will takes a lot of time.

To save time and money which is needed for the initial diagnostic process, this program suggests an automatic diagnostic system based on user input. The program takes

symptoms as a input from the user and provides accurately predicted diseases as a output in within the text box.

The system will predict the disease in within which symptoms are given as an infection or anything adverse happening to you which makes you uncomfortable. The disease will be predicted using the Naive Bayesian algorithm. In keeping with a literature study, this algorithm results in high accuracy of enormous datasets. The GUI contains symptoms of all possible diseases as labels, symptoms are chosen accordingly and then prediction will be don. 70% of the dataset will be used as training and 30% will be used for data testing. Training and testing will be done on the GUI and therefore the results obtained are available.

- Diseases are predicted using the Naïve Bayesian algorithm which works mainly with Multinomial NB as many symptoms will be there.
- According to the literature, this algorithm leads to the complete accuracy of an outsized database.
- The GUI contains diseases like labels gives symptoms for each disease.

V. CONCLUSION

The project is built so far that the system takes symptoms from the user as input and generates output i.e. disease prediction. The user can select at least one to five symptoms. Less accuracy will be obtained if only one input is selected. The greater the number of input, the greater the accuracy. In this paper we have proposed a learning model for a compact novel machine Algorithm of Naïve bayes. We also tried to reduce the number of features from the dataset. In this process we were able to obtain sufficient accuracy for all data sets using our machine learning model. We found the best accuracy of the most disease approx.(78.6%).

In upcoming Work, the event of complex ML algorithms is extremely necessary for enhancement of disease prediction. In additionally, data sets should be expanded to different demographics to avoid overcrowding and to extend the accuracy of the models used. Finally we'll try and put all the medical report into this specially last 10-20 medical records in order that everything should be smooth and our system will curtain the burden of medical staff and lots of aspect.

VI. ACKNOWLEDGMENT

It's an honor to work with Professor SPS Chauhan sir (Program Chair), he guided us in all the obstacles, thankful for the good support regarding paper and their all views. We are really thankful to our reviewer who encourages us to figure more on this and showing their keen interest in our paper. I also place on record, my sense of gratitude to one and all, who directly and indirectly, have lent their hand during this journey.

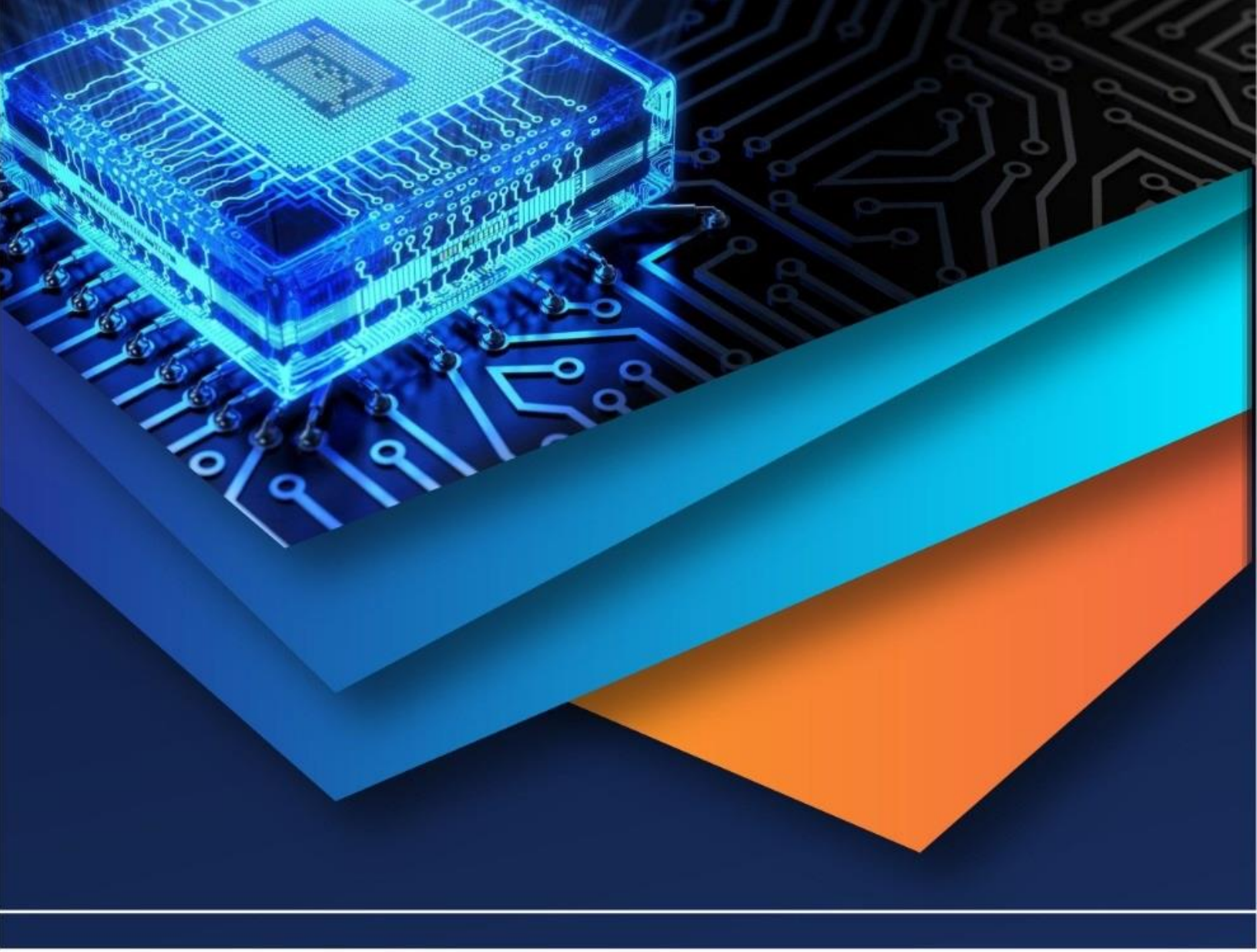
.

©IJRASET: All Rights are Reserved

REFERENCES

[1] M. Marimuthu, M. Abinaya, K. S., K. Madhankumar, and V. Pavithra, "A Review on cardiovascular disease Prediction using Machine Learning and Data Analytics Approach," International Journal of Computer Applications, vol. 181, no. 18, pp. 20–25, 2018.
[2] Joshi J, Rinal D, Patel J, Diagnosis And Prognosis of Breast Cancer Using Classification Rules, International Journal of Engineering Research and General Science,2(6):315-323, October 2014.
[3] V.Chaurasia and S. Pal, "Data Mining Approach to Detect Heart Diseases", International Journal of Advanced applied Science and Information Technology (IJACSIT), Vol. 2, No. 4, 2013, Page 56-66.
[4] K. Dwivedi, "Performance evaluation of various machine learning techniques for prediction of heart condition," Neural Computing and Applications, vol. 29, no. 10, pp. 685– 693, 2018.
[5] SolankiA.V., Data Mining Techniques using WEKA Classification for Sickle Cell Disease, International Journal of Computer Science and Information Technology,5(4): 58575860,2014.
[6] Ahmed F. Otoom, Emad E. Abdallah, Yousef Kilani, Ahmed Kefaye and Mohammad Ashour.(2015) .Effective Diagnosis and Monitoring of Heart Disease. International Journal of Software Engineering and Its Applications. Vol. 9, No. 1, pp. 143-156.
[7] P. P. Sengar, M. J. Gaikwad, and A. S. Nagdive, "Comparative study of machine learning algorithms for breast cancer prediction," Proceedings of the 3rd International Conference on Smart Systems and Inventive Technology, ICSSIT 2020, pp. 796–801, 2020.
[8] Chen, Yixue Hao, Kai Hwang , Lu Wang, and Lin Wang(2017). "Disease Prediction by Machine Learning Over Big Data From Healthcare Communities".IEEE Access. Vol. 5, pp. 8869-8879.
[9] Michael L. Raymer, William F. Punch, Erik D. Goodman, Leslie A. Kuhn, and Anil K. Jain(2000). "Dimensionality Reduction Using Genetic Algorithms". IEEE Transactions on Evolutionary Computation. Vol. 4, Issue 2, pp. 164-171.
[10] D. Asir Antony Gnana Singh,E.JebamalarLeavline R. Priyanka and P. Padma Priya(2016)."Dimensionality Reduction using Genetic Algorithm for Improving Accuracy in Medical Diagnosis", I.J. Intelligent Systems and Applications MECS. No. 1, pp.67-73.
[11] S. Jadhav, R. Kasar, N. Lade, M. Patil, and S. Kolte, "Disease Prediction by Machine Learning from Healthcare Communities," International Journal of Scientific Research in Science and Technology, pp. 29–35, 2019.

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)