e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:04/Issue:05/May-2022 Impact Factor- 6.752 www.irjmets.com

DISEASE PREDICTION WEB APP USING MACHINE LEARNING

Mohammad Ali Nawab*1, Tavisi Malpani*2, Tushar Kaundal*3,
Ms. Divya Soni*4
*1,2,3,4Department Of Computer Science And Engineering, HMRITM New Delhi, India.

ABSTRACT

The extensive adoption of computer-based technology in the health-care industry has resulted in the accumulation of electronic data. Due to vast amounts of data, medical specialists are having problems efficiently analysing symptoms and recognising diseases at an early stage. Supervised machine learning (ML) algorithms, on the other hand, have showed potential in outperforming traditional sickness diagnosis methods and enabling doctors in the early detection of high-risk conditions. Through the evaluation of performance indicators, the goal of this paper is to recognise trends across various types of supervised ML models in disease prediction. Django is used to create a disease prediction system based on the best accuracy model.

Keywords: Health Care, Supervised Machine Learning, Diseases Prediction, Django.

I. INTRODUCTION

Nowadays, advanced technologies are employed to make life simpler in practically every sector. Technological improvements have also had a significant influence in the medical industry. The healthcare area is providing better therapies and developing more effective methods to discover and further diagnose disorders. A significant demand for complex technologies such as AI, ML, and Big Data has come from the growing range of illnesses and maladies, as well as the rapid expansion of patients. The implementation of such technology will save a lot of money in the healthcare area while also improving patient care.

Machine learning is the process of programming computers to optimise their performance based on previous data or examples. The study of computer systems that learn from data and experience is known as machine learning. There are two tracks in the machine learning algorithm: training and testing. Prediction of a disease based on the symptoms and medical history of the patient Machine learning technology has been improving for decades. Machine Learning technology provides an immeasurable platform in the medical arena for quickly resolving healthcare challenges.

We are using machine learning to keep entire healthcare data. Physicians can utilise machine learning algorithms to improve patient care by making better decisions regarding diagnostic and treatment options. The most obvious use of machine learning in the medical field is healthcare.

The existing will be done on Support Vector Classifier, Naive Bayes Classifier, and Random Forest Classifier for disease prediction. The prediction system built using Django uses the trained model based on these models.

OBJECTIVE

The purpose of this study is to see if the hypothesis that supervised machine learning algorithms can improve health care through accurately and early disease detection is correct. In this paper, we look into studies that use multiple supervised machine learning models for each disease recognition issue.

Because evaluating the effectiveness of a single method across multiple research settings introduces bias, resulting in imprecise results, this approach provides better comprehensiveness and precision. Machine learning models will be used to analyse a dataset of diverse illness systems. Several techniques will be tested for disease identification, including on Support Vector Classifier, Naive Bayes Classifier, and Random Forest Classifier. The highest performing ML models for each disease will be determined for the Disease Prediction System.

II. EXISTING SYSTEM

Doctors have continued to use technology in a number of ways since the emergence of modern computing, such as surgical representation and x-ray photography, but the technology has mostly stayed unaltered. The technique still requires the doctor's knowledge and experience due to a range of factors such as medical records, weather conditions, atmosphere, blood pressure, and a variety of other considerations. Despite the fact that a large number of variables are supplied as complete variables that are essential to understand the entire

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:04/Issue:05/May-2022 Impact Factor- 6.752 www.irjmets.com

working process, no model has been adequately assessed. Medical decision support technology must be employed to overcome this challenge. This system can help clinicians make the best decision possible.

Machine learning is being used to keep entire hospital data. Machine learning technology allows clinicians to make more informed decisions about patient diagnosis and treatment options. The most prominent example of machine learning in the medical industry is healthcare.

III. PROPOSED SYSTEM

The Disease Prediction System will be used to forecast disease based on symptoms. The best model with the highest confidence and accuracy score is used in this system. Support Vector Classifier, Naive Bayes Classifier, and Random Forest Classifier are used to evaluate the model. This system is used by end users. The technology will use symptoms to anticipate disease. Machine Learning Technology is used in this system. The decision tree classifier method is used to forecast diseases.

This system is known as 'HealthSure.' This system is for people who are always concerned about their health; as a result, we've included various features that recognise them while also increasing their mood.

The system will require the user to first register before being able to utilise the various functions. If the user is a patient, the features include checking the disease via displaying symptoms, consultation information, and a feedback option. Using a machine learning model, the system forecasts the disease. There is also the option of learning more about the ailment. Based on the disease indicated, a list of doctors is also presented. This will allow the patient to receive the essential consultation and assistance sooner. This can be accomplished by speaking with the doctor. As a result, a 'Disease Predictor' tool for health awareness has been developed, which detects disease based on symptoms.

IV. DATASET AND MODEL DESCRIPTION

Data preparation is the initial step in any machine learning project. A dataset will be used in this Disease Prediction System. The dataset has 133 columns, with 132 representing symptoms and the final column showing prognosis. The ML technology will use symptoms to anticipate disease.

The Support Vector Classifier, Naive Bayes Classifier, and Random Forest Classifier algorithms are used to generate the Machine Learning Model. To predict sickness, the Aggregate of the Predictions of Different Classifiers is used. The quality of the models will be determined using a confusion matrix. In the Disease Prediction Web App, the best model from the aforementioned classifier is employed.

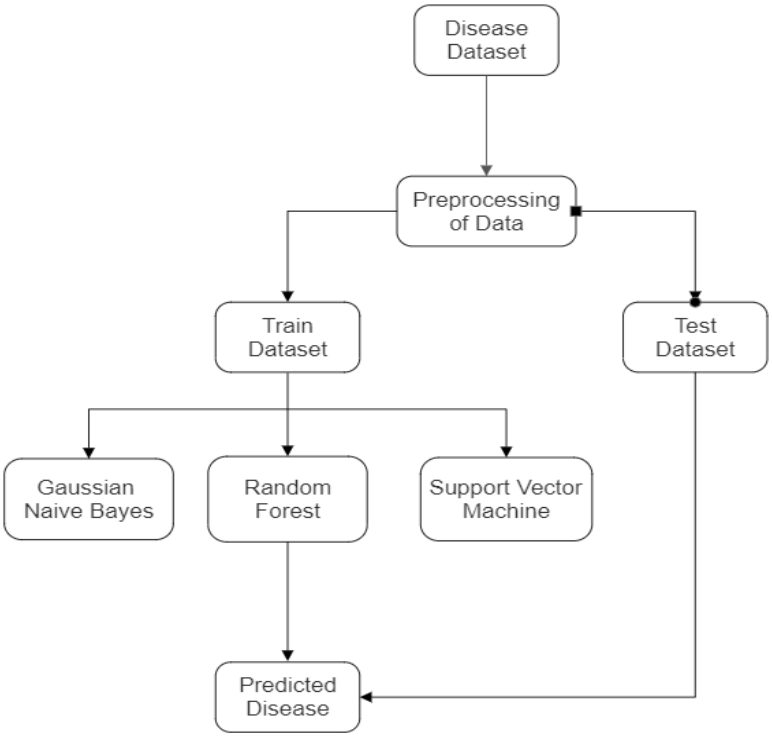


Fig: Model Architecture

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:04/Issue:05/May-2022 Impact Factor- 6.752 www.irjmets.com

V. METHODOLOGY

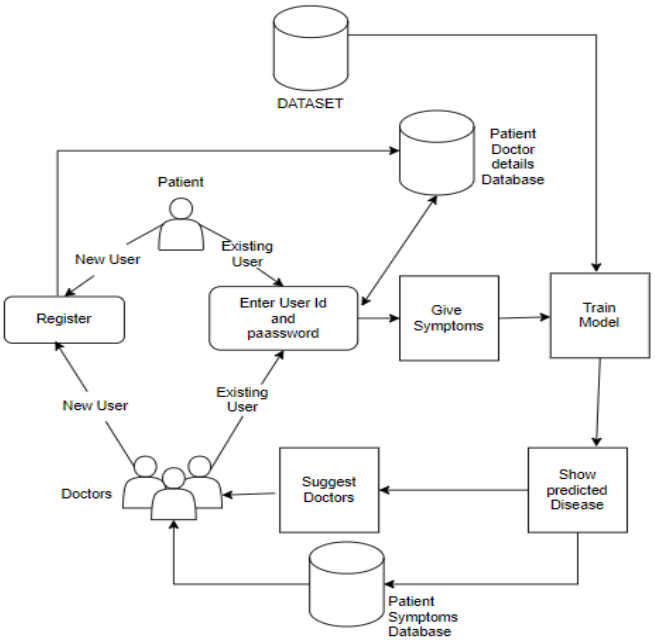


Fig: Process of HealthSure WebApp to predict disease and get consultation from doctor

There are three Modules in our system: Admin, User (Patient), and Doctor. Every new user must first register. After successfully registering, the user must enroll before logging in. Users will only need to enroll once. The disease prediction system comprises three users: a doctor, a patient, and an administrator.

- The system authenticates each user of the system.
- There is a role-based access to the system.
- The Web App allows the patient to provide symptoms, and the machine predicts a condition based on those symptoms.
- The Web App suggests doctors for predicted diseases.
- Patients can consult with the system online.
- The Web App allows people to consult the doctor at their leisure while sitting at home.

VI. ALGORITHM USED

a) Naïve Bayes Classifier

The supervised machine learning classification strategy is represented by the Naive Bayes algorithm. It uses a probabilistic model to determine the likelihood of various outcomes/outputs. It is employed in analytical and predictive tasks. The Naive Bayes algorithm is resistant to noise in the input dataset. Below is an example of a Naive Bayes implementation.

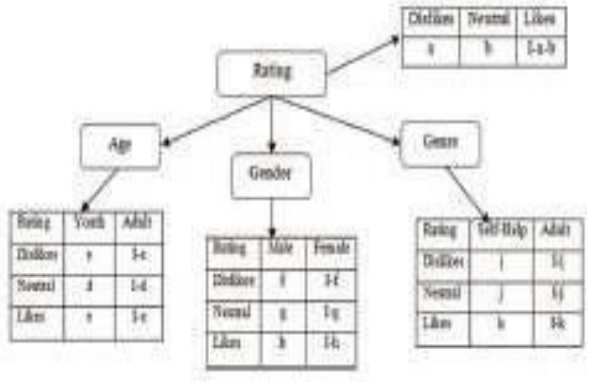


Fig: Naïve Bayes Classifier

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:04/Issue:05/May-2022 Impact Factor- 6.752 www.irjmets.com

b) Random Forest Algorithm

The Random Forest algorithm is modelled by the tree's algorithm and the bagging method. The algorithm's creators discovered that it has the potential to increase categorization accuracy. It can also handle a huge number of input variables in a data collection. As shown in figure below, the method begins by constructing a network of trees, each of which will vote for a class. The model for the Random Forest is shown in the diagram.

Assume that a data set has N data and M input variables, and that the real data utilised in this paper is made up of both data and input variables. Let k represent the number of sampling groups, and ni and mi represent the number of data and variables in group I where I is 1, 2,..., and k.

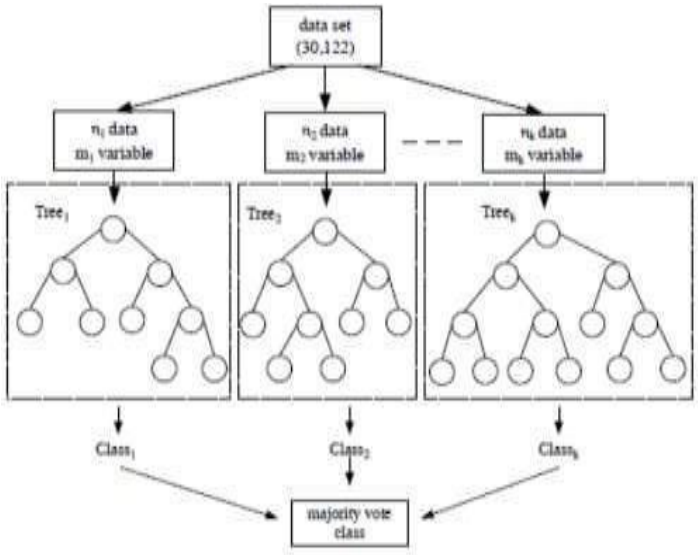


Fig: Random Forest

c) Support Vector Machine

SVM is a supervised machine learning technique that may be used for both classification and regression. Though we might also argue regression difficulties, categorization is the best fit. In an N-dimensional space, the SVM algorithm aims to find a hyperplane that clearly categorises data points. The dimension of the hyperplane depends upon the number of features. If there are only two input characteristics, the hyperplane is just a line. If there are only two input characteristics, the hyperplane is just a line. When there are three input features, the hyperplane becomes a two-dimensional plane. When the number of features exceeds three, it becomes impossible to imagine. Below figure show

Linear SVM classifier.

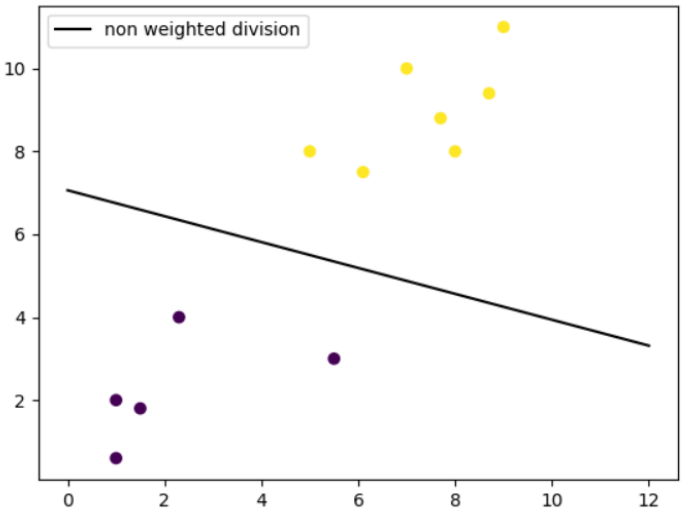
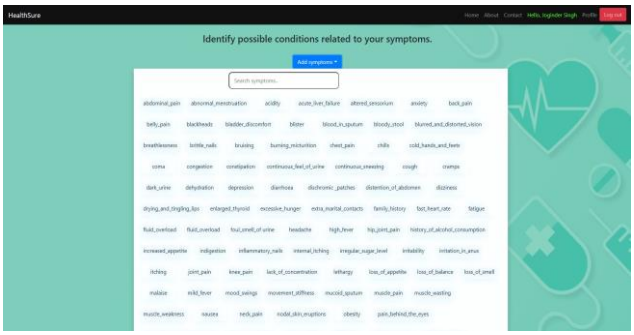


Fig: Linear SVM classifier

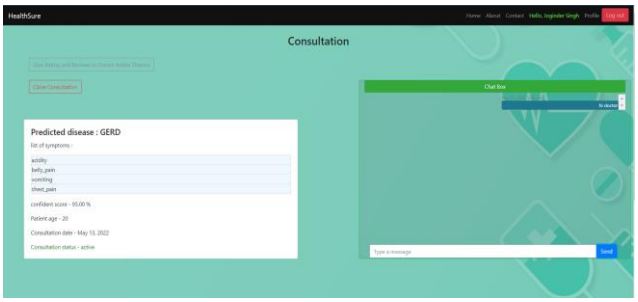
e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:04/Issue:05/May-2022 Impact Factor- 6.752 www.irjmets.com

VII. SCREENSHOTS



Patient will select the symptoms from the dropdown.

System will predict the disease with respect to the symptoms provided by user.



Patients may have a live chat or consultation with doctors whenever it is convenient for them in order to have their health concerns properly diagnosed and treated with correct care and prescriptions.

VIII. RESULT ANALYSIS

K fold cross validation is used to compare the performance of all the algorithms. It shows that the Nave Bayes and Random Forest approaches are slightly more accurate than the other two. The accuracy results after testing the model of 41 diseases are as follows:

0.95 for Naive Bayes

0.95 for Random Forest

0.96 for Support Vector Machine

Confusion matrix revealed that out of 41 diseases, 38 were correctly classified by Random Forest and Naive bayes and 39 were correctly classified by Support Vector Machine.

The percentage accuracy can reach 96 percent. As a result, the system employs the most accurate trained model of Support Vector Machine. The final results and confidence score are shown.

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:04/Issue:05/May-2022 Impact Factor- 6.752 www.irjmets.com

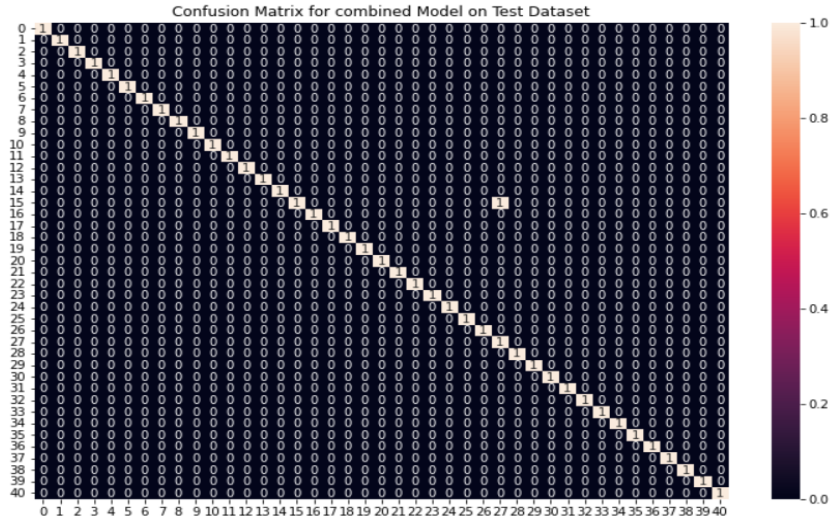


Fig: confusion matrix for combined model

IX. CONCLUSION

The accuracy, confusion matrix, and other metrics are used to assess performance. A comparison of the accuracies of each algorithm, such as Random Forest, Support Vector Machine, and Naive Bayes, was made, and it was found to be around 96 percent.

As a result, we came to the conclusion that our method had a higher accuracy in disease prediction. Comparisons show that SVM has slightly higher accuracy. As a result, the model has been appropriately trained, and its use in this system will benefit those patients who are constantly concerned about their health and want to know what is going on with their body.

The primary goal of developing this system is to assist such persons with their health. This system may also be utilised by small-scale doctors or dispensaries to forecast sickness and lessen the rush at hospital emergency rooms, as well as the stress on medical staff. The makes available the doctor's list of that specific projected condition in order to obtain immediate appointments, which benefits both patients and hospitals. This guarantees that the system does not interfere with the doctor's profession and maintains patient safety while increasing patient gain in the prediction system.

X. FUTURE WORK

Future work will primarily focus on delivering medical support and proper dose to patients as quickly as possible in order to establish the greatest infrastructure and the quickest and easiest path in the medical profession.

ACKNOWLEDGEMENT

The authors would like to thank the Department of computer science and engineering of HMRITM for supporting this research.

XI. REFERENCES

[1] B. Nithya and V. Ilango,"Predictive analytics in health care using machine learning tools and techniques," 2018

[2] Vijaya Shetty, Kathik GA, M Ashwin, "Symptom Based Health Prediction using Data Mining," 2019

[3] P. S. Kohli and S. Arora, "Application of Machine Learning in Disease Prediction," 2018 4th International Conference on Computing Communication and Automation (ICCCA), Greater Noida, India, 2018, pp. 1-4, doi: 10.1109/CCAA.2018.8777449.

[4] S. Uddin, A. Khan, M. E. Hossain, and M. A. Moni, "Comparing different supervised machine learning algorithms for disease prediction," BMC Medical Informatics and Decision Making, vol. 19, no. 1, pp. 1–16, 2019.

[5] A. Gavhane, G. Kokkula, I. Pandya, and K. Devadkar, "Prediction of heart disease using machine learning," in 2018 Second International Conference on Electronics, Communication and Aerospace Technology (ICECA), 2018, pp. 1275–1278.

[6] R. Saravanan and P. Sujatha, "A state of art techniques on machine learning algorithms: A perspective of supervised learning approaches in data classification," in 2018 Second International Conference on Intelligent Computing and Control Systems (ICICCS), 2018, pp. 945– 949.

[7] Aravindkumar.S, Ajith.A, "Prediction of Chronic Disease by Machine Learning," 2019.

[8] Sneha Grampurohit, Chetan Sagarnal, "Disease Prediction using Machine Learning Algorithms," 2020.

[9] Dhiraj Dahiwade, Gajanan Patle, Ektaa Meshram, "Designing Disease Prediction Model Using Machine Learning Approach," 2019.

[10] R.Jane Preetha Princy, Saravanan Parthasarathy, P. Subha Hency Jose, " Prediction of Cardiac Disease using Supervised Machine Learning Algorithms," 2020.