iJRASET
International Journal For Research in Applied Science and Engineering Technology



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 10 **Issue:** XI **Month of publication:** November 2022

DOI: https://doi.org/10.22214/ijraset.2022.47684

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 10 Issue XI Nov 2022- Available at www.ijraset.com

Smart System for Disease Prognostication in Healthcare using multi-classifier-based Prediction Model

Sakshi Tekale[1], Saee Kulkarni[2], Ramdas Patil[3], Shreya Diwan[4], Prof. Anita Vikram Shinde[5]
Computer Engineering Department, Marathwada Mitra Mandal's College of Engineering, SPPU, Pune, India

Abstract: ***In today's world, machine learning and Artificial Intelligence are playing a crucial role. We can find use cases of ML and AI everywhere. Starting from self-driving cars like Autopilot-Tesla to fields like medical, AI and ML having many advances. There are approaches which are generally helpful for healthcare and biomedical sectors for predicting diseases. The proposed work is to develop and deploy such a model for prognostication of diseases like Diabetes, Chronic, Heart, Liver, Malaria, Pneumonia. The proposed model is divided into 3 phases: 1) Data Normalization 2) Feature Extraction and (c) Prediction. For the process of model building, the datasets are taken from Kaggle datasets and UCI ML repository. By following all the three phases, to make the attribute's range at a certain level, normalization of data takes place. Then, the other two phases take place. Lower accuracy of the model can be fatal for patients sometimes so here the model is designed to predict the outcomes accurately. The application of the model includes entering parameters or symptoms in the system and getting the outcome whether the result is positive or negative. This model will help people to monitor their health regularly.***
Keywords: Healthcare, Prognostication, multi-classifier, Heart, Diabetes, Cancer, Prediction, Model

I. INTRODUCTION

Nowadays, human life is full of uncertainties. Many people suffer from health issues due to the late identification of diseases. The “Disease Prediction using Machine Learning” system is developed to identify general diseases in earlier stages. People don’t give attention to their health and ignore the symptoms, which leads to various health problems. According to research, 40% of people ignore the symptoms, because of fear of facing financial problems or other reasons. Many cannot afford the treatment or some are very busy with a tight schedule, but ignoring the constantly reflecting symptoms for a long period may have dangerous consequences for their health. According to research, 70% of Indians suffer from common diseases, and among them, 25%, ignore them in the early stages. The main motive to develop this project is that a user can conveniently have a check-up of their health if they have any of the symptoms [21].

Machine Learning has been one of the most used tools in recent times. It has helped further the progress of the healthcare industry smoothly. With such a tool, medical professionals and researchers have been able to diagnose and detect diseases with much accuracy and in less time. It has contributed to saving many lives [14].

Logistic regression (LR) analysis has become an increasingly used statistical tool in medical research [20]. Liver disease is one of the diseases which cause millions of deaths every year. Viral hepatitis leads to 1.34 million deaths every year. Problems with liver patients are not easily diagnosed in an early stage as it will be functioning normally even when it is partially damaged. Diagnosis of liver problems at the starting phase will increase the patient’s survival rate. Liver failures are at a high rate of risk among Indians [01]. Another fatal disease is pneumonia which is caused by a bacterium and is a very dangerous disease that infects the lungs and in turn affects the normal respiration of the human. This infection is life-threatening, especially to infants and aged people over 60. In India the cases recorded with pneumonia are 10 million per year. Hence, the early prediction of the disease is very important [06] The most dangerous problem in this decade is heart-related diseases. It mainly occurs due to the consumption of products such as tobacco and alcohol that have a damaging effect on the heart and overall health [14].

Diabetes is a chronic disease and it is now a worldwide healthcare crisis. According to the International Diabetes Federation, 382 million people are suffering from diabetes across the whole world. By 2035, this will be doubled and go to 592 million. Diabetes is caused due to an increased level of blood glucose. This high blood glucose level in the blood produces the symptoms of frequent urination, increased thirst, and increased hunger. Diabetes is one of the major causes of blindness, kidney failure, amputations, heart failure, and stroke.

©IJRASET: All Rights are Reserved

The basic outline of the paper includes different sections. Section one describes the related work. Section two highlights the proposed system architecture which further includes detailed description of the system design and analysis of algorithm. The paper ends with section of conclusion followed by references.

II. RELATED WORK

Sr. No.	Author	Title	Description
1	KM Jyoti Rani [2]	Diabetes Prediction Using Machine Learning	The algorithms like KNN, LR, RF, SVM and DT are used for prediction of diabetes.
2	Prof. Swati Dhabarde, Rohit Mahajan, Satyam Mishra, Sanjiv Chaudhari, Satish Manelu, Prof. Dr. N S Shelke [11]	Disease Prediction using Machine Learning Algorithms	Structured and unshaped data were used for disease prediction.
3	Harshit Jindal, Sarthak Agrawal, Rishabh Khera, Rachna Jain and Preeti Nagrath 13]	Heart disease prediction using machine learning algorithms	Prediction of heart diseases is done by using KNN, RF and LR algorithms
4	Aadar Pandita, Sarita Yadav, Siddharth Vashisht, Aryan Tyagi [14]	Review Paper on Prediction of Heart Disease using Machine Learning Algorithms	It gives an overview on recent research in the prediction of heart disease using different machine learning algorithms.
5	Noreen Fatima, Li Liu, Hong Shai and Haroon Ahmad [22]	Prediction of Breast Cancer, Comparative Review of Machine Learning Techniques and their Analysis	It provides the comparative analysis of ML, DL and data mining algorithms used for breast cancer prediction

Table 1: Related Work

[2] This paper proposed complete study of the diabetes dataset with comparative study of all algorithms like K-nearest neighbour, Logistic Regression, Support Vector Machine, Decision Tree and Random Forest on diabetes. The paper further gives training and testing accuracies of comparative algorithms.
[11] This paper has introduced the disease prediction system using machine learning algorithms like Decision Tree, Logistic Regression, Naive Bayes, Random Forest, SVM. From these algorithms Decision Tree had given the maximum accuracy.
[13] This project predicts cardiovascular disease by extracting the patient's medical history that leads to a fatal heart disease from a dataset that includes patients' medical history. This project focuses on mainly three data mining techniques namely: (1) Logistic regression, (2) KNN and (3) Random Forest Classifier.
[14] This paper proposed the complete study of Heart Disease and comparative study of algorithms like Generalised Linear Model, Decision Tree, Random Forest, Support Vector Machine, Neural Networks and K-nearest-Neighbour.
[22] This paper provides the comparative analysis of different machine learning, deep learning and data mining algorithms for the prediction of breast cancer. The main focus of the paper is to find out the most suitable algorithm that can predict the occurrences of breast cancer more effectively.

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

III. PROPOSED SYSTEM ARCHITECTURE

As the model is of multiple disease prediction, it is possible to predict more than one disease at a time. Users do not need to search for more than one website or application for diagnosing multiple diseases. With the model, the user is also getting the facility of chatbot and doctor recommendations.

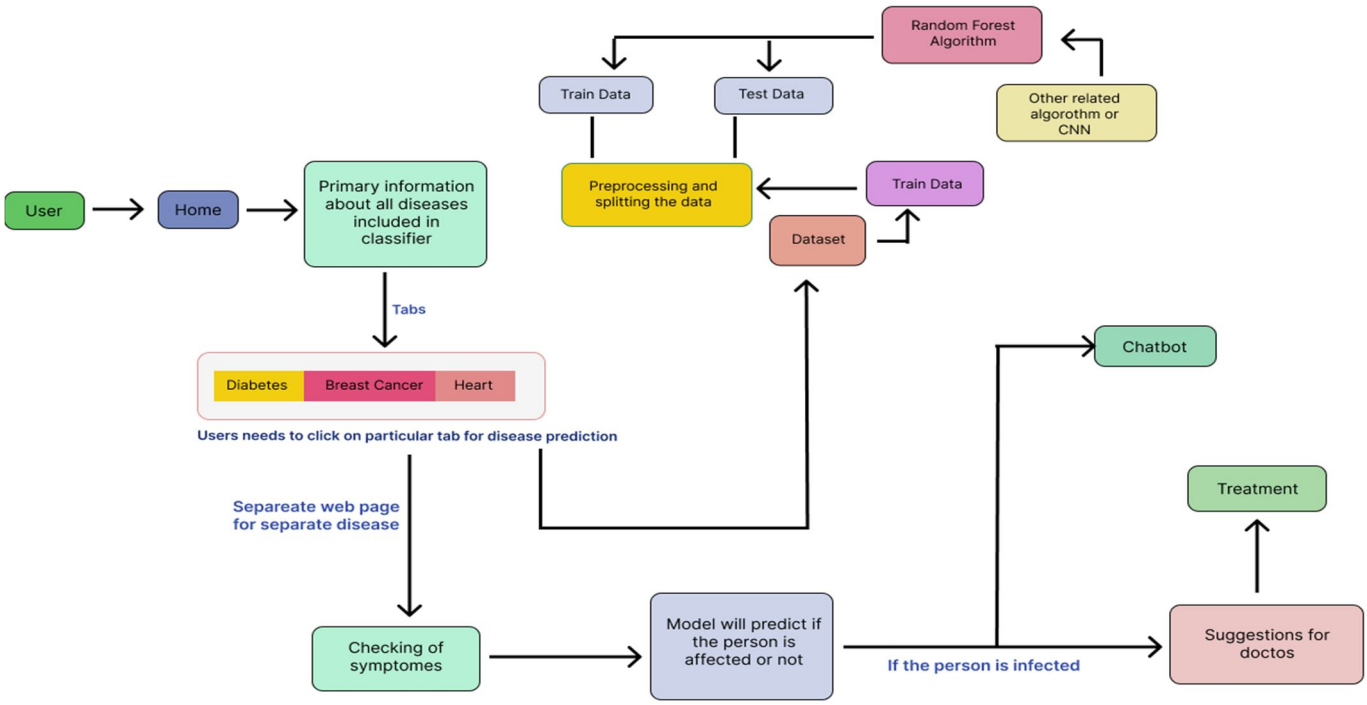


Fig 1. Proposed System Architecture

The general flow of model includes home screen which contains the user login functionality and primary information of the diseases with the tabs for each and every disease. Every tab is linked with a separate page for the user to put all the symptoms and necessary info in the form structure. The tab contains About page, Diabetes, Cancer and Kidney named components which are further linked with separate HTML page. In the backend model will work with all the pre-processing, splitting data, classification and regression steps and predict if the person is affected or not accurately. If the person wants any help related to doctors, there is functionality of doctor's suggestions with chatbot.

Here, for the predictions, we have chosen the best fit algorithm for each disease. We have worked with algorithms like Logistic Regression (LR), Random Forest (RF), K-Nearest Neighbours (KNN), Decision Tree (DT) and Support Vector Machine (SVM).

Accuracy of each disease with each algorithm is given by:

	Diabetes	Heart Disease	Cancer
k-Nearest Neighbours	83.33%	86.03%	93.57%
Logistic Regression	88.16%	88.31%	95.1%
Decision Tree	89.47%	97.72%	92.98%
Random Forest	92.54%	98.70%	95.32%
SVM	83.33%	71.75%	94.66%

Table 2 : Comparison of Algorithms

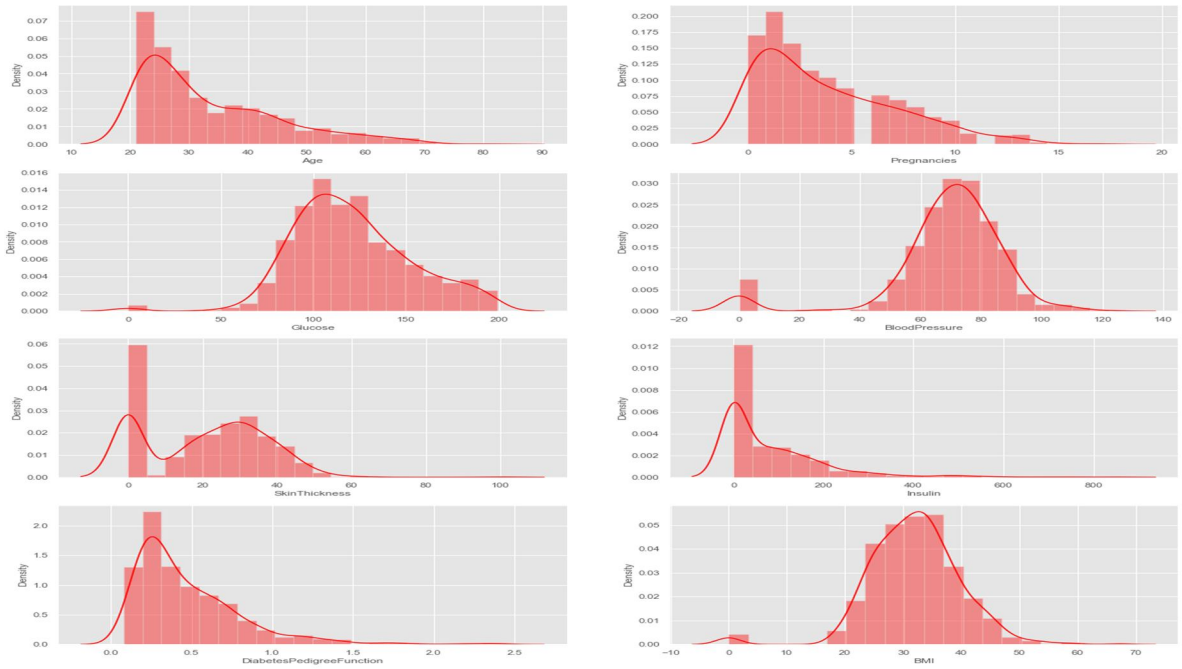


Fig 2 : Density graph of all variables for Diabetes

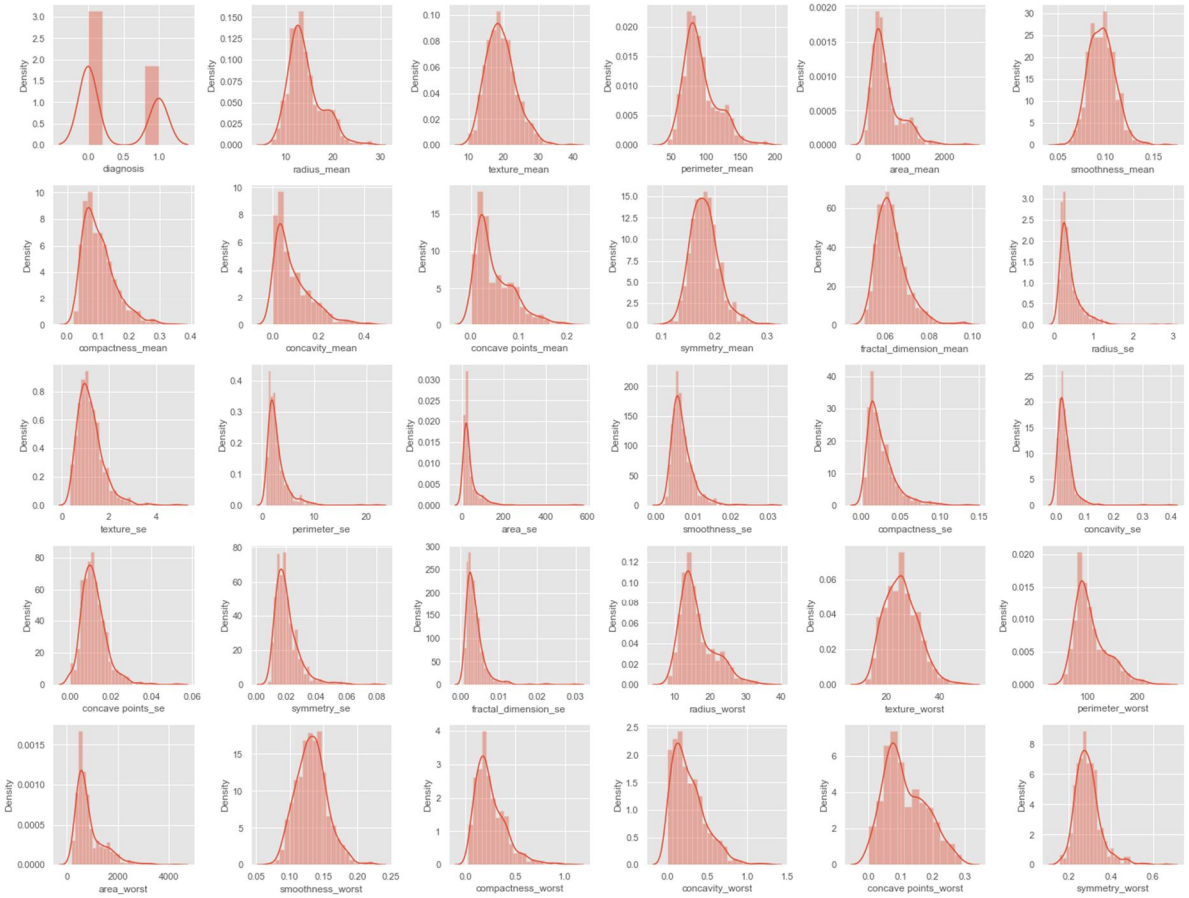


Fig 3 : Density graph of all variables for Cancer

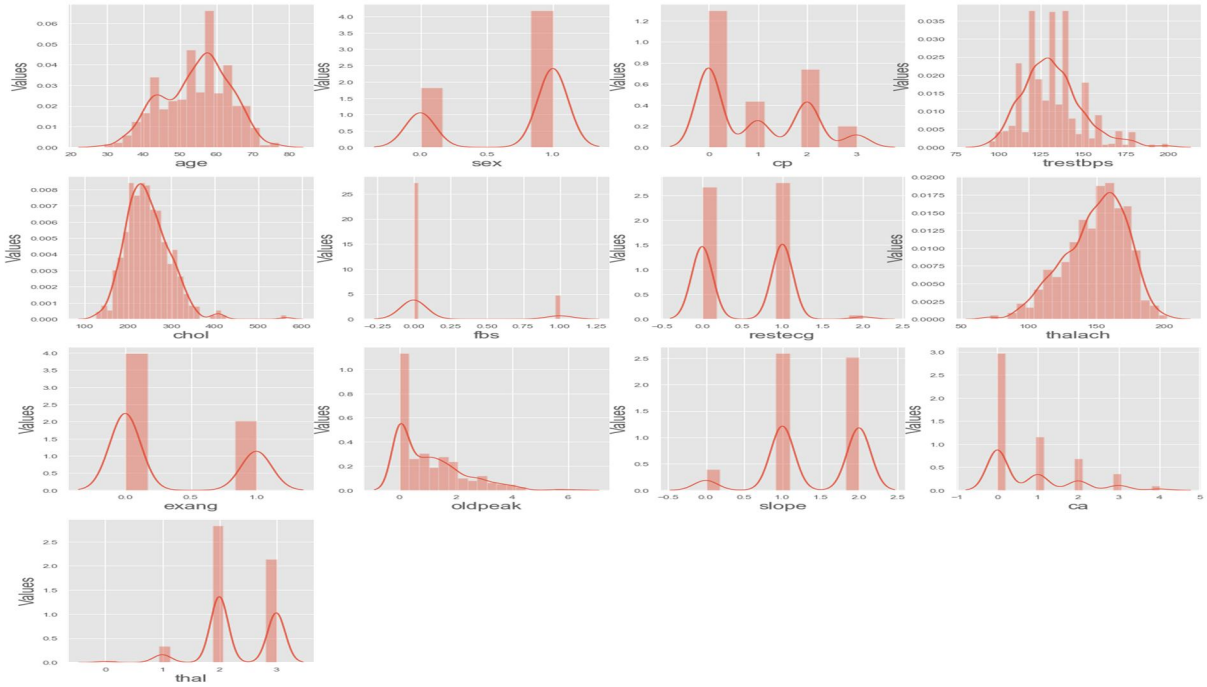


Fig 4 : Density graph of all variables for Cancer

IV. ANALYSIS OF ALGORITHM

Random Forest is a supervised machine learning algorithm. From the name itself we are able to know that, the Random Forest technique produces multiple decision trees before giving an output. This technique is based on the thing that more number of trees would tend to the right decision. For classification, it uses a voting system and then decides the class whereas in regression it takes the mean of all the outputs of the decision trees. It works well with large datasets with high dimensionality [15].

Random forest algorithm steps

1) Step-1: Load the respective datasets.
2) Step 2: Select random K data points from the training set.
3) Step 3: Build the decision trees by using the selected data points (Subsets).
4) Step 4: Choose N decision trees that we want to build.
5) Step 5: Repeat Steps 2 & 3.
6) Step-6: For new data points, find the predictions of each decision tree, and assign the new data points to the respective category.

Random forest is an ensemble algorithm which makes it very popular. Random forests predict the class of the dataset by combining multiple trees. It takes the average of all the trees and gives us the correct output. It takes less time in training and gives us high accuracy even when the dataset is large. It can maintain accuracy when a large proportion of data is missing

To visualize the speed tests, we need to understand the variables on which the speed depends:

a) *n_jobs:* Keeping n-jobs as -1 results in training many trees simultaneously.
b) *n_estimators:* It refers to the number of a tree to be trained. Its default value is 100 but it can be lowered to improve the final results
c) *Bootstrap:* Bootstrapping refers to creating various subsets of tree training samples. If this step is ignored, it results in low performance.
d) *Max_features:* Features could be limited and still have flexibility in a dataset.

Overall, we can conclude that random forest is a useful and easy training algorithm. We can achieve accuracy and performance for smaller datasets and larger datasets in a restricted format. To get absolute accuracy, we cannot go wild with the random forest, but with few restrictions, we get better results than any other algorithms out there. Also, decision trees have linear complexity, hence are pretty fast and accurate in nature.

The time complexity of the random forest algorithm is O(tf*nlogn)
where f=no.of features ,t=no of trees. This is polynomial in nature.

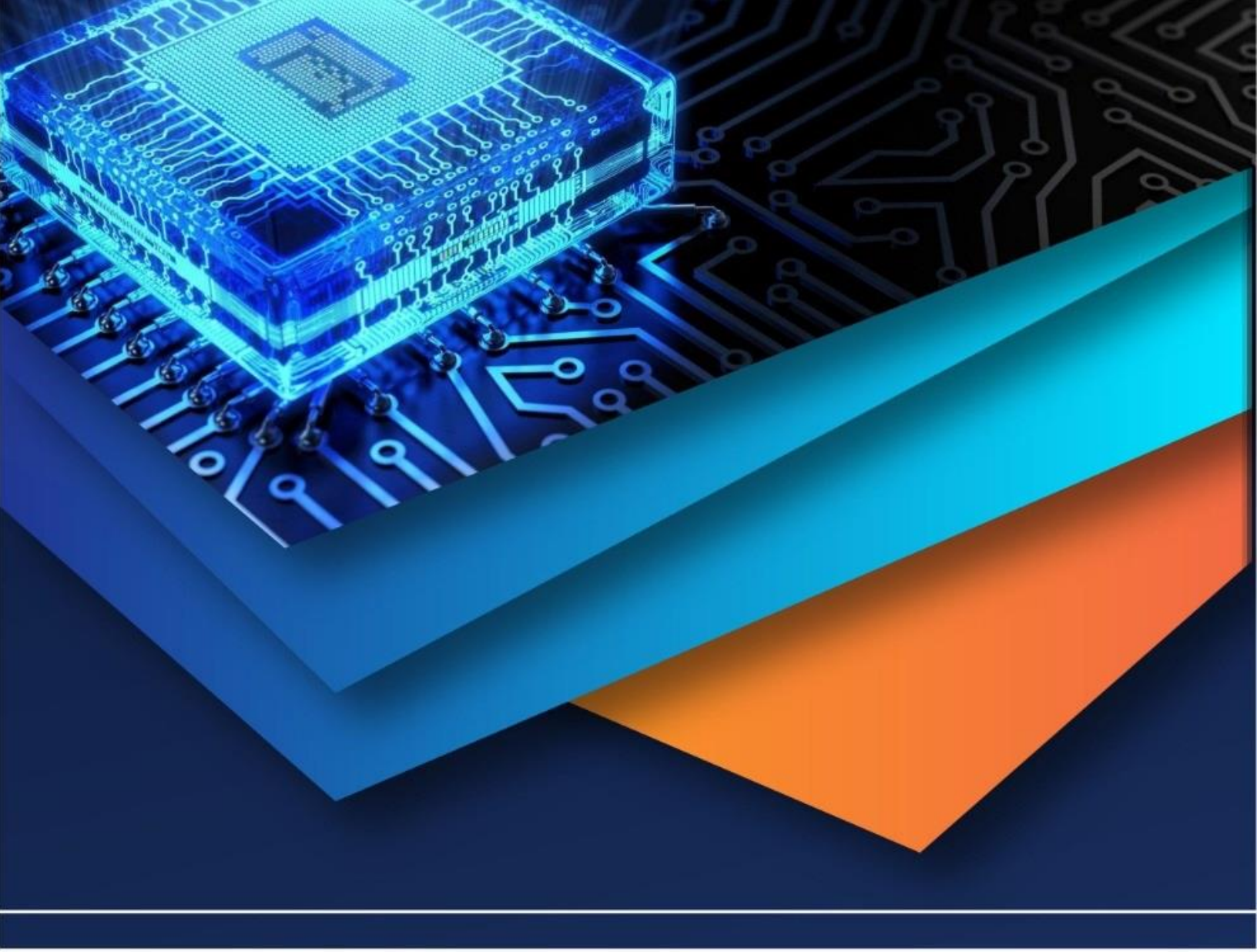
V. CONCLUSION

The main goal of this paper is to predict the disease in accordance with symptoms put down by the patients with proper implementation of Machine Learning algorithm. The system analyses the symptoms provided by the user as input parameters and predicts the disease as a result. Disease Prediction is done by implementing the Decision tree Classifier. Decision tree Classifier calculates the probability of the disease. The proposed system can help to reduce the risk of diseases by diagnosing them earlier and also reduces the cost for diagnosis, treatment, and doctor consultation. As the system is web based, users can access it from anywhere and anytime.

In future, developing more effective Machine Learning algorithms is much needed to increase the efficiency of disease prognostication. Learning models should be calibrated often after the training for a better performance. To improve the performance of learning models more relevant feature selection methods should be used.

REFERENCES

[1] Rakshith D B, Mrigank Srivastava, Ashwani Kumar, Gururaj S P, liver disease prediction system using machine learning techniques, doi:10.17577/ijertv10is060460,ijertv10is060460, volume 10, issue 06 (june 2021)
[2] Km Jyoti Rani, diabetes prediction using machine learning, doi:10.32628/cseit206463, july 2020, international journal of scientific research in computer science engineering and information technology
[3] c. boukhatem, h. y. youssef and a. b. nassif, "heart disease prediction using machine learning," 2022 advances in science and engineering technology international conferences (aset), 2022, pp. 1-6, doi: 10.1109/aset53988.2022.9734880.
[4] imesh udara ekanayake,damayanthi herath, chronic kidney disease prediction using machine learning methods, doi: 10.1109/mercon50084.2020.9185249, 2020 moratuwa engineering research conference (mercon)
[5] kumar bmh, srikanth pc, vaibhav am. a novel computation method for detection of malaria in rbc using photonic biosensor. int j inf technol. 2021;13(5):2053-2058. doi: 10.1007/s41870-021-00782-z. epub 2021 sep 2. pmid: 34493995; pmcid: pmc8412864..
[6] amulya m p,niranjan murthy m, prediction of pneumonia using big data, deep learning and machine learning techniques, doi: 10.1109/icces51350.2021.9489188, 2021 6th international conference on communication and electronics systems (icces)
[7] s. grampurohit and c. sagarnal, "disease prediction using machine learning algorithms," 2020 international conference for emerging technology (incet), 2020, pp. 1-7, doi: 10.1109/incet49848.2020.9154130.
[8] l. athota, v. k. shukla, n. pandey and a. rana, "chatbot for healthcare system using artificial intelligence," 2020 8th international conference on reliability, infocom technologies and optimization (trends and future directions) (icrito), 2020, pp. 619-622, doi: 10.1109/icrito48877.2020.9197833.
[9] marouane ferjani,disease prediction using machine learning,doi:10.13140/rg.2.2.18279.47521
[10] bhanuteja talasila,saipoornachand kolli,symptoms based multiple disease prediction model using machine learning approach,august 2021 international journal of innovative technology and exploring engineering ,doi: 10.35940/ijitee.i9364.0710921
[11] prof. swati dhabarde, rohit mahajan, satyam mishra, sanjiv chaudhari, satish manelu, prof. dr. n s shelke , disease prediction using machine learning algorithms , volume:04/issue:03/march-2022 ,international research journal of modernization in engineering technology and science
[12] a. kp and j. anitha, "plant disease classification using deep learning," 2021 3rd international conference on signal processing and communication (icpsc), 2021, pp. 407-411, doi: 10.1109/icspc51351.2021.9451696.
[13] harshit jindal, sarthak agrawal, rishabh khera, rachna jain and preeti nagrath,heart disease prediction using machine learning algorithms,doi 10.1088/1757-899x/1022/1/01207,1st international conference on computational research and data analytics (iccrda 2020) 24th october 2020, rajpura, india
[14] aadar pandita,sarita yadav,siddharth vashisht,aryan tyagi,review paper on prediction of heart disease using machine learning algorithms,doi: 10.22214/ijraset.2021.35626,june 2021
[15] baban uttamrao rindhe,nikita ahire,rupali patil,heart disease prediction using machine learning,doi: 10.48175/ijarsct-1131,international journal of advanced research in science, communication and technology (ijarsct) volume 5, issue 1, may 2021
[16] revathy ramesh,chronic kidney disease prediction using machine learning models,doi: 10.35940/ijeat.a2213.109119 , may 2020 international journal of engineering and advanced technology 9(1):6364
[17] thirunavukkarasu kannapiran,ajay shanker singh,abhishek chowdhury , prediction of liver disease using classification algorithms , doi: 10.1109/ccaa.2018.8777655 , conference: 2018 4th international conference on computing communication and automation (iccca)
[18] v. sirish kaushik,gaurav kataria,anand nayyar,rachna jain , pneumonia detection using convolutional neural networks (cnns) , doi: 10.1007/978-981-15-3369-3_36 , conference: proceedings of first international conference on computing, communications, and cyber-security (ic4s 2019)
[19] sounak chakraborty,mihail popescu, predicting disease risks from highly imbalanced data using random forest , doi:https://doi.org/10.1186/1472-6947-11-51 , bmc medical informatics and decision making volume 11, article number: 51 (2011)
[20] daniel a. abaye ,ernest yeboah boateng , a review of the logistic regression model with emphasis on medical research , doi: 10.4236/jdaip.2019.74012 , january 2019 journal of data analysis and information processing 07(04):190-207
[21] anjali bhatt, shruti singasane, neha chaube, disease prediction using machine learning , volume:04/issue:01/january-2022 , international research journal of modernization in engineering technology and science
[22] noreen fatima, li liu, hong shai and haroon ahmad, prediction of breast cancer, comparative review of machine learning techniques and their analysis doi: 10.1109/access.2020.3016715, ieee access, august 2020



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)