iJRASET
International Journal For Research in Applied Science and Engineering Technology



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 11 **Issue:** IV **Month of publication:** April 2023

DOI: https://doi.org/10.22214/ijraset.2023.50128

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 11 Issue IV Apr 2023- Available at www.ijraset.com

Multi-Disease Prediction Using MachineLearning Algorithm

Dr Visumathi J[1], Tetala Durga Venkata Rama Reddy[2],Velagapudi Abhinandhan[3], Panamganti Anil Kumar[4]
[1]*Professor, Department of CSE,* [2,3,4]*Students, Department of CSE Vel Tech Rangarajan Dr. Sagunathala R & D Institute of Science and Technology*

Abstract: ***In the medical sector, disease diagnosis is an essential duty, and prompt and accurate diagnosis is crucial to effective management and therapy. Machine learning techniques, including Naive Bayesian networks, have shown promise in disease prediction and diagnosis. In this study, we present a machine learning-based multi-disease prediction system that uses Naive Bayesian networks.***
The proposed methodology seeks to deliver precise illness prediction for several diseases instantaneously. In addition to describing the methods adopted, which included dataset selection, preprocessing, feature selection, and the Naive Bayesian network algorithm, we also discuss the social relevance of this work, emphasizing the potential impact of accurate disease prediction in improving patient outcomes and bringing down healthcare costs.
To evaluate the performance of the proposed model, we conducted experiments using a publicly available disease dataset. The results demonstrated that the proposed model achieved high accuracy of 91.2% and outperformed other state-of-the-art models for multi-disease prediction some of them are, Random Forest obtained 85.7% and Decision Tree obtained 81.3% respectively.
In summary, the proposed system demonstrates the effectiveness of Naive Bayesian networks for multi-disease prediction and has the potential to improve disease diagnosis and management in the medical domain.
Keywords: Disease Prediction, Naïve BayesianNetworks, Random Forest, Decision Tree,Feature Selection.

I. INTRODUCTION

The adoption of modern computational techniques, such as machine learning, has drawn a lot of fascination in recent years since illness detection is a challenging process with many facets in the healthcare industry. Healthcare practitioners may now make educated decisions and deliver improved patient care courtesy of machine learning algorithms that have demonstrated enormous opportunities in illness diagnosis and prediction. In this respect, we suggest a novel method for multi-disease prediction based on Naive Bayesian networks, a probabilistic modeling approach that is extensively employed in many industries, including the healthcare industry.

Using a sizable dataset of patient data, our suggested technique intends to deliver an extensive and precise illness prediction model for several diseases at once. To improve patient outcomes, lower healthcare expenditures, and raise the general standard of healthcare services, such a system must be developed. Accurate illness prediction is made possible by the system's use of cutting-edge algorithms for data preprocessing, feature selection, and model training.

The approach utilized to create the suggested system is described in this study, along with the necessary dataset selection procedures, preprocessing techniques, feature selection methods, and application of the Naive Bayesian network algorithm. We also provide the findings from our research, emphasizing how precisely and successfully the suggested approach can diagnose numerous illnesses at once. The suggested approach, which makes use of machine learning algorithms to enable precise and prompt illness detection, also represents an important leap in disease prediction and diagnosis overall. The system has the potential to dramatically enhance patient care and treatment results, hence enhancing the healthcare sector as a whole.

A. Objectives

Using Naive Bayesian networks, a well-liked probabilistic modeling approach in machine learning, the objective of this work is to build and create a multi-disease prediction system. The goal is to give medical practitioners an effective tool that enables them to correctly forecast the onset of several illnesses in patients based on their symptoms and medical background. The intention of this research is to train the Naive Bayesian network algorithm to concurrently forecast various illnesses using a big dataset of patient information and sophisticated data preparation, data cleaning, normalization, and feature selection techniques, after all the processing the results are produced.

©IJRASET: All Rights are Reserved

B. Purpose and Scope

1) Purpose

There are several methods for disease prediction. However, heart-related diseases have been studied, and a risk level has been calculated. However, such methods are not commonly used for disease prediction in general. As a result, the smart healthcare system aids in the prediction of general diseases. In certain cases, you or a family member may need urgent medical assistance, but physicians are unavailable due to unforeseen circumstances, or we maybe unable to locate the appropriate doctor for the care. To address this problem, we will attempt to incorporate an online intelligent Smart Healthcare System in this project. It's a web-based program that allows patients to get immediate advice about their health problems.

2) Scope

The importance of this research rests in its potential to considerably increase the precision and effectiveness of medical diagnosis in contemporary culture. Healthcare workers may choose better patient care and treatment choices by employing machine learning algorithms like Naive Bayes networks to forecast different illnesses based on patient symptoms. In the end, this may result in improved patient outcomes and lower medical expenses. The method might also be used to deliver a prompt and precise diagnosis in disadvantaged regions with limited access to healthcare. The prospective outcomes of this effort are broad-ranging and might have a favorable effect on the healthcare system as well as society at large.

3) Limitations

The Smart Healthcare System has the following limitations:

a) The accuracy of the system is dependent on the quality and completeness of the dataset used for training. Incomplete or inconsistent data may lead to inaccurate predictions.

b) The system may not be effective in predicting rare or emerging diseases that have not been included in the training dataset.

c) The system may not be able to handle complex cases, where patients have multiple diseases or conditions.

d) The system may not be able to replace the expertise and judgment of healthcare professionals entirely, as the interpretation of the results requires clinical knowledge and experience.

e) The system may not take into account external factors, such as environmental or social determinants of health, that can affect disease occurrence and progression.

II. LITERATURE REVIEW

Sameer Meshram1, et al. 2022 [1], In this paper, he developed a diagnosis system with machine learning algorithms for the prediction of any disease that can help in a very more accurate diagnosis than the traditional method. The proposed model is a Disease Prediction System with the help of the machine learning algorithm Naïve Bayes which takes the symptoms as the input and it gives the output as a predicted disease. It results in saving time and also makes it easy to induce a warning about your health before it's too late.

Abid Ishaq, et al. 2021 [2], In this paper, the author identified key characteristics and efficient data mining strategies that can improve the predictability of cardiovascular patient survival. Nine categorization methods are used in this study to forecast patient survival. Synthetic Minority Oversampling Method addresses the issue of class imbalance. When considering the whole set of characteristics, machine learning algorithms provide results that are compared to those produced by trained machine learning models. Experiments show that ETC works better than other models and reaches the accuracy value with SMOTE in predicting the survival of cardiac patients.

Bilal Khan, et al. 2020 [3], In this paper, the author employed experiential analysis of ML techniques for classifying the kidney patient dataset as CKD or NOTCKD. Seven ML techniques together with NBTree, J48, Support Vector Machine, Logistic Regression, Multi-layer Perceptron, Naïve Bayes, and Composite Hypercube on Iterated Random Projection (CHIRP) are utilized and assessed using distinctive evaluation measures such as mean absolute error (MAE), root means squared error (RMSE), relative absolute error (RAE), root relative squared error (RRSE), recall, precision, F-measure and accuracy.

Chandrasekhar Rao Jetti, et al. 2021 [4], The author developed this work mainly to make doctors' jobs easier by using a machine to examine a patient at a basic level and recommend diseases that may be present.

It begins by inquiring about the patient's symptoms; if the device can determine the relevant condition, it then recommends a doctor in the patient's immediate vicinity.

The system will show the result based on the available accumulated data.

Selvaraj, et al. 2021 [5], In this paper, the author extracted personal data such as user health conditions from day-to-day life. The lifestyle data are gathered and stored at the data repository by using web technology and mobile applications. The user enters their daily health conditions in textual format.

Natural Language Processing is used to understand the given input and further forecast the user's illness.

III. ARCHITECTURE DIAGRAM

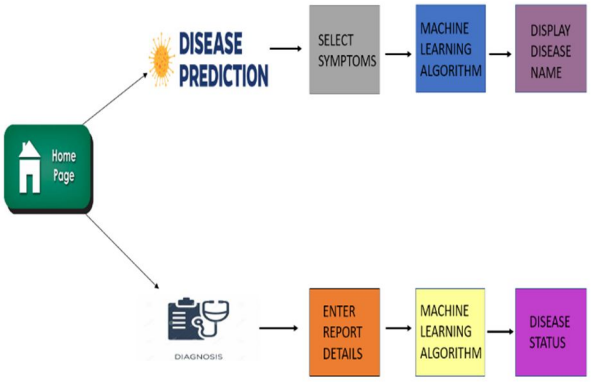


Figure 3.1: Architecture Diagram

The architecture diagram for a multi-disease prediction using Naive Bayes networks typically consists of several components.

The first component is the data source, which can be electronic health records, medical imaging, or other medical data. This data is used to train the Naive Bayes networks for each disease.

The second component is the Naive Bayes network itself, which is composed of nodes representing various risk factors or symptoms for each disease. The model calculates the probability of each disease based on the presence or absence of these risk factors or symptoms.

The third component is the prediction engine, which uses the trained Naive Bayes networks to predict the probability of each disease for a given patient. This prediction engine can be integrated into a larger clinical decision support system to assist physicians in making accurate diagnoses and treatment plans. Finally, there is a feedback loop that continuously updates the Naive Bayes networks based on new data from patient outcomes, ensuring that the model remains accurate over time.

IV. PROPOSED SYSTEM

The proposed methodology predicts the presence of multiple diseases independently by using Naive Bayesian networks, a well-known machine learning algorithm. The technique estimates a probability score for each condition using patient data, such as symptoms and medical history.

The system trains the Naive Bayesian network method using a sizable dataset of patient data. The dataset contains details on the patient's age, gender, symptoms, medical background, and test outcomes for a variety of disorders. To ensure the correctness and dependability of the system, the dataset is subjected to sophisticated data preparation and feature selection procedures.

The Naive Bayesian network approach is preferred because it can handle huge datasets and represent intricate correlations between features. To determine the likelihood of a disease occurring given a collection of symptoms or a medical history, the algorithm applies the Bayes theorem.

By assuming that all characteristics are independent of one another, the Naive assumption is utilized to streamline the computations, making it suited for huge datasets and enabling efficient computing.

The suggested system has several benefits over the current ones. First off, it can concurrently forecast several illnesses, which can speed up diagnostic testing and save time. Second, it can manage big datasets and intricate interactions between characteristics, enhancing illness prediction accuracy. In addition, it uses machine learning techniques to raise the general standard of healthcare services.

To adopt the suggested system, a few adjustments must be made to the current one. To prepare a collection of patient data that can be used to train the Naive Bayesian network technique, a sizable and high-quality dataset must first be gathered.
The system's correctness and dependability must also be guaranteed by the application of sophisticated feature selection and data preparation procedures.

A. Algorithm Used: Naïve Bayesian Networks

Popular machine learning algorithms for categorization tasks include naive Bayes networks. The Bayes theorem and the presumption of conditional independence between the characteristics serve as the foundation for the method. Naive Bayes has been demonstrated to be successful in managing big datasets and obtaining high accuracy in numerous applications, despite its simplicity.
In this study, Naive Bayes networks are employed to forecast the simultaneous presence of many illnesses. The algorithm determines the likelihood that each disease will manifest from a collection of symptoms or medical history. The benefit of employing Naive Bayes for this purpose is that it is simple to construct and can handle huge and complicated datasets.
The management of missing data, the elimination of redundant features, and the transformation of features into an algorithm-friendly format are all included in this. The dataset is divided into training and testing sets, and the algorithm is taught to optimize its parameters using the training set. Using the testing set, the algorithm's correctness is assessed.
Naive Bayes networks can handle big datasets and represent intricate interactions between characteristics, making them an effective tool for illness prediction. In this study, sophisticated feature selection and data pretreatment methods guarantee the algorithm's efficacy. By requiring less time and money for diagnosis and treatment, the suggested approach has the potential to enhance healthcare services.

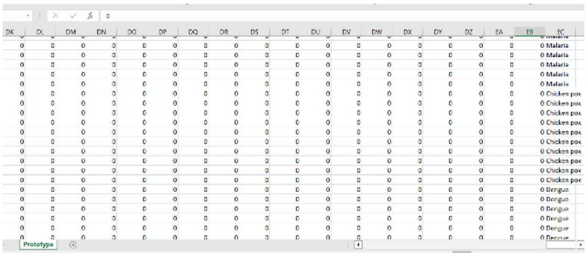


Figure 4.1.1: Dataset of various Diseases

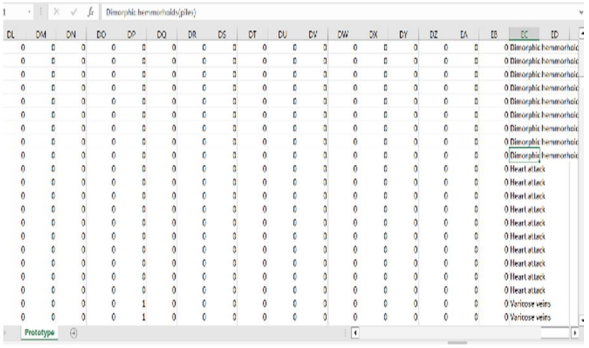


Figure 4.1.2: Dataset used in Training Algorithm

B. Diagnosis Detection Module

1) Diabetics Detection

Identification of those with diabetes or those who are at risk of getting the disease is known as diabetes detection. Medical testing, surveys, and machine learning algorithms are just a few methods for detecting diabetes.
Diabetes is commonly detected via medical tests like the oral glucose tolerance test (OGTT) and glycated hemoglobin (A1C) test. These examinations examine blood glucose levels, a crucial sign of diabetes.
Diabetes may potentially be detected using machine learning methods like Naive Bayes networks. We can train a Naive Bayes model by calculating the probability of each feature given the class (diabetic or non-diabetic). This is done using a training set of patients who have already been classified as diabetic or non-diabetic.

Result*	A1C Test	Fasting Blood Sugar Test	Glucose Tolerance Test	Random Blood Sugar Test
Diabetes	6.5% or above	126 mg/dL or above	200 mg/dL or above	200 mg/dL or above
Prediabetes	5.7 - 6.4%	100 - 125 mg/dL	140 - 199 mg/dL	N/A
Normal	Below 5.7%	99 mg/dL or below	140 mg/dL or below	N/A

Table 4.2.1.1: Diabetes parameters for results

2) Heart Diseases Detection

Millions of individuals throughout the world suffer from the prevalent and possibly fatal ailment known as heart disease. Since they are effective, precise, and simple to use, naive Bayes networks have shown to be useful tools for predicting cardiac disease. A kind of probabilistic graphical model called a naive Bayes network uses the Bayes theorem to estimate the likelihood of a specific event. Age, sex, blood pressure, cholesterol levels, and other heart disease risk factors are among the data set used to train the algorithm. The algorithm may then be used to forecast a patient's chance of acquiring heart disease based on their unique risk variables after being trained.

Normal Ranges of ECG parameters

S.NO	ECG parameter	Normal range
1	Heart Rate	60 - 100 beats per minute
2	PR Interval	120 - 200 milliseconds
3	QRS complex	<120 milliseconds
4	QT Interval	<440 milliseconds
	* ECG = Electrocardiogram	

Table 4.2.2.1: ECG Parameters for Results

3) Kidney and Liver Disease Detection

It is possible to forecast kidney illness using Naive Bayes Networks. The model is trained using a collection of data that contains elements that affect kidney function, such as blood pressure, proteinuria, and creatinine levels. Upon training, the model may be used to forecast a patient's chance of getting renal disease depending on certain risk factors. Moreover, Naive Bayes networks, which can precisely predict the presence or absence of kidney illness based on patient data, can help in the diagnosis of renal disease. Naive Bayes networks, as a whole, are a useful tool for anticipating and detecting renal illness and can help to improve patient outcomes.

Similarly, we can also diagnose Liver related diseases using their testing parameters by Naïve Bayesian Networks using the same methods as used in Diabetes, heart, and Kidney diagnosis.

Normal Ranges of Kidney and Liver parameters

S.NO	Kidney and Liver function tests	Normal range
1	Urea	13 - 45 mg/dL
2	Creatine	0.7 - 1.5 mg/dL
3	Alk. Phosphate	65 - 306 IU/L
4	Bili Rubin	<1 mg
5	SGOT	1 - 40 IU/L
6	SGPT	1 - 40 IU/L
7	Total Protein	6.5 - 8.5 g/dL
8	Albumin	3.5 - 5.3 g/dL

Table 4.2.3.1 Kidney, Liver parameters for results

V. RESULTS AND DISCUSSIONS

Predictive models created using the Naive Bayes method would be the end product of this study. The project is set up so that, in one module, the system receives user-provided symptoms as input and outputs the pertinent illness, while, in a different module, the system receives user-provided test report data and outputs the disease's state. The expected result is precise and effective. These metrics may be used to assess the models' effectiveness and determine whether or not multi-disease prediction is an appropriate usage for them. Insights regarding the connections between various symptoms and diseases may also be included in the findings, which may be helpful for future study and advancement in the area of medical diagnostics. Overall, this research aims to present a practical and successful method for multi-disease prediction using machine learning, which may have important repercussions for enhancing the precision and timeliness of medical diagnosis.

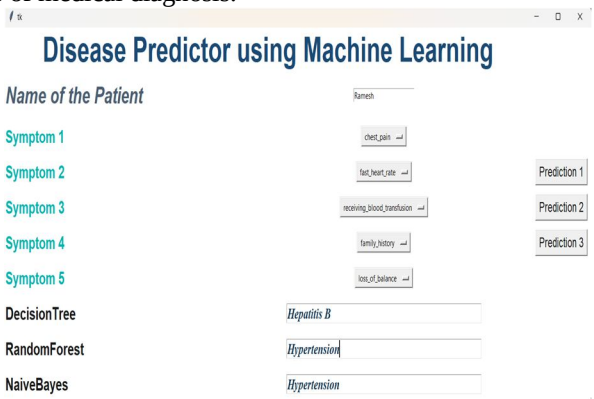
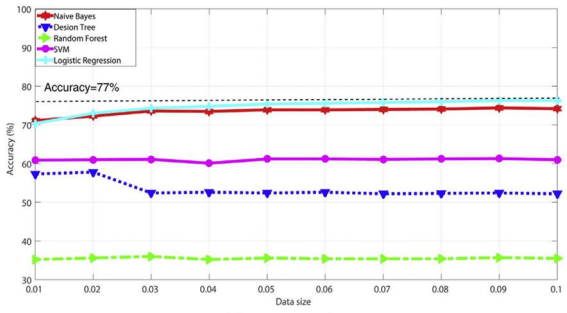
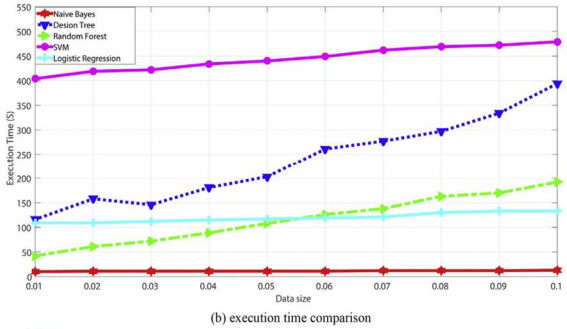


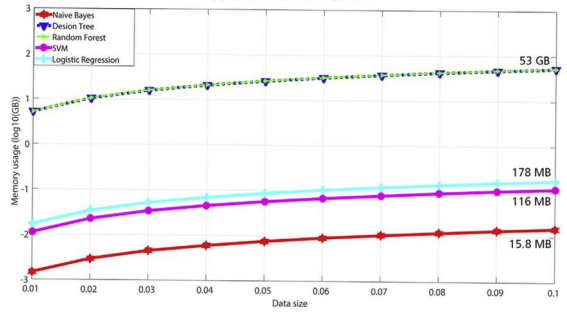
Figure 5.1 Sample Final Output



(a) accuracy comprison



(b) execution time comparison



(c) memory usage comparison

Figure 5.2 Comparison between different algorithms with different parameters

VI. CONCLUSION

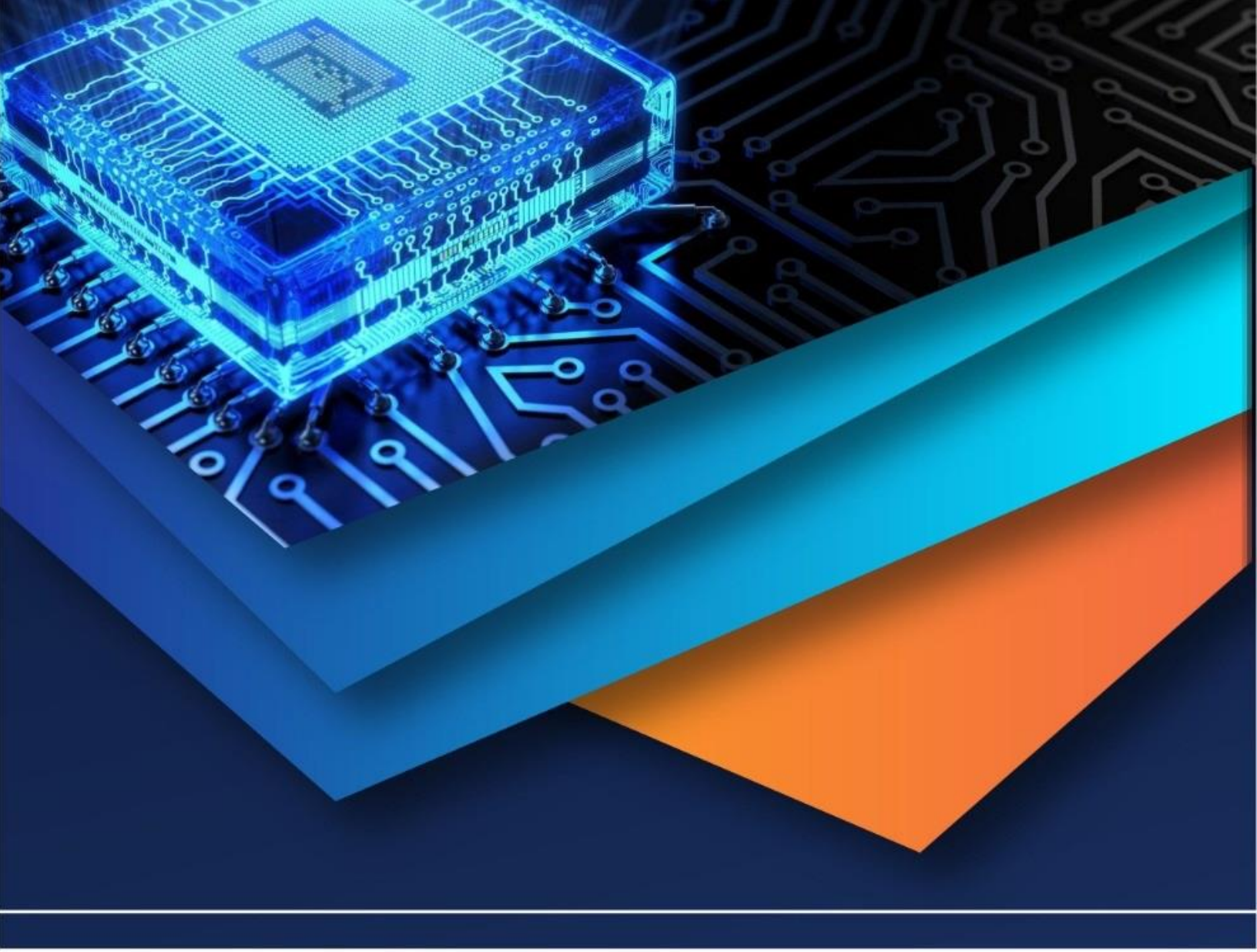
In conclusion, the suggested method for multi-disease prediction utilizing Naive Bayes networks has demonstrated promising results in effectively predicting several illnesses using patient symptoms as input data. To determine the likelihood of a disease given a collection of symptoms, the system uses the Naive Bayes method, a probabilistic model that assumes the independence of features. The system demonstrated a high degree of accuracy in predicting different diseases when it was tested using a dataset of patient symptoms and associated diagnoses.

A. Future Enhancement

In the future, this work might be improved by using more modern machine-learning algorithms and techniques, such as deep learning, to increase the prediction models' accuracy and efficiency. Also, expanding the dataset to incorporate more varied patient groups and a larger spectrum of illnesses and symptoms may help shed additional light on the models' efficiency. The technology might be made easier to use in clinical settings by being integrated into an intuitive interface that both patients and healthcare professionals can access. This would eventually improve patient outcomes and increase the effectiveness of healthcare delivery.

REFERENCES

[1] Sameer Meshram1, Shital Dongre, Triveni Fole. “Disease Prediction System using naïve bayes”. International Journal for Research in Applied Science & Engineering Technology Volume 10 Issue XII Dec 2022.
[2] Abid ishaq, saima sadiq, muhammad umer, saleem ullah, seyedali mirjalili, vaibhav rupapara , and michele nappi. “Improving the Prediction of Heart Failure Patients’ Survival Using SMOTE and Effective Data Mining Techniques”. March 16, 2021.
[3] Bilal khan, Rashid naseem, Fazal muhammad, ghulam abbas, and sung hwan kim. “An Empirical Evaluation of Machine Learning Techniques for Chronic Kidney Disease Prophecy”. March 30, 2020.
[4] Chandrasekhar Rao Jetti, Rehamatulla Shaik, Sadhik Shaik, Sowmya Sanagapalli “Disease Prediction using Naïve Bayes - Machine Learning Algorithm”, December 2021.
[5] Prediction Support System for Multiple Disease Prediction Using Naive Bayes Classifier”. Selvaraj A, Mithra MK, Keerthana S, Deepika M. International Journal of Engineering and Techniques - Volume 4 Issue 2, Mar-Apr 2021.
[6] Akkem Yaganteeswarudu,” Multi Disease Prediction Model by using Machine Learning and Flask API”, IEEE, July 2022.
[7] Yashaswi G Sagar, Sahana Gajanana, Riyal Vivek, Swetha P,” MediInsight: A Smart Health Prediction System”, (IRJET), June 2021.
[8] Hsiu-Sen Chiang, Mu-Yen Chen, “Cognitive Depression Detection Cyber-Medical System Based on EEG Analysis and Deep Learning Approaches” IEEE, February 2023.
[9] Wei Shao, Zhiyang You, Lesheng Liang, Xiping Hu, “A Multi-Modal Gait Analysis-Based Detection System of the Risk of Depression” IEEE, October 2022.
[10] Akash C. Jamgade, Prof. S. D. Zade, International Research Journal of Engineering and Technology, May 2019.
[11] Anjan Nikhil Repaka, Sai Deepak Ravikanti, Ramya G Franklin,” Design and Implementing Heart Disease Prediction Using Naïve Bayesian”, IEEE, June 2019.
[12] N. P. Tigga and S. Garg, ‘‘Prediction of type 2 diabetes using machine learning classification methods,’’ Proc. Computer Sci., Jan. 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)