



e-ISSN: 2319-8753 | p-ISSN: 2347-6710

# IJRSET

International Journal of Innovative Research in  
**SCIENCE | ENGINEERING | TECHNOLOGY**

# INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

Volume 13, Issue 4, April 2024

**ISSN** INTERNATIONAL  
STANDARD  
SERIAL  
NUMBER  
INDIA

Impact Factor: 8.423

9940 572 462

6381 907 438

[ijirset@gmail.com](mailto:ijirset@gmail.com)

[www.ijirset.com](http://www.ijirset.com)

# Identification and Prediction of Multiple Chronic Diseases Using Machine Learning Approach

**Mr. Amol C. Wani, Disha Salunke, Vaidehai Mukhedkar, Vaishnavi Baviskar, Yogita Chaudhari**

Assistant Professor at Shram Sadhana Bombay Trust's College of Engineering And Technology, Bambhori,  
Jalgaon, India

U.G. Students, Department of Computer Engineering, Shram Sadhana Bombay Trust's College of Engineering And  
Technology, Bambhori, Jalgaon, India

**ABSTRACT:** Machine Learning technology has revolutionized the healthcare industry by uncovering new, efficient ways to diagnose diseases. Chronic diseases pose a significant health burden globally in the 21st century. Healthcare expenditures are overwhelming national and corporate budgets due to asymptomatic infections, including kidney, cardiovascular, Parkinson's, and diabetic ailments. Consequently, there is an urgent requirement for early detection and treatment of such diseases. The information gathered through data analysis of hospitals is utilized by applying different blends of calculations and algorithms for the early-stage prediction of such ailments. Additionally, Machine Learning facilitates disease prediction based on symptoms, providing a valuable tool for healthcare professionals to identify potential illnesses in patients. It emerges as one of the trending innovations utilized in numerous circles worldwide, including healthcare applications for predicting illnesses.

**KEYWORDS :** Machine Learning, Healthcare, Chronic diseases, Diagnosis, Data analysis, Algorithms, Early detection, Prediction, Kidney disease, Cardiovascular disease, Parkinson's disease, Diabetes, Symptoms, Healthcare applications, Disease detection.

## I. INTRODUCTION

Multiple Disease Prediction System is an end-to-end machine learning system designed to analyze various medical parameters and predict the likelihood of a patient acquiring a particular disease. The system not only predicts disease likelihood based on medical parameters but also provides disease names by inputting symptoms. The goal is to revolutionize healthcare by harnessing the power of machine learning to accurately predict and diagnose diseases. Leveraging cutting-edge technology, we aim to develop a predictive model capable of accurately forecasting the likelihood of various diseases in individuals. By analyzing comprehensive medical data and employing advanced machine-learning algorithms, we can provide timely and precise predictions. Our research focuses on developing a robust model that can analyze various patient parameters and historical records to forecast the likelihood of specific diseases. Through this project, we aspire to enhance healthcare outcomes, empower medical professionals, and improve the overall well-being of individuals worldwide.

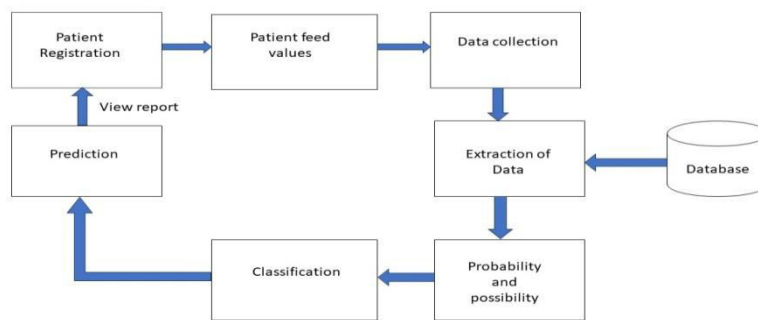
## II. RELATED WORK

In existing research on machine learning models in healthcare, there is a predominant focus on analyzing one disease per analysis. This often leads to fragmented systems, requiring users to navigate multiple platforms for predicting different diseases, thereby posing inconvenience and inefficiency. Moreover, the limitations of some models, characterized by lower accuracy rates, raise concerns regarding potential risks to patient health outcomes. Organizations tasked with analyzing patient health reports are further challenged by the necessity to deploy numerous models, resulting in increased costs and time consumption. Additionally, certain existing systems rely on a limited set of parameters, heightening the likelihood of inaccurate predictions and false results.



### III. ARCHITECTURE

The system architecture of a Multiple Disease Prediction System involves the organization and arrangement of various components, modules, and technologies to achieve the system’s objectives efficiently.



### IV. METHODOLOGY

#### 1. Data collection and data analysis

Chronic Diseases	Total Entries In Dataset	Classification		Required Test Reports
		0	1	
Heart Disease	303	138	165	Comprehensive Cardiac Health Assessment Test
Diabetes Disease	768	500	268	Oral Glucose Tolerance Test
Kidney Disease	400	250	150	Comprehensive Kidney Health Assessment Test
Parkinson's Disease	195	48	147	Parkinson's Voice Analysis Diagnostic Test

#### 2. Algorithm :

**Algorithm 1:** The Steps of Implementing Random Forest (RF)

**Input:** The n-dimensional data,  $X \in R^n$  and target outcome,  $Y \in R$

**Output:** The posterior probability,  $P \in [0, 1]$  of unseen test data,  $x$ , where  $X C_{i=1}^n P_i = 1$  and  $C = 2$  (diabetes present (C1) or not (C2))

- for  $b = 1$  to  $N$  ( $n\_Bagging$ ) do
- Draw a bootstrap sample,  $(X_b, Y_b)$  from given  $(X \in R^n, Y \in R)$
- Grow a random-forest tree  $T_b$  using  $X_b$  and  $Y_b$  by repeating recursively using the following steps until the minimum node size is  $n_{min}$ .
- Randomly select  $m$  variables from the given  $n$  variables
- Pick the best variable or split-point among the  $m$  variables
- Split the node into two daughter nodes Output the ensemble of trees will be  $\{T_b\}_{N=1}$
- The posterior probability,  $P^N_{RF}(x) = \text{Voting}\{P^k(x)\}_{N=1}$ , where  $P^k(x)$  is the class prediction of the  $k$ th random-forest.

**Algorithm 2:** The Steps of Implementing Naive Bayes (NB)

**Input:** The n-dimensional data,  $X \in R^n$  and target outcome,  $Y \in R$

**Output:** The posterior probability,  $P \in [0, 1]$  of unseen test data,  $x$ , where  $X C_{i=1}^n P_i = 1$  and  $C = 2$  (diabetes present (C1) or not (C2))



1. Compute the prior probabilities for each of the class [49],  $P(Y = C1) = \frac{NC1}{N}$  and  $P(Y = C2) = \frac{NC2}{N}$ , where N is the number of sample
2. The output posterior probability of class for the given predictor (attributes),  $P(Ci | X) = \frac{P(X|Ci) \times P(Y=Ci)}{P(X)}$ , where  $P(X|Ci)$  is the likelihood of the predictor for a given class and  $P(X)$  is the prior probability of predictor.

**3. Parameter used for calculating accuracy :**

**Accuracy :**

The classification accuracy is described as the ratio of correct predicted values to the total predicted values and is depicted mathematically as follows:

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} * 100$$

**Precision :**

The precision or positive predictive value (PPV) is described as the ratio of correct prediction to the total correct values including the true and false predictions and is depicted mathematically as follows:

$$\text{Precision} = \frac{TP}{TP + FP}$$

**Recall :**

The recall or sensitivity or true positive rate (TPR) is described as the ratio of correct predicted values to the sum of correct positive predictions and the incorrect negative predicted values and is depicted mathematically as follows:

$$\text{Recall} = \frac{TP}{TP + FN}$$

**F1-Score :**

The F-measure ( $F\beta$ ) is described as the weighted average of the values obtained from the calculation of precision and recall parameters. Whenever the distribution of class is not even, then the value of F1 – Score is highly important than the accuracy value. And whenever the values of false positives and negatives are dissimilar, the value of F1 – Score is highly suitable. 'e F1 – Score is depicted mathematically as follows:

$$F\beta = \frac{1 + \beta^2}{\beta^2} \frac{\text{Precision} * \text{Recall}}{\text{Precision} + \text{Recall}}$$

By simplifying using  $\beta = 1$ ,

$$F1 - \text{Score} = \frac{2 * (\text{Precision} * \text{Recall})}{(\text{Precision} + \text{Recall})}$$

**Classification report:-**

The classification report shows a representation of the main classification metrics on a per-class basis. This gives a deeper intuition of the classifier behavior over global accuracy which can mask functional weaknesses in one class of a multiclass problem.

**Heart Disease**

Classification Report:				
	precision	recall	f1-score	support
0	0.83	0.83	0.83	29
1	0.84	0.84	0.84	32
accuracy			0.84	61
macro avg	0.84	0.84	0.84	61
weighted avg	0.84	0.84	0.84	61





**Diabetes Disease**

Classification Report:

	precision	recall	f1-score	support
0	0.82	0.84	0.83	206
1	0.66	0.64	0.65	102
accuracy			0.77	308
macro avg	0.74	0.74	0.74	308
weighted avg	0.77	0.77	0.77	308

**Kidney Disease**

Classification Report:

	precision	recall	f1-score	support
0	1.00	1.00	1.00	9
1	1.00	1.00	1.00	23
accuracy			1.00	32
macro avg	1.00	1.00	1.00	32
weighted avg	1.00	1.00	1.00	32

**Parkinson's Disease**

Classification Report:

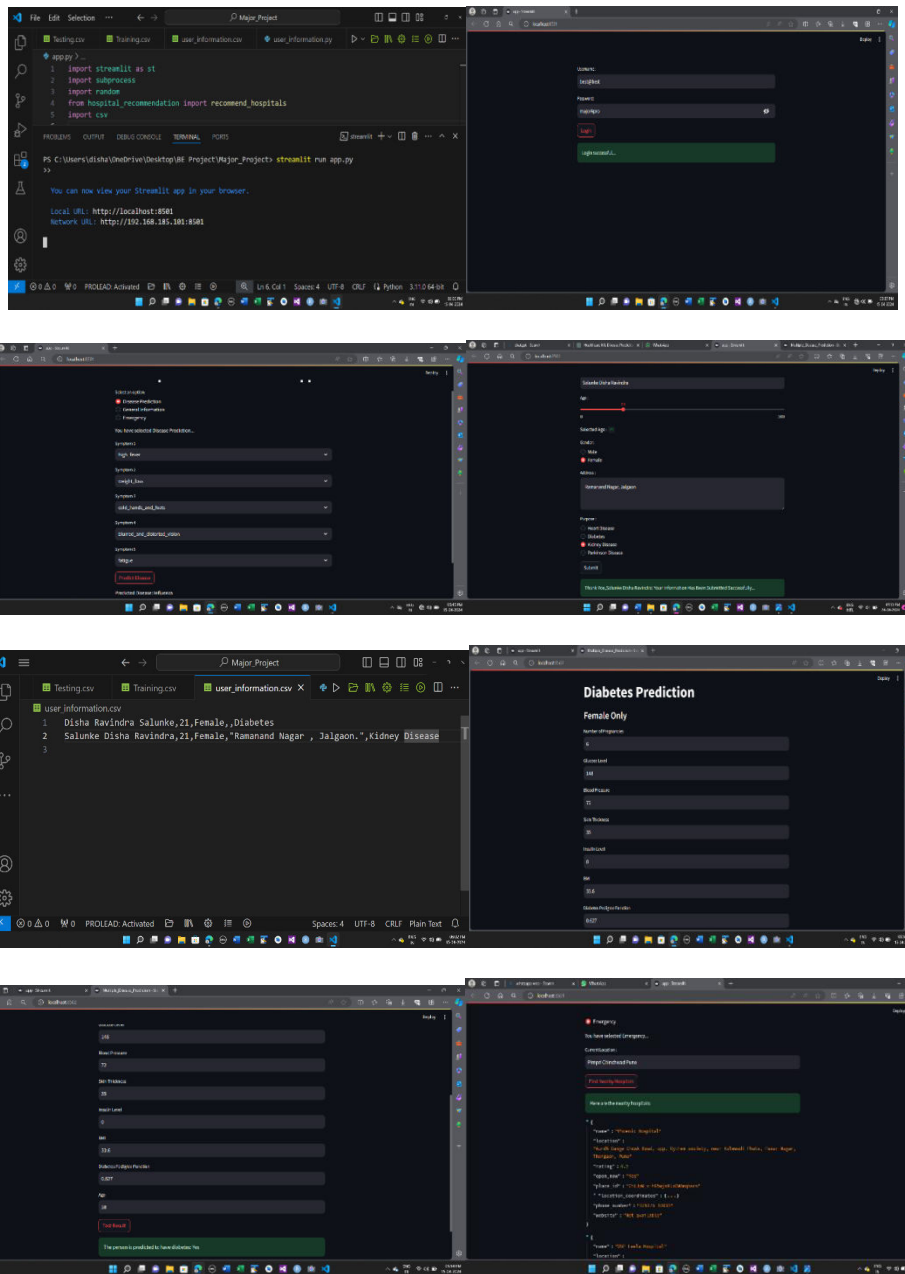
	precision	recall	f1-score	support
5	1.00	1.00	1.00	56
accuracy			1.00	56
macro avg	1.00	1.00	1.00	56
weighted avg	1.00	1.00	1.00	56

**Performance Evaluation Table :**

Diseases	Ideal Splitting Ratios With Accuracy of Random Forest Classifier			
	80-20	70-30	60-40	90-10
Heart	<b>0.84</b>	0.82	0.84	0.74
Parkinsons	0.88	0.92	<b>1.0</b>	0.78
Diabetes	0.73	0.75	<b>0.77</b>	0.71
Kidney	0.98	0.97	<b>1.0</b>	0.97

### V. EXPERIMENTAL RESULT

Fig shows the results of multiple chronic disease prediction from provided data by using Random forest classifier and Naïve Bayes.



### VI. CONCLUSION

Our proposed system encompasses a Disease Prediction System where users select a Chronic Disease and input parameter values from the respective disease report. The system then outputs whether the person has the disease or not. Additionally, our system offers the capability to predict multiple common diseases based on symptoms input by the user. Looking ahead, the future scope of this proposed system lies in addressing the limitations of traditional diagnosis methods, particularly in the case of serious ailments. Providing accurate and timely analysis of chronic disease-related



problems is crucial for early prevention and treatment. By leveraging machine learning technology, our system aims to contribute to early disease detection and improve healthcare outcomes.

#### REFERENCES

- [1]. MD. Kamrul Hasan, MD. Ashraful Alam, Dola Das, Eklas Hossain, (Senior Member, IEEE), and Mahmadul Hasan, "Diabetes Prediction Using Ensembling of Different Machine Learning Classifier". vol.8, April 23, 2020.
- [2]. Senthilkumar Mohan, Chandrasegar Thirumalai, and Gautam Srivastava, (Member, IEEE), "Effective Heart Disease Prediction Using Hybrid Machine Learning Techniques". vol.8, June 19, 2019.
- [3]. Bilal Khan, Rashid Naseem, Fazal Muhammad, Ghulam Abbas, (Senior Member, IEEE), and Sunghwam Kim\7 "An Empirical Evaluation of Machine Learning Techniques for Chronic Kidney Disease Prophecy". vol.8, March 18, 2020.
- [4]. Kunal Takke, Rameez Bhajjee, Avanish Singh and Mr.Abhay Patil published a paper as "Medical Disease Prediction Using Machine Learning Algorithms", May 2022.
- [5]. Marouane Fethi Ferjani published a paper named "Disease Prediction Using Machine Learning", December 2020.
- [6]. K.Arumugam, M.Naved, P.P.Shinde published a paper named "Multiple Disease Prediction Using Machine Learning Algorithms", August 2021.



INTERNATIONAL  
STANDARD  
SERIAL  
NUMBER  
INDIA



# INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

 9940 572 462  6381 907 438  [ijirset@gmail.com](mailto:ijirset@gmail.com)



[www.ijirset.com](http://www.ijirset.com)

Scan to save the contact details