Exploresearch ISSN: 3048-815X Vol. 03, No. 02, April-June 2026, 79-87

Original Article | **Peer Reviewed** | **Open Access**

Exploresearch

Impact Factor (Cosmos: 6.262& I2OR: 3.585)

www.mgmpublications.com

Crossref

Symptoms Based Smart Disease Predictor System

Dimpy Sharma[1] & Tanvi Rustagi[2*]

[1,2]Department of Computer Science and Engineering, World College of Technology and Management, Gurgaon, HR, India.

***Corresponding author:** tanvirustagi123@gmail.com*

***Citation**: Sharma, D. & Rustagi, T. (2026). Symptoms Based Smart Disease Predictor System. Exploresearch, 03(01), 79–87. https://doi.org/10.62823/ExRe/2026/03/02.222*

Article History:

Received: 04 May 2026
Revised: 18 May 2026
Accepted: 23 May 2026
Published: 26 May 2026

Keywords:

Machine Learning, Symptoms based disease prediction, Python, Decision Tree, Random Forest, KNN, Naïve Bayes.

Abstract:

The Disease Prediction System is a Machine Learning-based system that functions on the basis of the symptoms inputted by a user. In disease prediction, an algorithmic analysis is performed through the comparison of various sets of datasets on the basis of symptoms entered by the user. Nevertheless, it has been established that Machine Learning technology is highly effective as far as providing a limitless platform within the medical field is concerned, hence allowing healthcare problems to be resolved with ease. It is due to reliance on technological systems that many e-datasets are available in the healthcare industry. This has led to health care professionals facing complex situations in researching the signs and symptoms of diseases in order to diagnose illness effectively. In the situation of a severe illness, the conventional diagnosis process may not suffice. A medical diagnosis system can, therefore, be developed using machine learning (ML) algorithms for prediction purposes. Disease prediction model is constructed with the help of different ML algorithms. Training and Testing data sets are used for this purpose. According to the symptoms that individual has, the diagnosis system produces its output in form of diseases which an individual could have. Disease prediction with machine learning can be termed as a system that predicts the illness according to the input entered by individual into the system, which also suggests the cure of that particular disease. It is a tool that helps the user know how to maintain his health system through knowledge and tips on how to detect illness predictions. Using this method, by just asking questions of the symptoms experienced by the user and then diving into the software, in a few seconds, they will be able to diagnose precisely the exact diseases. The disease Predictions using machine learning have been completed using machine learning and python programming language with tkinter interface (UI). Data in the form of tabulated data is stored in SQLite Studio.

Introduction

Healthcare and medicine are one of the most important aspects of the economy as well as of human being's life. Due to the increasing numbers of patients per year, the medical system becomes overloaded with time and becomes very expensive. Every patient needs consultation from doctors in order to treat him/her properly. Having enough data, the prediction of the disease using algorithms can be quite easy and inexpensive. The prediction of the disease by considering symptoms is essential for the proper treatment of the person affected by disease. We have observed massive changes between the world before and today. It has turned gruesome and divergent. In such situations, where everything becomes virtual, doctors and nurses work hard to save the lives of patients, even risking theirs. In some areas, there is a lack of necessary medical equipment. Virtual doctors are board-certified doctors who prefer practicing through online consultations using videos and phones instead of consulting in clinics or hospitals, but not during emergencies.

It is said that machines always are better than humans as there is no room for human errors; therefore, the machines will work in a more efficient way with a similar level of precision. Disease predictors can also be seen as virtual doctors, which can predict any patient's disease without any human errors. Further, in cases of diseases such as COVID-19 and Ebola, the disease predictor can be very helpful since it can detect disease without coming into contact with anyone physically. Virtual doctors have existed in some form; however, they do not have the level of accuracy required because all the parameters needed have not been taken into consideration. The main aim was to build various models so that we could determine which one gives the best predictions. Though different in size, complexity, and scope, all Machine Learning projects share a common structure. Various methods of Rule-Based systems used in ML were recalled regarding the prediction model. Some models have been introduced by applying different machine learning (ML) algorithms on data which had been collected in raw form and had then been split on the basis of gender, age group, and symptoms. These data sets have been processed through several different ML models, such as Decision Trees, Naive Bayes, KNN, and Random Forest. The level of accuracy differed with respect to the use of different ML models. While processing this data, data set was provided to all ML models as input parameter, and disease came out as the output in different accuracies.

Disease Prediction Using Machine Learning is a project designed for overcoming diseases generally prevailing in the early stage within the competitive arena of economic growth. The basic purpose behind developing this project is that users can sit at their convenient place and do a test of their health conditions. The UI is designed in such a way that everyone can use it easily and get a checkup done.

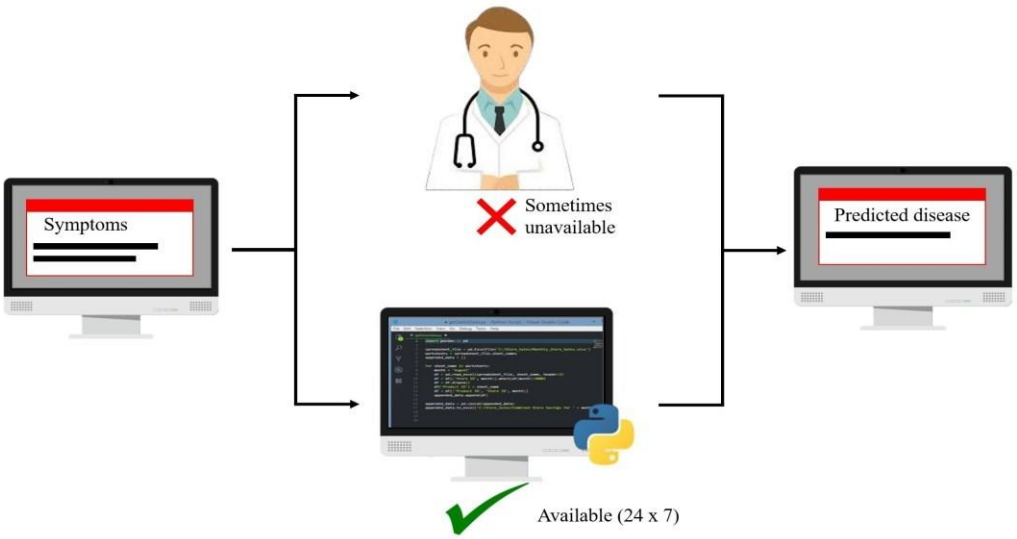


Fig. 1 Proposed model for predicting the disease. The doctor is not always there whenever required. However, in the present time scenario, at any point of time, as per the requirement, one can use this prediction system. The input provided here will be the symptoms, age, and gender of the individual to the machine learning algorithm.

Methodology

The overall proposed system is classified into five modules.

- Importing Libraries
- Collection of Clinical Data
- Data Pre-Processing
- Model Building
- Smart Disease Predictor System
- Database Creation

Importing Libraries

- **Numpy:** The numpy python library is used in the code to include any kind of mathematical operation in it. In Python, it can be imported as follows:

Import Numpy as np

Where, np is the short form of Numpy, and it will be used throughout the program.

- **Matplotlib:** This module is used for making charts for any kind of data in Python program. It can be imported as follows:

Import Matplotlib.pyplot as plt

Here, plt acts as an abbreviation of the above library.

- **Pandas:** It is an open-source data manipulation and analysis library. It will be imported as below:

Import Pandas as pd

Here, pd used as a short name for this library.

- **Tkinter:** Tkinter is the standard GUI library for Python. Tkinter provides a powerful object- oriented interface to the Tk GUI toolkit.

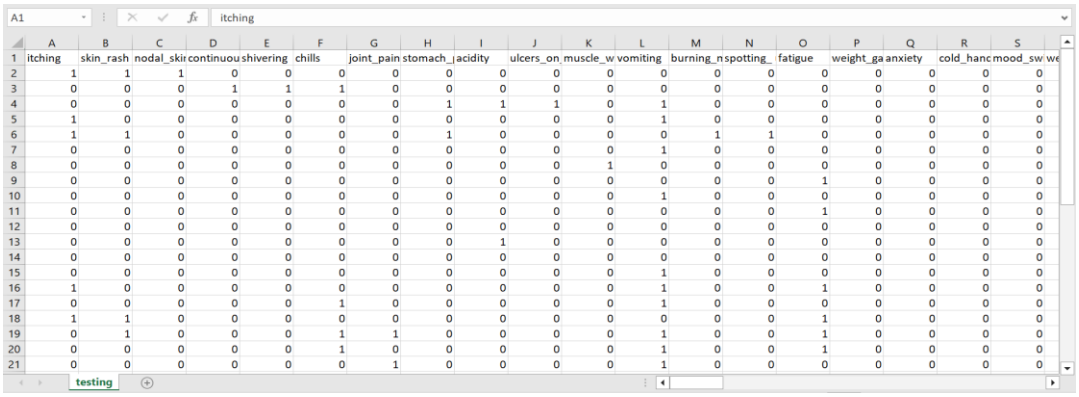
Consider the below image:

```
#Importing Libraries from matplotlib to visualize the data
from mpl_toolkits.mplot3d import Axes3D
from sklearn.preprocessing import StandardScaler
import matplotlib.pyplot as plt

#Importing Libraries to create GUI
from tkinter import *

#Importing Libraries to perform calculations
import numpy as np
import pandas as pd
import os
```

Fig. 2: Importing Libraries

- **Collection of Clinical Data**

The dataset could be a database of information on disease symptom relationship that was generated via automation using patient discharge summaries at New York Presbyterian Hospital as shown in Fig.3.



	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S
1	itching	skin_rash	nodal_skir	continuou	shivering	chills	joint_pain	stomach_i	acidity	ulcers_on	muscle_w	vomiting	burning_n	spotting_	fatigue	weight_ga	anxiety	cold_hanc	mood_sw
2	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
3	0	0	0	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0
4	0	0	0	0	0	0	0	1	1	1	0	1	0	0	0	0	0	0	0
5	1	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0
6	1	1	0	0	0	0	0	1	0	0	0	0	1	1	0	0	0	0	0
7	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0
8	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0
9	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0
10	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0
11	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0
12	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
13	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0
14	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
15	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0
16	1	0	0	0	0	0	0	0	0	0	0	1	0	0	1	0	0	0	0
17	0	0	0	0	0	1	0	0	0	0	0	1	0	0	0	0	0	0	0
18	1	1	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0
19	0	1	0	0	0	1	1	0	0	0	0	1	0	0	1	0	0	0	0
20	0	0	0	0	0	1	0	0	0	0	0	1	0	0	1	0	0	0	0
21	0	0	0	0	0	0	1	0	0	0	0	1	0	0	0	0	0	0	0

Fig. 3: Clinical Data

- **Data Pre-Processing**

Real-world data may contain noise, be incomplete, and possibly may be in a non-compatible format that cannot be directly fed into the machine learning model. The process of data pre-processing helps clean data and makes it ready to be inputted to the machine learning algorithm and also improves its efficiency.

The process of data pre-processing includes below activities:

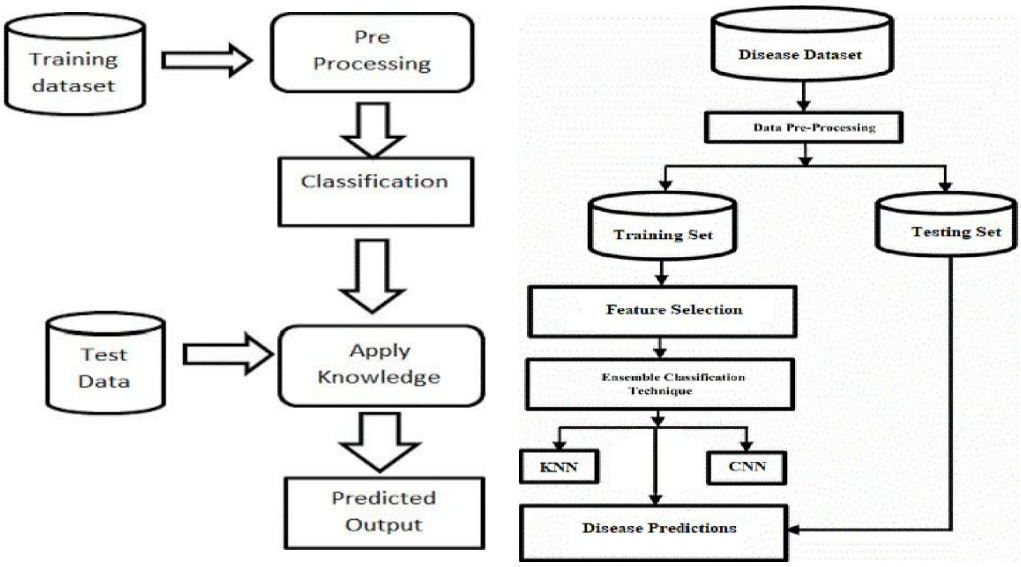


Fig. 4: Data Pre-Processing

- **Model Building**

Models of Predictive Classification were created to precisely classify the Disease entered by the user. The predictive algorithm that was used to classify Disease is Random Forest (RF), Decision Tree, K Nearest Neighbour, and Naïve Bayes Algorithm.

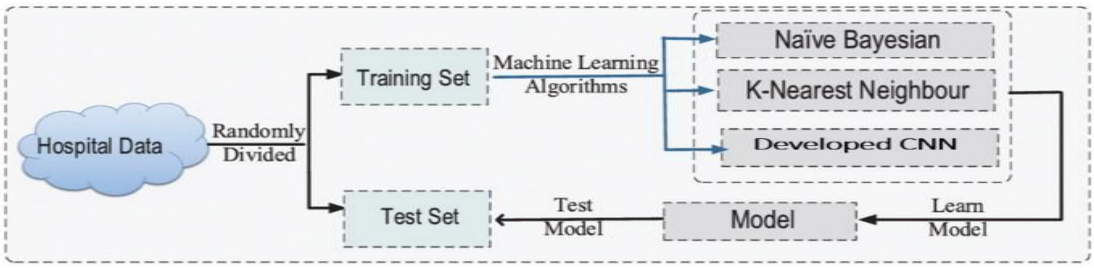


Fig. 5: Machine Learning Algorithms

- **Decision Tree**

Decision tree induction is the process of inducing decision trees from a set of training examples that are labelled by class. A decision tree is a tree shaped flow chart, Decision trees are very robust to noise, especially after overfitting avoidance techniques.

```
root = Tk()
pred1=StringVar()
def DecisionTree():
```

Fig. 6: Decision Tree Function

DecisionTreeClassifier() method is applied to fit and test the model with disease prediction based on symptoms provided by the user. Predicted disease with Decision Tree is stored in variable "pred1", as indicated in Fig.6.

- **Random Forest**

A random forest belongs to the ensemble classification algorithm in which there are numerous decision trees. This algorithm works on multiple input attributes without deleting any. The accuracy and importance of variables will also be provided with the results. Random forest means classifier that is made up of a set of tree classifiers k, where k is identical random trees and each random tree contributes votes in classifying the input.

```
pred2=StringVar()
def randomforest():
```

Fig. 7: Random Forest Function

Definition of randomforest() Function. The name "pred2" is assigned for storing the predicted disease through random forest algorithm as presented in Fig.7. The machine learning classifier RandomForestClassifier() is employed for training the model and predicting the disease from testing data based upon the entered symptoms by the user. "Pred2" is the final predicted disease by the random forest algorithm.

K-nearest Neighbour

The K-Nearest Neighbour is the most successful algorithm for machine learning when it comes to supervised learning. The K-NN algorithm stores all the previous data and then makes classification based on similarity. That means that when new data comes up, then using K-NN algorithm, it can be easily classified within an appropriate category. K-NN may be used not only for Classification but also for Regression. However, it is primarily used for solving Classification problems. It finds a pattern within the data, which relates the data to output and with every cycle enhances the pattern recognition.

```
pred4=StringVar()
def KNN():
    if len(NameEn.get()) == 0:
```

Fig. 8: K-nearest Neighbour Function

Definition of KNN() function. “pred4” is used to store the predicted disease using kNearestNeighbour algorithm shown in Fig.8.

Naive Bayes Algorithm

Naïve Bayes can be used to predict the class label, which is categorical. The category information is classified by Naïve Bayes based on the training set as well as values in the classifying attribute. The algorithm involves two steps, which include the model construction step and model utilization step. The Naive Bayes algorithm was named after Thomas Bayes, and its technique is used for classification and supervised learning techniques.

```
pred3=StringVar()
def NaiveBayes():
```

Fig. 9: Naïve Bayes Function

Definition of NaiveBayes() function. “pred3” is used to store the predicted disease using Naïve Bayes algorithm shown in Fig.9.

Smart Disease Predictor System

```
#Tk class is used to create a root window
root.configure(background='Ivory')
root.title('Smart Disease Predictor System')
root.resizable(0,0)

#taking first input as symptom
Symptom1 = StringVar()
Symptom1.set("Select Here")

#taking second input as symptom
Symptom2 = StringVar()
Symptom2.set("Select Here")

#taking third input as symptom
Symptom3 = StringVar()
Symptom3.set("Select Here")

#taking fourth input as symptom
Symptom4 = StringVar()
Symptom4.set("Select Here")

#taking fifth input as symptom
Symptom5 = StringVar()
Symptom5.set("Select Here")
Name = StringVar()
```

Fig. 10: GUI Sample Code

Tkinter makes it easy for programmers to develop GUI elements through the use of widgets that exist within the Tk toolkit. These widgets may be used to create buttons, menus, data fields, among others, in a program written in Python. The graphics objects created can then be linked up with or connected to functionality.

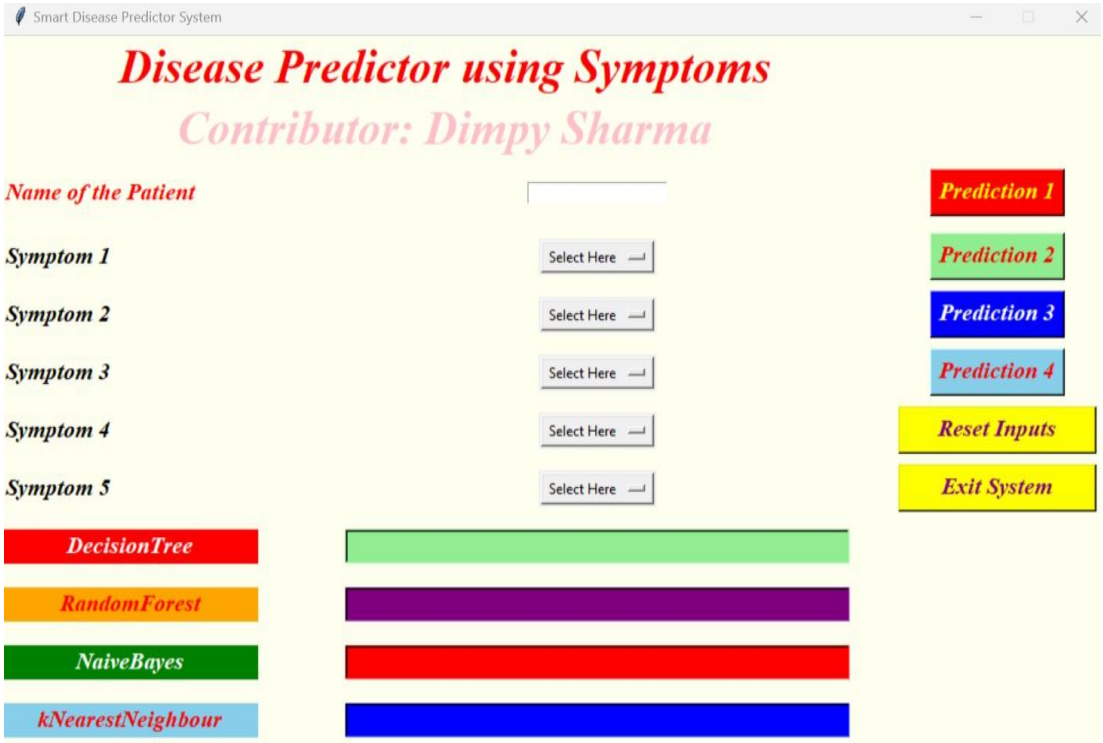


Fig. 11: Smart Disease Predictor System

Database Creation

The database is created by using SQLite to store the details entered by the user in the GUI page shown in Fig.12.

```
#Creating the database if not exists named as dimpy-data.db and creating table if not exists named as RandomForest using SQli
import sqlite3
conn = sqlite3.connect('dimpy-data.db')
c = conn.cursor()
c.execute("CREATE TABLE IF NOT EXISTS RandomForest(Name StringVar,Symtom1 StringVar,Symtom2 StringVar,Symtom3 StringVar,Symtor
c.execute("INSERT INTO RandomForest(Name,Symtom1,Symtom2,Symtom3,Symtom4,Symtom5,Disease) VALUES(?,?,?,?,?,?,?)",(NameEn.get(
conn.commit()
c.close()
conn.close()
#printing scatter plot of disease predicted vs its symptoms
scatterplt(pred2.get())
```

Fig. 12: Database Code

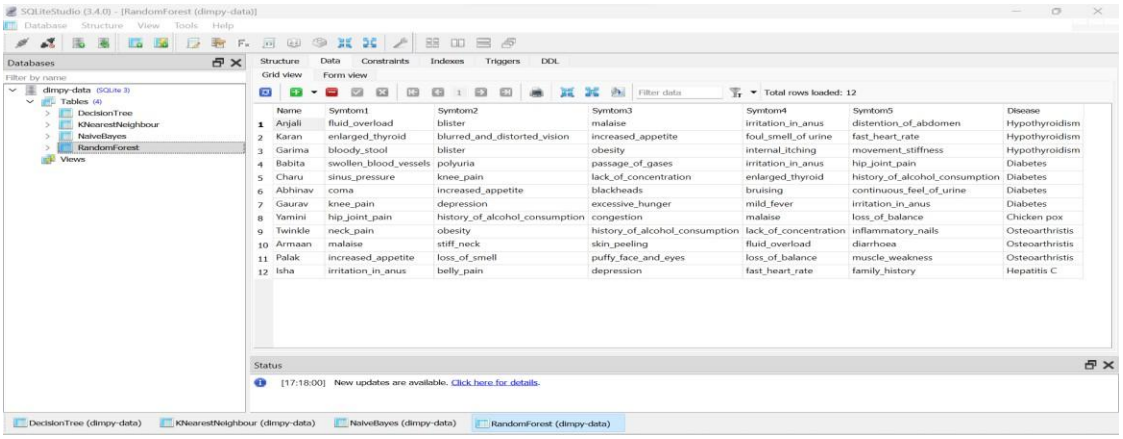


Fig. 13: Database Storage

Conclusion

The findings of this research demonstrate that machine learning algorithm can be used in predicting and detecting early signs of diseases. According to our best knowledge, the designed model according to the described methodology demonstrates higher efficiency compared to the currently used models.

In conclusion, the importance of the machine learning disease prediction victimization cannot be underestimated in anyone's daily routine, but at the same time, it is especially crucial for the health care sector, because those are people who constantly use these models to detect diseases of patients based on their symptoms and provide necessary medication for it.

The further research will mainly concentrate on providing medical assistance and medication to patients to make sure that all the processes within the health care facilities can be carried out in the most efficient manner.

Acknowledgement

This work could be used in the field of Medicine, as well as basic level in all corners of the world. Hence making the medical industry stronger. The author wishes to express his heartfelt thanks to the anonymous reviewers for their valuable criticisms and suggestions regarding the improvement of the paper.

References

1. 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.
2. B. D. Kanchan and M. M. Kishor, "Study of machine learning algorithms for special disease prediction using principal of component analysis," 2016 International Conference on Global Trends in Signal Processing, Information Computing and Communication (ICGTSPICC), Jalgaon, 2016,pp.510,doi:10.1109/ICGTSPICC.2016.7955260.
3. R.D.H.D.P. Sreevalli, K.P.M. Asia, Prediction of diseases using random forest classification algorithm
4. S. Mohan, C. Thirumalai, G. Srivastava, Effective heart disease prediction using hybrid machine learning techniques, IEEE Access 7, 81542 (2019)
5. Y. Khourdifi, M. Bahaj, Heart disease prediction and classification using machine learning algorithms optimized by particle swarm optimization and ant colony optimization, Int. J. Intell. Eng. Syst. 12(1), 242 (2019)
6. Dhenakaran, K. Rajalakshmi Dr SS. "Analysis of Data mining Prediction Techniques in Healthcare Management System." International Journal of Advanced Research in Computer Science and Software Engineering 5.4 (2015)

7. Jin Ma, Sung Chan Park, Jung Hun Shin, Nam Gyu Kim, Jerry H. Seo, Jong Suk Ruth Lee, Jeong Hwan Sa. "AI based intelligent system on the EDISON platform", Proceedings of the 2018 Artificial Intelligence and Cloud Computing Conference on ZZZ - AICCC '18, 2018
8. Sayantan Saha, Argha Roy Chowdhuri et al., "Web Based Disease Detection System", IJERT, vol. 2, no. 4, April 2013, ISSN 2278-0181.
9. Andrew Alikberov, Stephan Broadly et al., "The Learning Machine", [online] Available: https://www.thelearningmachine.ai.
10. M. Chen, Y. Hao, K. Hwang, L. Wang and L. Wang, "Disease Prediction by Machine Over Learning Over Big Data From Healthcare Communities," in IEEE Access, vol. 5, pp. 8869-8879, 2017, doi: 10.1109/ACCESS.2017.2694446.

❑❑❑