Disease Detection System (DDS) Using Machine Learning Technique

Sumana De and Baisakhi Chakraborty

Abstract In this world, a human being suffers from many different diseases. Diseases can have a physical, but also a psychological impact on people. Mainly for four reasons, diseases are formed: (i) infection, (ii) deficiency, (iii) heredity and (iv) body organ dysfunction. In our society, doctors or medical professionals have the responsibility to detect and diagnose appropriate disease and provide medical therapies or treatments to cure or restrain the disease. Some diseases are cured after treatment, but chronic diseases are never cured despite the treatment; treatment can prevent chronic diseases to be worse over time. So, it is always important to detect and treat disease in early stage. To help doctors or medical professionals, this chapter proposes Disease Detection System (DDS) that can be used by doctors or medical professionals to detect diseases in patients using Graphical User Interface (GUI) of DDS. DDS is developed to detect some diseases such as Liver disorders, Hepatitis, Heart disease, Diabetes, and Chronic Kidney disease. Each of the diseases has different signs and symptoms among the patients. Different datasets are obtained from the Kaggle machine learning database to implement DDS. For the classification calculation, Adaboost Classifier Algorithm is used in DDS to detect diseases. This is a machine learning algorithm that results in the identification of referred diseases in DDS with 100% accuracy, precision and recall. The DDS GUI was created with the support of python as a screening tool so that doctors or medical professionals can easily detect patients with disease.

Keywords Diseases detection · Disease detection system (DDS) · Adaboost classifier algorithm · Query from user · Result from system

---

S. De (✉) · B. Chakraborty
Department of Computer Science and Engineering, National Institute of Technology, Durgapur, India
e-mail: sumanade@gmail.com

B. Chakraborty
e-mail: baisakhichak@yahoo.co.in

V. Jain and J. M. Chatterjee (eds.), *Machine Learning with Health Care Perspective*, Learning and Analytics in Intelligent Systems 13,
https://doi.org/10.1007/978-3-030-40850-3_6

1 Introduction

Human disease is an abnormal condition that affects the human organism negatively that may cause pain, uncomfortable, dysfunction, and even death. Diseases can have a physical, but also a psychological impact on people. And there are many different kinds of diseases in the world. According to [1] mainly for four reasons, diseases are formed: (i) infection, (ii) deficiency, (iii) heredity and (iv) body organ dysfunction. Infectious diseases are occurring for viruses, bacteria, fungus, worms, arthropods, etc. And it can be treated using antibiotic, antiviral, antifungal, antiprotozoal or anthelmintic medicines. Deficiency diseases are caused by long time nutrition deficiency in the human body. Hereditary diseases are caused either by genetic diseases or non-genetic diseases. Physiological diseases are caused when the proper functioning of the body malfunction because of malfunctioning of body organs. For example, asthma, glaucoma, diabetes etc.

In our society, doctors or medical professionals have the responsibility to detect and diagnose appropriate disease and provide medical therapies or treatments to cure the disease. However, chronic diseases continue to persist which may worsen over time. So, it is always important to detect and treat disease in early stage. To help doctors or medical professionals, this chapter proposes a Disease Detection System (DDS) that can be used by doctors or medical professionals to detect diseases in patients using Graphical User Interface (GUI) of DDS. DDS is developed to detect some diseases such as Liver disorders, Hepatitis, Heart disease, Diabetes, and Chronic kidney disease. Liver disorder is a damaging condition of the liver. It can be inherited or caused by other factors like viruses or excessive use of alcohol. There are many symptoms of liver disorder such as skin and eyes turning yellowish, abnormal abdomen pain, pale stool color, etc. Doctors diagnose the disorder testing CT scan, MRI reports, testing blood and analyzing tissue. Hepatitis is an inflammatory condition of the liver. There may be many different causes for hepatitis. The most common cause is a viral infection. Doctor diagnose hepatitis testing liver function and testing blood. Heart disease is a disorder of heart. There may be many types of heart disease such as diseases in blood vessels, in a coronary artery, in heart rhythm and inherited disease in the heart. A Patient can feel chest pain, chest discomfort, breathing difficulty. Doctors diagnose the disease checking physical condition, family history and testing blood, chest x-ray, ECG report, etc. Diabetes is a disorder of metabolism that increases sugar in blood. Type 1 and Type 2 can be classified by diabetes. Enough insulin is not generated in body in type 1 and for type 2, the body is able to generate sufficient insulin, but it cannot properly use the insulin. Doctors diagnose diabetes testing blood sugar level, insulin level, fasting plasma glucose, etc. Chronic kidney disease is indeed the kidney's destructive situation that can become more severe over the period. When the kidneys are very severely damaged, they will stop working. So, early diagnosis of chronic kidney disease is very significant. Doctors diagnose Chronic kidney disease testing patient's urine and blood. So, all these diseases have different symptoms and different diagnosis process. To help doctors in case of diagnosis of these diseases, Disease Detection System (DDS) has been developed as an assistant tool of a doctor. To implement DDS, different datasets for different diseases are collected from the database of machine learning in Kaggle. The Liver disorders dataset contains 583 instances of observations that have total 11 attributes, the dataset of hepatitis contains 155 instances of observations that have total 20 attributes, the dataset of Heart disease contains 303 instances of observations that have total 14 attributes, The dataset of Diabetes contains 768 instances of observations that have total 9 attributes, and the dataset of chronic kidney disease contains 400 instances of observations that have total 25 attributes. The machine learning algorithm, AdaBoost Classifier Algorithm is used in DDS to detect diseases. In machine learning, boosting is a method that merges comparatively weak and incorrect laws to create a law of prediction that is highly accurate. AdaBoost Classifier Algorithm provides the results of detection of these diseases with 100 percent accuracy, precision and recall. The DDS Graphical User Interface (GUI) was created with the support of python as a screening tool so that doctors or medical professionals can easily detect patients with any disease. The chapter's remaining details are as described: Sect. 2 concentrates on earlier relevant works, Sect. 3 on the methodology and implementation of the system, Sect. 4 deals with the proposed DDS model, Sect. 5 provides a correlation of accuracy between previous works related to DDS, Sect. 6 describes the outcome of the simulation, and then, Sect. 7 comes to a conclusion.

2 Previous Related Works

In today's world disease diagnosis through the computer becomes a very popular topic in the research area. In this research case, machine learning algorithms have a very important role. Because machine learning algorithms are fast and accurate to detect any diseases. Many machine learning diagnostic applications of AI have been successfully made in the present world for the diagnosis of diseases. Much research was done in the field of medicine in the earlier years where researchers utilized machine learning algorithms to identify diseases. This section of the chapter is going to discuss about some recent previous works in which algorithms have been utilized for machine learning to identify diseases. The journal paper [2] provides a survey that analyses different machine learning algorithms to diagnose different diseases like a disease in heart, diabetes, disease in the liver, dengue and hepatitis. This paper focused on the different machine learning algorithms and the most utilized tools to analyze different diseases.

To help doctors and patients to detect disease in early stage, a project based on machine learning algorithms has been discussed in [3]. The dataset has been used in this project is purely text based. This paper has designed a system where a doctor can enter the symptoms of patients and diagnoses common diseases. The paper [4] shows a review on disease diagnosis using machine learning techniques. This paper shows that machine learning is used for the high dimensional and the multi-dimensional data and concludes some limitations of machine learning algorithms. Another study on machine learning algorithms for medical diagnosis is done in paper [5]. This paper focused on the use of different machine learning algorithms for accurate medical diagnosis. A deep study of machine learning algorithms for disease diagnosis is done in paper [6]. This paper concentrates on latest developments in machine learning that have significantly affected the detection and diagnosis of different diseases. Many researches are done to predict heart diseases such as in [7], the research paper used various machine learning algorithms such as SVM, RF, KNN and ANN classification algorithms were used to identify early-stage heart disease. A diagnostic system was developed in the research paper [8] using NN, capable of predicting reliably the level of risk of heart problem. In this world, congestive heart failure has become one of the primary causes of death. So, to prevent congestive heart failure, the research paper [9] has designed system using algorithms like Boosted Decision Tree, CNN. The research [10] aimed to establish a rapid and accurate automatic detection of ischemic heart disease. The machine learning algorithm, XGBoost classifier is used here to establish the task. Liver disease is one of the serious diseases and it should be diagnosed in the early stage so many researches are done to predict liver diseases such as in [11], liver disease is detected using machine learning algorithms where five different phases are used and it is seen that for feature selection, J48 algorithm works better.

The paper [12] proposes an electromagnetic system that includes an antenna as a data capture tool and a supervised Machine Learning (ML) system to learn directly from gathered data an inferring model for FLD.

This paper aims to get a better diagnosis of liver diseases the paper [13] proposes 2 methods of identification, one is patient parameters and second is genome expression. The machine learning technique is used here for the diagnosis.

Four classification models are developed in [14] to diagnose fatty liver disease accurately. Among the four models this paper shows that the random forest model performs better.

The project [15] aims at improving the diagnosis of liver disease using approaches to machine learning. This research's main goal is to use classification algorithms to classify healthy individuals' liver patients.

Hepatitis disease has been predicted in [16] using machine learning models. Two ML algorithms are developed and compared in this paper; to predict the development of cirrhosis in a large cohort infected with CHC.

To predict Hepatitis B, the research article [17] has developed four machine learning models based on four different algorithms, that includes the decision tree (DCT), extreme gradient boosting (XGBoost), logistic regression (LR) and random forest (RF).

One of the popular techniques of machine learning is the Neural Network (NN). The paper [18] shows diagnosis of hepatitis disease using different NN architectures.

In paper [19] logistic regression, random forest, decision tree, C4.5 and multilayer perceptron classifier algorithms are used to predict Hepatitis—Infectious Disease. In this analysis random forest classifier algorithm predicts accurately in a minimum time.

In the paper [20], for the smart prediction of chronic kidney disease (CKD), DFS with D-ACO algorithm has been used to develop a smart health monitoring system.

Data mining algorithms are used in [21] to predict kidney disease stages. Among those PNN algorithm provides better performance in classification and prediction.

To predict chronic kidney disease, two different machine learning algorithms like DT and SVM are used in [22], among which SVM performs better.

For predicting the chronic kidney disease in [23] four machine learning methods are used, out of which SVM classifier gives highest accuracy.

A model based on XGBoost is developed with better accuracy for the prediction of Chronic Kidney Disease in [24].

Type-2 Diabetes is predicted using a machine learning algorithm on paper [25]. Support vector machine is used to implement the prediction model.

The experiment results in the research paper [26] show that the random forest algorithm is very efficient to develop a powerful machine learning model to predict diabetes.

To predict diabetic mellitus among the adult population, the paper [27] used four different machine learning algorithms. Among these algorithms only C4.5 decision tree provides higher accuracy.

Different tree classifiers for machine learning are evaluated based on their True Positive Rate (TPR) and accuracy in [28]. Higher precision was achieved in this study using Logistic Model Tree (LMT) for predicting diabetes mellitus.

3 System Implementation and Disease Detection Methodology

Figure 1 shows the steps to implement DDS and to diagnose diseases. Detail about the steps is discussed in following:

1. Datasets collection: Different datasets are obtained from the machine learning database of Kaggle to implement DDS. The dataset of liver disorders contains

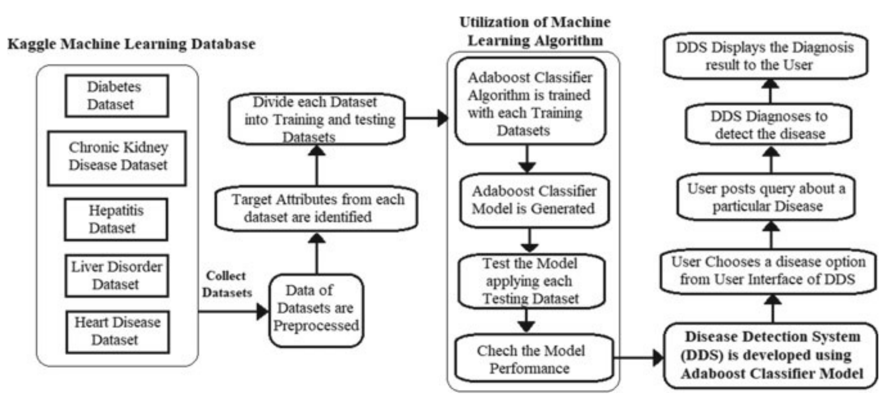

Fig. 1 DDS implementation and detection of diseases

Table 1 Attributes-values in liver disease dataset

Input attributes	*Values*
Age(in years)	4–90
Gender	Male, Female
Total_Bilirubin	0.4–75
Direct_bilirubin	0.1–19.7
Alkaline_Phosphate	63–2110
Alamine_Aminotransferase	10–2000
Aspartate_Aminotransferase	10–4929
Total_Protiens	2.7–9.6
Albumin	0.9–5.5
Albumin_and_Globulin_Ratio	0.3–2.8
Output attribute	*Values*
Dataset	1 = liver disease; 2 = no liver disease

583 instances of observations that have total 11 attributes, the dataset of hepatitis contains 155 instances of observations that have total 20 attributes, the dataset of heart disease contains 303 instances of observations that have total 14 attributes, the dataset of diabetes contains 768 instances of observations that have total 9 attributes, and the dataset of chronic kidney disease contains 400 instances of observations that have total 25 attributes. The following tables show the datasets detail (Tables 1, 2, 3, 4 and 5).

2. Data of datasets are preprocessed: Pre-processing data strategy is used to turn imperfect actual data into a valuable and usable form. For example, convert all nominal data to numerical data; the missing attribute values are substituted by the attribute's calculated average value.
3. Target attributes from each datasets are identified: Target attribute is identified from each dataset, such as liver disorder has "Dataset" attribute as the target attribute that contains two classes: "liver disease yes" and "liver disease no", hepatitis has "Class" attribute as the target attribute that contains two classes: "Die" and "alive", heart disease has "num" attribute as the target attribute that contains two classes: "heart disease yes" and "heart disease no", diabetes dataset contain 'outcome' attribute as the target attribute, that contains two classes: "Diabetic" and "Not-Diabetic", and Chronic kidney disease has "Classification" attribute as the target attribute that contains two classes: "Ckd" and "Not-ckd".
4. Divide each dataset into training and testing datasets: Excluding the column of the target attribute, the datasets are categorized into 2 sets with 7:3 proportions, 70% of it is utilized to train the model of machine learning, while 30% is utilized to test the model's precision, accuracy and recall tests.
5. Adaboost classifier algorithm is trained with each training datasets: AdaBoost classifier Algorithm is used to detect diseases. In machine learning boosting is a method that merges comparatively weak and incorrect laws to create a law of

Table 2 Attributes-values in hepatitis disease dataset

Input attributes	*Values*
Age (in years)	10–80
Sex	1 = male, 2 = female
Steroid	1 = no, 2 = yes
Antivirals	1 = no, 2 = yes
Fatigue	1 = no, 2 = yes
Malaise	1 = no, 2 = yes
Anorexia	1 = no, 2 = yes
Liver_big	1 = no, 2 = yes
Liver_firm	1 = no, 2 = yes
Spleen_palable	1 = no, 2 = yes
Spiders	1 = no, 2 = yes
Ascites	1 = no, 2 = yes
Varices	1 = no, 2 = yes
Bilirubin	0.39–4.0
Alk_phosphate	33–250
Sgot	13–500
Albumin	2.1–6.0
Protime	10–90
Histology	1 = no, 2 = yes
Output attribute	*Values*
Class	1 = die, 2 = live

prediction that is highly accurate. The classification of adaBoost is an iterative assembly process. After merging many poorly performing classifiers it creates a strong classifier to improve classifier accuracy. The main idea of adaboost is just to define the classifier weights and train the data set for each iteration to assure that uncommon events are predicted accurately. The Training Dataset trains the adaboost classifier algorithm and the Model of Machine Learning is produced afterwards. The main idea is to be to construct a machine learning model that is capable of receiving inputs and using statistical analysis detect an outcome. The model's performance is checked with respect to accuracy, precision and recall, where Testing Dataset is applied as the inputs to the Machine Learning Model. So, adaboost Classifier Algorithm is trained with each different disease's training dataset.

6. Adaboost classifier model is generated: Whenever Adaboost Classifier Algorithm is trained with the different training datasets, then Adaboost Classifier Model is generated for each of the disease detection. This Machine Learning Model is able to take input for different diseases and diagnose the specific disease appropriately.

Table 3 Attributes-values in heart disease dataset

Input attributes	*Values*
Age (in years)	29–77
Sex	(1 = male; 0 = female)
cp (Chest pain type)	1 = typical angina 2 = atypical angina 3 = non-anginal pain 4 = asymptomatic
Trestbps [Resting blood pressure (in mm Hg on admission to the hospital)]	94–200
Chol (Serum cholestoral in mg/dl)	126–564
Fbs (Fasting blood sugar > 120 mg/dl)	(1 = true; 0 = false)
Restecg (Resting electrocardiographic results)	0 = normal 1 = abnormal ST-T wave 2: probable or definite left ventricular hypertrophy
Thalach (Maximum heart rate achieved)	71–202
Exang (Exercise induced angina)	1 = yes; 0 = no
Oldpeak (ST depression induced by exercise relative to rest)	0–6.2
Slope (The slope of the peak exercise ST segment)	1 = not sloping 2 = flat 3 = down sloping
ca (Number of major vessels colored by flourosopy)	0–3
Thal	1 = normal; 2 = fixed defect; 3 = reversible defect
Output attribute	*Values*
Target [Diagnosis of heart disease (angiographic disease status)]	Value 0: <50% diameter narrowing Value 1: >50% diameter narrowing

7. Test the model applying each testing dataset: Once the Adaboost Classifier Model is generated. To test its performance the testing datasets for each disease are applied on the Model.
8. Check the model Performance: The quality of the model is tested for accuracy, precision, and the outcome of the recall. The adaboost classifier model delivers 100% accuracy, precision and recall outcomes in detection of each disease.

Table 4 Attributes-Values in diabetes disease dataset

Input attributes	*Values*
Pregnancies	0–17
Glucose	0–199
Blood pressure	0–122
Skin thickness	0–99
Insulin	0–846
BMI	0–67.1
Diabetes pedigree function	0.078–2.42
Age (in years)	21–81
Output attribute	*Values*
Outcome	0–1

9. Disease Detection System (DDS) is developed using adaboost classifier model: The adaboost classifier model is used to develop the Disease Detection System (DDS). The Graphical User Interface (GUI) of the DDS is designed using Jupyter Notebook in Python Environment.
10. User chooses a disease option from the User Interface (UI) of the system: The GUI of the system shows five options for disease selection. User can select any of the diseases for further diagnosis.
11. User posts query about a particular disease: After selecting the option, user can post queries about a particular disease through another form provided by the system.
12. DDS diagnoses to detect the disease: After getting input from the user, DDS diagnose the disease with the help of adaboost classifier model to detect the disease.
13. DDS displays the diagnosis result to the user: After completing the diagnosis calculation, the final result is to be displayed to the user through the GUI of the DDS.

4 Proposed DDS Model

4.1 Architecture of DDS

DDS is the request-response system in which the user needs to submit a question as a request form and the DDS offers a response with the detection result. Figure 2 shows DDS model has following architectural components.

1. Data Entry Module or User Interface (UI): It is the responsibility of the UI to connect the user to the system. A user can view the question form via the UI that

Table 5 Dataset attributes-values of chronic kidney disease

Input attributes	*Values*
Age (in years)	2–90
bp [Blood pressure (mm/Hg)]	50–100
sg (Specific gravity)	1.005–1.025
al (Albumin)	0–4
su (Sugar degree)	0–4
rbc (Red blood cells)	Normal, Abnormal
pc (Pus cell)	Normal, Abnormal
pcc (Pus cell clumps)	Present, Not present
ba (Bacteria)	Present, Not present
bgr [Blood glucose random (mgs/dl)]	22–490
bu [Blood urea (mgs/dl)]	1.5–391
sc [Serum creatinine (mgs/dl)]	0.4–7.6
pcv (Packed cell volume)	0–54
pot [Potassium (mEq/L)]	0–4.7
sod [Sodium (mEq/L)]	0–163
hemo [Hemoglobin (gms)]	0–17.8
wbcc [White blood count (cells/cumm)]	0–26,400
rbcc [Red blood cell count (millions/cmm)]	0–8
htn (Hypertension)	Yes, No
dm (Diabetes mellitus)	Yes, No
cad (Coronery artery disease)	Yes, No

(continued)

Table 5 (continued)

appet (Appetite)	Good, Poor
pe (Pedal edema)	Yes, No
ane (Anemia)	Yes, No
Output attribute	*Values*
Classification	ckd, notckd

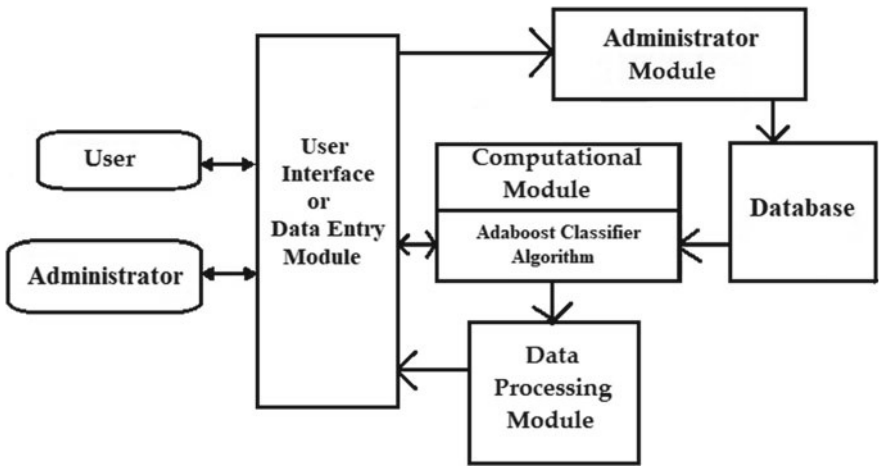

Fig. 2 Components of architecture in DDS

is provided by the system, as well as the user fills out and sends the form to the system with values.

2. User: The user is meant the doctors or medical professionals. They log into the system, provide data on the health status of the patient to the system via the UI. The user can supply the system with any of the five disease-related data.
3. Computational Module: The responsibility of Computational Module is to classify whether the recently posted data indicates disease or not. For example, if the user provides data for checking hepatitis, then The Computational module is responsible for classifying whether or not the posted input data belong to the class of hepatitis. In this way, similarly, other diseases may be detected. Computational Module uses adaboost Classifier model to calculate the classification. And it estimates the accuracy, precision, and recall outcome of the system as well.
4. Data Processing Module: Once the Computational Module performs the classification the Data Processing Module tests the outcome of the diagnosis. When it determines that the patient has the disease, it will send the user a message that the patient is suffering from that particular disease, otherwise it will demonstrate

that the patient has no particular disease. The system accuracy is also displayed to the user by this module.

5. Administrator Module: This module supports DDS administration, administrators. Just administrators can insert, remove, upgrade, and alter the database dataset information.
6. Administrator: doctors or medical professionals should be the Administrator, who will have adequate knowledge of the diseases. With valid data, they can upgrade the datasets or erase unwanted records from Datasets.
7. Database: The database stores all of the Datasets of diseases (Liver disorders, Hepatitis, Heart disease, Diabetes, and Chronic Kidney disease).

4.2 Use Case Diagram of DDS

Multiple consecutive actions performed by the DDS are shown in Fig. 3 use case diagram. In the use case diagram, the actor is DDS. The detail about the actions, performed by DDS is discussed in below:

Check User Input: User can choose a disease option from the user interface of DDS. Whenever the user selects the disease, to diagnose the disease, the system provides a form and asks users to fill up the form with required data. Then through the user interface of the system user submits the required data and fills up the form.

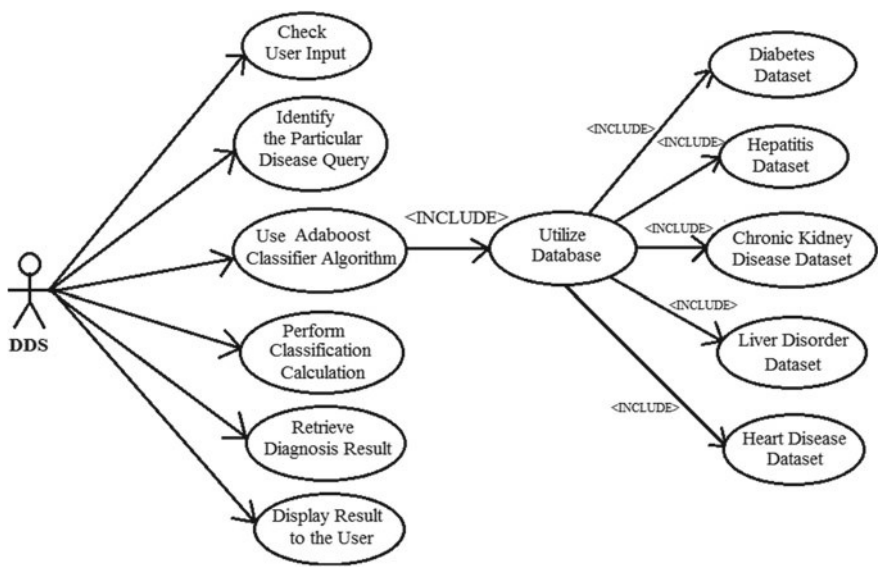

Fig. 3 DDS use case diagram

- Identify the Particular Disease Query: Whenever a user submits the form, the system receives the set of data and can identify which particular disease it should diagnose.
- Use Adaboost Classifier Algorithm: DDS utilizes the system database that includes the different datasets. For a particular disease diagnosis, 70% of that particular dataset is used to train the adaboost Classifier algorithm. Then adaboost classifier model is generated that is able to classify the new input.
- Perform Classification Calculation: Now, DDS uses the adaboost classifier model to classify that according to the user data for particular disease the patient has the disease or not. For example, if the user wants to diagnose heart disease and he submits data according to system demands, then system utilize 70% of heart disease dataset to train the adaboost classifier algorithm then an adaboost classifier model is generated which can classify that according to the user inputted data the patient has the heart disease or not.
- Retrieve Diagnosis Result: After classification calculation is done, the system gets the diagnosis result.
- Display Result to the User: The diagnostic result will then be displayed to the user through the system's UI module.

4.3 Context Diagram of the DDS

Figure 4 depicts the DDS context diagram. The user enters into the Disease Detection System and the system displays a form that includes five diseases options such as Liver disorders, Hepatitis, Heart disease, Diabetes, and Chronic Kidney disease options. The user must select one of the diseases from the options that he/she wants to diagnose. Then the system checks the selected option submitted by the user and provides another form according to that particular selected disease. If the user does not select any option and submits the form, then system finds an error and asks the user to select an option again. When the system finds what disease, the user has selected, then it displays a form that asks some questions about that selected disease. The query form asks the user to provide some values of the disease input attributes. Whenever the user completes the form and sends it to the system, the system must approve the form otherwise the system will show the user's error message and the user will have to resubmit the form. After accepting the form, the computation module of the system uses the adaboost classifier algorithm to calculate the classification. The computational module generates adaboost classifier model and then feed the new inputted set of data to the model. The model will classify that according to the inputted data the patient has the disease or not. For example, if the user has provided Hepatitis data, then the computational model uses Hepatitis dataset from the database to train the adaboost classifier algorithm and the adaboost classifier model will classify that according to that inputted hepatitis data the patient has hepatitis or not. After classification is done the Data Processing Module will obtain

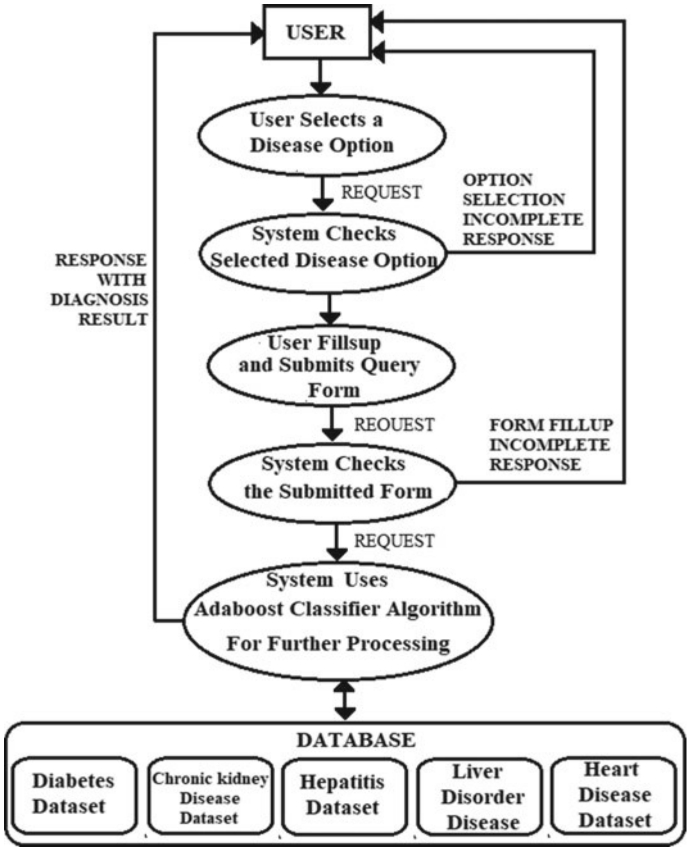

Fig. 4 DDS context diagram

the outcome of the diagnosis and show it to the user. If the diagnosis result is '1' then the Data Processing Module shows the user a message that 'The patient is suffering from Hepatitis' otherwise, for the diagnosis result '0', the user gets the message 'The patient is not suffering from Hepatitis'.

5 Accuracy Comparison

In the past many researches were done in the medical field where researchers used different machine learning algorithms to detect different diseases. This segment of the chapter shows the best accuracy of machine learning algorithms from various previous works. And compares their performances with our proposed machine learning model's performance.

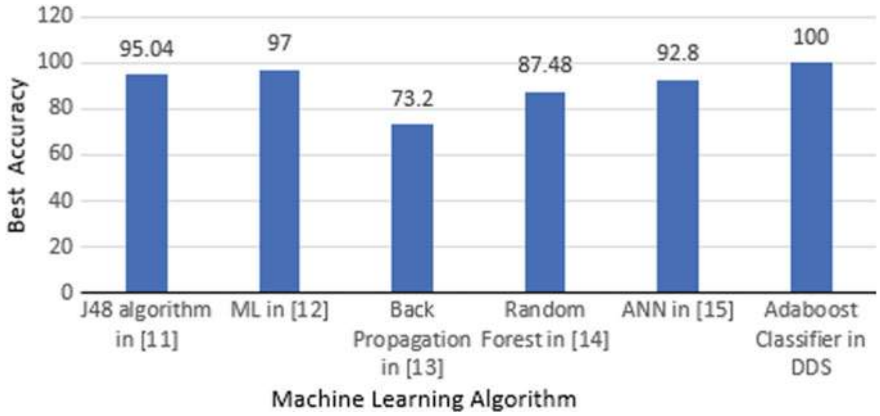

Fig. 5 Accuracy comparison for liver disease detection

5.1 Accuracy Comparison of DDS with Previous Works

5.1.1 Liver Disease Detection

From the different previous works, mentioned in Sect. 2 in this chapter, to diagnose liver disease, different machine learning algorithms are used. And different machine learning algorithms provide different best accuracies to detect the disease. Figure 5 shows the comparison graph where the accuracy of the adaboost classifier algorithm in DDS is compared with other best accuracies from different previous works those diagnose liver Disease.

5.1.2 Hepatitis Disease Detection

From the different previous works, mentioned in Sect. 2 in this chapter, to diagnose Hepatitis disease, different machine learning algorithms are used. And different machine learning algorithms provide different best accuracies to detect the disease. Figure 6 shows the comparison graph where the accuracy of the adaboost classifier algorithm in DDS is compared with other best accuracies from different previous works those diagnose Hepatitis Disease.

5.1.3 Heart Disease Detection

From the different previous works, mentioned in Sect. 2 in this chapter, to diagnose heart disease, different machine learning algorithms are used. And different machine learning algorithms provide different best accuracies to detect the disease. Figure 7 shows the comparison graph where the accuracy of the adaboost classifier algorithm

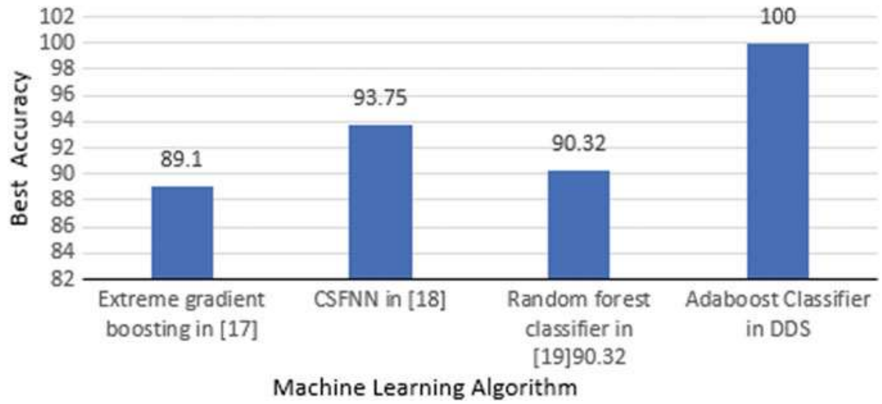

Fig. 6 Accuracy comparison for hepatitis disease detection

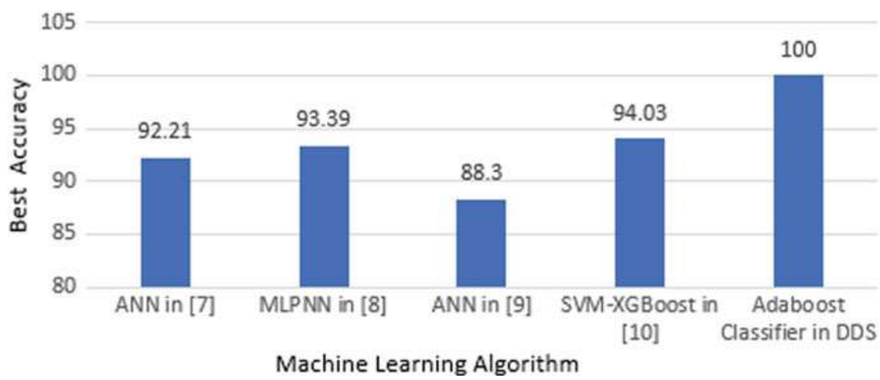

Fig. 7 Accuracy comparison for heart disease detection

in DDS is compared with other best accuracies from different previous works those diagnose Heart Disease.

5.1.4 Diabetes Disease Detection

From the different previous works, mentioned in Sect. 2 in this chapter, to diagnose Diabetes disease, different machine learning algorithms are used. And different machine learning algorithms provide different best accuracies to detect the disease. Figure 8 shows the comparison graph where the accuracy of the adaboost classifier algorithm in DDS is compared with other best accuracies from different previous works those diagnose Diabetes Disease.

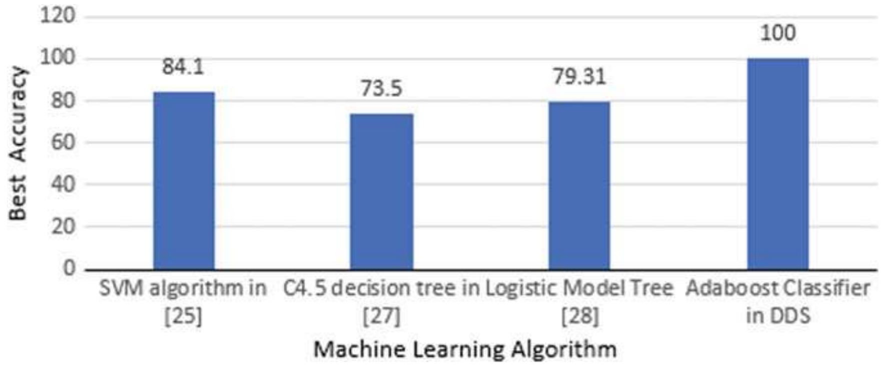

Fig. 8 Accuracy comparison for diabetes disease detection

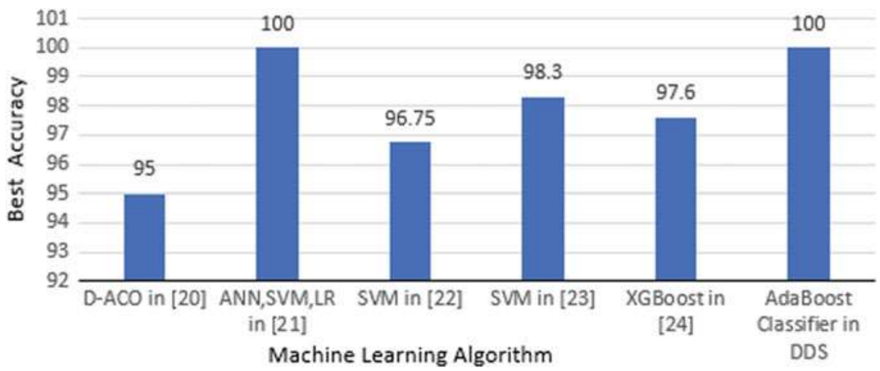

Fig. 9 Accuracy comparison for chronic kidney disease detection

5.1.5 Chronic Kidney Disease Detection

From the different previous works, mentioned in Sect. 2 in this chapter, to diagnose Chronic Kidney disease, different machine learning algorithms are used. And different machine learning algorithms provide different best accuracies to detect the disease. Figure 9 shows the comparison graph where the accuracy of the adaboost classifier algorithm in DDS is compared with other best accuracies from different previous works those diagnose Chronic Kidney Disease.

6 Simulation for Result

In this chapter, to design the Graphical User Interface (GUI) of the DDS, Jupyter Notebook is used as the simulation tool in Python Environment. As the python library Scikit Learn module has been used that supports different classifications, regression and clustering algorithms and includes Python numerical and scientific libraries NumPy and SciPy respectively. Figure 10 shows the first display form of the DDS that is used by the doctors as the gateway to detect different diseases.

Whenever the doctor wants to detect a disease, then he has to click the appropriate button for diagnosis the disease and then the Following forms will appear. Figures 11 and 12 show the liver disease diagnosis results. Figures 13 and 14 show the diagnosis results for Hepatitis Disease. Figures 15 and 16 show the diagnosis results for Heart Disease. Figures 17 and 18 show the diagnosis results for Diabetes Disease. Figures 19 and 20 show the diagnosis results for Chronic Kidney Disease.

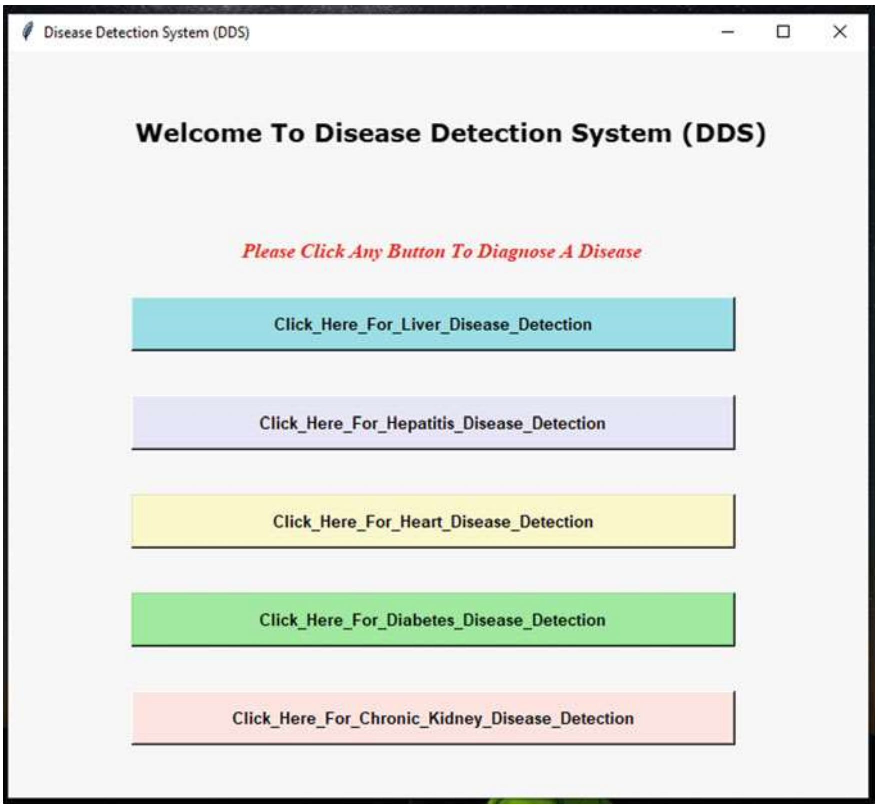

Fig. 10 First display form of DDS

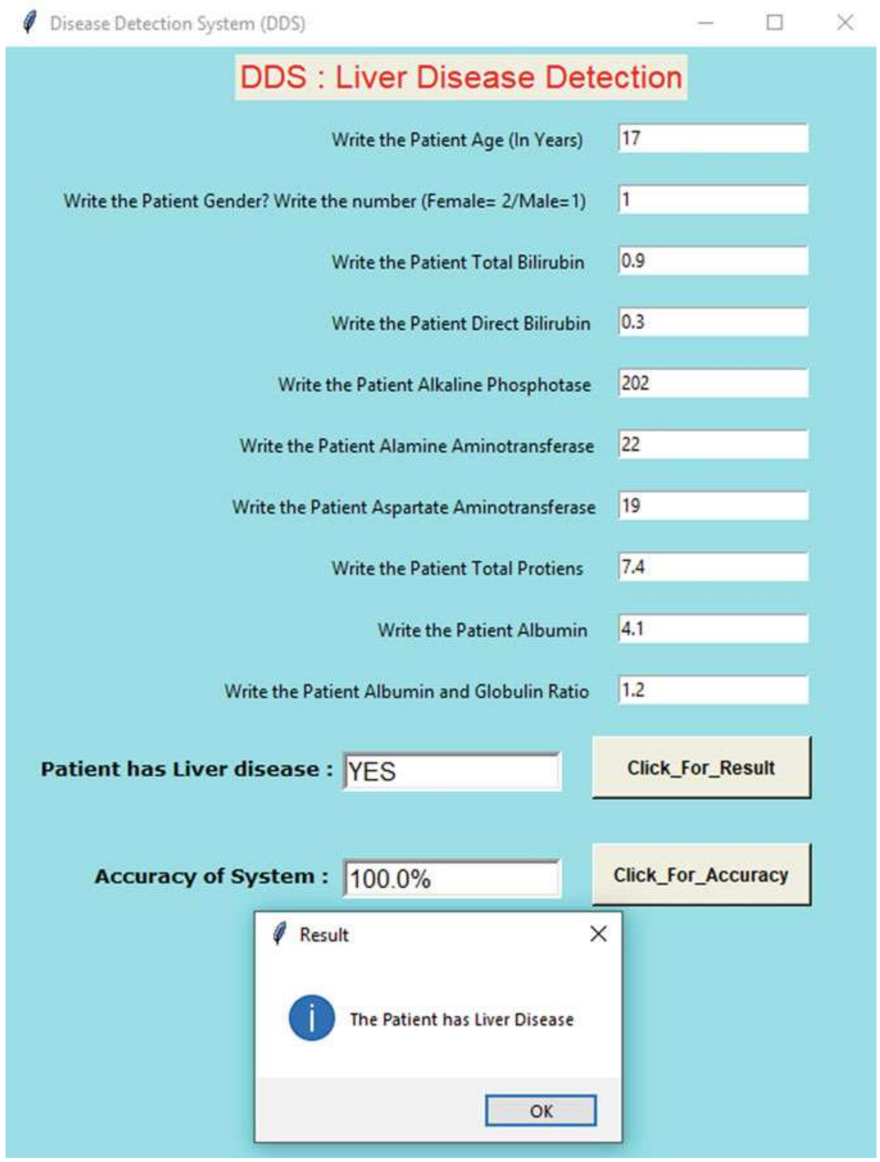

Fig. 11 Diagnosis result is positive for liver disease

7 Conclusion

This chapter proposes a Disease Detection System (DDS) that uses Adaboost Classifier Algorithm as a machine learning algorithm. DDS is designed to help doctors or medical professionals to detect five diseases such as Liver disorders, Hepatitis,

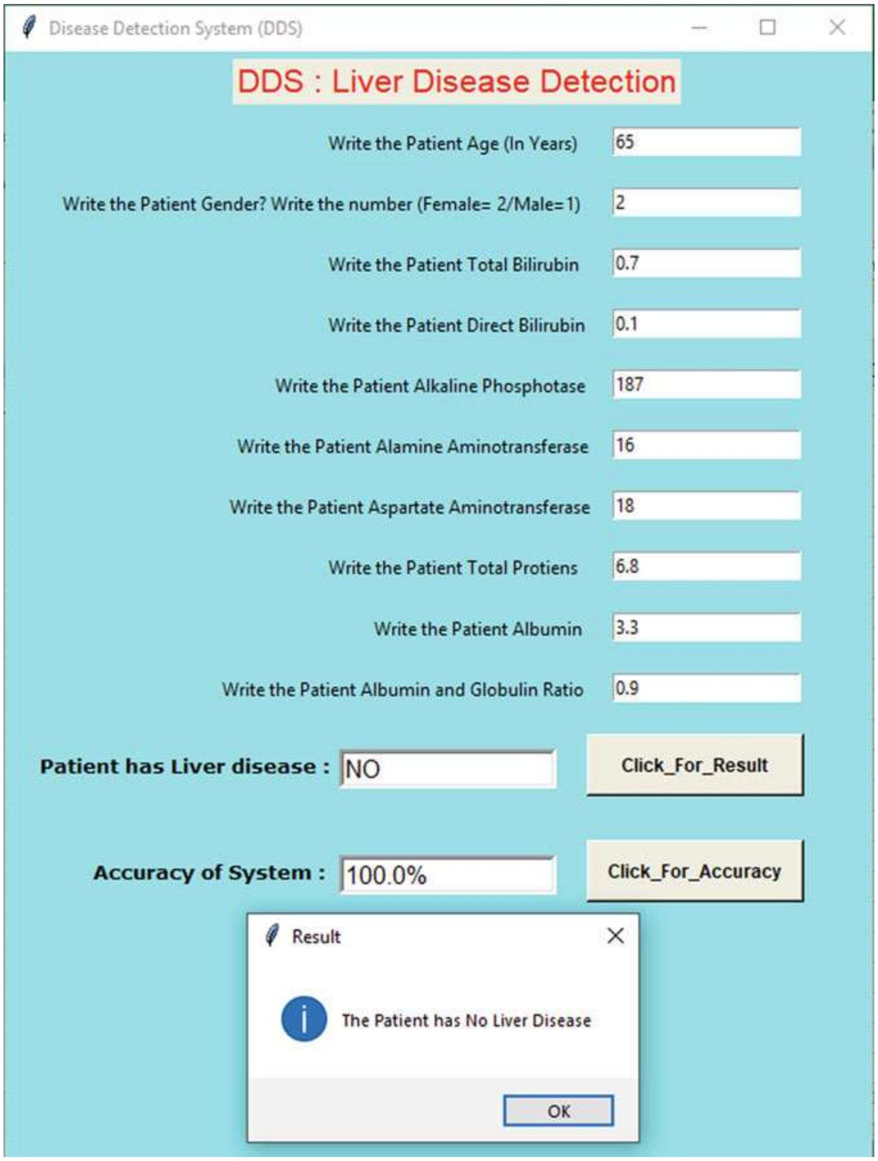

Fig. 12 Diagnosis result is negative for liver disease

Heart disease, Diabetes, and Chronic Kidney disease. For the implementation of the DDS and detect diseases this chapter follows some steps such as: from the machine learning database of Kaggle, different datasets for different diseases are obtained, pre-processing data strategy is used to turn imperfect actual data into a valuable and usable form, target attribute is identified from each dataset, excluding the column

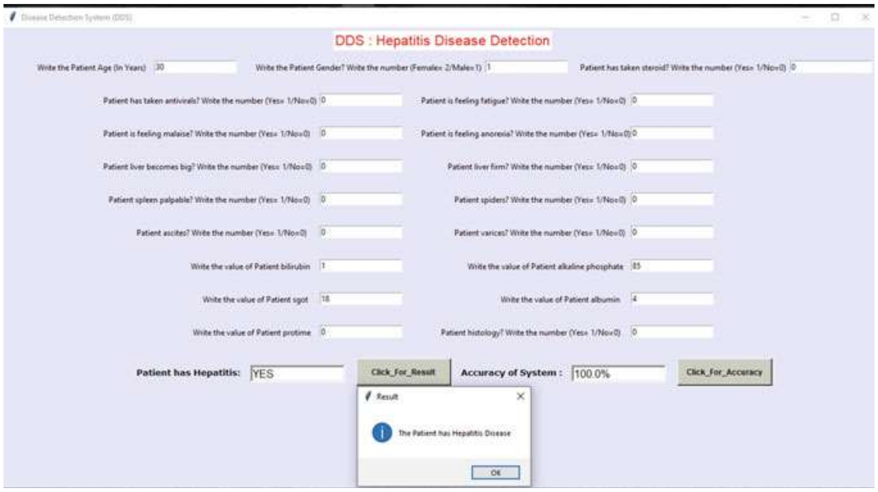

Fig. 13 Diagnosis result is positive for hepatitis disease

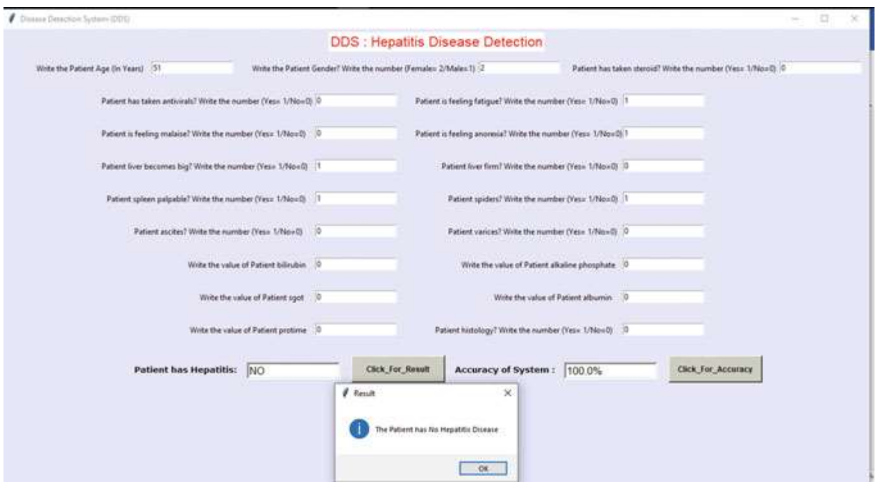

Fig. 14 Diagnosis result is negative for hepatitis disease

of the target attribute, the datasets are categorized into 2 sets with 7:3 proportions, adaboost classifier algorithm is trained with different disease's training dataset and adaBoost classifier model is generated for each of the disease detection that is able to detect different diseases. To test the Performance of adaboost classifier model, different disease's testing dataset are applied to their adaboost classifier model. It is found that adaboost classifier model provides 100% accuracy, precision and recall results in detection of each disease in the DDS, finally Disease Detection System (DDS) is implemented using the adaboost classifier model. In Python Environment,

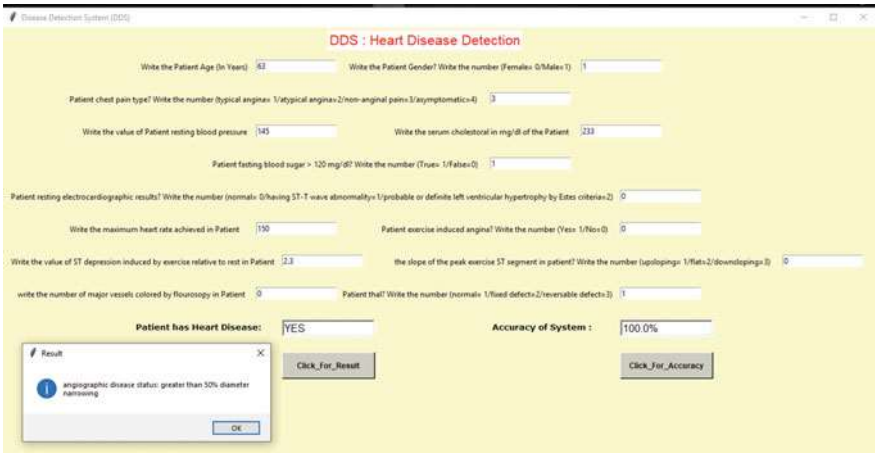

Fig. 15 Diagnosis result is positive for heart disease

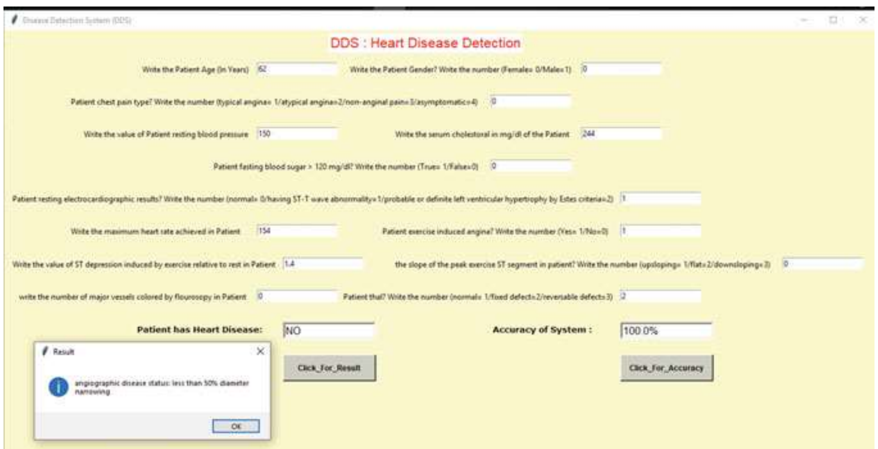

Fig. 16 Diagnosis result is negative for heart disease

DDS Jupyter Notebook's Graphical User Interface (GUI) is used as a simulation tool so that doctors or professionals in medicine can easily diagnose any disease among patients. Once DDS is developed, doctors or medical professionals select one disease from the five disease options in the GUI of the DDS for detection. Then through the GUI of the system he/she submits the required patient's data and fills up the form. Lastly system uses the adaboost Classifier model to detect whether the patient has the specified disease or not and displays the result to the doctor or medical professional. Accuracies of various machine learning algorithms from various previous related works were compared in this chapter and it is shown that DDS works better.

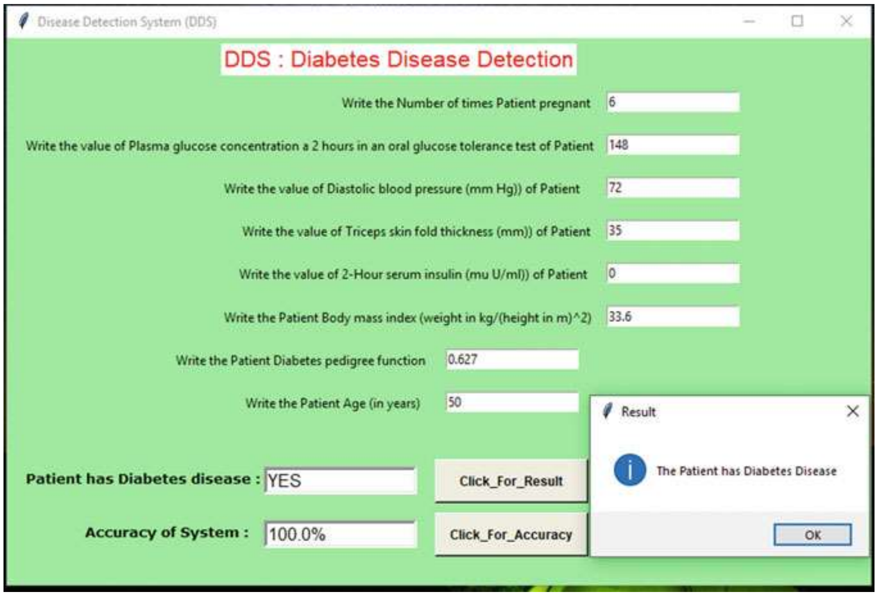

Fig. 17 Diagnosis result is positive for diabetes disease

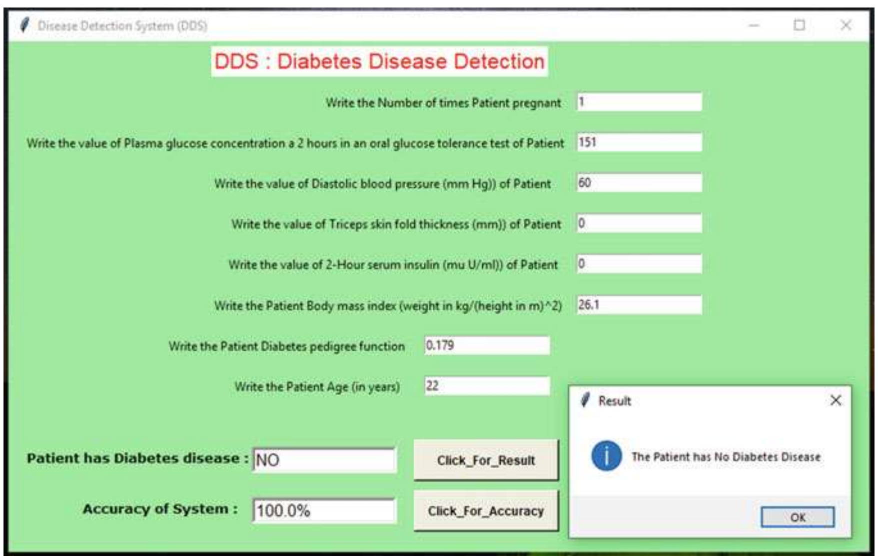

Fig. 18 Diagnosis result is negative for diabetes disease

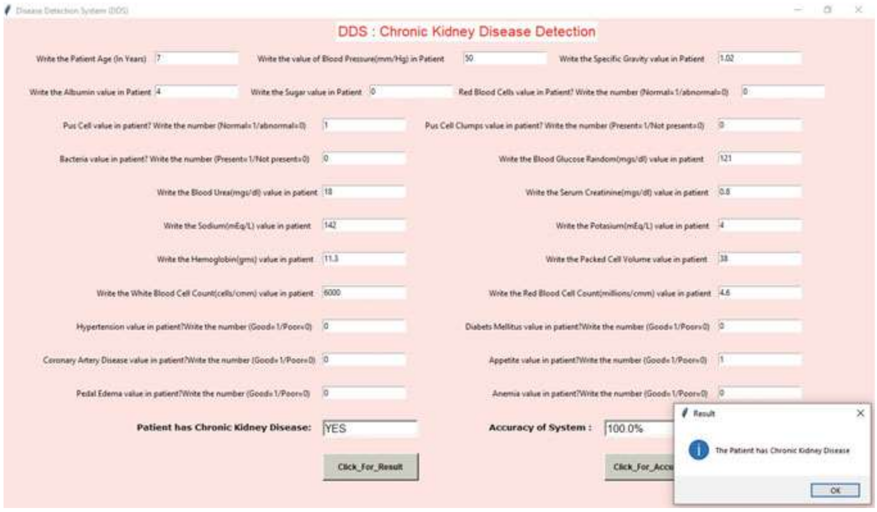

Fig. 19 Diagnosis result is positive for chronic kidney disease

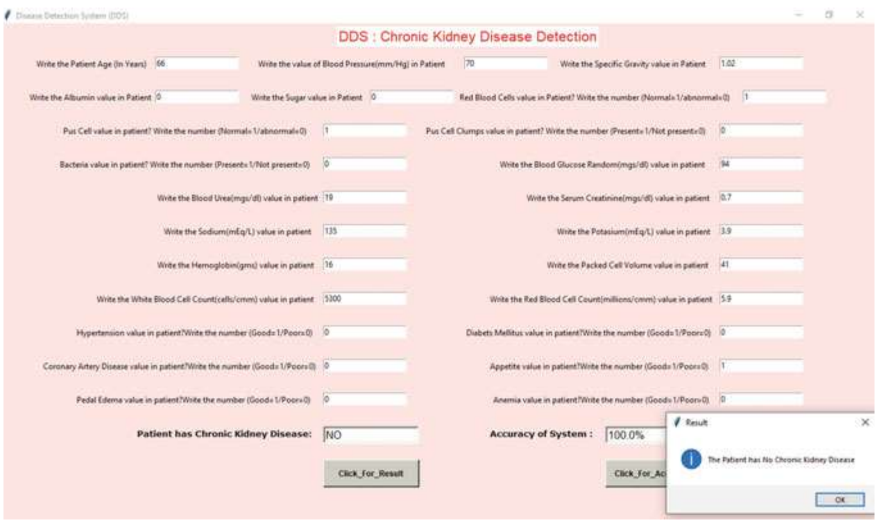

Fig. 20 Diagnosis result is negative for chronic kidney disease

References

1. Disease. Available: https://en.wikipedia.org/wiki/Disease. Last accessed 2019
2. M. Fatima, M. Pasha, Survey of machine learning algorithms for disease diagnostic. J. Intell. Learn. Syst. Appl. **9**, 1–16 (2017)

3. A.D. Sunny, S. Kulshreshtha, S. Singh, Ba M. Srinabh, H. Sarojadevi, Disease diagnosis system by exploring machine learning algorithms. Int. J. Innov. Eng. Technol. (IJIET) **10**(2), 14–21 (2018)
4. S. Razia, P. Swathi Prathyusha, N. Vamsi Krishnan, S. Sumana, A review on disease diagnosis using machine learning techniques. Int. J. Pure Appl. Math. **117**(16), 79–85 (2017)
5. D. Pavithra, A.N. Jayanthi, A study on machine learning algorithm in medical diagnosis. Int. J. Adv. Res. Comput. Sci. **9**(4), 42–46 (2018)
6. A.J. Dinu, R. Ganesan, F. Joseph, V. Balaji, A study on deep machine learning algorithms for diagnosis of diseases. Int. J. Appl. Eng. Res. **12**(17), 6338–6346 (2017)
7. P. Mamatha Alex, S P Shaji, Prediction and diagnosis of heart disease patients using data mining technique, in *International Conference on Communication and Signal Processing* (2019), pp. 0848–0852
8. K. Subhadra, B. Vikas, Neural network based intelligent system for predicting heart disease. Int. J. Innov. Technol. Exploring Eng. (IJITEE) **8**(5), 484–487 (2019)
9. S. Khade, A. Subhedar, K. Choudhary, T. Deshpande, U. Kulkarni, A system to detect heart failure using deep learning techniques. Int. Res. J. Eng. Technol. (IRJET) **6**(6), 384–387 (2019)
10. R. Tao, S. Zhang, X. Huang, M. Tao, J. Ma, S. Ma, C. Zhang, T. Zhang, F. Tang, J. Lu, C. Shen, X. Xie, Magnetocardiography based ischemic heart disease detection and localization using machine learning methods. IEEE Trans. Biomed. Eng. **66**(6), 1658–1667 (2019)
11. V. Durai, S. Ramesh, D. Kalthireddy, Liver disease prediction using machine learning. Int. J. Adv. Res. Ideas Innov. Technol. **5**(2), 1584–1588 (2019)
12. A. Brankovic, A. Zamani, A. Abbosh, Electromagnetic based fatty liver detection using machine learning, in *13th European Conference on Antennas and Propagation (EuCAP 2019)* (2019), pp. 1–3
13. S. Sontakke, J. Lohokare, R. Dani, Diagnosis of liver diseases using machine learning, in *2017 International Conference on Emerging Trends & Innovation in ICT (ICEI)* (2017), pp. 129–133
14. C.-C. Wu, W.-C. Yeh, W.-D. Hsu, MdM Islam, P.A. Nguyen, T.N. Poly, Y.-C. Wang, H.-C. Yang, Y.-C. Li, Prediction of fatty liver disease using machine learning algorithms. Comput. Methods Programs Biomed **170**, 23–29 (2019)
15. J. Jacob, J.C. Mathew, J. Mathew, E. Issac, Diagnosis of liver disease using machine learning techniques. Int. Res. J. Eng. Technol. (IRJET) **5**(4), 4011–4014 (2018)
16. M.A. Konerman, L.A. Beste, T. Van, B. Liu, X. Zhang, J. Zhu, S.D. Saini, G.L. Su, B.K. Nallamothu, G.N. Ioannou, A.K. Waljee, Machine learning models to predict disease progression among veterans with hepatitis C virus. PLoS ONE **14**(1), 1–14 (2019)
17. X. Tian, Y. Chong, Y. Huang, P. Guo, M. Li, W. Zhang, Z. Du, X. Li, Y. Hao, Using machine learning algorithms to predict hepatitis B surface antigen seroclearance. Comput. Math. Methods Med. **2019**, 1–7 (2019)
18. L. Ozyilmaz, T. Yildirim, Artificial neural networks for diagnosis of hepatitis disease, in *Proceedings of the International Joint Conference on Neural Networks*, vol 1 (2003), pp. 586–589
19. K.K. Napa, V. Dhamodaran, Hepatitis-infectious disease prediction using classification algorithms. Res. J. Pharm. Technol. **12**(8), 2019 (2003)
20. M. Elhoseny, K. Shankar, J. Uthayakumar, Intelligent diagnostic prediction and classification system for chronic kidney disease. Sci. Rep **9**(1), 1–4 (2019)
21. E.-H.A. Rady, A.S. Anwar, Prediction of kidney disease stages using data mining algorithms. Inf. Med. Unlocked **15**, 100178 (2019)
22. S. Tekale, P. Shingavi, S. Wandhekar, A. Chatorikar, Prediction of chronic kidney disease using machine learning algorithm. Int. J. Adv. Res. Comput. Commun. Eng. **7**(10), 92–96 (2018)
23. A. Charleonnan, T. Fufaung, T. Niyomwong, W. Chokchueypattanakit, S. Suwannawach, N. Ninchawee, Predictive analytics for chronic kidney disease using machine learning techniques, in *2016 Management and Innovation Technology International Conference (MITicon)* (2016), pp. 80–83
24. A. Ogunleye, Q. Wang, Enhanced XGBoost-based automatic diagnosis system for chronic kidney disease, in *2018 IEEE 14th International Conference on Control and Automation (ICCA)* (2018), pp. 805–810

25. H. Abbas, L. Alic, M. Rios, M. Abdul-Ghani, K. Qaraqe, Predicting diabetes in healthy population through machine learning, in *2019 IEEE 32nd International Symposium on Computer-Based Medical Systems (CBMS)* (2019), pp. 567–570
26. S. Benbelkacem, B. Atmani, Random forests for diabetes diagnosis, in *2019 International Conference on Computer and Information Sciences (ICCIS)* (2019), pp. 1–4
27. M.F. Faruque, Asaduzzaman, I.H. Sarker, Performance analysis of machine learning techniques to predict diabetes mellitus, in *2019 International Conference on Electrical, Computer and Communication Engineering (ECCE)* (2019), pp. 1–4
28. D. Vigneswari, N.K. Kumar, V. Ganesh Raj, A. Gugan, S.R. Vikash, Machine learning tree classifiers in predicting diabetes mellitus, in *2019 5th International Conference on Advanced Computing & Communication Systems (ICACCS)* (2019), pp. 84–87

Sumana De received the B.Tech. and M.Tech. degree in Computer Science and Engineering from West Bengal University of Technology (WBUT), Kolkata, West Bengal India, in 2009 and 2012 respectively. She is at present a part-time Ph.D. research scholar in Computer Science and Engineering in the National Institute of Technology (NIT), Durgapur, West Bengal, India. Her research area includes Development of Knowledge Management Systems (KMS): System for Problem detection and diagnosis.

Baisakhi Chakraborty received the Ph.D. degree in 2011 from National Institute of Technology, Durgapur, India in Computer Science and Engineering. Her research interest includes knowledge systems, knowledge engineering and management, database systems, data mining, natural language processing applications and software engineering. She has several research scholars under her guidance. She has more than 30 international publications. She has a decade of industrial and 14 years of academic experience.