DESIGN AND IMPLEMENTATION IN WEB APPLICATION OF EFFECTING HEART DISEASE AND DIABETES FOR PREDICTING COVID-19

A PROJECT REPORT

Submitted by

MANGAMMAGARI PANITHAVYA
[REGISTER NO:211417104142]

MEENAKSHI.G [REGISTER NO:211417104145]

SANDHIYA.K [REGISTER NO:211417104236]

in partial fulfillment for the award of the degree

of

BACHELOR OF ENGINEERING

IN

COMPUTER SCIENCE AND ENGINEERING

PANIMALAR ENGINEERING COLLEGE, CHENNAI-600123.

ANNA UNIVERSITY: CHENNAI 600 025

AUGUST 2021

BONAFIDE CERTIFICATE

Certified that this project report **"DESIGN AND IMPLEMENTATION IN WEB APPLICATION OF EFFECTING HEART DISEASE AND DIABETES FOR PREDICTING COVID-19 "** is the bonafide work of **"MANGAMMAGARI PANITHAVYA [REG NO:211417104142], MEENAKSHI.G [REG NO:211417104145], SANDHIYA.K [REG NO:211417104236]"** who carried out the project work under my supervision.

SIGNATURE

Dr.S.MURUGAVALLI,M.E.,Ph.D.
HEAD OF THE DEPARTMENT

DEPARTMENT OF CSE,
PANIMALAR ENGINEERING COLLEGE,
NAZARATHPETTAI,
POONAMALLEE,
CHENNAI-600 123.

SIGNATURE

Mrs.A.KANCHANA M.E.
SUPERVISOR
ASSISTANT PROFESSOR

DEPARTMENT OF CSE,
PANIMALAR ENGINEERING COLLEGE
NAZARATHPETTAI,
POONAMALLEE,
CHENNAI-600 123.

Certified that the above candidate(s) was/ were examined in the Anna University Project Viva-Voce Examination held on **05-08-2021**

INTERNAL EXAMINER

EXTERNAL EXAMINER

ACKNOWLEDGEMENT

We express our deep gratitude to our respected Secretary and Correspondent **Dr.P.CHINNADURAI, M.A., Ph.D.** for his kind words and enthusiastic motivation, which inspired us a lot in completing this project.

We would like to extend our heartfelt and sincere thanks to our Directors**Tmt.C.VIJAYARAJESWARI**,**Dr.C.SAKTHIKUMAR,M.E.,Ph.D.** and **Tmt. SARANYASREE SAKTHIKUMAR B.E.,M.B.A.,** for providing us with the necessary facilities for completion of this project.

We also express our gratitude to our Principal **Dr.K.MANI, M.E., Ph.D.** for his timely concern and encouragement provided to us throughout the course.

We thank the HOD of CSE Department,**Dr.S.MURUGAVALLI ,M.E.,Ph.D.** for thesupport extended throughout the project.

We would like to thank my **Project Guide Mrs. A. KANCHANA , M.E.** and all the faculty members of the Department of CSE for their advice and suggestions for the successful completion of the project.

MANGAMMAGARI PANITHAVYA
MEENAKSHI.G
SANDHIYA.K

ABSTRACT

- Heart disease and Diabetes is one among the foremost vital causes of mortality within the world nowadays. Prediction of COVID -19 could be a critical challenge within a space of clinical knowledge analysis.
- Machine learning (ML) has been shown to be effective in aiding in creating selections and predictions from the big amount of knowledge made by the health care trade.
- We've conjointly seen ML techniques being employed in recent developments in numerous areas of Internet of Things (IoT). Numerous studies offer solely a glimpse into predicting COVID-19 with ML techniques.
- For better accuracy 4 algorithms were analyzed namely:

  Support Vector Machine (SVM)

  Decision Tree (DT)

  K-Nearest Neighbor Algorithm (KNN)

  Random Forest Classifier
- Performance and accuracy of the applied algorithms is mentioned and compared. Comparison of the various machine learning techniques employed in this study reveals which algorithm is best fitted for prediction of COVID-19.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF ABBREVIATIONS

CVDs	-	Cardiovascular Diseases
COVID-19	-	Corona virus Disease-19
SARS-CoV-2	-	Severe Acute Respiratory Syndrome Corona Virus 2
SARS	-	severe Acute Respiratory Syndrome
BMI	-	Body Mass Index
ANN	-	Artificial Neural Networks
ML	-	Machine Learning
IOT	-	Internet Of Things
HTML	-	Hyper Text Markup Language
CSS	-	Cascading Style Sheets
JS	-	Java Script
UCI	-	University of California, Irvine
SRS	-	System Requirement Specification
SVM	-	Support Vector Machine
DT	-	Decision Tree
KNN	-	K-Nearest Neighbor Algorithm
RFC	-	Random Forest Classifier
MERS	-	Middle East Respiratory Syndrome

INTRODUCTION

CHAPTER-1

INTRODUCTION

1.1 Overview

Heart is one of the most vital organs for the proper functioning of our body. According to a survey by WHO, 31% of the worldwide deaths every year occurs due to Cardiovascular Diseases (CVDs). Also, more than 75% of these deaths occur in low and middle income countries including India.

The main challenge is to accurately predict the existence of CVDs inside human body. The older techniques have not been very successful in efficiently predicting the heart diseases. Many medical instruments are available in the market for the prediction of heart diseases but there are some drawbacks of these instruments like they are very costly, they are not efficient enough for predicting heart diseases.

Age, Sex, Blood Pressure, Cholesterol, Blood Sugar, Diabetes, etc. and some lifestyle factors like obesity, eating unhealthy food, less physical activity, smoking, consumption of alcohol ,etc. are some of the major risk factors that leads to heart diseases. Most of the lifestyle risk factors are controllable. In the last few decades, medical science has used the technological advancements very well to improve the quality of healthcare. These advancements in technology have paved ways for accurate diagnosis and prediction of diseases. Machine learning could be a very good choice to achieve high accuracy for predicting heart diseases a sit is able to analyse large amounts of data and identifying patterns & trends. Moreover, machine learning provides much faster and reliable results. There are other soft computing approaches as well such as Artificial Neural Networks (ANN), Deep Learning, Fuzzy Logic, Data Mining, Genetic Algorithm, etc. that can put into effect for predicting heart diseases.

Diabetes is a condition where person’s body is unable to balance glucose-insulin level after various prandial modes. The patients of diabetes have increased exponentially from past few years due to their unbalanced diet and unhealthy lifestyle. There are around 463 million diabetic people around the world who can be benefited by smart healthcare technologies to improve their quality of life. Diabetes may lead to heart disease, kidney infection, blindness and nerve damages. It has been observed that severe acute respiratory syndrome corona virus 2 (SARS-CoV-2) has

infected the diabetic patients largely among the others. People with diabetes were also more susceptible to previous pandemic known as Middle East respiratory syndrome (MERS) and severe Acute Respiratory Syndrome (SARS) type of corona virus along with H1N1 type of severe influenza during 2009 . SARS-CoV-2 has affected mainly to old-age people and persons having underneath health conditions. The various case studies reported diabetes as the major pre-existing comorbidity among COVID-19 patients. The challenges for diabetes patient to control their the blood glucose levels after the infection due to following:

1) The fluctuation of blood sugar affects the immunity of person which expose him against COVID-19 and unbalanced glycemic profile may lead to longer time of recovery for the patient.

2) The high blood glucose allows the virus to infect the human body easily.

1.2 Problem Definition

Cardiovascular and Diabetes disease is considered as one of the main causes of death around the globe. It is very difficult to be predicted by the medical practitioners as it is a complex task which requires expertise and a lot of experience for prediction. An automated system in medical diagnosis would increase medical efficiency and would also help in reducing costs. We will design a system that can efficiently discover the rules to predict heart and diabetes disease in patients based on the given parameters about their health. The goal is to find the hidden patterns by employing machine learning algorithms, which are quick and reliable enough to predict the Covid-19 by finding the presence of heart and diabetes disease in users and patients.

LITERATURE SURVEY

CHAPTER-2

LITERATURE SURVEY

A literature Survey is an objective, crucial outline of printed analysis literature relevant toa subject into account for analysis. Its purpose is to form familiarity with current thinking and analysis on a selected topic, and should justify future analysis into previously not noted or understudied space. it's the foremost vital a part of the report because it offers a direction within the space of analysis. It helps to line a goal for the analysis so giving out downside statement. A literature review in respect of the project, the researches created by varied analysts – their methodology (which is essentially their abstract) and therefore the conclusions they need found. Itadditionally offers an account of however this analysis has influenced the thesis.

[1] TITLE: Prediction of Heart Disease Using Machine Learning Algorithms
DESCRIPTION : This paper typically explains the Health care field has a vast amount of data, for processing those data certain techniques are used. Data mining is one of the techniques often used. Heart disease is the Leading cause of death worldwide. This System predicts the arising possibilities of Heart Disease. The outcomes of this system provide the chances of occurring heart disease in terms of percentage. The datasets used are classified in terms of medical parameters. This system evaluates those parameters using data mining classification technique. The datasets are processed in python programming using two main Machine Learning Algorithm namely Decision Tree Algorithm and Naive Bayes Algorithm which shows the best algorithm among these two in terms of accuracy level of heart disease.

[2] TITLE: A Hybrid Machine Learning Approach for Prediction of Heart Diseases
DESCRIPTION : This paper typically explains the Heart diseases are the chief cause of death all over the world over the last few decades. To avoid heart disease or coronary illness and discover indications early, individuals over 55 years must have a total cardiovascular checkup.

Researchers and specialists developed various intelligent techniques to improve capacity of the health care professionals in recognition of cardiovascular disease. In cardiovascular disease finding and treatment, single data mining strategies are giving the reasonable precision and accuracy. Nevertheless the usage data mining procedure be capable of reducing the number of test that is required to be carried out. In order to decrease the Figure of deaths from heart diseases there has to be a quick and efficient detection technique providing better accuracy and precision. The aim of this paper is to present an efficient technique of predicting heart diseases using machine learning approaches. Hence we proposed a hybrid approach for heart prediction using Random forest classifier and simple k-means algorithm machine learning techniques. The dataset is also evaluated using two other different machine learning algorithms, namely, J48 tree classifier and Naive Bayes classifier and results are compared. Results attained through Random forest classifier and the corresponding confusion matrix shows robustness of the methodology.

[3] TITLE : An Investigation of Heart Disease Prediction Techniques

DESCRIPTION : Heart disease indicates the type of condition which leads to heart malfunction. Not all people with coronary artery disease have chest pain as a symptom. There are other factors such as R-Blood Pressure, S-Cholesterol, F-Blood Sugar, R-ECG, Thalach, Ex-Ang, Number of major Vessels blocked, Thallium Scan which also results in heart disease. At an initial stage the prediction of heart disease can save human lives. Data mining techniques such as Naive Bayes, Decision Tree, K-nearest neighbor, etc., are used in the Heart Disease prediction based on the parameters / factors. The objective of this research is to investigate about the various factors and its significance in identifying the heart disease. Further it also explores about the existing techniques and models used for prediction.

[4]TITLE : A Cloud Based Four-Tier Architecture for Early Detection of Heart Disease with Machine Learning Algorithms.

DESCRIPTION : Heart disease prediction and detection has long been considered as a critical issue. Early detection of heart disease is an important issue in health care services (HCS). In growing amount of health care systems, patients are offered expensive therapies and operation that is quiet expensive for developing countries. Recently, heart disease is a prominent public chronic disease, ex. it's a growing concern in the US. The main reason of these diseases are tobacco consumption, bad life style, lack of physical activity and the intake of alcohol. Therefore, there is a need for the cloud based architecture that can efficiently predict and track health information. Recently, machine learning techniques have already been established to solve clinical problem and medical diagnosis. In this study, we proposed a cloud-based 4- tier architecture that can significantly improve the prediction and monitoring of patient's health information. Hence, we used five popular supervised learning based machine learning technique for early detection of heart disease. The major purpose of this study is to examine the performance of the selected classification techniques. In addition, we use prominent evaluation criteria to observe the best performance of these machine learning techniques. Moreover, we used the ten-fold cross validation technique to evaluate the performance of the five classifiers. The analysis results indicate that the Artificial Neural Network (ANN) achieved the highest performance of all. However, health care researchers and practitioners can obtain independent understanding from this work while selecting machine learning techniques to apply in their area.

[5]TITLE : A comprehensive investigation and comparison of Machine Learning Techniques in the domain of heart disease

DESCRIPTION : This paper typically explains the This paper aims to investigate and compare the accuracy of different data mining classification schemes, employing Ensemble Machine Learning Techniques, for the prediction of heart disease. The Cleveland data set for heart diseases, containing 303 instances, has been used as the main database for the training and testing of the developed system. 10-Fold Cross-Validation has been applied in order to increase the amount of data, which would otherwise have been limited. Different classifiers,

namely Decision Tree (DT), Naïve Bayes (NB), Multilayer Perceptron (MLP), K-Nearest Neighbor (K-NN), Single Conjunctive Rule Learner (SCRL), Radial Basis Function (RBF) and Support Vector Machine (SVM), have been employed. Moreover, the ensemble prediction of classifiers, bagging, boosting and stacking, has been applied to the dataset. The results of the experiments indicate that the SVM method using the boosting technique outperforms the other aforementioned methods.

[6]TITLE : Smart Healthcare for Diabetes during COVID-19

DESCRIPTION : This paper typically explains the The diabetic patients are at higher risk from novel coronavirus disease 2019 (COVID-19) that spreads through Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2). There are around 20% to 50% of COVID-19 cases had diabetes across the different re-gions from the world. This article discusses recommendations and associated risk for diabetic patients to balance glyncemic profile during COVID-19 outbreak. It also discusses the case study of various countries with impact of COVID-19 for diabetic patients. It presents emerging smart healthcare that can potentially safeguard against COVID-19 .

[7] TITLE : Diabetes disease prediction using data mining

DESCRIPTON : This paper typically explains the Data mining is a subfield in the subject of software engineering. It is the methodical procedure of finding examples in huge data sets including techniques at the crossing point of manufactured intelligence, machine learning, insights, and database systems. The goal of the data mining methodology is to think data from a data set and change it into a reasonable structure for further use. Our examination concentrates on this part of Medical conclusion learning design through the gathered data of diabetes and to create smart therapeutic choice emotionally supportive network to help the physicians. The primary target of this examination is to assemble Intelligent Diabetes Disease Prediction System that gives analysis of diabetes malady utilizing diabetes patient's database.

In this system, we propose the use of algorithms like Bayesian and KNN (K-Nearest Neighbor) to apply on diabetes patient's database and analyze them by taking various attributes of diabetes for prediction of diabetes disease.

[8]TITLE : An intelligent system for diabetes prediction.

DESCRIPTION : This paper typically explains the emerging increase of diabetes, that recently affects around 346 million people, of which more than one-third go undetected in early stage, a strong need for supporting the medical decision-making process is generated. A number of researches have focused either in using one of the algorithms or in the comparisons of the performances of algorithms on a given, usually predefined and static datasets that are accessible through the Internet. This paper focuses on the joint implementation of the support vector machine (SVM) and Naïve Bayes statistical modelling, in the dataset acquired from the medical examinations of 402 patients, in order to improve the computer-supported diagnosis reliability. The dataset contains some attributes that have not been previously used in computer-based evaluations. The results show that the joint implementation of two algorithms improves significantly the overall reliability of the system outcome, which is crucial in the computersupported diabetes diagnostic process.

[9] TITLE : Prediction of Diabetes Risk based on Machine Learning Techniques
DESCRIPTION : The explosive population growth and health maintenance is an extremely crucial matter worldwide. Many lethal diseases are causing threats at a high peak in recent years. Introducing machine learning technologies into healthcare for early prognosis and diagnosis need to be more accurate based on the parameters and frames selected from the available clinical databases. The objective of this paper is to analyze, explore various research outcomes of machine learning methodologies used in diabetes mellitus and how the efficiencies obtained could be helpful in future perspective of a predictive diabetes model designing. The exploration inferred that more variables and hybrid disciplines should be

considered for an accuracy of result which can overcome the existing limitations.

[10] TITLE : A computation analysis to predict diabetes based on data mining **DESCRIPTION :** This paper typically explains the report suggested from different health organization shows the alarming condition due to diabetes worldwide. Different researchers around the globe have investigated it on different parameters and the investigation is going on for the early stage detection. The main objective of this paper is explore and elaborate the methodological prospective to predict the diabetes based on the dataset presented. This exploration is helpful in providing us a direction to find out the research gaps, so that future efficient diabetes framework can be developed. This computation analysis also provides us the parametric exploration along with the knowledge of the attributes and the way of implementing the classification framework.

SYSTEM ANALYSIS

CHAPTER 3

SYSTEM ANALYSIS

3.1 Existing System

Before, the Doctors only view the report to convey the result to the patients. There are some problems appeared while seeing the laboratory details, they can't predict it properly. There is some difficulty in existing project, they created for some other purpose to test for different disease prediction , but According to covid 19, Diabetes and heart disease result is very important to predict , whether the person will be affected by covid-19 or not.

3.2 Proposed System

The prime objective of this project is to find the factors impacting the health of the society with the greater accuracy level and giving the good result.

- This software is used to predict the covid by checking the heart disease and diabetes report.
- We have used python and pandas operations to perform heart disease and Diabetes classification of the Cleveland UCI repository.
- It provides an easy-to-use visual representation of the dataset, working environment and building the predictive analytics.
- ML process starts from a pre- processing data phase followed by feature selection based on data cleaning, classification of modeling performance evaluation, and the results with improved accuracy.

3.3 Requirements Analysis and Specifications

The system Requirement specification (SRS) document describes all data, functional and behavioral requirements of the software under production or development. It is produced at the culmination of the analysis task. The software and hardware requirements of this project are listed below,

3.3.1 Input requirements

Data Set, User Data

3.3.2 Output requirements

Predicted Result

3.3.3. Functional Requirements

Format of data plays an important role in this application. When the user enters the medical details, it should be in correct format and also within the specified range, else ERROR dialog box will be prompted. The following four algorithms will be analysed:

- Support Vector Machine (SVM)
- Decision Tree (DT)
- *K*-Nearest Neighbour Algorithm (KNN)
- Random Forest Classifier

The working of these algorithms has been explained in the sections ahead. The algorithms have been trained using the UCI (University of California, Irvine) Cleveland data set. 75% of the entries in the data set have been used for training and the remaining 25% for testing the accuracy of the algorithm. Furthermore, some steps have been taken for optimizing the algorithms thereby improving the accuracy. These steps include cleaning the dataset as well as data pre-processing. The algorithms were judged based on their accuracy and it was observed that the *K*-Nearest Neighbour Algorithm (KNN) was the most accurate out of the four with 87.0% efficiency. Hence, it was selected for implementation of the main application.
The main application is a web application which accepts the various parameters from the user as input and computes the result.

3.4 Technology Stack

HARDWARE REQUIREMENTS

- Hard Disk : 500GB and Above

- RAM : 4GB and Above
- Processor : I3 and Above

SOFTWARE REQUIREMENTS

- Operating System: Windows 7, 8, 10 (64 bit)
- Software : Python
- Tools : Anaconda (Jupiter Note Book IDE)

SYSTEM DESIGN

CHAPTER 4

SYSTEM DESIGN

4.1 ER-Diagram

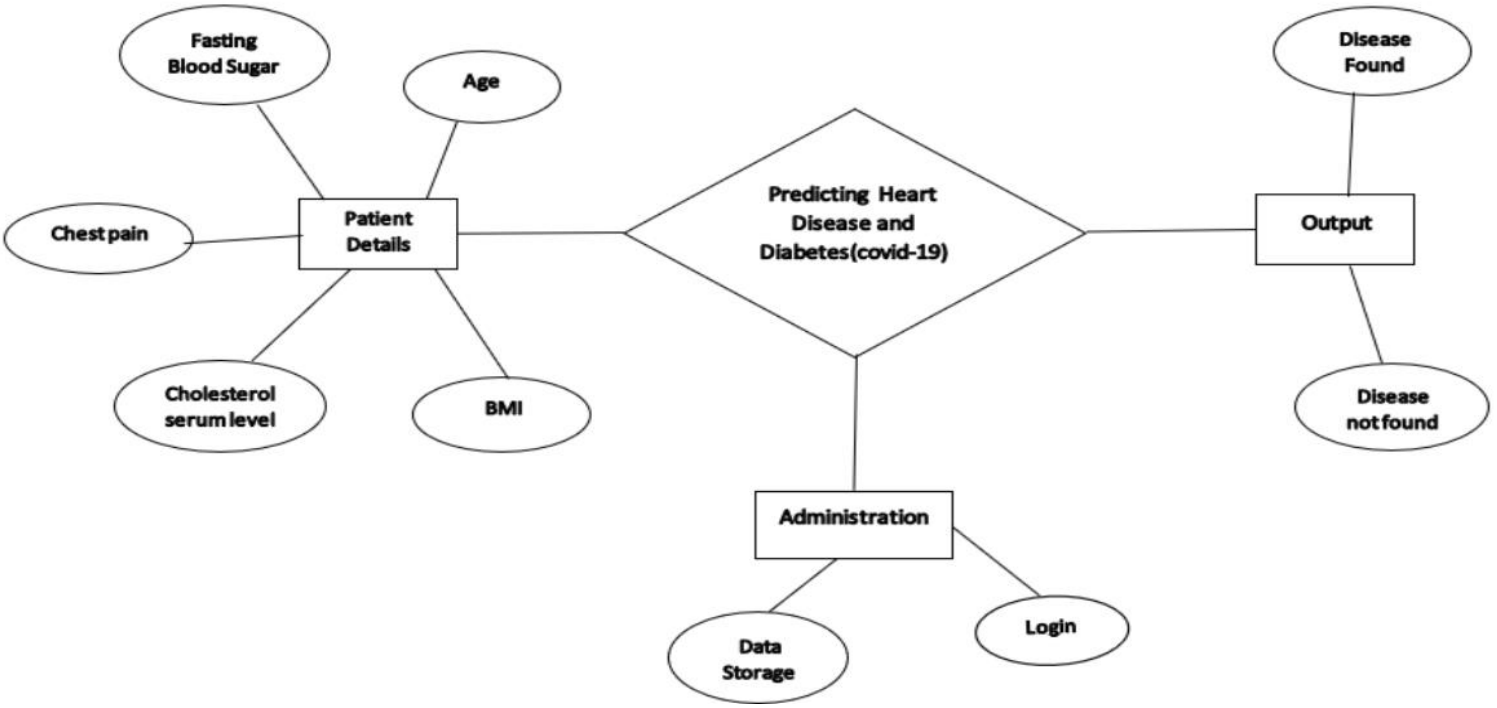


Fig- 4.1: ER Diagram4

4.2Data Flow Diagram

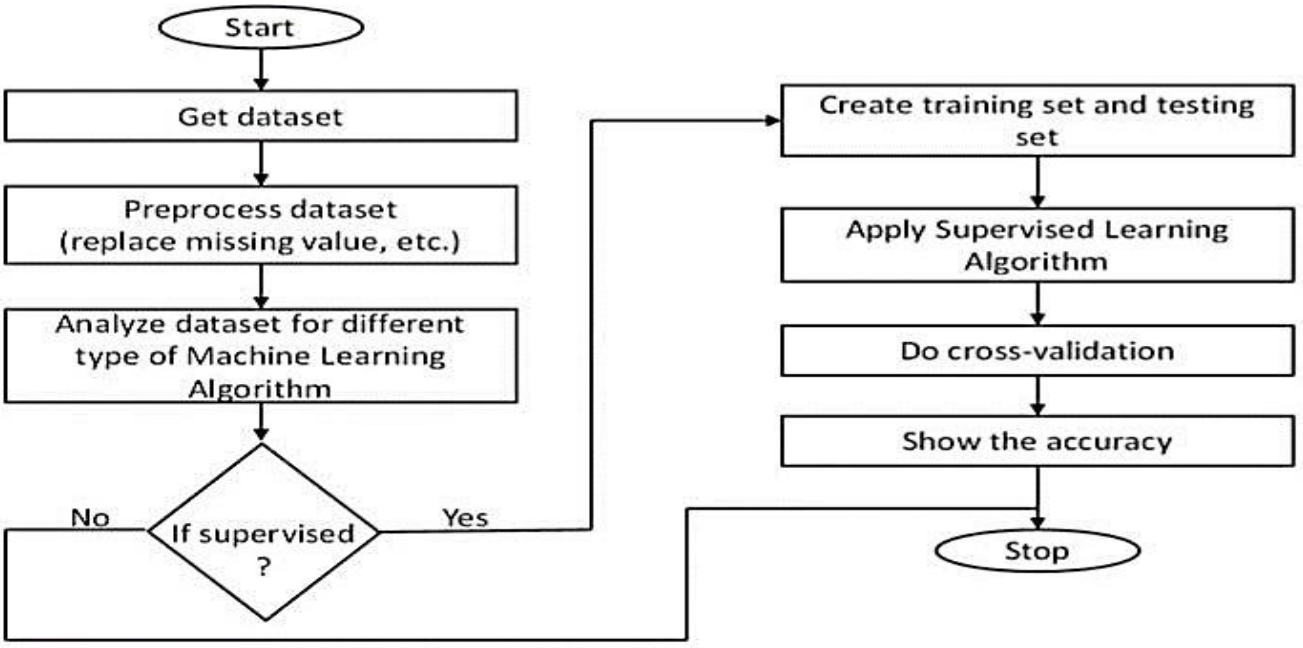


Fig- 4.2: Data Flow Diagram

4.3UML Diagram

4.3.1Use Case Diagram

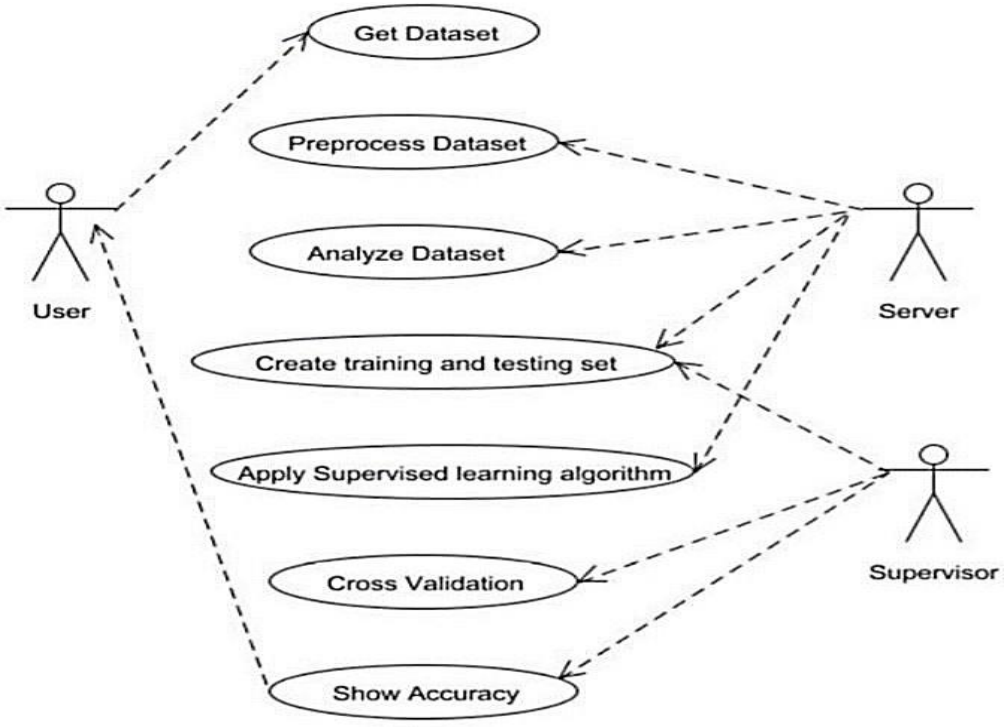


Fig.4.3.1: Use Case Diagram

4.3.2Class Diagram

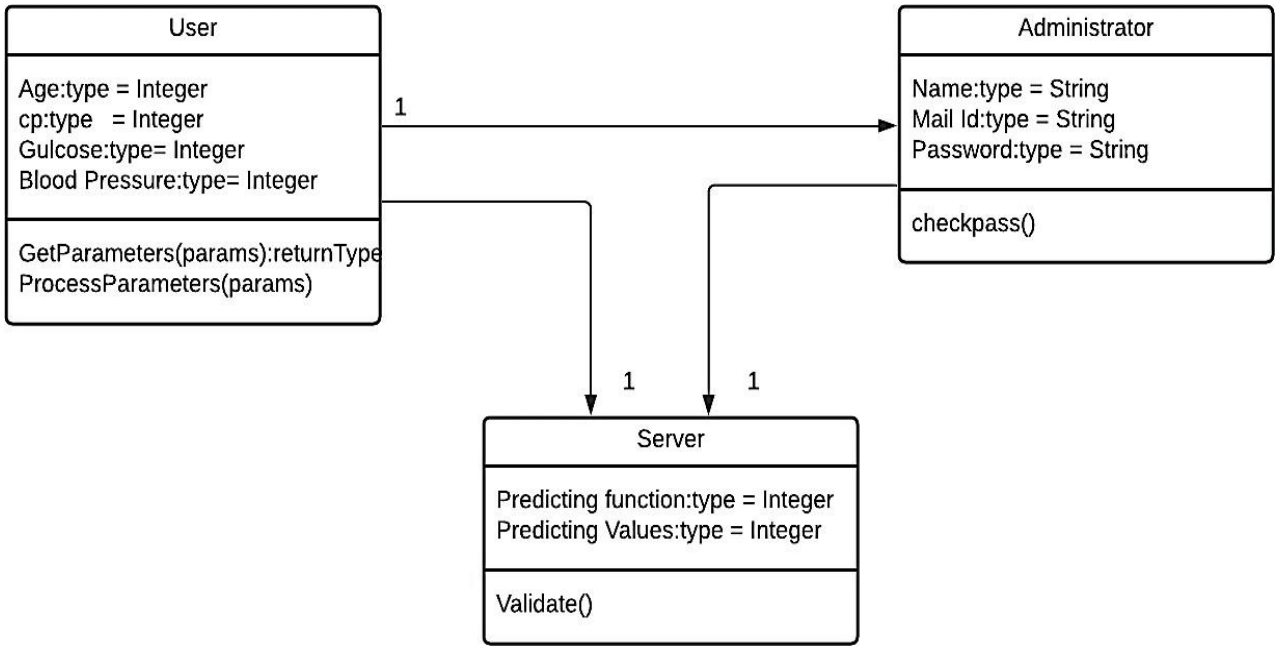


Fig.4.3.2: Class Diagram

4.3.3 Activity Diagram

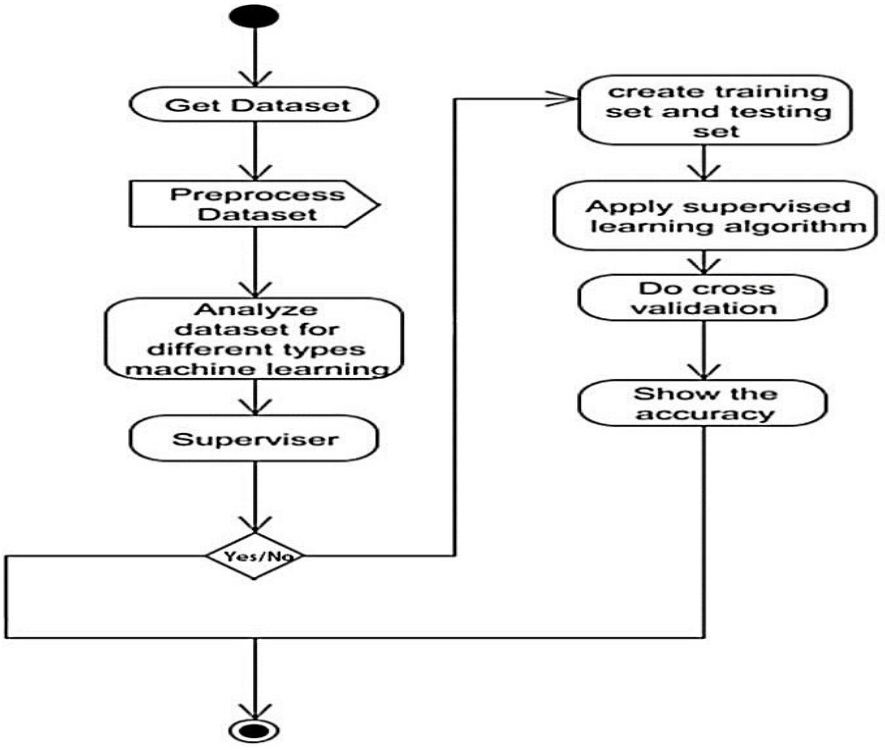


Fig.4.3.3: Activity Diagram

4.3.4 Sequence Diagram

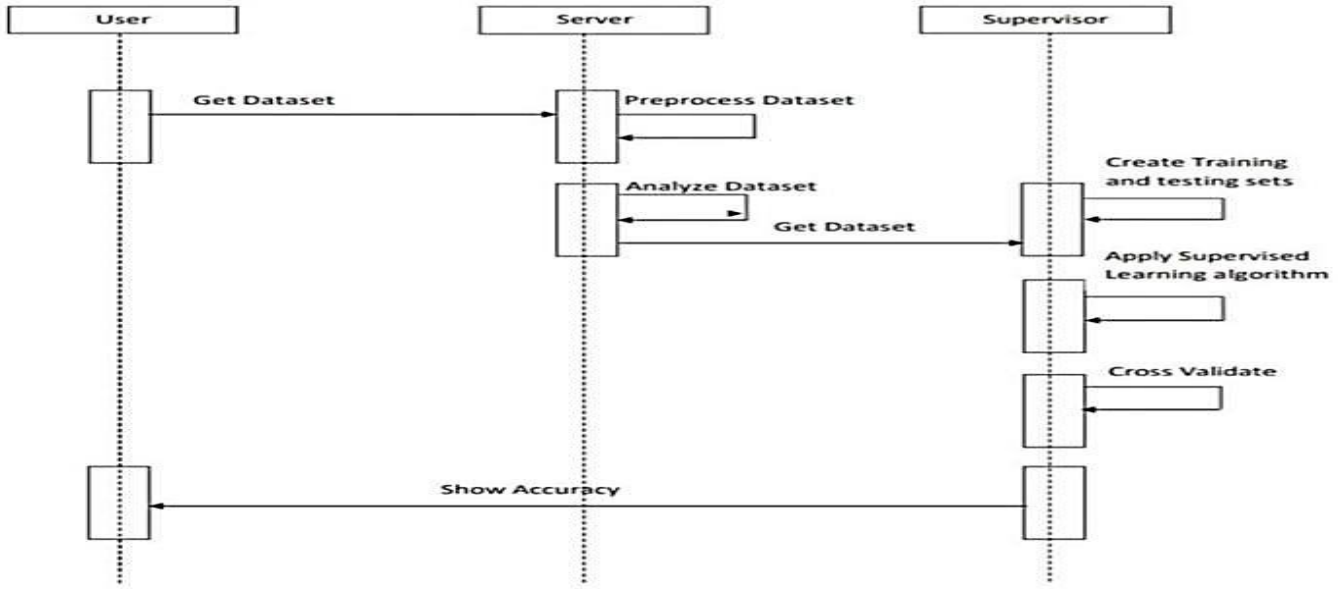


Fig.4.3.4: Sequence Diagram

4.3.5 Collaboration Diagram

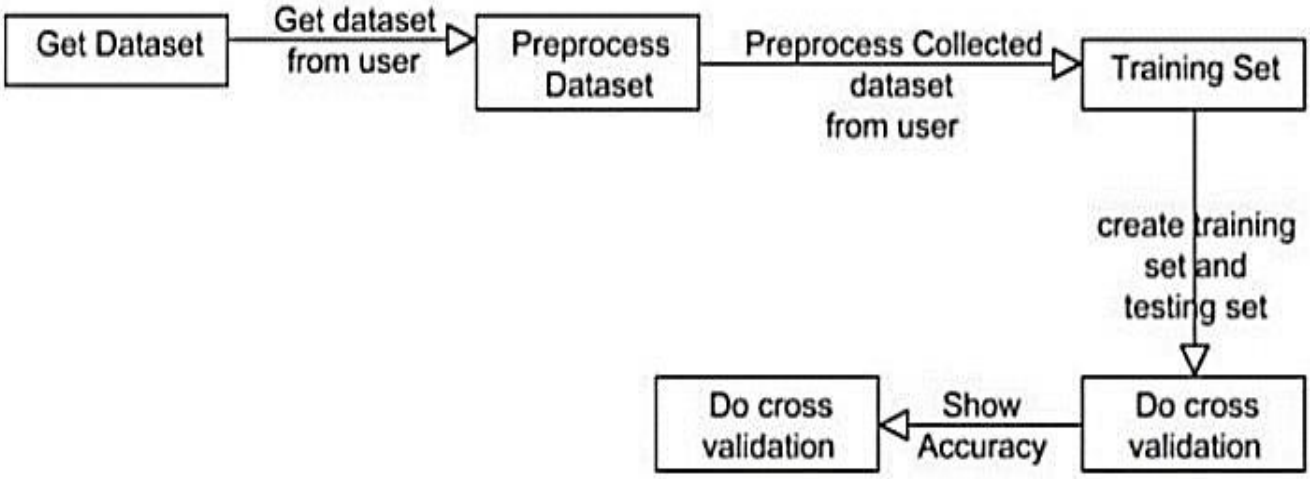


Fig.4.3.5: Collaboration Diagram

4.3.6 Interface Diagram

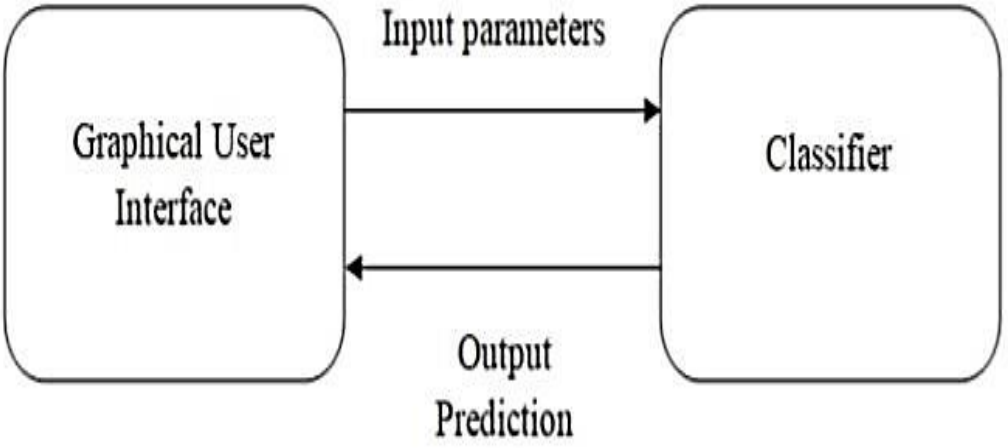


Fig.4.3.6: Interface Diagram

4.3.7 State Machine Diagram

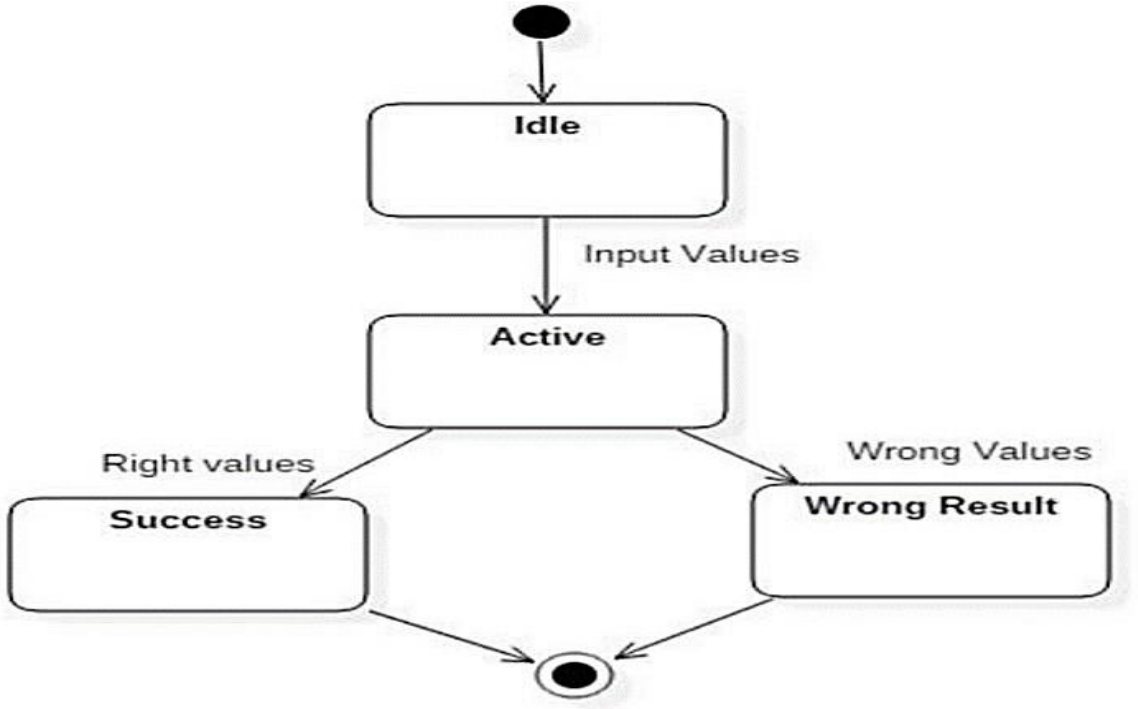


Fig.4.3.7: State Machine Diagram

4.3.8 DB Design

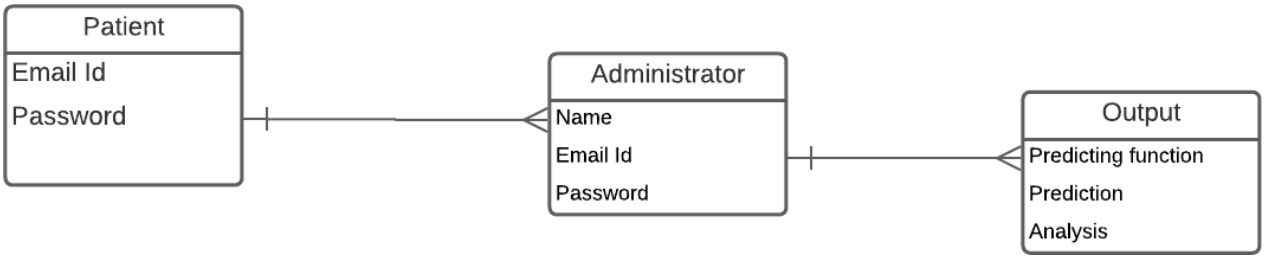


Fig.4.3.8: DB Design

SYSTEM ARCHITECTURE

CHAPTER-5

SYSTEM ARCHITECTURE

5.1 Architecture Diagram

A system architecture is the conceptual model that defines the structure, behavior, and more views of a system. An architecture description is a formal description and representation of a system, organized in a way that supports reasoning about the structures and behaviors of the system.

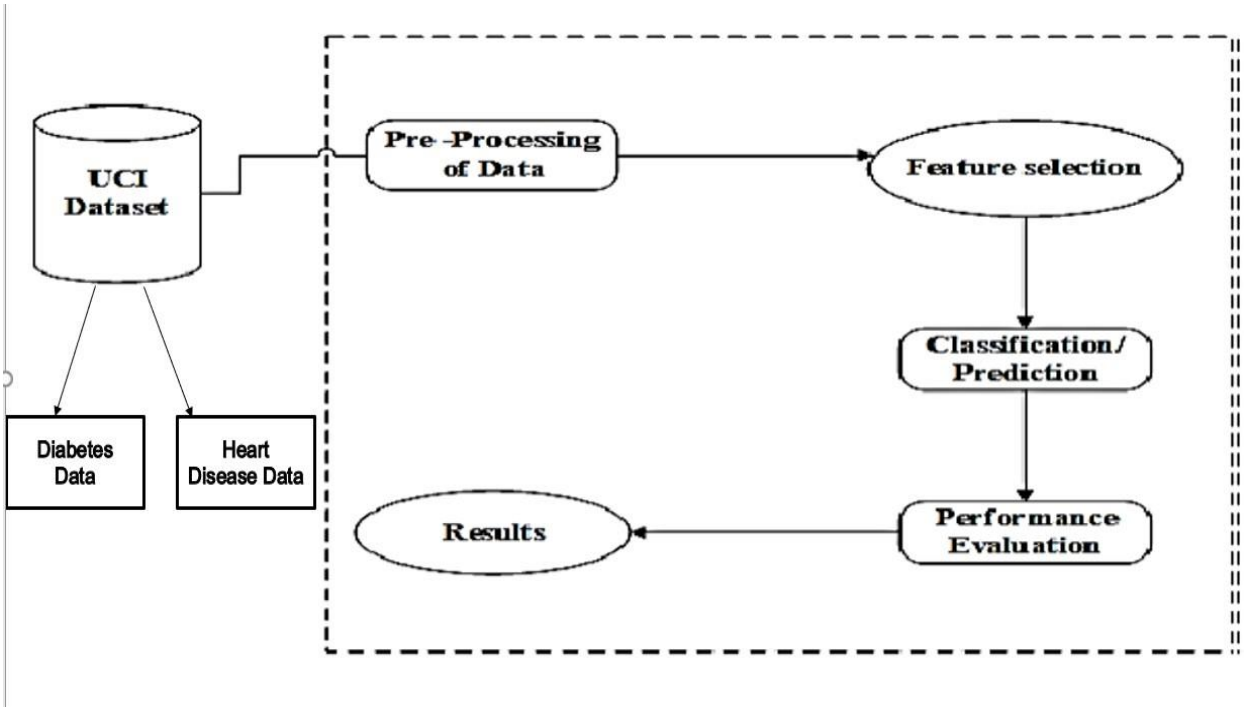


Fig.5.1: Architecture Diagram

At first, we getting the data of Diabetes and Heart Disease from UCI Dataset, then we are Pre-Processing of Data. After that we are doing the Feature Selection for classification / Prediction. Then the Performance Evaluation Occurs for the result.

5.2 Module Design Specification

5.2.1 Dataset Description

✓ The objective of the dataset is to predict whether or not a patient has diabetes, based on certain diagnostic measurements included in the dataset.

✓ The datasets consists of several medical predictor variables and one target variable, Outcome. Predictor variables include the number of pregnancies the patient has had, their BMI, insulin level, age, chest Pain, Cholesterol, FBP, ECG and so on.

- Pregnancies: Number of times pregnant
- Glucose: Plasma glucose concentration a 2 hours in an oral glucose tolerance test
- Blood Pressure: Diastolic blood pressure (mm Hg)
- Skin Thickness: Triceps skin fold thickness (mm)
- Insulin: 2-Hour serum insulin (mu U/ml)
- BMI: Body mass index (weight in kg/(height in m)^2)
- Diabetes Pedigree Function: Diabetes pedigree function
- Age: Age (years)
- Outcome: Class variable (0 or 1) 268 of 768 are 1, the others are 0.
- Chest Pain: 1-4 (1: Typical Angina, 2: Atypical Angina, 3: Non-anginal, 4: Asymptotic)
- Cholesterol: Serum Cholesterol in mg/dl
- Fasting Blood Sugar: Blood sugar content before food intake if >120 mg/dl.
- ECG: Resting Electrocardiographic results , 0-1 (0: Normal, 1: Having ST-T wave)
- Max Heart Rate: Maximum heart beat rate, Beats/min
- Exercise Induced Angina: Has pain been induced by exercise, 0-1 (0: No, 1: Yes).
- Old Peak: ST depression induced by exercise relative to rest,0-4.
- Slope of Peak Exercise: Slope of the peak exercise ST segment, 1-3 (1: Up sloping, 2: Flat, 3: Down sloping)
- Ca: Number of vessels colored by fluoroscopy, 0-3
- Thala: Displays the thalassemia.
- Num: Diagnostics of Heart Disease.

5.2.2 Data Pre-Processing

- ✓ Diabetes disease and Heart Disease data is pre-processed after collection of various records.
- ✓ The dataset contains a total of 769 patient records, where 6 records are with some missing values.
- ✓ Those 6 records have been removed from the dataset and the remaining 763 patient records are used in pre-processing.

5.2.3 Feature Section

- ✓ From among the 8 attributes of the data set, one attributes pertaining to age is used to identify the personal information of the patient.
- ✓ The remaining 7 attributes are considered important as they contain vital clinical records.
- ✓ Clinical records are vital to diagnosis and learning the severity of diabetes disease.

5.2.4 Classification Modeling

The clustering of datasets is done on the basis of the variables and criteria of Decision Tree (DT) features. Then, the classifiers are applied to each clustered dataset in order to estimate its performance. The best performing models are identified from the above results based on their low rate of error.

- ❖ Support Vector Machine
- ❖ Decision Trees
- ❖ K Nearest neighbors
- ❖ Random Forest Classifier

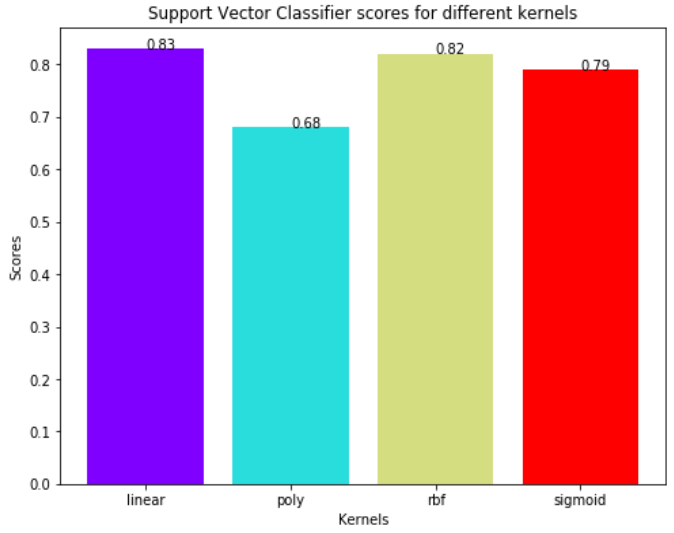
Support vector machine

The accuracy score of SVM was found to be **83.0%**.

In [17]:
```
svc_scores = []
kernels = ['linear', 'poly', 'rbf', 'sigmoid']
for i in range(len(kernels)):
    svc_classifier = SVC(kernel = kernels[i])
    svc_classifier.fit(X_train, y_train)
    svc_scores.append(svc_classifier.score(X_test, y_test))
```

In [18]:
```
colors = rainbow(np.linspace(0, 1, len(kernels)))
plt.bar(kernels, svc_scores, color = colors)
for i in range(len(kernels)):
    plt.text(i, svc_scores[i], svc_scores[i])
plt.xlabel('Kernels')
plt.ylabel('Scores')
plt.title('Support Vector Classifier scores for different kernels')
```

Out[18]:
```
Text(0.5, 1.0, 'Support Vector Classifier scores for different kernels')
```



In [19]:
```
print("The score for Support Vector Classifier is {}% with {} kernel.".format(svc_scores[0]*100, 'linear'))
```

```
The score for Support Vector Classifier is 83.0% with linear kernel.
```

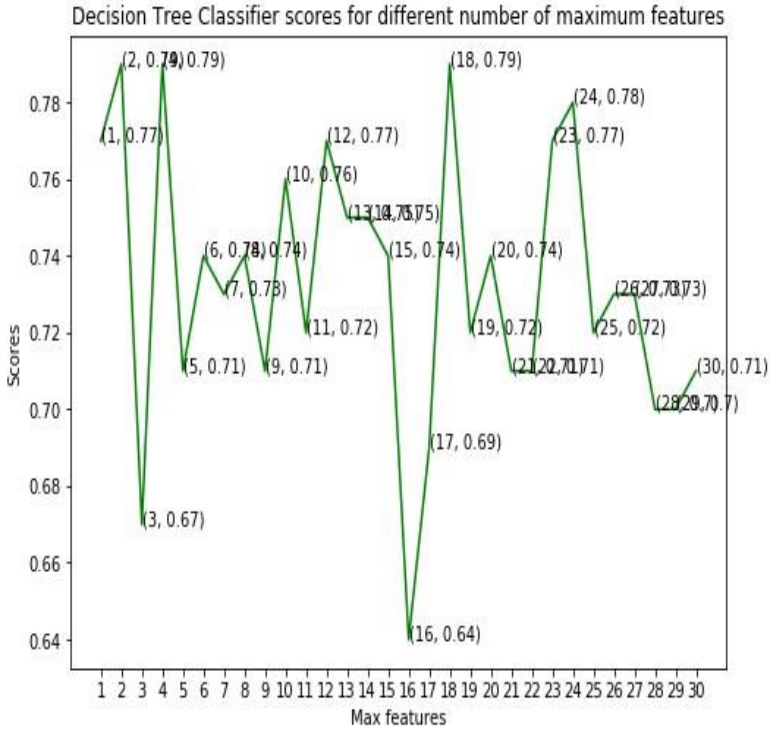
Decision Tree Classifier

The accuracy score of Decision Tree was found to be **79.0%**.

In [20]:
```
dt_scores = []
for i in range(1, len(X.columns) + 1):
    dt_classifier = DecisionTreeClassifier(max_features = i, random_state = 0)
    dt_classifier.fit(X_train, y_train)
    dt_scores.append(dt_classifier.score(X_test, y_test))
```

In [21]:
```
plt.plot([i for i in range(1, len(X.columns) + 1)], dt_scores, color = 'green')
for i in range(1, len(X.columns) + 1):
    plt.text(i, dt_scores[i-1], (i, dt_scores[i-1]))
plt.xticks([i for i in range(1, len(X.columns) + 1)])
plt.xlabel('Max features')
plt.ylabel('Scores')
plt.title('Decision Tree Classifier scores for different number of maximum features')
```

Out[21]:
```
Text(0.5, 1.0, 'Decision Tree Classifier scores for different number of maximum features')
```



In [22]:
```
print("The score for Decision Tree Classifier is {}% with {} maximum features.".format(dt_scores[17]*100, [2,4,18]))
```

```
The score for Decision Tree Classifier is 79.0% with [2, 4, 18] maximum features.
```

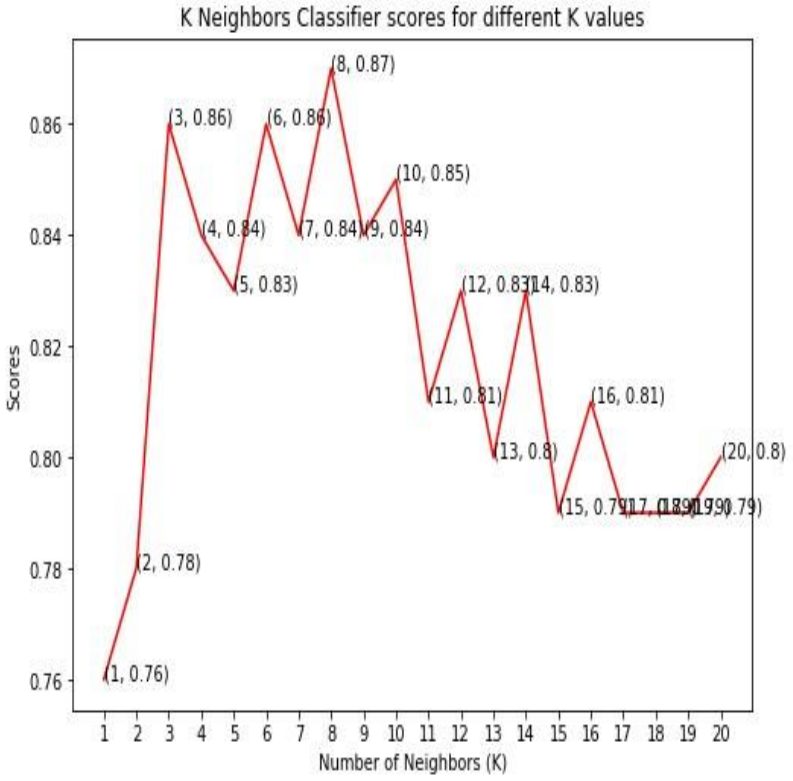
K-Nearest Neighbour

The accuracy score of KNN was found to be **87.0%**.

```
In [14]: knn_scores = []
         for k in range(1,21):
             knn_classifier = KNeighborsClassifier(n_neighbors = k)
             knn_classifier.fit(X_train, y_train)
             knn_scores.append(knn_classifier.score(X_test, y_test))
```

```
In [15]: plt.plot([k for k in range(1, 21)], knn_scores, color = 'red')
         for i in range(1,21):
             plt.text(i, knn_scores[i-1], (i, knn_scores[i-1]))
         plt.xticks([i for i in range(1, 21)])
         plt.xlabel('Number of Neighbors (K)')
         plt.ylabel('Scores')
         plt.title('K Neighbors Classifier scores for different K values')
```

```
Out[15]: Text(0.5, 1.0, 'K Neighbors Classifier scores for different K values')
```




```
In [16]: print("The score for K Neighbors Classifier is {}% with {} nieghbors.".format(knn_scores[7]*100, 8))
```

```
The score for K Neighbors Classifier is 87.0% with 8 nieghbors.
```

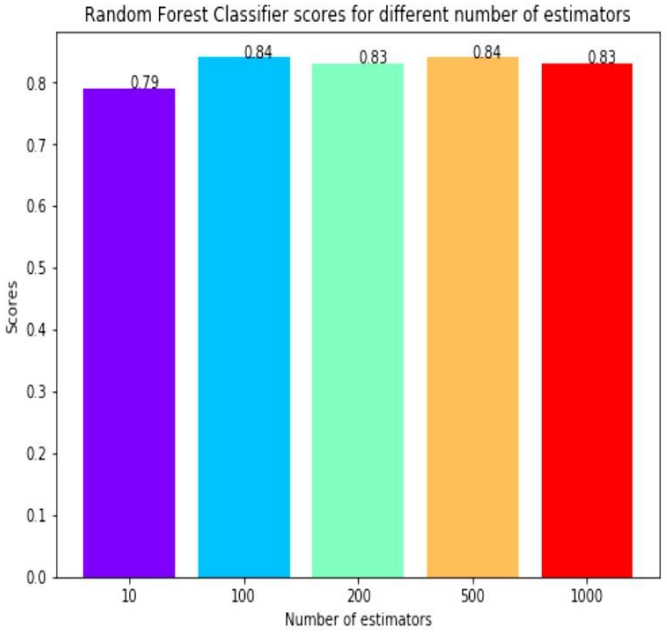
Random Forest Classifier

The accuracy of Random Forest Classifier was found to be **84.0%**.

In [23]:
```
rf_scores = []
estimators = [10, 100, 200, 500, 1000]
for i in estimators:
    rf_classifier = RandomForestClassifier(n_estimators = i, random_state = 0)
    rf_classifier.fit(X_train, y_train)
    rf_scores.append(rf_classifier.score(X_test, y_test))
```

In [24]:
```
colors = rainbow(np.linspace(0, 1, len(estimators)))
plt.bar([i for i in range(len(estimators))], rf_scores, color = colors, width = 0.8)
for i in range(len(estimators)):
    plt.text(i, rf_scores[i], rf_scores[i])
plt.xticks(ticks = [i for i in range(len(estimators))], labels = [str(estimator) for estimator in estimators])
plt.xlabel('Number of estimators')
plt.ylabel('Scores')
plt.title('Random Forest Classifier scores for different number of estimators')
```

Out[24]:
```
Text(0.5, 1.0, 'Random Forest Classifier scores for different number of estimators')
```



In [25]:
```
print("The score for Random Forest Classifier is {}% with {} estimators.".format(rf_scores[1]*100, [100, 500]))
```

```
The score for Random Forest Classifier is 84.0% with [100, 500] estimators.
```

5.2.5 Performance Measure

- ✓ Several standard performance metrics such as accuracy, precision and error in classification have been considered for the computation of performance efficiency of this model.
- ✓ Logistic Regression : 71.42857142857143
- ✓ K Nearest neighbors : 78.57142857142857
- ✓ Support Vector Classifier : 73.37662337662337
- ✓ Naive Bayes : 71.42857142857143
- ✓ Decision tree : 68.18181818181817
- ✓ Random Forest : 75.97402597402598

5.3 Program design language

About Python

Python is a free, open-source programming language. Therefore, all you have to do is install Python once, and you can start working with it. Not to mention that you can contribute your own code to the community. Python is also a cross- platform compatible language.

Python is also a great visualization tool. It provides libraries such as Matplotlib, seaborn and bokeh to create stunning visualizations. In addition, Python is the most popular language for machine learning and deep learning. As a matter of fact, today, all top organizations are investing in Python to implement machine learning in the back-end.

Advantages of Python

- Universal Language Construct
- Support both High Level and Low-Level Programming
- Language Interoperability
- Fastest Development life cycle therefore more

productive coding environment

- Less memory used because a single container hold multiple data types and each type doesn't require its own function
- Learning Ease and open source development
- Speed and user-friendly data structure
- Extensive and extensible libraries.
- Simple & support IoT
- And many more

Based on the 3 languages that are described above, we decided to use Python as programming language for developing this e-Voting web based application. The important motives are

- Easy to learn, even non experienced programmers can use it. Ex: spacing and tabbing instead of extra syntax
- Interactive mode
- Large and comprehensive standard libraries
- Python programs resemble to that of pseudo-code. This makes it a basis and a must have for beginner programmers due to its extreme ease a difficulty when compared to C++ Java, Perl, and so forth

Selection of Integrated Development Environment

An Integrated Development Environment (IDE) is a product application that gives a programming domain to streamline creating and troubleshooting programming. Instead of playing out every one of the means required to make an executable program as disconnected individual assignments, it brings every one of the devices required into one application and workspace. Every one of the devices has a familiarity with the earth, and they cooperate to display a consistent improvement set for the designer.

HTML, CSS And JavaScript

- Hypertext Markup Language, a standardized system for tagging text files to achieve font, color, graphic, and hyperlink effects on World Wide Web pages.
- **CSS** stands for Cascading Style Sheets. **CSS** describes how HTML elements are to be displayed on screen, paper, or in other media. **CSS** saves a lot of work.
- **JavaScript** is a programming language commonly used in web development. It was originally developed by Netscape as a means to add dynamic and interactive elements to websites. Like server-side scripting languages, such as PHP and ASP, **JavaScript** code can be inserted anywhere within the HTML of a webpage.

Machine Learning

Machine learning is an application of artificial intelligence (AI) that provides systems the ability to automatically learn and improve from experience without being explicitly programmed. **Machine learning focuses on the development of computer programs** that can access data and use it to learn for themselves. The process of learning begins with observations or data, such as examples, direct experience, or instruction, in order to look for patterns in data and make better decisions in the future based on the examples that we provide. **The primary aim is toallow the computers learn automatically** without human intervention or assistance and adjust actions accordingly.

SYSTEM IMPLEMENTATION

CHAPTER 6

SYSTEM IMPLEMENTATION

Based on the analysis, K-Nearest Neighbor (KNN) was found to be most accurate and reliable. Therefore, KNN was used for the final implementation of the project. Python 3 was used for modelling and classification. The dataset was split into training and testing data in the ratio of 3:1 i.e., 75% of the dataset was used for training purpose & the remaining 25% was used for testing and validation. Front-end is based on HTML5, CSS and JS. Python's micro web-framework Flask is also used for database connection.

6.1 Diabetes Sample Code

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
import warnings
warnings.filterwarnings('ignore')
dataset = pd.read_csv('diabetes.csv')
# Preview data
dataset.head()
# Features data-type
dataset.info()
dataset.describe().T
dataset.isnull().sum()
# # Step 2: Data Visualization
sns.countplot(x = 'Outcome',data = dataset)
```

```
# Histogram of each feature
import itertools
col = dataset.columns[:8]
plt.subplots(figsize = (20, 15))
length = len(col)
for i, j in itertools.zip_longest(col, range(length)):
plt.subplot((length/2), 3, j + 1)
plt.subplots_adjust(wspace = 0.1,hspace = 0.5)
dataset[i].hist(bins = 20)
plt.title(i)
plt.show()
# Scatter plot matrix
from pandas.tools.plotting import scatter_matrix
scatter_matrix(dataset, figsize = (20, 20));
sns.pairplot(data = dataset, hue = 'Outcome')
plt.show()
# Heatmap
sns.heatmap(dataset.corr(), annot = True)
plt.show()
# # Step 3: Data Preprocessing
dataset_new = dataset
dataset_new[["Glucose", "BloodPressure", "SkinThickness", "Insulin", "BMI"]] =
dataset_new[["Glucose", "BloodPressure", "SkinThickness", "Insulin",
"BMI"]].replace(0, np.NaN)
dataset_new.isnull().sum()
# Replacing NaN with mean values
dataset_new["Glucose"].fillna(dataset_new["Glucose"].mean(), inplace = True)
```

```
dataset_new["BloodPressure"].fillna(dataset_new["BloodPressure"].mean(),
inplace = True)
dataset_new["SkinThickness"].fillna(dataset_new["SkinThickness"].mean(),
inplace  = True)
dataset_new["Insulin"].fillna(dataset_new["Insulin"].mean(), inplace = True)
dataset_new["BMI"].fillna(dataset_new["BMI"].mean(), inplace = True)
dataset_new.describe().T
# Feature scaling using MinMaxScaler
from sklearn.preprocessing import MinMaxScaler
sc = MinMaxScaler(feature_range = (0, 1))
dataset_scaled = sc.fit_transform(dataset_new)
dataset_scaled = pd.DataFrame(dataset_scaled)
# Selecting features - [Glucose, Insulin, BMI, Age]
X = dataset_scaled.iloc[:, [1, 4, 5, 7]].values
Y = dataset_scaled.iloc[:, 8].values
import train_test_split
X_train, X_test, Y_train, Y_test = train_test_split(X, Y, test_size = 0.20,
random_state = 42, stratify = dataset_new['Outcome'] )
#  Checking dimensions
print("X_train shape:", X_train.shape)
print("X_test shape:", X_test.shape)
print("Y_train shape:", Y_train.shape)
print("Y_test shape:", Y_test.shape)
# # Step 4: Data Modelling
from sklearn.linear_model import LogisticRegression
logreg = LogisticRegression(random_state = 42)
logreg.fit(X_train, Y_train)
```

```
# Plotting a graph for n_neighbors
from sklearn import metrics
from sklearn.neighbors import KNeighborsClassifier
X_axis = list(range(1, 31))
acc = pd.Series()
x = range(1,31)
for i in list(range(1, 31)):
knn_model = KNeighborsClassifier(n_neighbors = i)
knn_model.fit(X_train, Y_train)
prediction = knn_model.predict(X_test)
acc = acc.append(pd.Series(metrics.accuracy_score(prediction, Y_test)))
plt.plot(X_axis, acc)
plt.xticks(x)
plt.title("Finding best value for n_estimators")
plt.xlabel("n_estimators")
plt.ylabel("Accuracy")
plt.grid()
plt.show()
print('Highest value: ',acc.values.max())
# K nearest neighbors Algorithm
from sklearn.neighbors import KNeighborsClassifier
knn = KNeighborsClassifier(n_neighbors = 24, metric = 'minkowski', p = 2)
knn.fit(X_train, Y_train)
# Support Vector Classifier Algorithm
from sklearn.svm import SVC
svc = SVC(kernel = 'linear', random_state = 42)
svc.fit(X_train, Y_train)
```

```
# Naive Bayes Algorithm
from sklearn.naive_bayes import GaussianNB
nb = GaussianNB()
nb.fit(X_train, Y_train)
# Decision tree Algorithm
from sklearn.tree import DecisionTreeClassifier
dectree = DecisionTreeClassifier(criterion = 'entropy', random_state = 42)
dectree.fit(X_train, Y_train)
# Random forest Algorithm
from sklearn.ensemble import RandomForestClassifier
ranfor = RandomForestClassifier(n_estimators = 11, criterion = 'entropy',
random_state = 42)
ranfor.fit(X_train, Y_train)
# Making predictions on test dataset
Y_pred_logreg = logreg.predict(X_test)
Y_pred_knn = knn.predict(X_test)
Y_pred_svc = svc.predict(X_test)
Y_pred_nb = nb.predict(X_test)
Y_pred_dectree = dectree.predict(X_test)
Y_pred_ranfor = ranfor.predict(X_test)
# # Step 5: Model Evaluation
from sklearn.metrics import accuracy_score
accuracy_logreg = accuracy_score(Y_test, Y_pred_logreg)
accuracy_knn = accuracy_score(Y_test, Y_pred_knn)
accuracy_svc = accuracy_score(Y_test, Y_pred_svc)
accuracy_nb = accuracy_score(Y_test, Y_pred_nb)
accuracy_dectree = accuracy_score(Y_test, Y_pred_dectree)
```

```
accuracy_ranfor = accuracy_score(Y_test, Y_pred_ranfor)
# Accuracy on test set
print("Logistic Regression: " + str(accuracy_logreg * 100))
print("K Nearest neighbors: " + str(accuracy_knn * 100))
print("Support Vector Classifier: " + str(accuracy_svc * 100))
print("Naive Bayes: " + str(accuracy_nb * 100))
print("Decision tree: " + str(accuracy_dectree * 100))
print("Random Forest: " + str(accuracy_ranfor * 100))
# Confusion matrix
from sklearn.metrics import confusion_matrix
cm = confusion_matrix(Y_test, Y_pred_knn)
sns.heatmap(pd.DataFrame(cm), annot=True)
# Classification report
from sklearn.metrics import classification_report
print(classification_report(Y_test, Y_pred_knn))
```

6.2 Heart Disease Sample code

```
from flask import Flask ,render_template,request,jsonify,session
from flask import Flask, render_template, url_for, request
import pandas as pd
import pickle
from sklearn.feature_extraction.text import CountVectorizer
from sklearn.naive_bayes import MultinomialNB
from sklearn.externals import joblib
import sqlite3 as sql
import base64
import pandas as pd
```

```
from sklearn.preprocessing import LabelEncoder
#from flask_bootstrap import Bootstrap
import numpy as np
from sklearn.utils import shuffle
import os
from flask import Flask, render_template, request, url_for,send_from_directory
import os
#from geo import getTweetLocation
app = Flask(__name__)
app.secret_key = 'any random string'
PEOPLE_FOLDER = os.path.join('static', 'people_photo')
@app.route('/', methods=['GET', 'POST'])
def home():
return render_template('index.html')
def validate(username,password):
con = sql.connect('static/chat.db')
completion = False
with con:
cur = con.cursor()
cur.execute('SELECT * FROM persons')
rows = cur.fetchall()
for row in rows:
dbuser = row[1]
dbpass = row[2]
if dbuser == username:
completion = (dbpass == password)
return completion
```

```
@app.route('/login', methods=['GET', 'POST'])
def login():
error = None
if request.method == 'POST':
username = request.form['username']
password = request.form['password']
completion = validate(username,password)
if completion == False:
error = 'invalid Credentials. please try again.'
else:
session['username'] = request.form['username']
return render_template('index.html')
return render_template('first.html', error=error)
@app.route('/register', methods = ['GET','POST'])
def register():
if request.method == 'POST':
try:
name = request.form['name']
username = request.form['username']
password = request.form['password']
with sql.connect("static/chat.db") as con:
cur = con.cursor()
cur.execute("INSERT INTO persons(name,username,password) VALUES
(?,?,?)",(name,username,password))
con.commit()
msg = "Record successfully added"
except:
```

```
con.rollback()
msg = "error in insert operation"
finally:
return render_template("index.html",msg = msg)
con.close()
return render_template('register.html')
@app.route('/first',methods = ['POST'])
def first():
return render_template('first.html')
@app.route('/form',methods = ['POST'])
def form():
return render_template('form.html')
#prediction function
def ValuePredictor(to_predict_list):
to_predict =  np.array(to_predict_list).reshape(1,13)
loaded_model = pickle.load(open("model/heart_model.pkl","rb"))
result = loaded_model.predict(to_predict)
return result[0]
@app.route('/result',methods = ['POST'])
def result():
```

SYSTEM TESTING

CHAPTER 7

SYSTEM TESTING

7.1 Unit testing

Unit testing primarily focuses on verification on the fundamental element of the program model. By using the in depth design description most significant control paths are tested to find the errors within the limits of the module. In this system each sub module has been tested individually. Their input field validations are tested.

Test to be conducted	Input	Expected Output
Web app	13 Attributes entered by user	Should support HTML 5, CSS 3
GUI Test	On click event on predict button	All fields on the page

Table 1: Unit Testing

7.2 Integration testing

Integration testing is a well-structured approach for building the structure of the program and at the same time carrying out tests to spot errors within the program. Individual modules are very prone to interface errors that is why we should not assume that they will work perfectly when put together. Therefore, the major problem which arises is “putting them together”. Once all the individual unit have been tested there is need to test how they were put together to ensure no data is lost across the interface. The purpose is to make sure that each module does not have an unfavorable impact on another. After unit testing each and every sub module is tested with integrating each other.

7.3 Acceptance testing

Acceptance testing is a testing methodology that is used to ensure that the project software has met the requirement specifications or not. The sole purpose of this testing methodology is to examine the system's agreement with the requirements of business and then cross check if it is has met the set benchmarks for delivery to the client.
Acceptance test falls under black box system test. Basically every acceptance test represents an accepted result from the system. However, the correctness of the acceptance test has to be verified by the customers. Reviewing test scores are used to determine the priority of the failed tests. User acceptance testing of the system is crucial element for the success of any system. This is done by keeping the following points in mind -

- Input screen design.
- Output screen design.

Test ID	**Test Category**	**Test Description**
TC – 1	Network Connectivity	Check the connectivity of the network
TC – 2	Database Connectivity	Proper connection of front end and back end.

Table 2: Acceptance Testing

7.4 Program Testing

The main purpose of program testing is to eradicate logical and syntax errors in the program. If a program statement violates the rules of the language in which it is written, it is called a syntax error. These errors pop up as error messages which are generated by the computer system. On the other hand, a logical error includes incorrect data fields, invalid combination and out-off-range. The problem is that compiler is not going to deduct logical errors, so the programmer must examine the output as well.

7.5 Security Testing

Security testing tries to authenticate the protection mechanisms of the software. To be more specific it helps in protecting it from illegal penetration. The security of the system must be tested for invulnerability from frontal attack as well as from rear attack.

7.6 Performance Analysis

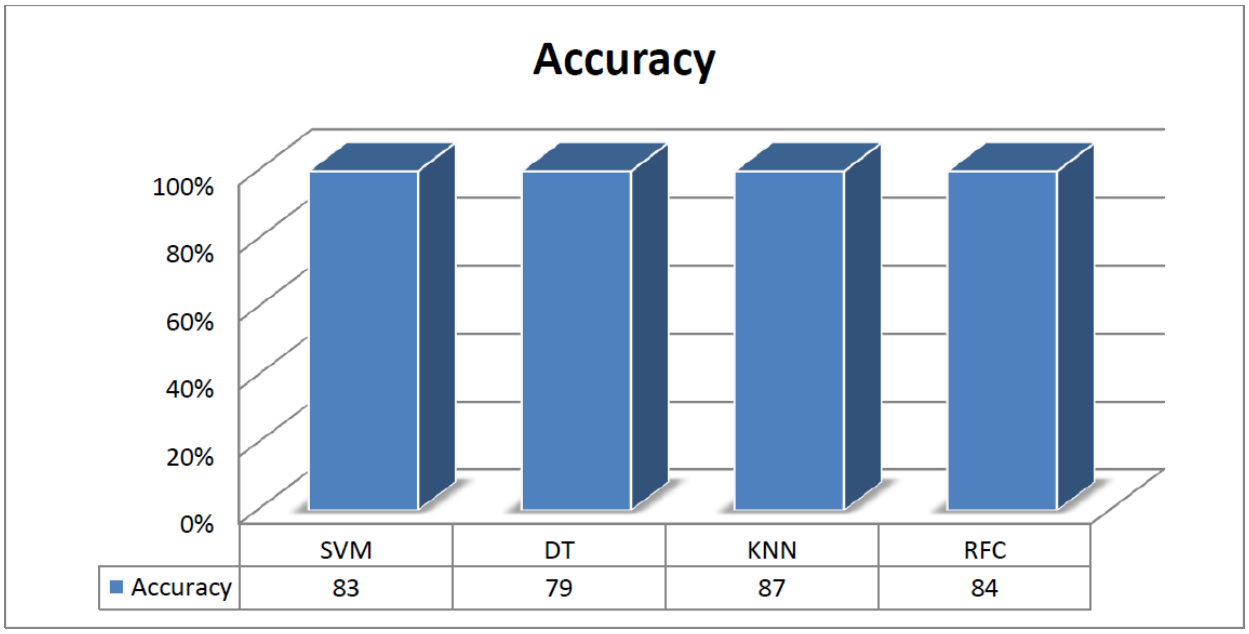


Fig . 7.6 Performance Analysis

SVM - Support vector machine
DT - Decision Tree
KNN - K-Nearest Neighbour
RF - Random Forest

CONCLUSION

CHAPTER 8

8.1 Conclusion

At first, the four algorithms were implemented. Datasets were trained for all the algorithms individually. After this, all of them were tested. The most efficient algorithm was to be selected based on various criteria. We found out that KNN algorithm was the most efficient out of the four algorithms with an accuracy of 87.0%. Decision tree, Support Vector Machine and Random Forest Classifier had accuracy of 79.0%, 83.0% and 84.0% respectively. Thus, KNN algorithm was further implemented using a better user interface in form of a web application. For this HTML5, CSS, JS and Flask (Python's micro web-framework) were used. This would help the end users get a preliminary prediction about the condition of their heart and Diabetes. Since heart diseases and Diabetes are a major killer in India and throughout the world, application of a promising technology like machine learning to the initial prediction of Covid-19 will have a profound impact on the society.
This will tell the user if they are at a risk and if they need to visit the doctor. This will help reduce the death rate due to covid. Hence by using the above approach successful analysis of heart diseases and Diabetes of the individual was performed and the result was obtained which predicted the risk of Covid-19 based on the parameters provided by the user.

8.2 Future Enhancement

By using the machine learning concept, newly trained dataset can be used for an even more accurate prediction system. Accounts can be created for each user and then by referring the past choice history of user's heart condition and Diabetes can be monitored to tell if there is any improvement or if the condition has deteriorated. Our future enhancement is, we going to develop it as Android app.

APPENDICES

A.1. SAMPLE SCREENSHOTS

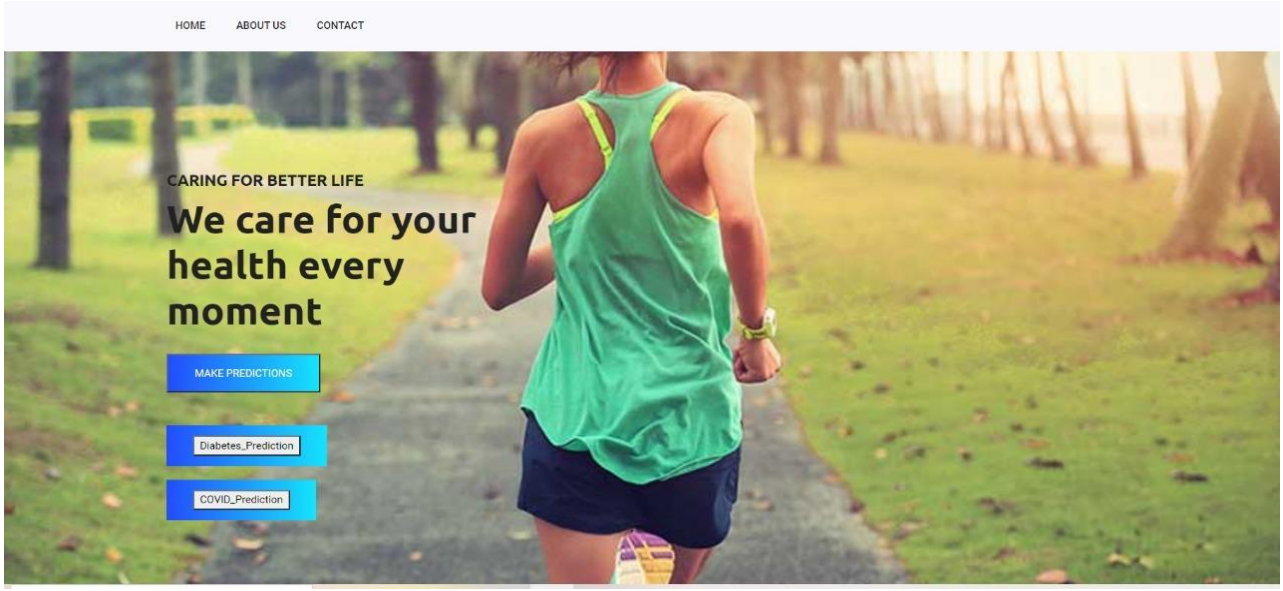


Fig. 1 Home Screen

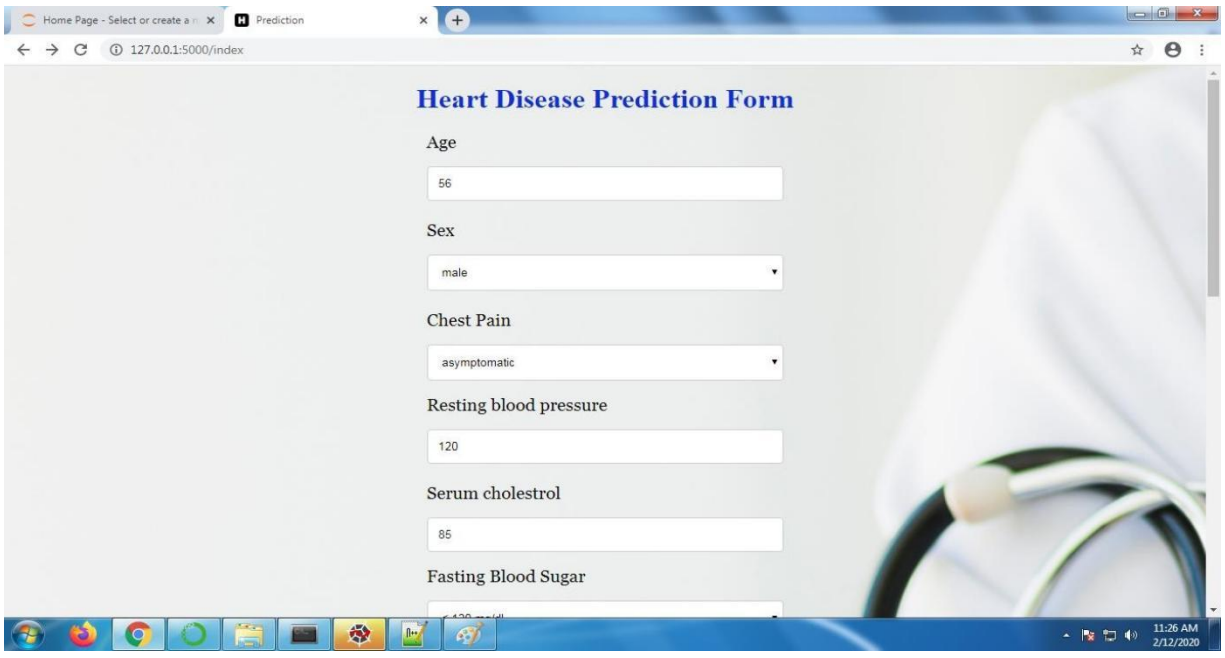


Fig.2 Input field of Heart Disease Form

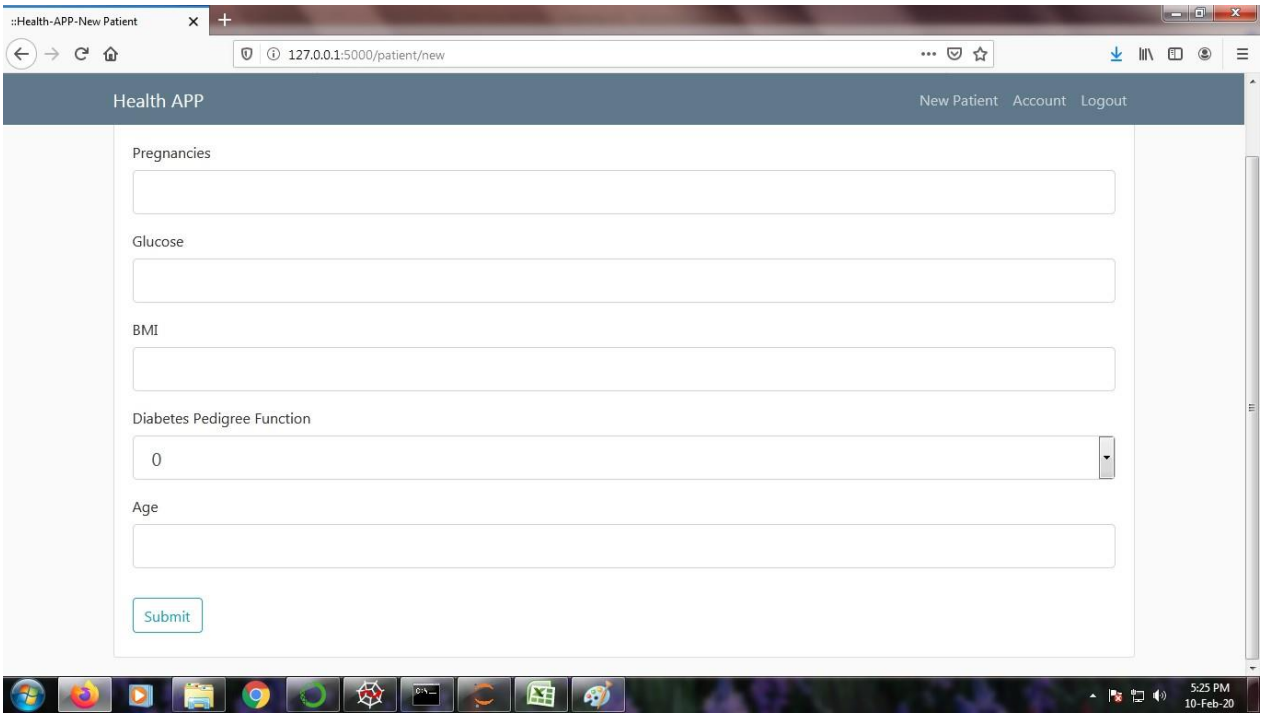


Fig.3 Input field of Diabetes Form

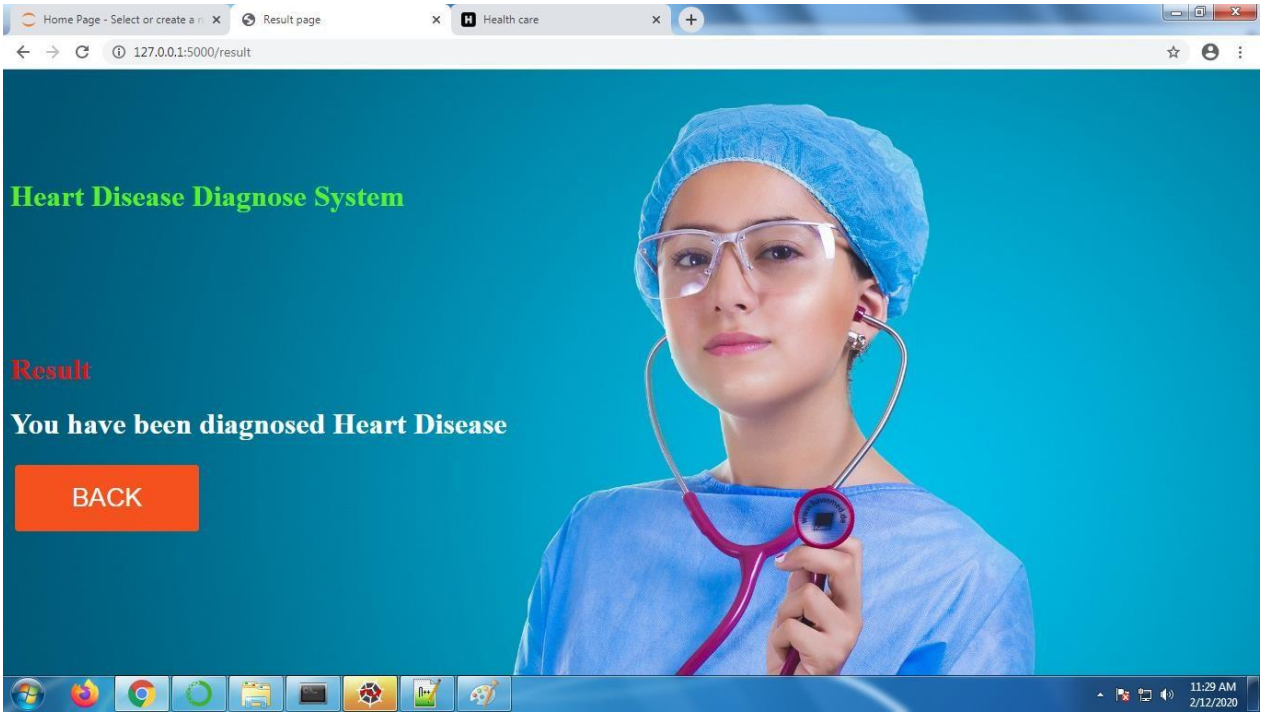


Fig.4 Output field of Heart Disease

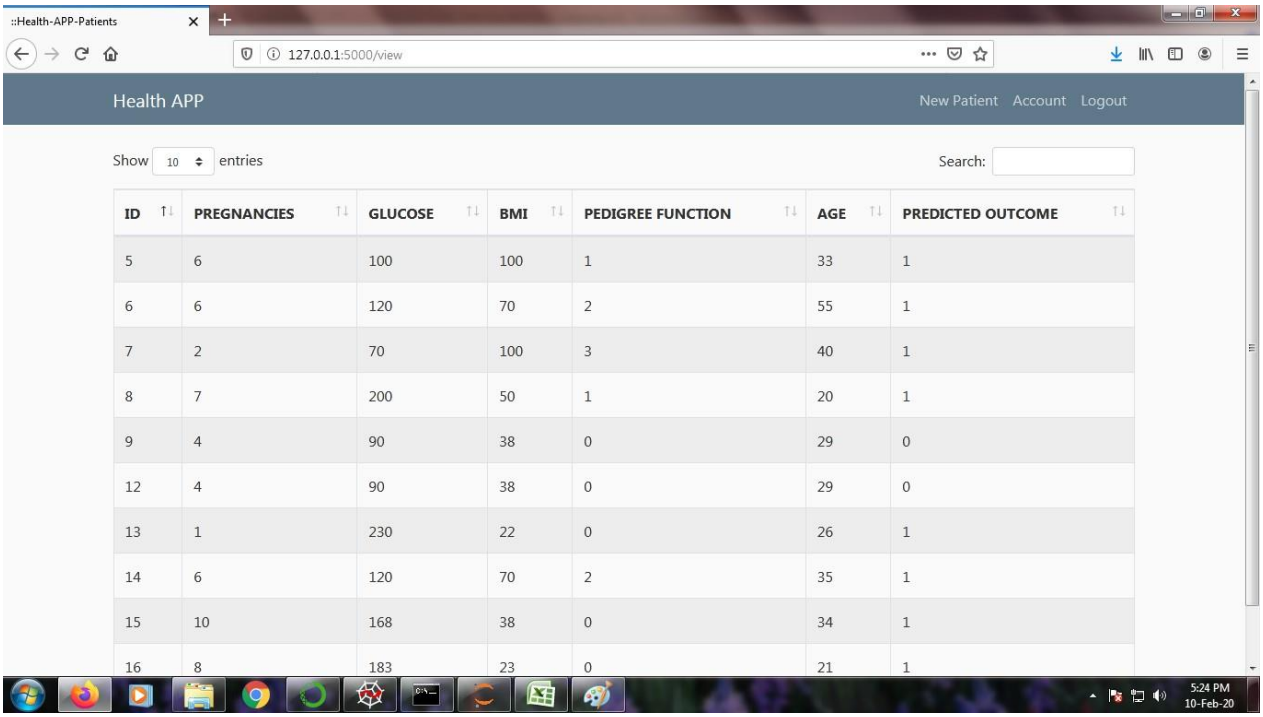


Fig.5 Output field of Diabetes

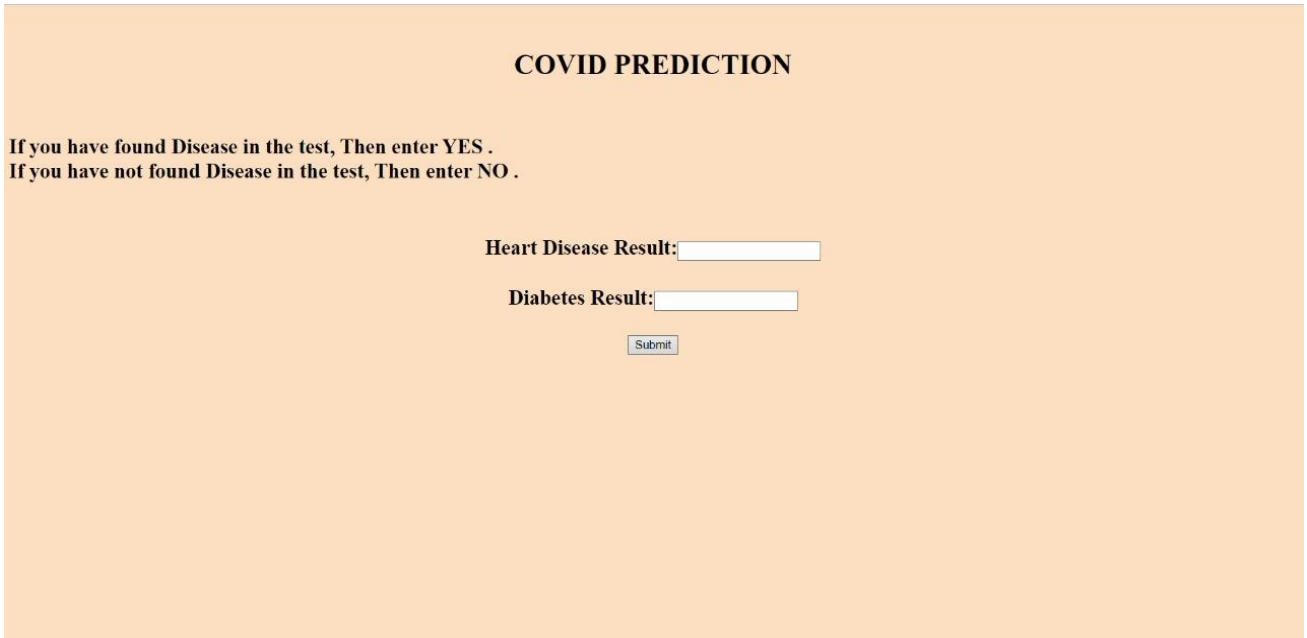


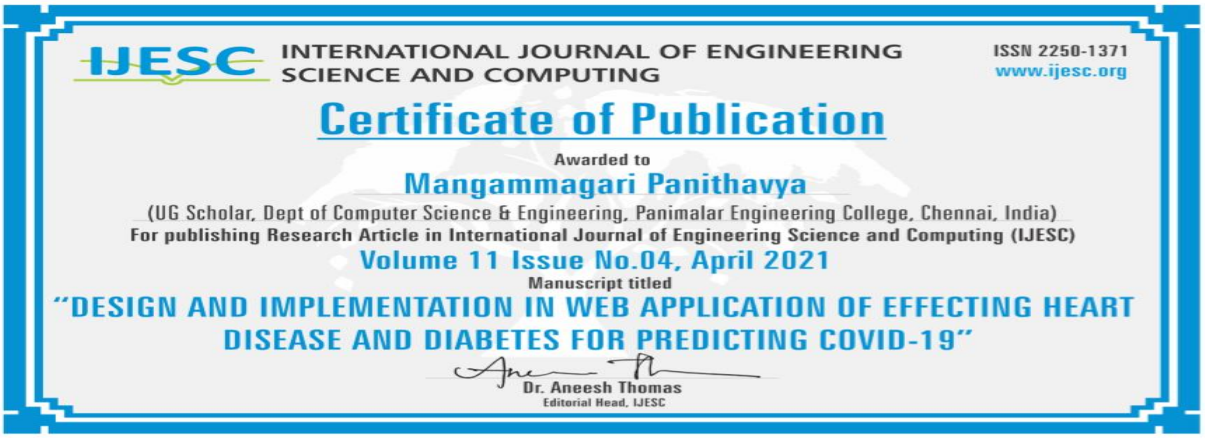
Fig.5 Covid Prediction

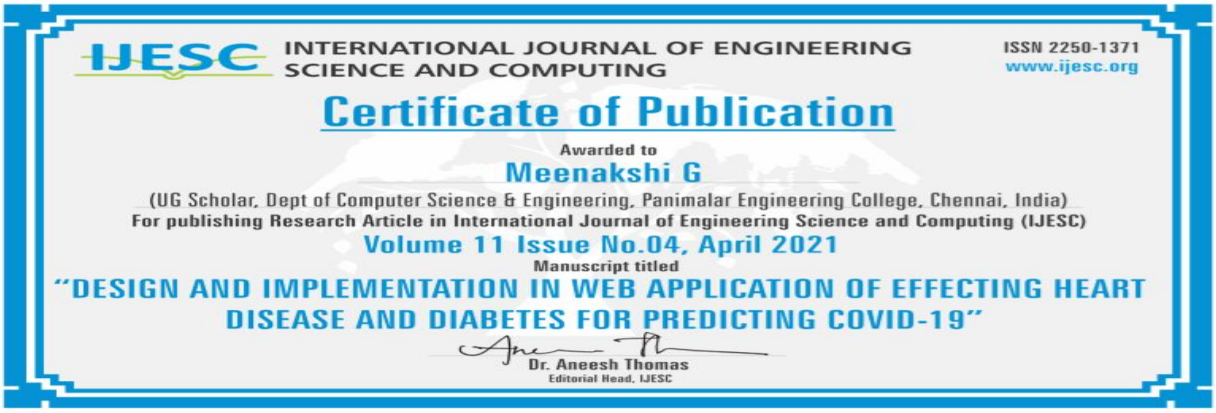
PUBLICATIONS

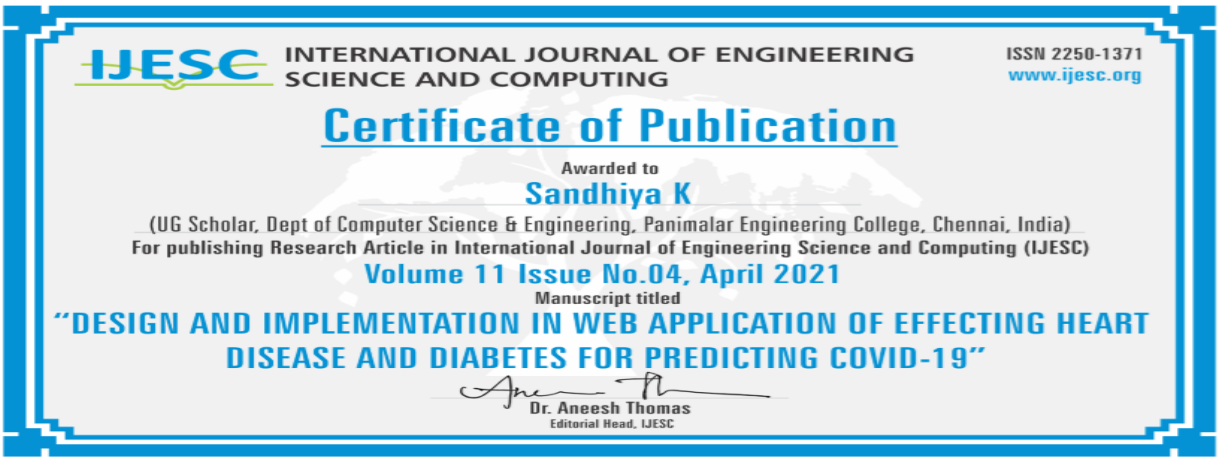
Journal Name : International Journal of Engineering Science and Computing

Paper Title : Design and Implementation in Web Application of Effecting Heart Disease and Diabetes for Predicting Covid-19

Publication Issue : Volume 11 Issue No.04,April-2021







ISSN 2321 3361 © 2021 IJESC

IJESC

Research Article Volume 11 Issue No.04

Design and Implementation in Web Application of Effecting Heart Disease and Diabetes for Predicting Covid-19

Kanchana. A [1], Manganmagari Panithavya[2], Meenakshi. G[3], Sandhiya. K[4]
Assistant Professor [1], UG Scholar [2, 3, 4]
Department of Computer Science and Engineering
Panimalar Engineering College, Chennai, India

Abstract:
Heart disease and Diabetes is one among the foremost vital causes of mortality within the world nowadays. Prediction of COVID -19 could be a critical challenge within a space of clinical knowledge analysis. Machine learning (ML) has been shown to be effective in aiding in creating selections and predictions from the big amount of knowledge made by the health care trade. We've conjointly seen ML techniques being employed in recent developments in numerous areas of Internet of Things (IoT). Numerous studies offer solely compared. Comparison of the various machine learning techniques employed in this study reveals which algorithm is best fitted for prediction of COVID-19.

Keywords: Heart Disease, Diabetes, ML(Machine Learning), Covid-19.

I. INTRODUCTION

Heart is one of the most vital organs for the proper functioning of our body. According to a survey by WHO, 31% of the worldwide deaths every year occurs due to Cardiovascular Diseases (CVDs). Also, more than 75% of these deaths occur in low and middle income countries including India. The main challenge is to accurately predict the existence of CVDs inside human body. The older techniques have not been very successful in efficiently predicting the heart diseases. Many medical instruments are available in the market for the prediction of heart diseases but there are some drawbacks of these instruments like they are very costly, they are not efficient enough for predicting heart diseases. Age, Sex, Blood Pressure, Cholesterol, Blood Sugar, Diabetes, etc. and some lifestyle factors like obesity, eating unhealthy food, less physical activity, smoking, consumption of alcohol ,etc. are some of the major risk factors that leads to heart diseases. Most of the lifestyle risk factors are controllable. In the last few decades, medical science has used the technological advancements very well to improve the quality of healthcare. These advancements in technology have paved ways for accurate diagnosis and prediction of Diabetes is a condition where person's body is unable to balance glucose-insulin level after various prandial modes. The patients of diabetes have increased exponentially from past few years due to their unbalanced diet and unhealthy lifestyle. There are around 463 million diabetic people around the world who can be benefited by smart healthcare technologies to improve their quality of life. Diabetes may lead to heart disease, kidney infection, blindness and nerve damages. It has been observed that severe acute respiratory syndrome corona virus 2 (SARS-CoV-2) has infected the diabetic patients largely among the others. People with diabetes were also more susceptible to previous pandemic known as Middle East respiratory syndrome (MERS) and severe Acute Respiratory Syndrome (SARS) type of corona virus along with H1N1 type of severe influenza during 2009 SARS-CoV-2 has affected mainly to old-age people and persons having underneath health conditions. The various case studies reported diabetes as the major pre-existing comorbidity among COVID-19 patients.

II. LITERATURE REVIEW

HImanshu Sharma, M A Rizvi.[1] has proposed a Prediction heart disease using Machine Learning. In this paper, typically explains the Health care field has a vast amount of data, for processing those data certain techniques are used. Data mining is one of the techniques often used. Heart disease is the Leading cause of death worldwide. This System predicts the arising possibilities of Heart Disease. The outcomes of this system provide the chances of occurring heart disease in terms of percentage. The datasets used are classified in terms of medical parameters. This system evaluates those parameters using data mining classification technique. The datasets are processed in python programming using two main Machine Learning Algorithm namely Decision Tree Algorithm and Naive Bayes Algorithm which shows the best algorithm among these two in terms of accuracy level of heart disease.

Amit M. Joshi, Urvashi P. Shukla, Saraju P. Mohanty,[2] has done a paper on Smart Healthcare for Diabetes duringCOVID-19. In this paper explains the The diabetic patients are at higher risk from novel coronavirus disease 2019 (COVID-19) that spreads through Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2). There are around 20% to 50% of COVID-19 cases had diabetes across the different re-gions from the world. This article discusses recommendations and associated risk for diabetic patients to balance glyncemic profile during COVID-19 outbreak. It also discusses the case study of various countries with impact of COVID-19 for diabetic

patients. It presents emerging smart healthcare that can potentially safeguard against COVID-19.

Sanchayita Dhar, Krishna Roy, Tanuree Dey, Pritha Datta, Ankur Biswas.[3] has proposed on A Hybrid Machine Learning Approach for Prediction of Heart Diseases.In this paper explains the Heart diseases are the chief cause of death all over the world over the last few decades. To avoid heart disease or coronary illness and discover indications early, individuals over 55 years must have a total cardiovascular checkup. Researchers and specialists developed various intelligent techniques to improve capacity of the health care professionals in recognition of cardiovascular disease. In cardiovascular disease finding and treatment, single data mining strategies are giving the reasonable precision and accuracy. Nevertheless the usage data mining procedure be capable of reducing the number of test that is required to be carried out. In order to decrease the Figure of deaths from heart diseases there has to be a quick and efficient detection technique providing better accuracy and precision. The aim of this paper is to present an efficient technique of predicting heart diseases using machine learning approaches. Hence we proposed a hybrid approach for heart prediction using Random forest classifier and simple k-means algorithm machine learning techniques. The dataset is also evaluated using two other different machine learning algorithms, namely, J48 tree classifier and Naive Bayes classifier and results are compared. Results attained through Random forest classifier and the corresponding confusion matrix shows robustness of the methodology.

Md. Razu Ahmed, S M Hasan Mahmud, Md. Altab Hossin, Hosney Jahan, Shek Rashed Haider Noori.[4] has done a paper on A Cloud Based Four-Tier Architecture for Early Detection of Heart Disease with Machine Learning Algorithms.In this paper, Heart disease prediction and detection has long been considered as a critical issue. Early detection of heart disease is an important issue in health care services (HCS). In growing amount of health care systems, patients are offered expensive therapies and operation that is quiet expensive for developing countries. Recently, heart disease is a prominent public chronic disease, ex. it's a growing concern in the US. The main reason of these diseases are tobacco consumption, bad life style, lack of physical activity and the intake of alcohol. Therefore, there is a need for the cloud based architecture that can efficiently predict and track health information.

GG Ladha.[5] has done a paper on A computation analysis to predict diabetes based on data mining.In this paper, explains the report suggested from different health organization shows the alarming condition due to diabetes worldwide. Different researchers around the globe have investigated it on different parameters and the investigation is going on for the early stage detection. The main objective of this paper is explore and elaborate the methodological prospective to predict the diabetes based on the dataset presented. This exploration is helpful in providing us a direction to find out the research gaps, so that future efficient diabetes framework can be developed. This computation analysis also provides us the parametric exploration along with the knowledge of the attributes and the way of implementing the classification framework.

D.Shetty.[6] has done a paper on Diabetes disease prediction using data mining. In this paper explains the Data mining is a subfield in the subject of software engineering. It is the methodical procedure of finding examples in huge data sets including techniques at the crossing point of manufactured intelligence, machine learning, insights, and database systems. The goal of the data mining methodology is to think data from a data set and change it into a reasonable structure for further use. Our examination concentrates on this part of Medical conclusion learning design through the gathered data of diabetes and to create smart therapeutic choice emotionally supportive network to help the physicians. The primary target of this examination is to assemble Intelligent Diabetes Disease Prediction System that gives analysis of diabetes malady utilizing diabetes patient's database. In this system, we propose the use of algorithms like Bayesian and KNN (K-Nearest Neighbor) to apply on diabetes patient's database and analyze them by taking various attributes of diabetes for prediction of diabetes disease.

III. EXISTING SYSTEM

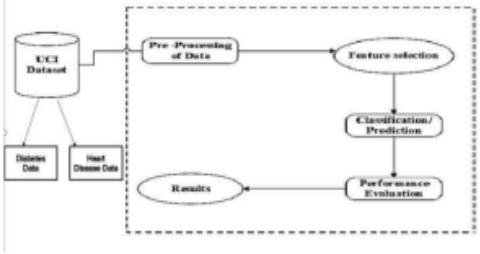
Before, the Doctors only view the report to convey the result to the patients. There are some problems appeared while seeing the laboratory details, they can't predict it properly. There is some difficulty in existing project, they created for some other purpose to test for different disease prediction , but According to covid 19, Diabetes and heart disease result is very important to predict , whether the person will be affected by covid-19 or not.

IV. PROPOSED SYSTEM

The prime objective of this project is to find the factors impacting the health of the society with the greater accuracy level and giving the good result. We have used python and pandas operations to perform heart disease and Diabetes classification of the Cleveland UCI repository. It provides an easy-to-use visual representation of the dataset, working environment and building the predictive analytics. ML process starts from a pre- processing data phase followed by feature selection based on data cleaning, classification of modeling performance evaluation, and the results with improved accuracy. This software is used to predict the covid by checking the heart disease and diabetes report.

V. SYSTEM ARCHITECTURE

A system architecture is the conceptual model that defines the structure, behavior, and more views of a system. An architecture description is a formal description and representation of a system, organized in a way that supports reasoning about the structures and behaviors of the system.



At first, we getting the data of Diabetes and Heart Disease from UCI Dataset, then we are Pre-Processing of Data. After that we

are doing the Feature Selection for classification / Prediction. Then the Performance Evaluation Occurs for the result.

VI. METHODOLOGY

1. **Dataset Description:** The objective of the dataset is to predict whether or not a patient has diabetes, based on certain diagnostic measurements included in the dataset. The datasets consists of several medical predictor variables and one target variable, Outcome. Predictor variables include the number of pregnancies the patient has had, their BMI, insulin level, age, chest Pain, Cholesterol, FBP, ECG and soon.

DATASET	DESCRIPTION
Pregnancies	Number of times pregnant
Glucose	Plasma glucose concentration a 2 hours in an oral glucose tolerance test
Blood Pressure	Diastolic blood pressure (mm Hg)
Skin Thickness	Triceps skin fold thickness (mm)
Insulin	2-Hour serum insulin (mu U/ml)
BMI	Body mass index (weight in kg/(height in m)^2)
Diabetes Pedigree Function	Diabetes pedigree function
Age	Age (years)
Outcome	Class variable (0 or 1) 268 of 768 are 1, the others are 0.
Chest Pain	1-4 (1: Typical Angina, 2: Atypical Angina, 3: Non-anginal, 4: Asymptotic)
Cholesterol	Serum Cholesterol in mg/dl
Fasting Blood Sugar	Blood sugar content before food intake if >120 mg/dl.
ECG	Resting Electrocardiographic results , 0-1 (0: Normal, 1: Having ST-T wave)
Max Heart Rate	Maximum heart beat rate, Beats/min
Exercise Induced Angina	Has pain been induced by exercise, 0-1 (0: No, 1: Yes).
Old Peak	ST depression induced by exercise relative to rest,0-4.
Slope of Peak Exercise	Slope of the peak exercise ST segment, 1-3 (1: Up sloping, 2: Flat, 3: Down sloping)
Ca	Number of vessels colored by fluoroscopy, 0-3
Thala	Displays the thalassemia
Num	Diagnostics of Heart Disease

2. **Data Pre-Processing**

Diabetes disease and Heart Disease data is pre-processed after collection of various records. The dataset contains a total of 769 patient records, where 6 records are with some missing values. Those 6 records have been removed from the dataset and the remaining 763 patient records are used in pre-processing.

Pregnanc	Glucose	BloodPres	SkinThickr	Insulin	BMI	DiabetesF	Age	Outcome
6	148	72	35	0	33.6	0.627	50	1
1	85	66	29	0	26.6	0.351	31	0
8	183	64	0	0	23.3	0.672	32	1
1	89	66	23	94	28.1	0.167	21	0
0	137	40	35	168	43.1	2.288	33	1
5	116	74	0	0	25.6	0.201	30	0
3	78	50	32	88	31	0.248	26	1
10	115	0	0	0	35.3	0.134	29	0
2	197	70	45	543	30.5	0.158	53	1
8	125	96	0	0	0	0.232	54	1
4	110	92	0	0	37.6	0.191	30	0
10	168	74	0	0	38	0.537	34	1
10	139	80	0	0	27.1	1.441	57	0
1	189	60	23	846	30.1	0.398	59	1
5	166	72	19	175	25.8	0.587	51	1
7	100	0	0	0	30	0.484	32	1
0	118	84	47	230	45.8	0.551	31	1
7	107	74	0	0	29.6	0.254	31	1
1	103	30	38	83	43.3	0.183	33	0

	A	B	C	D	E	F	G	H	I	J
1	age	sex	cp	trestbps	chol	fbs	restecg	thalach	exang	num
2	67	1	4	160	286	0	2	108	1	2
3	67	1	4	120	229	0	2	129	1	1
4	37	1	3	130	250	0	0	187	0	0
5	41	0	2	130	204	0	2	172	0	0
6	56	1	2	120	236	0	0	178	0	0
7	62	0	4	140	268	0	2	160	0	3
8	57	0	4	120	354	0	0	163	1	0
9	63	1	4	130	254	0	2	147	0	2
10	53	1	4	140	203	1	2	155	1	1
11	57	1	4	140	192	0	0	148	0	0
12	56	0	2	140	294	0	2	153	0	0
13	56	1	3	130	256	1	2	142	1	2
14	44	1	2	120	263	0	0	173	0	0
15	52	1	3	172	199	1	0	162	0	0
16	57	1	3	150	168	0	0	174	0	0
17	48	1	2	110	229	0	0	168	0	1
18	54	1	4	140	239	0	0	160	0	0
19	48	0	3	130	275	0	0	139	0	0
20	49	1	2	130	266	0	0	171	0	0
21	64	1	1	110	211	0	2	144	1	0
22	58	0	1	150	283	1	2	162	0	0
23	58	1	2	120	284	0	2	160	0	1

3. Feature Section

From among the 8 attributes of the data set, one attributes pertaining to age is used to identify the personal information of the patient. The remaining 7 attributes are considered important as they contain vital clinical records. Clinical records are vital to diagnosis and learning the severity of diabetes disease.

4. Classification Modeling

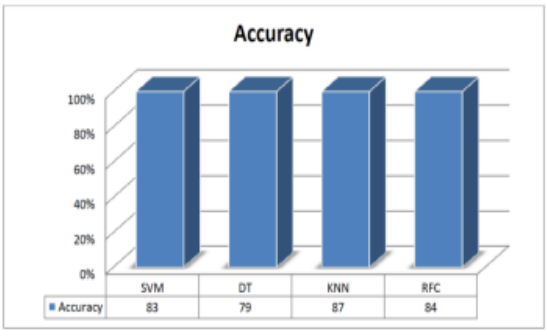
The clustering of datasets is done on the basis of the variables and criteria of Decision Tree (DT) features. Then, the classifiers are applied to each clustered dataset in order to estimate its performance. The best performing models are identified from the above results based on their low rate of error.

- Support Vector Machine
- Decision Trees
- K Nearest neighbors
- Random Forest Classifier

5. Performance Analysis

Several standard performance metrics such as accuracy, precision and error in classification have been considered for the computation of performance efficiency of this model.

✔ K Nearest neighbors : 87.571
✔ Support Vector Classifier : 83.376
✔ Decision tree : 79.181
✔ Random Forest : 84.974



6. Algorithms

1. Support Vector Machine

• Support Vector Machine (SVM) is a supervised learning methodology which is used to analyse data used for classification and regression analysis. The training data is fed into an SVM training algorithm which then builds a model that classifies new examples to one category or the other, that makes SVM a non- probabilistic binary linear classifier.
• The accuracy score of SVM was found to be 83.0%.

1. Decision Tree

• Decision Tree algorithm belongs to the family of supervised learning algorithms. The goal of using a Decision Tree is to create a training model that can use to predict the class or value of the target variable by learning simple decision rules inferred from prior data(training data)
•The accuracy score of Decision Tree was found to be 79.0.

2. K Nearest Neighbours

• During testing, KNN classification algorithm has to find K-nearest neighbours of a new instance. This is time consuming if we do exhaustive comparison. K-nearest neighbours use the local neighborhood to obtain a prediction.
• The accuracy score of KNN was found to be 87.0%.

3. Random Forest Classifier

• Select random samples from a given dataset.
• Construct a decision tree for each sample and get a prediction result from each decision tree.
• Perform a vote for each predicted result.
• Select the prediction result with the most votes as the final prediction
• The accuracy of Random Forest Classifier was found to be 84.0%.

VI. CONCLUSION

At first, the four algorithms were implemented. Datasets were trained for all the algorithms individually. After this, all of them

were tested. The most efficient algorithm was to be selected based on various criteria. We found out that KNN algorithm was the most efficient out of the four algorithms with an accuracy of 87.0%. Decision tree, Support Vector Machine and Random Forest Classifier had accuracy of 79.0%, 83.0% and 84.0% respectively. Thus, KNN algorithm was further implemented using a better user interface in form of a web application. For this HTML5, CSS, JS and Flask (Python's micro web-framework) were used. This would help the end users get a preliminary prediction about the condition of their heart and Diabetes. Since heart diseases and Diabetes are a major killer in India and throughout the world, application of a promising technology like machine learning to the initial prediction of Covid-19 will have a profound impact on the society. This will tell the user if they are at a risk and if they need to visit the doctor. This will help reduce the death rate due to covid. Hence by using the above approach successful analysis of heart diseases and Diabetes of the individual was performed and the result was obtained which predicted the risk of Covid-19 based on the parameters provided by the user.

VII. FUTURE ENHANCEMENT

By using the machine learning concept, newly trained dataset can be used for an even more accurate prediction system. Accounts can be created for each user and then by referring the past choice history of user's heart condition and Diabetes can be monitored to tell if there is any improvement or if the condition has deteriorated. Our future enhancement is, we going to develop it as Android app.

3. Input of Diabetes

4. Output of Heart Disease

5. Output of Diabetes

VIII. EXPERIMENTAL AND RESULTS

1. Home Screen

2. Input of Heart Disease

IX. REFERENCES

[1]. Himanshu Sharma, M A Rizvi. : "Prediction of Heart Disease using MachineLearning Algorithms: A Survey ", International Journal on Recent and Innovation Trends in Computing and Communication, Volume: 5 Issue: 8 , 2017.

[2].Amit M. Joshi, Urvashi P. Shukla, Saraju P. Mohanty, "Smart Healthcare for Diabetes during COVID-19", IEEE, 2020, https:// ieeexplore. ieee.org/document/9174644.

[3]. Sanchavita Dhar, Krishna Roy, Tanuree Dey, Pritha Datta, Ankur Biswas.: "A Hybrid Machine Learning Approach for Prediction of Heart Diseases", IEEE, 2018, https:// ieeexplore.ieee. org/document/8777531.

[4]. Md. Razu Ahmed, S M Hasan Mahmud, Md. Altab Hossin, Hosney Jahan, Shek Rashed Haider Noori.: "A Cloud Based

Four-Tier Architecture for Early Detection of Heart Disease with Machine Learning Algorithms", IEEE, 2018.

[5]. GG Ladha, "A computation analysis to predict diabetes based on data mining", IEEE, 2018.

[6]. Shetty, "Diabetes disease prediction using data mining", IEEE, 2017.

[7].G. Shanmugasundaram, V. Malar Selvam, R. Saravanan, S.Balaji.: "An Investigation of Heart Disease Prediction Techni ques", IEEE, 2018.

[8].Seyedamin Pouriyeh, Sara Vahid, Giovanna Sannino, Giuseppe De Pietro, Hamid Arabnia, Juan Gutierrez.: "A comprehensive investigation and comparison of Machine Learning Techniques in the domain of heart disease", IEEE, 2017.

[9]. Zhilbert Tafa, "An intelligent system for diabetes prediction IEEE, 2020.

[10]. M Rout, "Prediction of Diabetes Risk based on Machine Learning Techniques", IEEE, 2020.

REFERENCES

[1] Himanshu Sharma, M A Rizvi. : “Prediction of Heart Disease using Machine Learning Algorithms: A Survey ”, International Journal on Recent and Innovation Trends in Computing and Communication, Volume: 5 Issue: 8 , 2017.

[2] Sanchayita Dhar, Krishna Roy, Tanuree Dey, Pritha, Datta, Ankur Biswas.: "A Hybrid Machine Learning Approach for Prediction of Heart Diseases", IEEE, 2018.

[3] G. Shanmugasundaram, V. Malar Selvam, R. Saravanan, S.Balaji.: "An Investigation of Heart Disease Prediction Techniques", IEEE, 2018.

[4] Md. Razu Ahmed, S M Hasan Mahmud, Md. Altab Hossin, Hosney Jahan, Shek Rashed Haider Noori.: "A Cloud Based Four-Tier Architecture for Early Detection of Heart Disease with Machine Learning Algorithms", IEEE, 2018.

[5] Seyedamin Pouriyeh, Sara Vahid, Giovanna Sannino, Giuseppe De Pietro, Hamid Arabnia, Juan Gutierrez.: "A comprehensive investigation and comparison of Machine Learning Techniques in the domain of heart disease", IEEE, 2017.

[6] Amit M. Joshi, Urvashi P. Shukla, Saraju P. Mohanty, “Smart Healthcare for Diabetes during COVID-19”, IEEE, 2020.

[7] Shetty, “Diabetes disease prediction using data mining”, IEEE, 2017.

[8] Zhilbert Tafa, “An intelligent system for diabetes prediction”, IEEE, 2020.

[9] M Rout, “Prediction of Diabetes Risk based on Machine Learning Techniques”, IEEE, 2020.

[10] GG Ladha, “A computation analysis to predict diabetes based on data mining”, IEEE, 2018.