PREDICTION OF CORONARY ARTERY DISEASE USING MACHINE LEARNING

Ajaygouda Yallappagoudar[1], Shashank G[2], Vishal Agarwal[3], Arun C M[4], Anitha M [5]

[1]UG Student, Department of Computer Science & Engineering, DSCE, Bangalore, Karnataka, India
[2]UG Student, Department of Computer Science & Engineering, DSCE, Bangalore, Karnataka, India
[3]UG Student, Department of Computer Science & Engineering, DSCE, Bangalore, Karnataka, India
[4]UG Student, Department of Computer Science & Engineering, DSCE, Bangalore, Karnataka, India
[5]Assistant Professor, Department of Computer Science & Engineering, DSCE, Bangalore, Karnataka, India

ABSTRACT

In recent times, there has been a lot of implementation with respect to machine learning and its application methods in the Medical field and is often referred to be a valuable rich information. Coronary Artery Disease (CAD) is one of the major causes of death all around the world and early detection can prevent it. The aim of this project is to predict Coronary artery disease using historical medical data. The dataset is then trained using random forest algorithm and visualized important predictive analytics using confusion matrix. We also focus on a designing a webpage which receives attributes as input and give them an accurate result upto 93%.

Keyword : - *Machine Learning, Random Forest, CAD, Framingham Dataset.*

1. INTRODUCTION

Coronary artery disease (CAD) is the most common type of heart disease affecting millions worldwide. Present-days heart disease is treated as one of the major causes of human death in the world. 10 percent of the total death occurs in the world is due to heart disease only. Hence the disease has become one of the biggest concerns in various countries of the world. As per Japan death rate statistics, heart disease occupies the second position. Due to the improvement of technology and the availability of automation, people perform very less physical work and use the mental ability which makes them prone to get heart disease. Due to this people are getting addicted to smoking, alcohol which leads to have big bellies. As per University of Rochester's Medical centre view the major source for Heart disease are overweight, lack of physical activity, fatness, consumption of malnutrition and tobacco. As heart disease is widely accepted as the major source of death hence medical analysis of heart disease becomes a regular need for every human being. Due to the number of ambiguity and risk factor the prediction of the disease became a very tough task for every physician. If the heart attack can be identified earlier then the life of the patient can be saved through proper medication and also harm to the heart can be saved up to a large extent. Heart diseases are of various types like Coronary artery disease, Valvular heart disease, Cardiomyopathy. These diseases mainly affect the arteries of the heart, blood in and out valves of heart, heart muscle squeezing. Proper heart functioning is really a highly essential thing for a healthy life. In Coronary artery disease cholesterol, calcium and some other substance getting deposited in the veins through which blood circulation is done. As a result of which some blockage is being created against the smooth passage of blood. Due to this the heart muscles will not get an adequate amount of oxygen which creates discomfort in the patient's chest and results as chest pain with the patient. As per WHO report up to 2030 around 23.6 million people in the world will die due heart disease. So there is a need to take some preventive steps to minimize the threats of heart disease. Practitioners mainly view the symptoms, expressions and medical test to identify the occurrence of the disease with the patient. For coronary artery disease identification doctors mainly uses SPECT and ECG methods. In SPECT method radioactive tracers are being injected in to blood for generating images of heart which are used by the doctors to know about the identification of coronary artery disease and also the prediction of the heart attack. ECG reports are used to know about abnormality in heart beating. The diagnosis made by the doctors about any disease is not always 100 percent correct. Hence various computerised tools are used in the healthcare domain. These tools are used to identify critical parameters for the diagnosis of the disease. Improvement in the health condition of any patient can be known by analysing various critical parameters related to their disease. The main goal of the intelligent systems assisting medical diagnosis is to predict the presence of any disease accurately. A number of input symptoms are used to indicate the occurrence of Coronary artery disease. Out of all age, sex and family medical history cannot be changed. But others symptoms like smoking, blood pressure, cholesterol level, physical exercise can be changed to reduce the possibility of Coronary artery disease. As so many parameters are involved while diagnosing the disease so the practitioner uses the present medical report of the patient. After going through the report doctors adopt the same method which was used for any previous patient having a similar type of test report. Some modern genetic algorithms now days are trying to find out some important data which are mainly contributing towards the occurrence of the heart disease instead of analysing a number of data. Through this the algorithms trying to identify the disease with an optimal number of parameters and consuming very less time. Recent development in the field of ML contributes a lot to medical science towards the development of intelligent systems. In last few decades various computational systems have developed which were helpful for the physicians in improving their diagnosis decisions.

2. LITERATURE SURVEY

Authors: Alberto Palacios Pawlovsky ‘An Ensemble Based on Distances for a kNN Method for Heart Disease Diagnosis’
Description: Alberto Palacios Pawlovsky has used an ensemble based KNN method which employs distance-based heart disease prediction. Here a two-phase method has been introduced. In the first phase different K values are chosen. For each K point we put all the 5 different distance formula like Euclid, Manhattan, Chebyshev, Canberra and Mahala Nobis and obtained distance value from the test instance and noted the class of each k neighbor. We also found the classification accuracy by taking different cross validation size from 10 percent to 90 percent of the total records. In second phase of the algorithm, we put majority based voting algorithm to assign the class to an unknown instance as per the majority class is chosen.

Authors: Yeshvendra K. Singh, Nikhil Sinha., and Sanjay K. Singh., ‘Heart Disease Prediction System Using Random Forest’
Description: This paper talks about using Random Forest method for detecting the heart disease. Introduced system uses all the 13 important input features for the prediction of heart disease provided UCI repository. The prediction system is implemented by removing the features between whom no correlation can be established. The enhancement in the accuracy is achieved by tuning various linearly dependent variables of the random forest like randomness, the number of trees, the minimum number of splits and the minimum number of leaf nodes. Initially the number of trees and the minimum number of splits are considered to find a better correlation with accuracy. Highest accuracy obtained when numbers of splits are 20 and the number of trees is 75.

Authors: TanmayKasbe , Ravi Singh Pippal ‘Design of Heart Disease Diagnosis System using Fuzzy Logic
Description: This paper presents an expert heart disease prediction system using fuzzy based logic. Here a fuzzy indicator functions like triangular and trapezoid are introduced for the implementation of a fuzzy expert system and fuzzy rule base. The fuzzy expert system first does categorization of independent features and dependent feature. In this stage, the features are experimented to observe its value range and its equivalent class. In the next stage Fuzzy rules over data is employed by the different combination of single or several features with AND, OR operator. Here the system development is done using total 86 fuzzy based rules with all possible arrangements. In the rule base, all the possible input variables with different combinations and its corresponding output value are used for the output level calculation. A relationship is established between all independent features value and their corresponding dependent feature values are set. This relation will be helpful to find out the class for an unknown instance.

Authors: Purushottam,Prof. (Dr.) Kanak Saxena,Richa Sharma, ‘Efficient Heart Disease Prediction System
Description: This paper put forward a well-organized heart disease prediction system which uses a large dataset from Cleveland which contains data about coronary diseases. First, the database pre-processing is done using all

Possible-MV algorithms. It is used mainly to fill the missing data in the dataset. After that classification decision rules are generated to do the proper classification through pruned rules, original rules, classified rules and rules without duplicates.

Authors: Hongmei Yan and Jun Zheng, et al, 'Selecting critical clinical features for heart diseases diagnosis with a real-coded genetic algorithm.'
Description: In this paper, they have put forward a real genetic algorithm related system which selects the essential features for diagnosing heart disease. This system assists the diagnosing of hypertension, chronic pulmonale , coronary artery disease, congenital heart disease and rheumatic valvular heart disease. They considered the dataset which consisted of 352 entities and each entity recorded 40 attributes. Among the 352 entities, 24 major features were identified which helped majorly in diagnosing the heart disease and these features helped in securing high accuracy for diagnosing heart disease.

Authors: Resul Das and Ibrahim Turkoglu, et al, 'Diagnosis of valvular heart disease through neural networks ensembles'
Description: They put forward many methodologies and tools for developing medical decision support system. Many researchers tried helping the physicians by developing intelligent systems which would increase the chance of diagnosing a heart disease. They established a method which used Statically Analysis System software 9.1.3 to treat CAD. In these various neural networks were combined to practice on the matching task which used neural networks ensemble model and led to increase in performance. The results were amazing as the experimental results secured for diagnosing the heart disease.95.91 percent specificity values, 89 percent classification accuracy and 80.95 percent sensitivity.

Authors: Yoon-Joo Park and Se-Hak Chun, et al, 'Cost-sensitive case-based reasoning using a genetic algorithm: Application to medical diagnosis'
Description: In this paper they have put forward a CSCBR - Cost-Sensitive Case-Based Reasoning, extraction technique which includes different misclassification cost into conventional case-based reasoning. A GA - genetic algorithm was introduced to examine the absence of disease. By using number of neighbors and boundary point they tried to reduce the misclassification errors costs into CBR. CSCBR helped to overpower the constant number of nearest neighbors. The best neighbors were selected by modifying maximum cut-off distance point and cut-of classification point. This technique was put forward in 5 various pharmaceutical datasets and then collated with CART and C5.0. It was deduced that the total misclassification cost of this technique was much lesser than the other cost-responsive methods.

Author: Rahul C Deo. 'Machine Learning In Medicine'
Description: This paper has put forward 2 major disadvantages in the current diagnosing system that they take less number of factors into consideration in the dataset by each tool and impotency of these systems to deduce major information on the disease. They used a 2-phase strategy to reach this maxim. In the first phase, based on the model of Na¨ıve-Bayes classifier they tried to introduce a common representation procedure and applied to the set of current features. In the second phase, based on various features of Bayes probabilistic judging and conditional probabilities they introduced a combination programme which was optimized from the genetic algorithm.

Authors: K.Rajeswari and V.Vaithiyanathan, et al, Feature Selection in Ischemic Heart Disease Identification using Feed Forward Neural Networks
Description: This paper has addressed about the choosing the features when collecting data in order to reduce the number of inputs under assessment. They introduced a system which introduced the efficiency with respect to accuracy, time and cost. They used an artificial neural network to select important features from the dataset. They considered the IHD database which consisted 712 entities, initially they took 12 features and 17 attributes as input to neural networks. The predicted accuracy was around 82.25 for testing and 89.4 percent for training.

3. ARCHITECTURE AND DESIGN

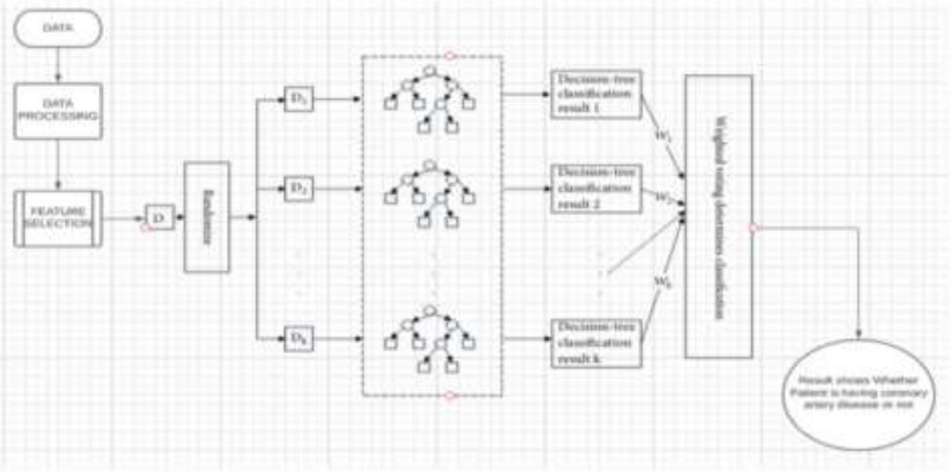

DATA
DATA PROCESSING
FEATURE SELECTION
D
Randomize
D_1
D_2
D_k
Decision-tree classification result 1
Decision-tree classification result 2
Decision-tree classification result k
W_1
W_2
W_k
Weighted voting determines classification
Result shows Whether Patient is having coronary artery disease or not

Figure 1: ML Model Training Diagram

The block diagram of Model Training in the proposed system is shown in above diagram • The main block diagram of the system which consist of main constituents they are Dataset, Data Processing, Feature Selection, Random Forest Algorithm. • The main architecture of the project is represented in above figure. • Input data is processed and then fed to the Pre-Processing and then given to random forest algorithm for prediction. • Output prediction result will be displayed in the UI.

Design

Following Steps were implemented:

- Input the data from the CSV file. The data from the Framingham dataset is imported into the system which is publicly available on Kaggle website.
- Cleaning the Data
- Handle missing values. Missing values can be handled by deleting the rows or columns having null values. If columns have more than half of rows as null then the entire column can be dropped. The rows which are having one or more columns values as null can also be dropped.
- Handling Duplicate Data. Duplicates are an extreme case of nonrandom sampling, and they bias your fitted model. Including them will essentially lead to the model overfitting this subset of points.
- Features that are prone to diseases are selected. Select K Best technique which is one of the feature selection method is used to select the features which are relatively important for the prediction variable and removing irrelavent features helps the model by increasing its accuracy.
- Create training and testing sets (using 75 percent of the data for the training and remaining for testing). Training and Testing is used to measure the accuracy of the model. The data is split into two parts, 75 percent for Training which is used to train the model and 25 percent for Testing which is used to test the trained model.
- Build a machine learning model to predict Coronary artery disease using Random forest algorithm. We have used Random Forest algorithm for our model and used Random forest classifier which creates a set of decision trees from randomly selected subset of training set. It then aggregates the votes from different decision trees to decide the final class of the test object. • The trained program is pickled to .pkl file. After training the model we dump that trained data into a pickle file(.pkl file) which will be later used for prediction using Flask.
- This .pkl file is unpickled in frontend. In frontend the pickle file is unpickled by deserializing the binary data into python object.
- Values of features are extracted from the frontend. The features are extracted from the form which is provided on the web page and put forward to prediction.

- Using predict method, results are obtained and appropriate web page is displayed through flask. The results are obtained by predict method and suitable web page is shown and appropriate measures are provided that the user should take in mere future.

4. PREDICTIVE MODELING

We have used the random forest machine learning algorithm to predict the coronary artery disease. Below diagram shows the random forest implementation model.

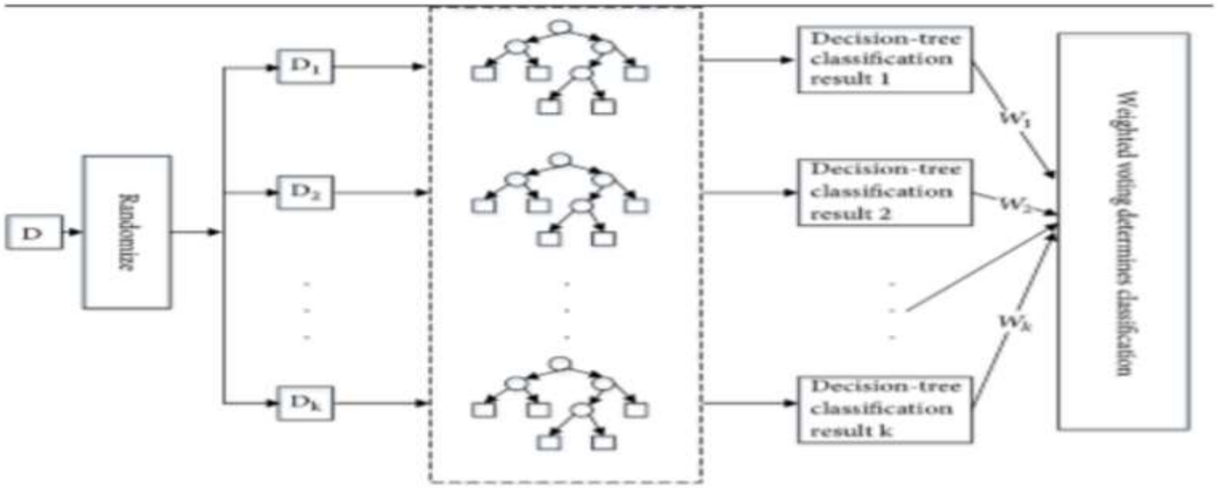

D
Randomize
D1
D2
Dk
Decision-tree classification result 1
Decision-tree classification result 2
Decision-tree classification result k
W1
W2
Wk
Weighted voting determines classification

Figure 2: Random Forest Implementation Model

Random forest is one of the supervised learning techniques. It is based on of ensemble learning, which is a process of combining multiple classifiers to solve a complex problem and to improve the performance of the model. It can be used both in regression and classification problem.

Random forest works by dividing dataset by row and column sampling where each set of these samples are deployed in different Decision Tree. More the number of trees in the random forest leads to higher accuracy and reduces the problem of overfitting. Each tree in this algorithm runs parallelly gets trained on dataset obtained through sampling.

Since the random forest joins different trees to anticipate the class of the dataset it is conceivable that some decision trees may anticipate the right yield, while others may not. But together, all the trees predict the correct output. Consequently, below are two suppositions for a superior random forest classifier:

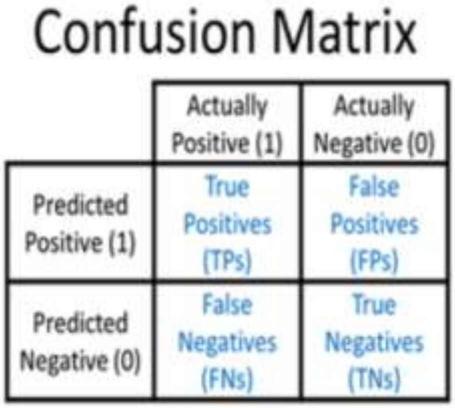
- There should be some actual values in the feature variable of the dataset so that the classifier can predict accurate results rather than a guessed result.
- The predictions from each tree must have very low correlations.

When we have to predict the input values are passed on to each tree in the forest and the results from each tree are clubbed since it is classification through majority voting technique the final result would be predicted for the inputs.

5. MODEL EVALUATION METRICS

Using the data from Kaggle, we have cleaned the data, pre-processed and then we have implemented it to a random forest machine learning algorithm. Now we can measure the effectiveness of our model. Better the effectiveness, better the performance. This is when Confusion matrix comes into the limelight. Confusion Matrix is a performance measurement for machine learning classification.

Confusion matrix, is a performance measurement for machine learning classification problem where output can be two or more classes. It is a table with 4 different combinations of predicted and actual values. It is extremely useful for measuring Recall, Precision, Specificity, Accuracy.



Confusion Matrix

	Actually Positive (1)	Actually Negative (0)
Predicted Positive (1)	True Positives (TPs)	False Positives (FPs)
Predicted Negative (0)	False Negatives (FNs)	True Negatives (TNs)

Figure 3: Confusion Matrix Table

- True Positive: Interpretation: Predicted positive and it’s true. In this case the patient actually suffers from coronary artery disease and model also predicted that patient is suffering from the disease.
- True Negative: Interpretation: predicted negative and it’s true. In this case the patient does not suffers from coronary artery disease and model also predicted that patient is not suffering from the disease.
- False Positive: Interpretation: predicted positive and it’s false. In this case the patient does not suffers from coronary artery disease but model predicted that patient is suffering from the disease. This sort of error is also called as Type 1 error.
- False Negative: Interpretation: predicted negative and it’s false. In this case the patient does suffers from coronary artery disease but model predicted that patient is not suffering from the disease . This sort of error is also called as Type 2 error. Note: predicted values as Positive and Negative and actual values as True and False.

Recall: Out of all the positive classes, how much we predicted correctly. It should be high as possible.

Precision: Out of all the positive classes we have predicted correctly, how many are actually positive.

Accuracy: Out of all the classes, how much we predicted correctly. It should be high as possible.

F1-measure: It is difficult to compare two models with low precision and high recall or vice versa. So to make them comparable, we use F-Score. F-score helps to measure Recall and Precision at the same time. It uses Harmonic Mean in place of Arithmetic Mean by punishing the extreme values more.

$$Acc = \frac{TP + TN}{TP + FP + TN + FN}$$

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

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

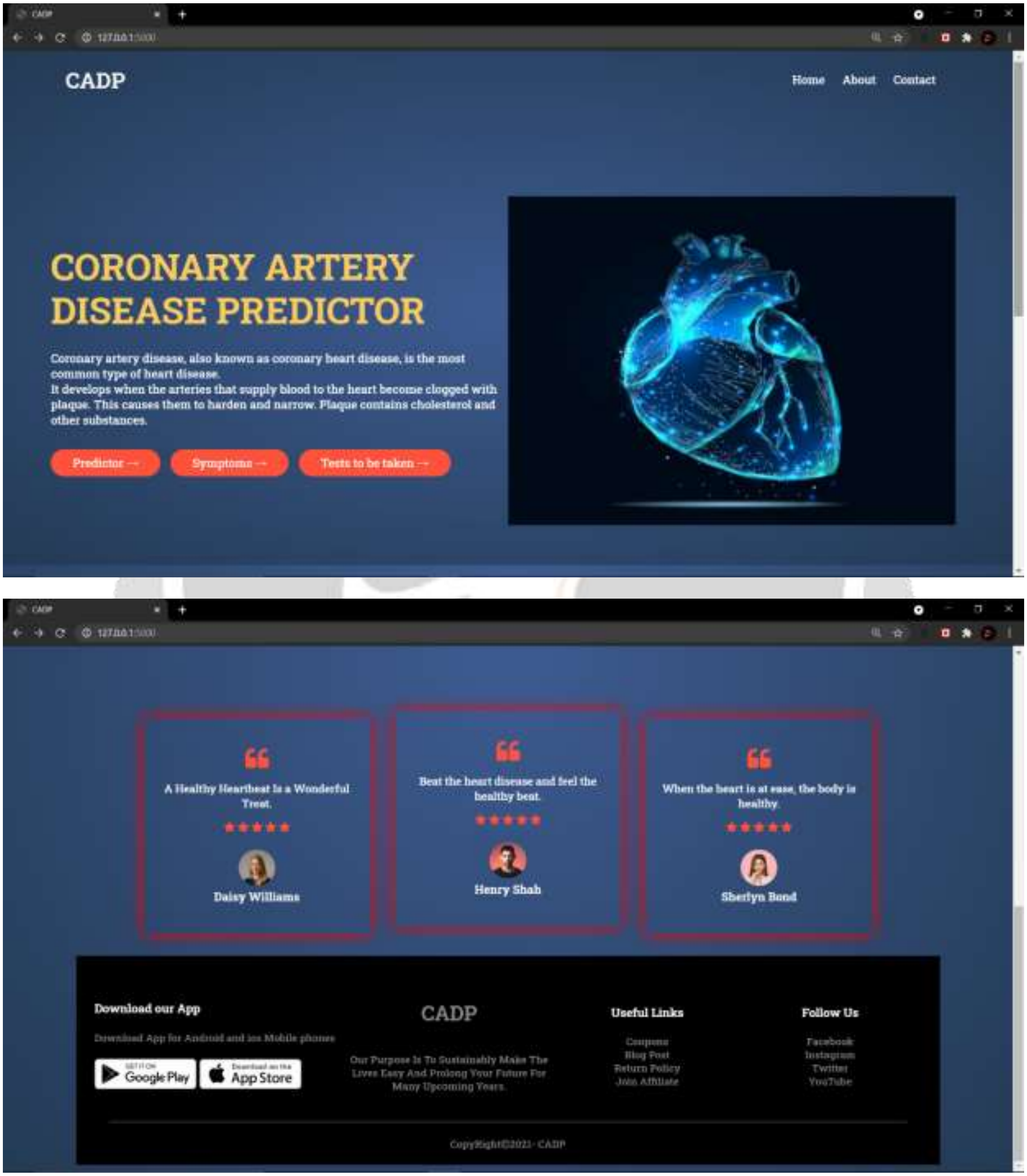
$$F1 = \frac{2 * (precision * recall)}{precision + recall}$$

Figure 4: Formulas

6. Results

We could predict the results with an accuracy of 93% using Random Forest algorithm.

This is our home page



CORONARY ARTERY DISEASE PREDICTOR

Coronary artery disease, also known as coronary heart disease, is the most common type of heart disease.
It develops when the arteries that supply blood to the heart become clogged with plaque. This causes them to harden and narrow. Plaque contains cholesterol and other substances.

A Healthy Heartbeat Is a Wonderful Treat.
★★★★★
Daisy Williams

Beat the heart disease and feel the healthy beat.
★★★★★
Henry Shah

When the heart is at ease, the body is healthy.
★★★★★
Sherlyn Bond

Download our App

Download App for Android and ios Mobile phones

CADP

Our Purpose Is To Sustainably Make The Lives Easy And Prolong Your Future For Many Upcoming Years.

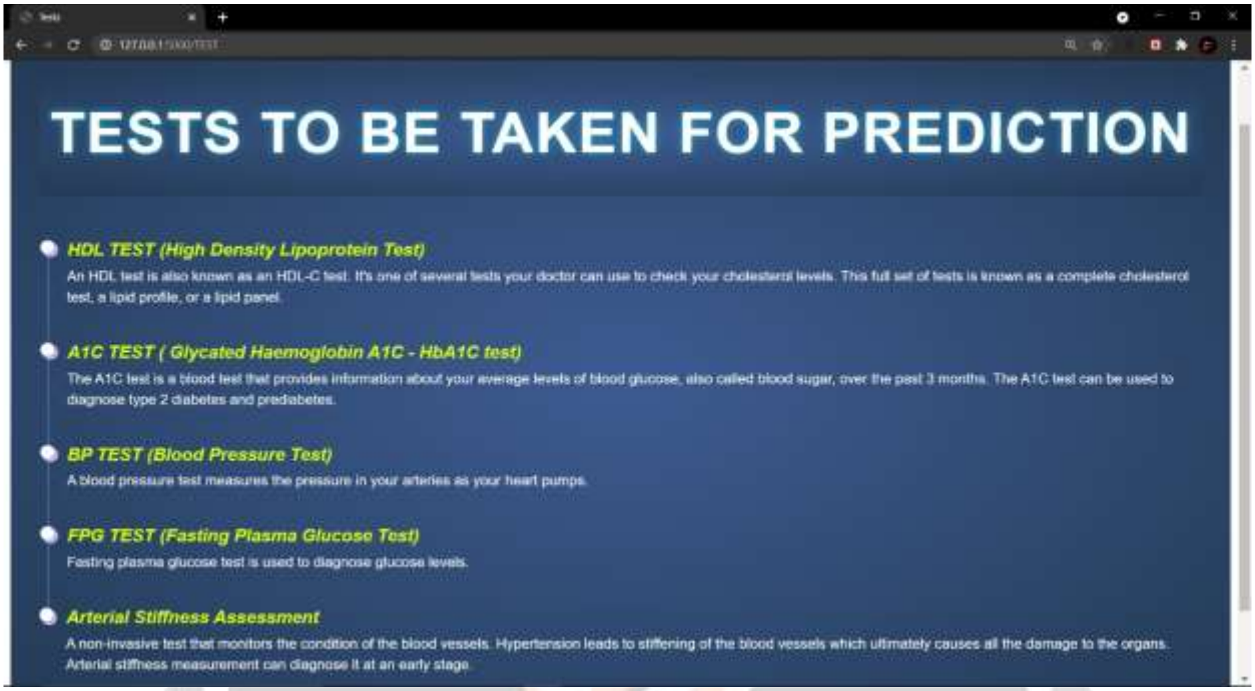
Useful Links

Follow Us

CopyRight©2021- CADP

Figure 5: Home page

These are the tests to be taken for getting the values which the user oughts to give in the form for prediction.

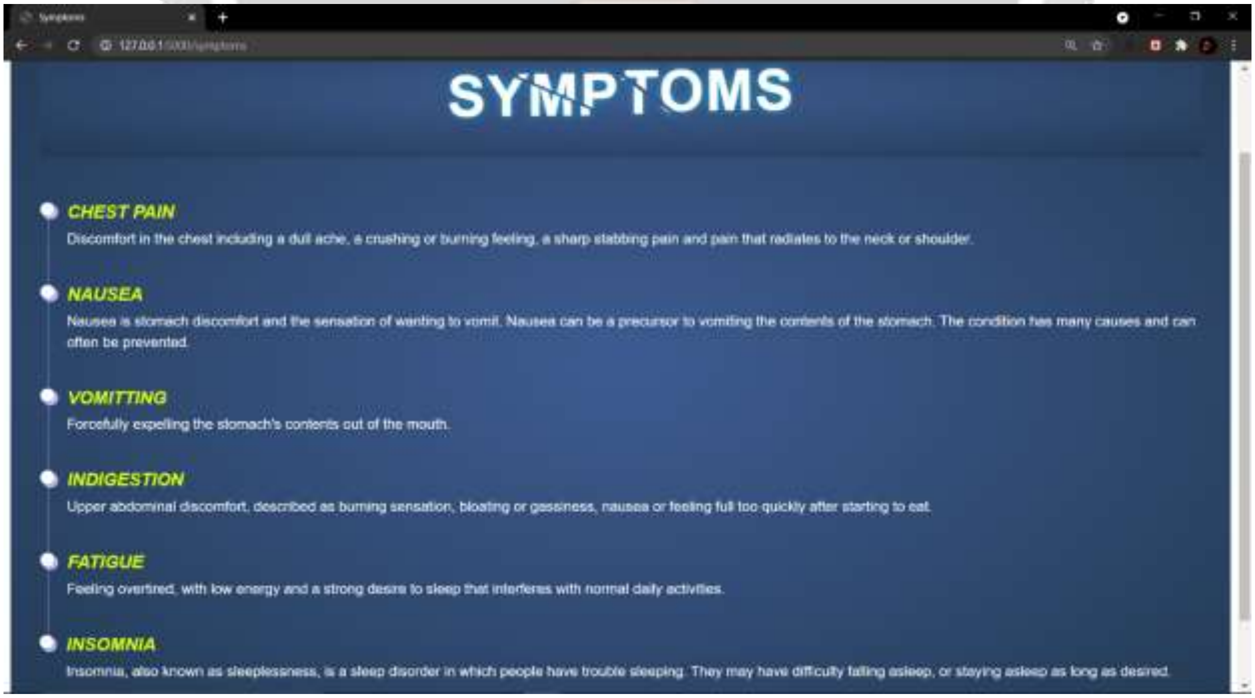


TESTS TO BE TAKEN FOR PREDICTION

- *HDL TEST (High Density Lipoprotein Test)*
  An HDL test is also known as an HDL-C test. It's one of several tests your doctor can use to check your cholesterol levels. This full set of tests is known as a complete cholesterol test, a lipid profile, or a lipid panel.
- *A1C TEST (Glycated Haemoglobin A1C - HbA1C test)*
  The A1C test is a blood test that provides information about your average levels of blood glucose, also called blood sugar, over the past 3 months. The A1C test can be used to diagnose type 2 diabetes and prediabetes.
- *BP TEST (Blood Pressure Test)*
  A blood pressure test measures the pressure in your arteries as your heart pumps.
- *FPG TEST (Fasting Plasma Glucose Test)*
  Fasting plasma glucose test is used to diagnose glucose levels.
- *Arterial Stiffness Assessment*
  A non-invasive test that monitors the condition of the blood vessels. Hypertension leads to stiffening of the blood vessels which ultimately causes all the damage to the organs. Arterial stiffness measurement can diagnose it at an early stage.

Figure 6: Test Page

These are the symptoms a likely CAD suffering patient may have, if the user has the following symptoms, we suggest to take the required tests and visit nearby hospital.

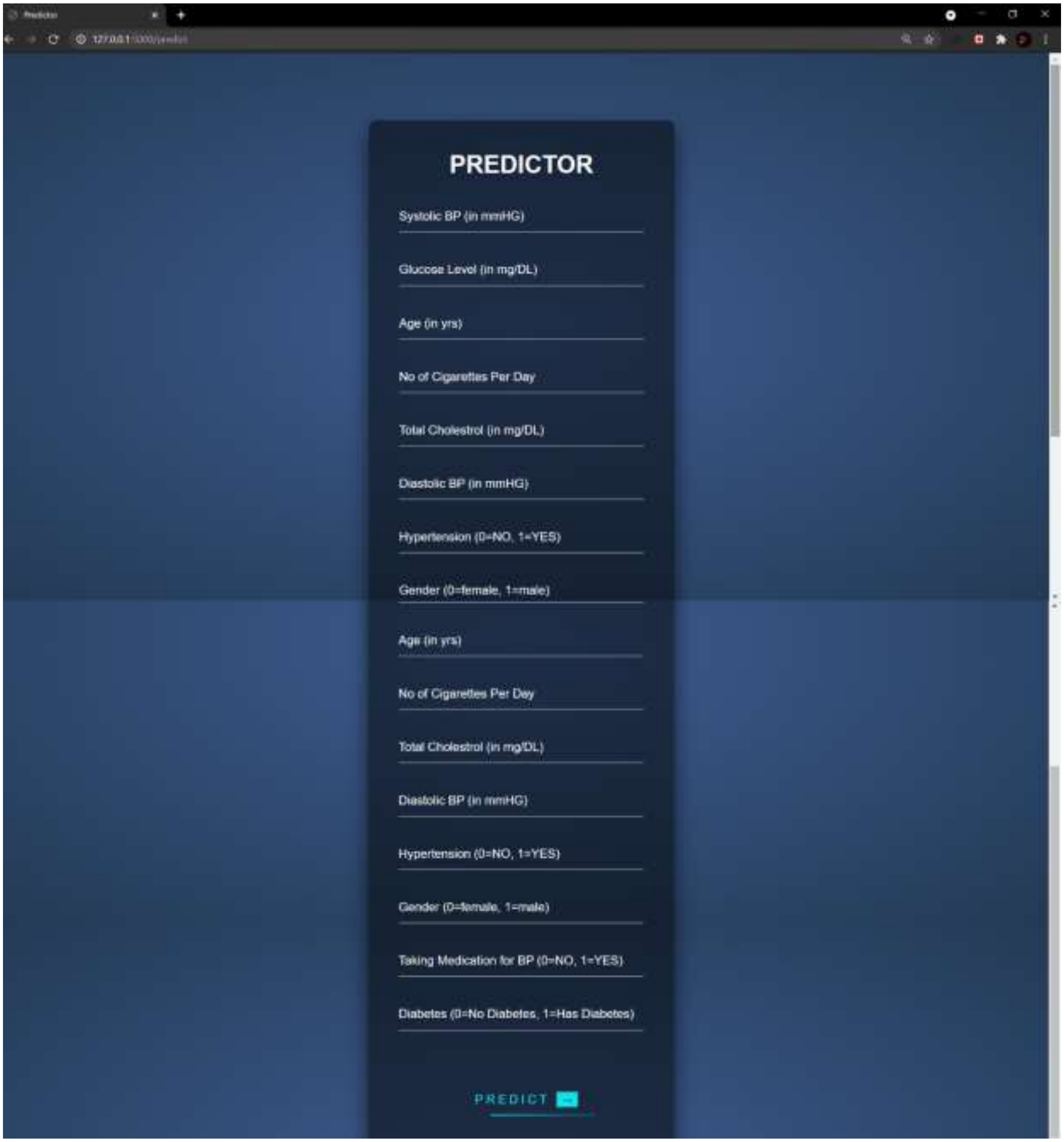


SYMPTOMS

- *CHEST PAIN*
  Discomfort in the chest including a dull ache, a crushing or burning feeling, a sharp stabbing pain and pain that radiates to the neck or shoulder.
- *NAUSEA*
  Nausea is stomach discomfort and the sensation of wanting to vomit. Nausea can be a precursor to vomiting the contents of the stomach. The condition has many causes and can often be prevented.
- *VOMITTING*
  Forcefully expelling the stomach's contents out of the mouth.
- *INDIGESTION*
  Upper abdominal discomfort, described as burning sensation, bloating or gassiness, nausea or feeling full too quickly after starting to eat.
- *FATIGUE*
  Feeling overtired, with low energy and a strong desire to sleep that interferes with normal daily activities.
- *INSOMNIA*
  Insomnia, also known as sleeplessness, is a sleep disorder in which people have trouble sleeping. They may have difficulty falling asleep, or staying asleep as long as desired.

Figure 7: Symptoms page

This is the prediction part of our web page where user has to enter the values according to the specified units for each feature and can proceed with the Prediction.



PREDICTOR

Systolic BP (in mmHG)

Glucose Level (in mg/DL)

Age (in yrs)

No of Cigarettes Per Day

Total Cholestrol (in mg/DL)

Diastolic BP (in mmHG)

Hypertension (0=NO, 1=YES)

Gender (0=female, 1=male)

Age (in yrs)

No of Cigarettes Per Day

Total Cholestrol (in mg/DL)

Diastolic BP (in mmHG)

Hypertension (0=NO, 1=YES)

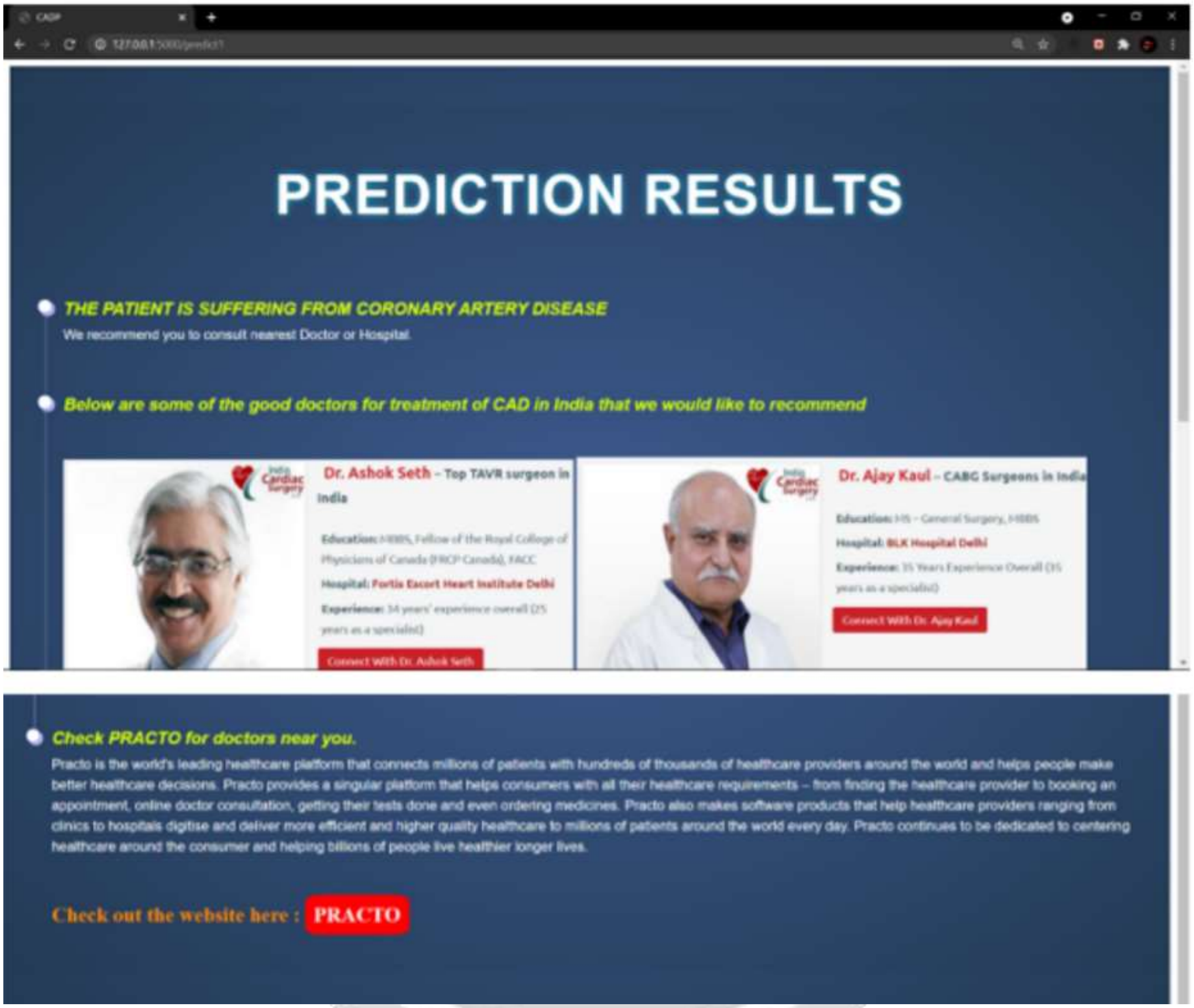
Gender (0=female, 1=male)

Taking Medication for BP (0=NO, 1=YES)

Diabetes (0=No Diabetes, 1=Has Diabetes)

Figure 8: Prediction page

This webpage shows that the user is suffering from Coronary Artery Disease and we have also recommended some of the best doctors in India who can cure CAD followed with a link to Practo which helps to find nearby good doctors.



PREDICTION RESULTS

THE PATIENT IS SUFFERING FROM CORONARY ARTERY DISEASE

We recommend you to consult nearest Doctor or Hospital.

Below are some of the good doctors for treatment of CAD in India that we would like to recommend

Dr. Ashok Seth – Top TAVR surgeon in India

Education: MBBS, Fellow of the Royal College of Physicians of Canada (FRCP Canada), FACC

Hospital: Fortis Escort Heart Institute Delhi

Experience: 34 years' experience overall (25 years as a specialist)

Dr. Ajay Kaul – CABG Surgeons in India

Education: MS – General Surgery, MBBS

Hospital: BLK Hospital Delhi

Experience: 35 Years Experience Overall (35 years as a specialist)

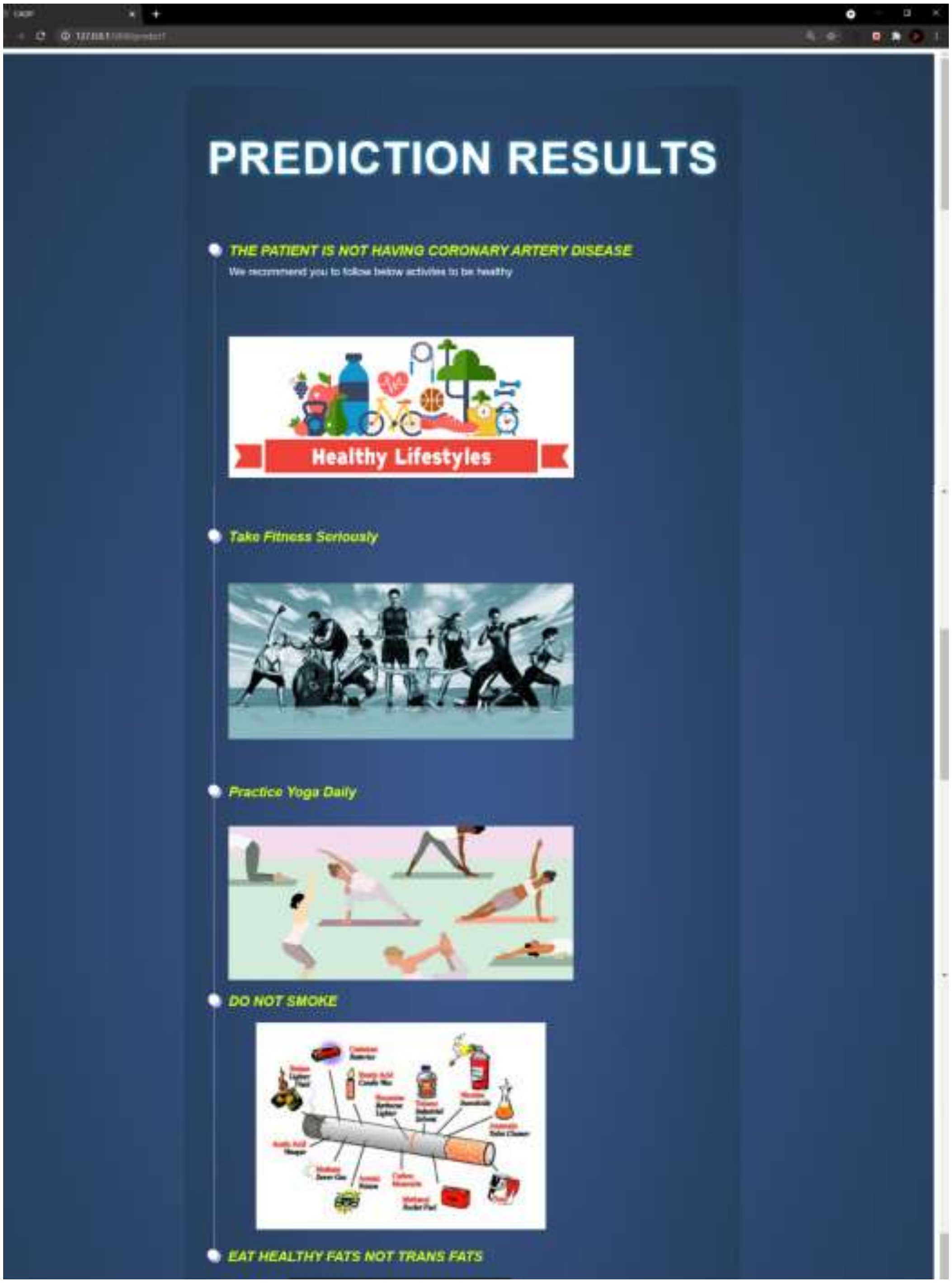
Check PRACTO for doctors near you.

Practo is the world's leading healthcare platform that connects millions of patients with hundreds of thousands of healthcare providers around the world and helps people make better healthcare decisions. Practo provides a singular platform that helps consumers with all their healthcare requirements – from finding the healthcare provider to booking an appointment, online doctor consultation, getting their tests done and even ordering medicines. Practo also makes software products that help healthcare providers ranging from clinics to hospitals digitise and deliver more efficient and higher quality healthcare to millions of patients around the world every day. Practo continues to be dedicated to centering healthcare around the consumer and helping billions of people live healthier longer lives.

Check out the website here :

Figure 9: Prediction result is Yes

This web page shows that the user is not suffering from Coronary Artery Disease and we have recommended some tips to keep the body healthy such as to practice Yoga, not to smoke, to eat healthy fats not trans fats and to take fitness seriously.



PREDICTION RESULTS

THE PATIENT IS NOT HAVING CORONARY ARTERY DISEASE

We recommend you to follow below activities to be healthy

Take Fitness Seriously

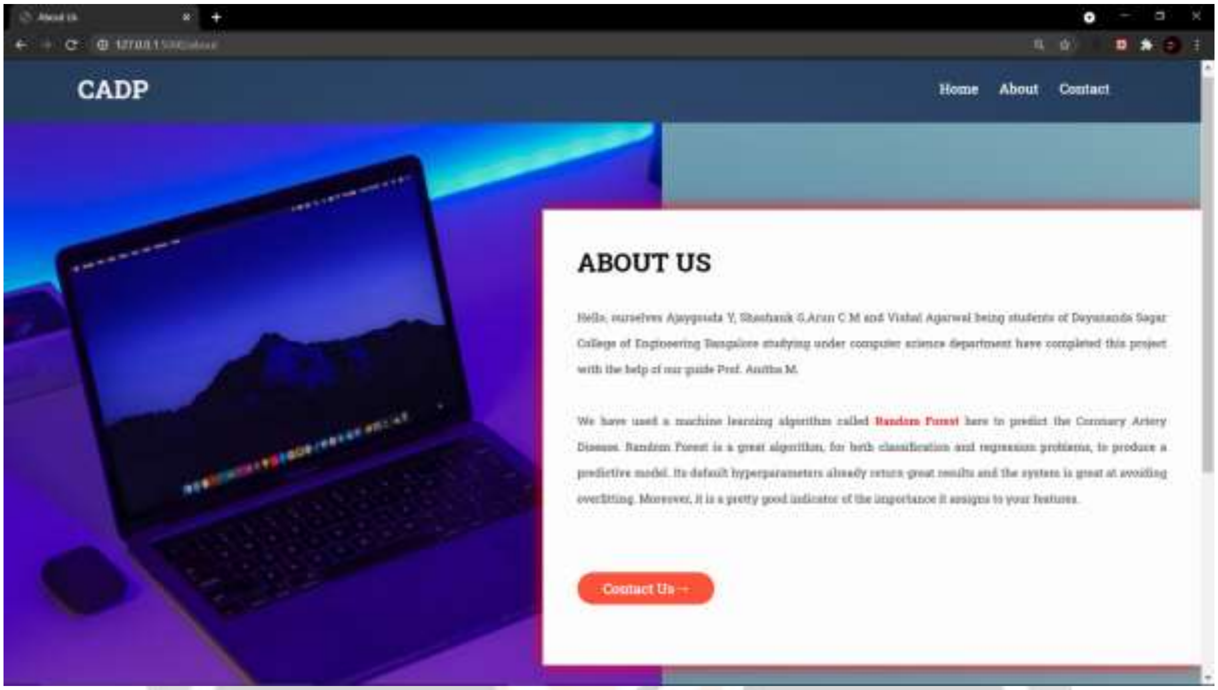
Practice Yoga Daily

DO NOT SMOKE

EAT HEALTHY FATS NOT TRANS FATS

Figure 10: Prediction result is No

This webpage shows information about us and the machine learning algorithm(Random Forest) which we used in this project and why that algorithm was best suited for this particular project.



ABOUT US

Hello, ourselves Ajaygouda Y, Shashank G,Arun C M and Vishal Agarwal being students of Dayananda Sagar College of Engineering Bangalore studying under computer science department have completed this project with the help of our guide Prof. Anitha M.

We have used a machine learning algorithm called Random Forest here to predict the Coronary Artery Disease. Random Forest is a great algorithm, for both classification and regression problems, to produce a predictive model. Its default hyperparameters already return great results and the system is great at avoiding overfitting. Moreover, it is a pretty good indicator of the importance it assigns to your features.

Figure 11: About Us page

This webpage shows various methods a user can contact us that is through mail id, Mobile number or directly coming to the location specified.

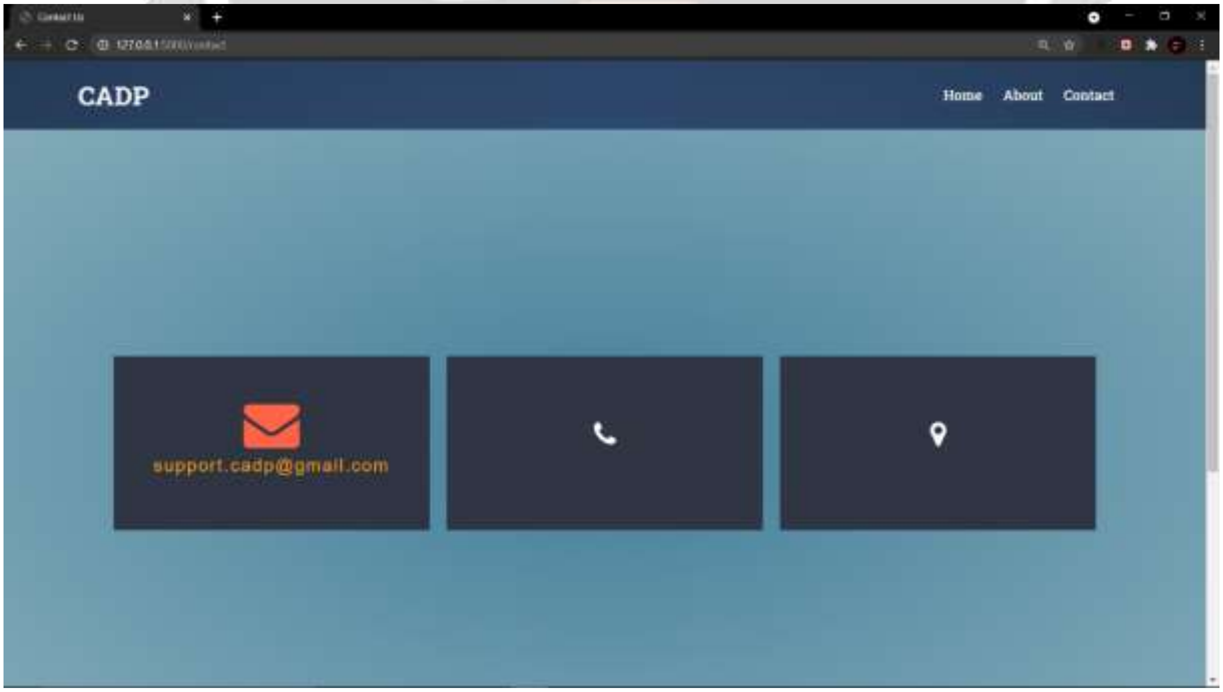


Figure 12: About Us page

7. CONCLUSION

In this project, we had implemented random forest machine learning algorithm for high accuracy and to detect the presence of coronary artery disease using a publicly available dataset in Kaggle which will help people to identify coronary artery disease and take prior precautions to prevent it and live a smooth and healthy life. We have also hosted this prediction of coronary artery disease using machine learning system in the public domain with hopes of contributing to an open-source community. Where people can get results faster by giving values for features that are necessary for prediction of coronary artery disease through web interface and result can also be observed through the web interface. We believe that with our current effort, we have achieved the best and we are hopeful that this algorithm can help for predicting the coronary artery disease. As we have gained 93 percent accuracy for the dataset.

Future Work

Implementation of even larger dataset in Machine Learning algorithm can be done to get higher accuracy.

8. REFERENCES

[1] Prediction of Cardiovascular Disease using Machine Learning Algorithms International Journal of Engineering and Advanced Technology (IJEAT) ISSN: 2249 – 8958, Volume-9 Issue-3, February, 2020 2404 Published By: Blue Eyes Intelligence Engineering Sciences Publication Retrieval Number: B3986129219/2020©BEIESP DOI: 10.35940/ijeat.B3986.029320

[2] An Efficient System for the Prediction of Coronary Artery Disease using Dense Neural Network with Hyper Parameter Tuning International Journal of Innovative Technology and Exploring Engineering (IJITEE) ISSN: 2278-3075, Volume-8 Issue6S, April 2019

[3] Artificial intelligence in cardiovascular imaging: state of the art and implications for the imaging cardiologist - K. R. Siegersma · T. Leiner · D. P. Chew · Y. Appelman · L. Hofstra · J. W. Verjans

[4] A Prototype for Second Generation Chronic Heart Disease Detection and Classification - DOI: 10.1007/978-981-15-5148-2_29 In book: International Conference on Innovative Computing and Communications (pp.321-329)

[5] Sajda P. Machine learning for detection and diagnosis of disease. Annu. Rev. Biomed. Eng. 2006; 8:537–65.pmid: 16834566 doi:10.1146/annurev.bioeng.8.061505.095802 3. Foster KR, Koprowski R, Skufca J D.

[6] Machine learning, medical diagnosis, and biomedical engineering research-commentary. Biomed Eng Online. 2014; 13(1): 94. pmid: 24998888 doi: 10.1186/1475-925X-13-94 4. Deo RC.

[7] Machine learning in medicine. Circulation. 2015; 132(20):1920–30. pmid: 26572668 doi: 10.1161/CIRCULATIONAHA.115.001593

[8] Alberto Palacios Pawlovsky 'An Ensemble Based on Distances for a kNN Method for Heart Disease Diagnosis' 2018 International Conference on Electronics, Information, and Communication (ICEIC).

[9] Yeshvendra K. Singh, Nikhil Sinha., and Sanjay K. Singh., 'Heart Disease Prediction System Using Random Forest'; First International Conference, ICACDS 2016 , November 11–12, 2016, pp 613-623.

[10] TanmayKasbe , Ravi Singh Pippal 'Design of Heart Disease Diagnosis System using Fuzzy Logic' 2017 International Conference on Energy, Communication, Data Analytics and Soft Computing (ICECDS-2017)

[11] Purushottam,Prof. (Dr.) Kanak Saxena,Richa Sharma, 'Efficient Heart Disease Prediction System'., Procedia, Computer Science 85 (2016) 962 – 969.

[12] Hongmei Yan and Jun Zheng, et al, https://www.researchgate.net/publication/220199641_Selecting_critical_clinical_features_forheart_diseases_diagnosis_with__real-coded_genetic_algorithm

[13] Resul Das and Ibrahim Turkoglu, et al, https://www.researchgate.net/publication/23414922_Diagnosis_of_valvular_heart_disease_through_neural_networks_ensembles

[14] Yoon-Joo Park and Se-Hak Chun, et al, https://www.researchgate.net/publication/49738633_Costsensitive_casebased_reasoning_using_a_genetic_algorithm_Application_to_medical_diagnosis

[15]K.Rajeswari and V.Vaithiyanathan, et al, https://www.researchgate.net/publication/271618752_Feature_Selection_in_Ischemic_Heart_Disease_Identification_using_Feed_Forward_Neural_Networks

[16]Deo RC. Machine learning in medicine. Circulation.2015; 132(20):1920–30.pmid: 26572668 doi: 10.1161/CIRCULATIONAHA.115.001593