ANALYSIS AND PREDICTION OF CARDIO VASCULAR DISEASE

MUNTHA SUPRIYA VANI[1], KAMBHAM SALIVAHANA REDDY[2]

[1]MTech Student, Department of **CSE**, **Global College of Engineering and Technology**, kadapa, AP.

[2]Assistant professor, Department of **CSE**, **Global College of Engineering and Technology**, kadapa, AP.

ABSTRACT:

Cardio Vascular Disease (CVD) is, for the most part, alluding to conditions that include limited or blocked veins that can prompt a heart attack, chest torment (angina) or stroke. The machine learning classifier predicts the ailment dependent on the state of the side effect endured by the patient. This paper intends to look at the presentation of the Machine learning tree classifiers in anticipating Cardio Vascular Disease (CVD). Machine learning tree classifiers, for example, Random Forest, Decision Tree, Logistic Regression, Support vector machine (SVM), K-nearest neighbors (KNN) were broke down dependent on their precision and AUC ROC scores. In this investigation of foreseeing Cardiovascular Disease, the Random woodland Machine learning classifier accomplished a higher precision of 85%, ROC AUC score of 0.8675 and execution time of 1.09 sec.

Keywords: SVM, KNN, CVD, ROC, AUC.

DOI Number: 10.14704/nq.2022.20.12.NQ77050 **NeuroQuantology 2022; 20(12): 668-677**

1. INTRODUCTION:

Cardio Vascular Disease (CVD) is the most well-known perilous infection around the world: the greater part of the populaces bites the dust every year from Cardio Vascular Disease (CVD) than from some other ailment. A degree of 17.9 million individuals passed on from Cardio Vascular Disease (CVD) in , thinking about 31% of every single worldwide demise. Of these deaths, 85% are because of heart stroke and heart failure. More than three-fourths of CVD deaths occur in dejected yield nations. Out of the 17 million less than ideal closures (younger than 70) due to noninfectious maladies in 2015, 82% are in discouraging yield nations and 37% are brought about via Cardio Vascular Disease(CVD). All most Cardio Vascular Disease(CVD) can be killed by tending to discernible hazard factors, for example, tobacco use, undesirable eating routine and heftiness, physical dormancy and destructive utilization of liquor utilizing populace wide situations. Individuals with Cardio Vascular Disease(CVD) or who are at high cardiovascular hazards (because of the nearness of at least one hazard factor, for example, hypertension, diabetes, hyperlipidemia or effectively settled sickness) need an early introduction and directorate utilizing brief prescriptions, as set apart. All in all, Cardio Vascular Disease (CVD) is winded up with a development of greasy stores inside the conduits (atherosclories) and development of blood clusters. It can likewise be related to harm to courses in organs, for example, the mind, heart, kidneys, and eyes. CVD is one of the fundamental drivers of death and incapacity in the UK, however, it can regularly to a great extent be avoided by driving a solid way of life. Coronary episodes and strokes are typically brought about by intense occasions and are for the most part brought about by a blockage that averts bloodstream to the heart or mind. The most widely recognized purpose behind this is the development of greasy stores most inward dividers of veins. The reason for cardiovascular failures and strokes is generally the nearness of a blend of hazard factors, for example, tobacco use, unfortunate eating regimen, and heftiness.

DATASET DESCRIPTIONS:

Heart disease describes a range of conditions that affect your heart. Diseases under the heart disease umbrella include blood vessel diseases, such as coronary artery disease, heart rhythm problems (arrhythmias) and heart defects you're born with (congenital heart defects), among others.

The term "heart disease" is often used interchangeably with the term "cardiovascular disease". Cardiovascular disease generally refers to conditions that involve narrowed or blocked blood vessels that can lead to a heart attack, chest pain (angina) or stroke. Other heart conditions, such as those that affect your heart's muscle, valves or rhythm, also are considered forms of heart disease. Heart disease is one of the biggest causes of morbidity and mortality among the population of the world. Prediction of cardiovascular disease is regarded as one of the most important subjects in the section of clinical data analysis. The amount of data in the healthcare industry is huge. Data mining turns the large collection of raw healthcare data into information that can help to make informed decisions and predictions.

According to a news article, heart disease proves to be the leading cause of death for both women and men. The article states the following :

About 610,000 people die of heart disease in the United States every year–that's 1 in every 4 deaths.1

Heart disease is the leading cause of death for both men and women. More than half of the deaths due to heart disease in 2009 were in men.1

Coronary Heart Disease(CHD) is the most common type of heart disease, killing over 370,000 people annually.

Every year about 735,000 Americans have a heart attack. Of these, 525,000 are a first heart attack and 210,000 happen in people who have already had a heart attack.

This makes heart disease a major concern to be dealt with. But it is difficult to identify heart disease because of several contributory risk factors such as diabetes, high blood pressure, high cholesterol, abnormal pulse rate, and many other factors. Due to such constraints, scientists have turned towards modern approaches like Data Mining and Machine Learning for predicting the disease.

Machine learning (ML) proves to be effective in assisting in making decisions and predictions from the large quantity of data produced by the healthcare industry.

2. LITERATURE SURVEY

1) Comparison Of Machine Learning Algorithms For Clinical Event Prediction (Risk Of Coronary Heart Disease)

AUTHORS: Beunza, Juan-Jose

The aim of this study is to compare the utility of several supervised machine learning (ML) algorithms for predicting clinical events in terms of their internal validity and accuracy. The results, which were obtained using two statistical software platforms, were also compared.The data used in this research come from the open database of the Framingham Heart Study, which originated in 1948 in Framingham, Massachusetts as a prospective study of risk factors for cardiovascular disease. Through data mining processes, three data models were elaborated and a comparative methodological study between the different ML algorithms – decision tree, random forest, support vector machines, neural networks, and logistic regression – was carried out. The global selection criterium for choosing the right set of hyperparameters and the type of data manipulation was the area under a curve (AUC).

2) Enhancing Detection Accuracy for Clinical Heart Failure Utilizing Pulse Transit Time Variability and Machine Learning

AUTHORS: Zhao, Lina

Physiological signal variability can offer important insight into cardiovascular activity and clinical cardiovascular diseases. Heart rate variability (HRV) and pulse transit time variability (PTTV) are two important time series variabilities. However, combining HRV and PTTV can enhance the classification accuracy for heart failure which is unknown. In this paper, a simultaneous analysis of HRV and PTTV performed on both normal subjects and heart failure patients, was carried out, aiming to investigate the improvement of HRV-based heart failure detection with the assistant of PTTV analysis. Forty normal subjects and forty heart failure patients were enrolled. Standard limb lead-II electrocardiogram and radial artery pressure waveforms were synchronously recorded. HRV and PTTV analysis were performed on the acquired RR and PTT time series using the standard time- (MEAN, SDNN, and RMSSD), frequency- (LF, HF, and LF/HF), and non-linear (SD1, SD2, sample entropy, and fuzzy measure entropy) domain indices. The results showed that all HRV indices except MEAN (P = 0.1) and LF/HF (P = 0.9) showed significant differences (all P <;0.01) between the two group, while only MEAN in PTTV significantly decreases in heart failure patients (P <; 0.01). Moreover, when combined the HRV, PTTV indices, and the predicted probabilities generated from the distance distribution matrix-based convolutional neural network models, the highest classification performances were achieved by a support vector machine classifier, outputting a sensitivity of 0.93, a specificity of 0.88, and an accuracy of 0.90. This paper demonstrated the potential of PTTV analysis for the detection of clinical heart failure.

3) Using Machine Learning to Predict One-year Cardiovascular Events in Patients with Severe Dilated Cardiomyopathy

AUTHORS : Chen, Rui

Dilated cardiomyopathy (DCM) is a common form of cardiomyopathy and it is associated with poor outcomes. A poor prognosis of DCM patients with low ejection fraction has been noted in the short-term follow-up. Machine learning (ML) could aid clinicians in risk stratification and patient management after considering the correlation between numerous features and the outcomes. The present study aimed to predict the 1-year cardiovascular events in patients with severe DCM using ML, and aid clinicians in risk stratification and patient management. The dataset used to establish the ML model was obtained from 98 patients with severe DCM (LVEF < 35%) from two centres. Totally 32 features from clinical data were input to the ML algorithm, and the significant features highly relevant to the cardiovascular events were selected by Information gain (IG). A naive Bayes classifier was built, and its predictive performance was evaluated using the area under the curve (AUC) of the receiver operating characteristics by 10-fold cross-validation.

4) A Novel Ontology And Machine Learning Driven Hybrid Cardiovascular Clinical Prognosis As A Complex Adaptive Clinical System.

AUTHORS : **Farooq, Kamran, and Amir Hussain**

This multidisciplinary industrial research project sets out to develop a hybrid clinical decision support mechanism (inspired by ontology and machine learning driven techniques) by combining evidence, extrapolated through legacy patient data to facilitate cardiovascular preventative care. The proposed cardiovascular clinical decision support framework comprises of two novel key components: (1) Ontology driven clinical risk assessment and recommendation system (ODCRARS) (2) Machine learning driven prognostic system (MLDPS). State of the art machine learning and feature selection methods are utilised for the prognostic modelling purposes. The ODCRARS is a knowledge-based system which is based on clinical expert’s knowledge, encoded in the form of clinical rules engine to carry out cardiac risk assessment for various cardiovascular diseases. The MLDPS is a non knowledge-based/data driven system which is developed using state of the art machine learning and feature selection techniques applied on real patient datasets. Clinical case studies in the RACPC, heart disease and breast cancer domains are considered for the development and clinical validation purposes. For the purpose of this paper, clinical case study in the RACPC/chest pain domain will be discussed in detail from the development and vaidation perspective.The proposed clinical decision support framework is validated through clinical case studies in the cardiovascular domain. This paper demonstrates an effective cardiovascular decision support mechanism for handling inaccuracies in the clinical risk assessment of chest pain patients and help clinicians effectively distinguish acute angina/cardiac chest pain patients from those with other causes of chest pain. The new clinical models, having been evaluated in clinical practice, resulted in very good predictive power, demonstrating general performance improvement over benchmark multivariate statistical classifiers. Various chest pain risk assessment prototypes have been developed and deployed online for further clinical trials.

5) A Real-time Arrhythmia Heartbeats Classification Algorithm using Parallel Delta Modulations and Rotated Linear-Kernel Support Vector Machines

AUTHORS: Tang, Xiaochen

Real-time wearable electrocardiogram monitoring sensor is one of the best candidates in assisting cardiovascular disease diagnosis. In this paper, we present a novel real-time machine learning system for Arrhythmia classification. The system is based on the parallel Delta modulation and QRS/PT wave detection algorithms. We propose a patient dependent rotated linear-kernel support vector machine classifier that combines the global and local classifiers, with three types of feature vectors extracted directly from the Delta modulated bit-streams. The performance of the proposed system is evaluated using the MIT-BIH Arrhythmia database. According to the AAMI standard, two binary classifications are performed and evaluated, which are supraventricular ectopic beat (SVEB) versus the rest four classes, and ventricular ectopic beat (VEB) versus the rest. For SVEB classification, the preferred SkP-32 method's F1 score, sensitivity, specificity, and positive predictivity value are 0.83, 79.3%, 99.6%, and 88.2%, respectively, and for VEB classification, the numbers are 0.92%, 92.8%, 99.4%, and 91.6%, respectively. The results show that the performance of our proposed approach is comparable to that of published research. The proposed low-complexity algorithm has the potential to be implemented as an on-sensor machine learning solution.

3. METHODOLOGY

User:

The User can register the first. While registering he required a valid user email and mobile for further communications. Once the user register then admin can activate the user. Once admin activated the user then user can login into our system. User can upload the dataset based on our dataset column matched. For algorithm execution data must be in int of float format. Here we took UCI repository dataset for testing purpose. User can also add the new data for existing dataset based on our Django application. User can click the Data Preparations in the web page so that the data cleaning process will be starts. The cleaned data and its required graph will be displayed.

Admin:

Admin can login with his login details. Admin can activate the registered users. Once he activate then only the user can login into our system. Admin can view the overall data in the browser. He can also check the algorithms ROC Curve, confusion matrix and accuracy. The comparison accuracy bar graph also displayed here. All algorithm execution complete then admin can see the overall accuracy in web page.

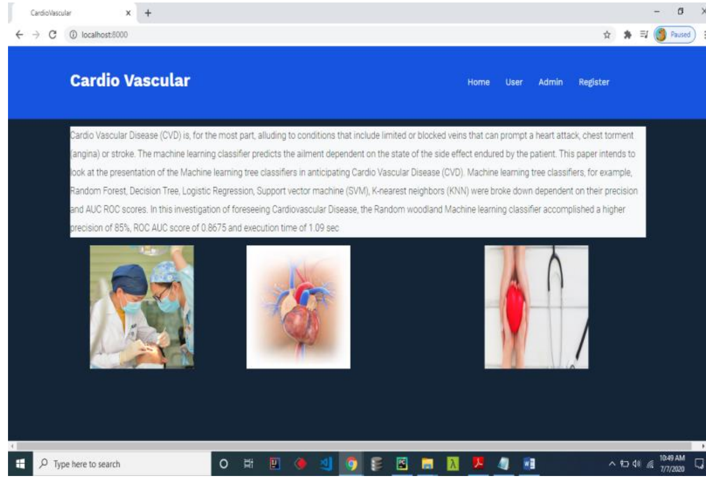
Data Preprocessing:

A dataset can be viewed as a collection of data objects, which are often also called as a records, points, vectors, patterns, events, cases, samples, observations, or entities. Data objects are described by a number of features that capture the basic characteristics of an object, such as the mass of a physical object or the time at which an event occurred, etc. Features are often called as variables, characteristics, fields, attributes, or dimensions. The data preprocessing in this forecast uses techniques like removal of noise in the data, the expulsion of missing information, modifying default values if relevant and grouping of attributes for prediction at various levels.

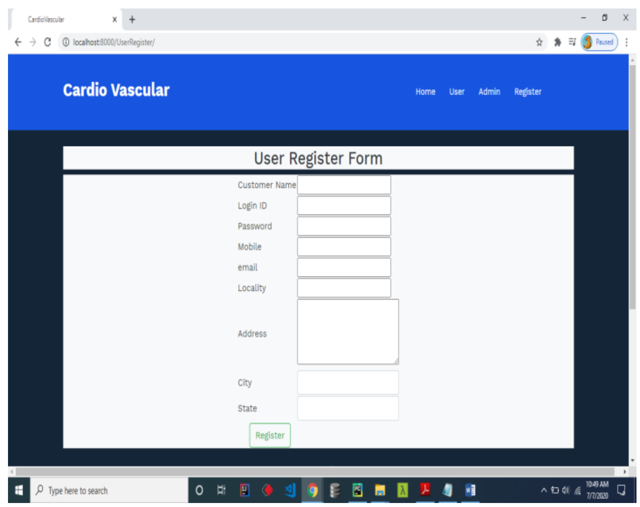
Machine learning:

Based on the split criterion, the cleansed data is split into 60% training and 40% test, then the dataset is subjected to five machine learning classifiers such as Logistic Regression (LR) with pipeline, Support Vector Machine (SVM), Decision Tree (DT), Random Forest (RF), K-Nearest Neighbors (KNN). The accuracy of the classifiers was calculated using the confusion matrix. The classifier which bags up the highest accuracy could be determined as the best classifier. For arch algorithm confusion matrix roc curve and accuracy has been calculated and displayed in my results.

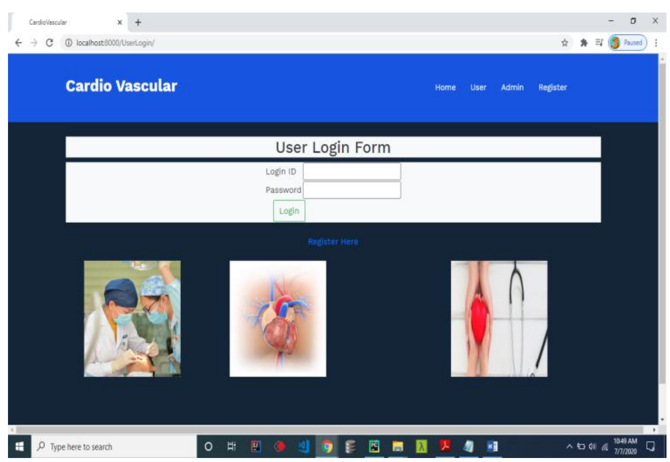
Home page



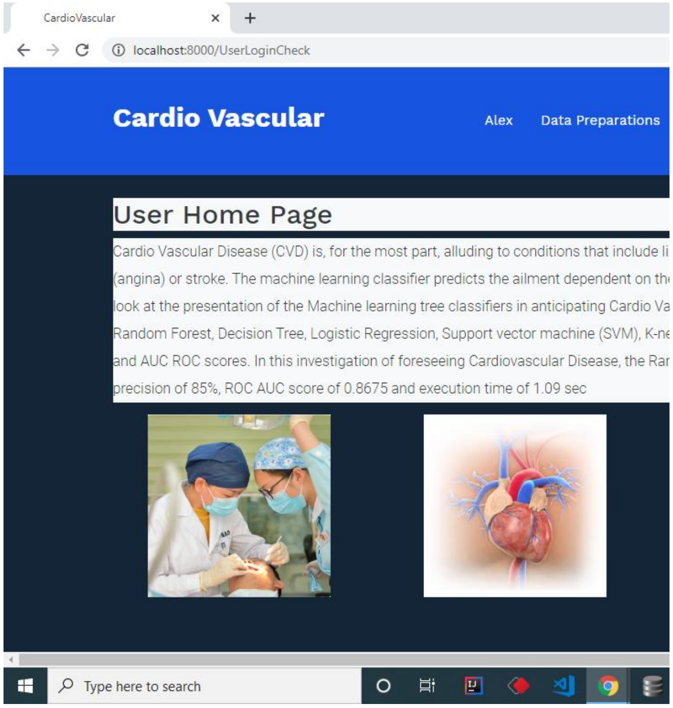
User Register Form



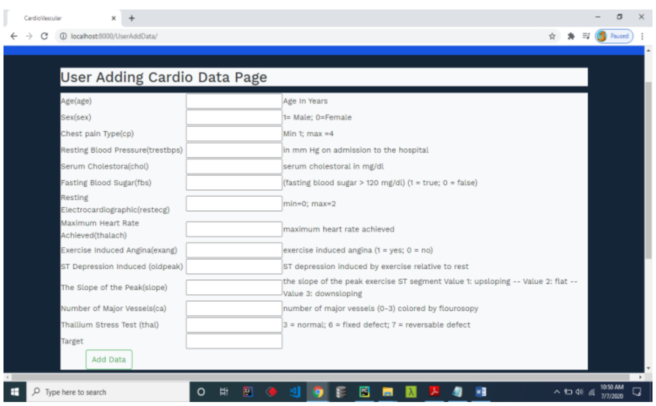
User Login Form



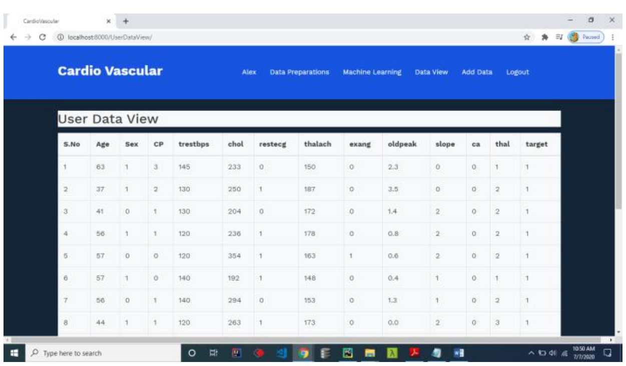
User Home Page:



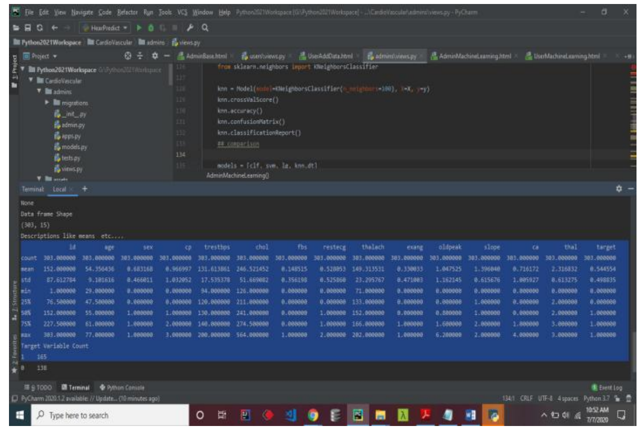
User Can Add the Data:



Data View :

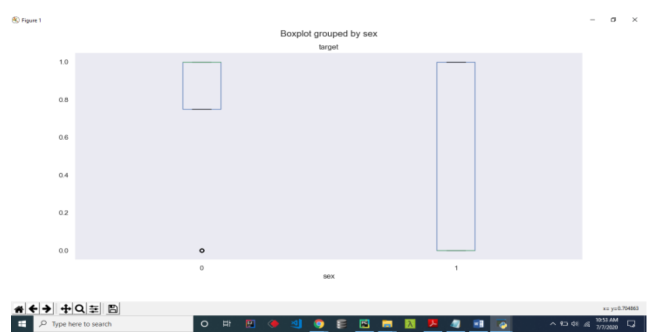


Data Descriptions



Correlation Matrix:

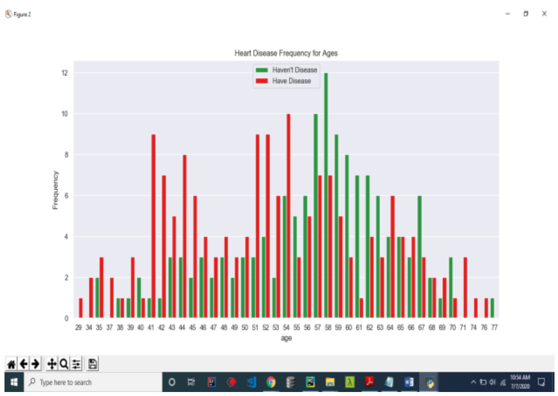
Box Plot:



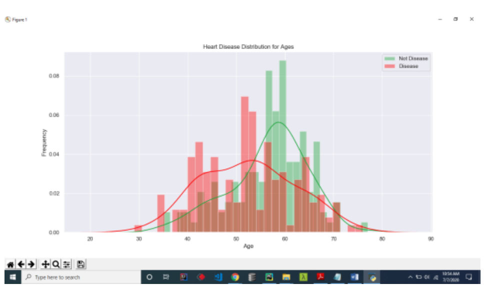
Target Count:

eISSN 1303-515

Heart Frequency:



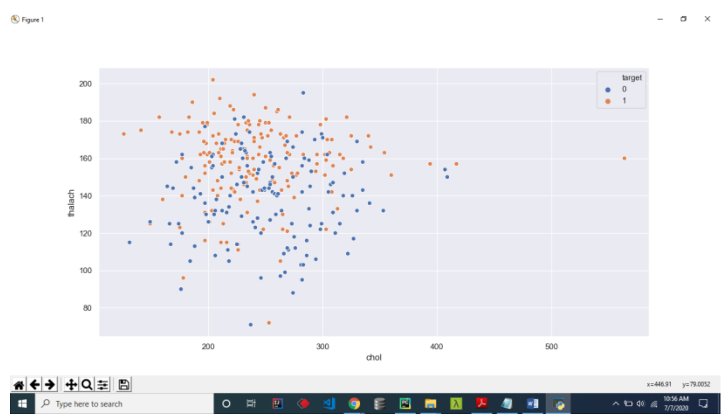
Distribution on Age:



Calculated Frequency:

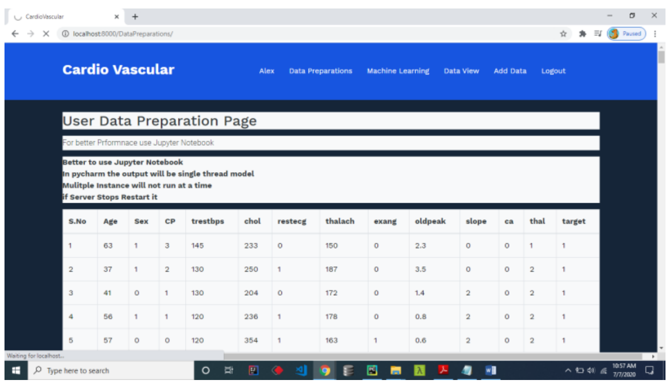
Scatter matrix:

Thalch anc Chol

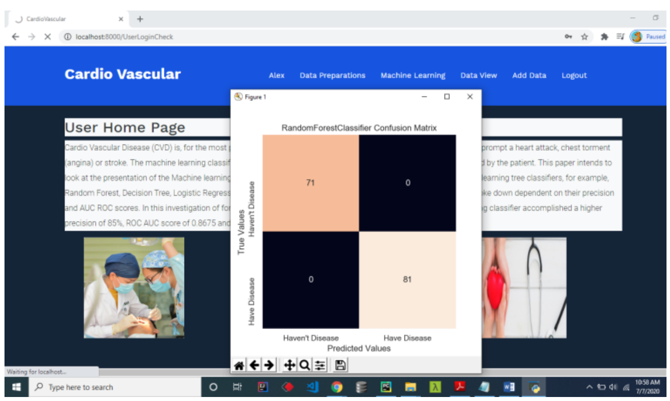


Age wise Disease scatter

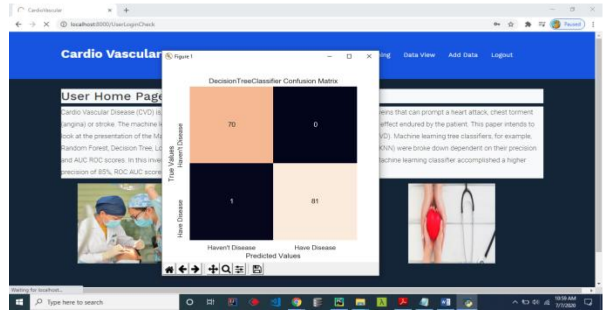
Cleaned Data:



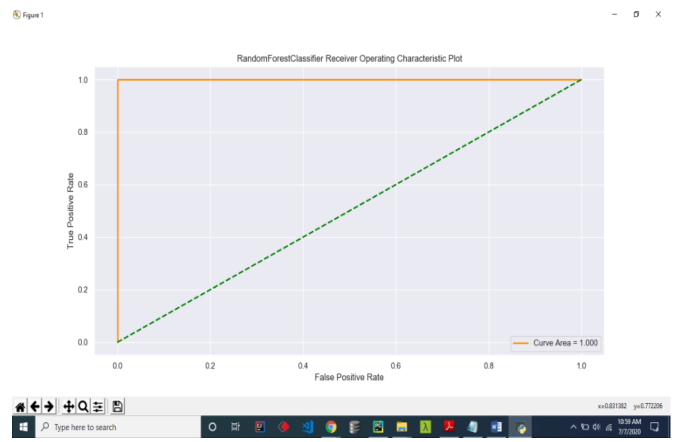
Randome Forest Confusion matrix:



Decision Tree Confusion Matrix:



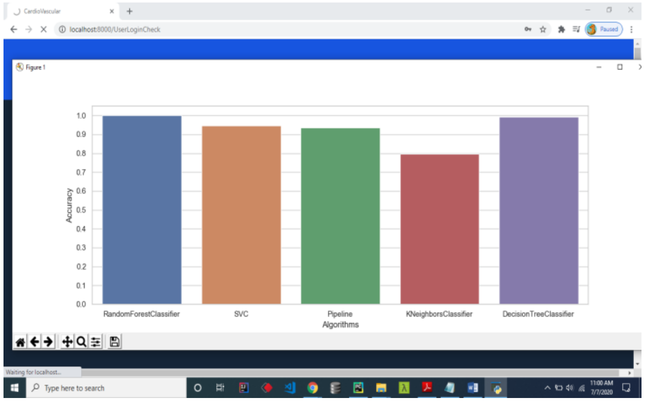
Randome Forest Roc:



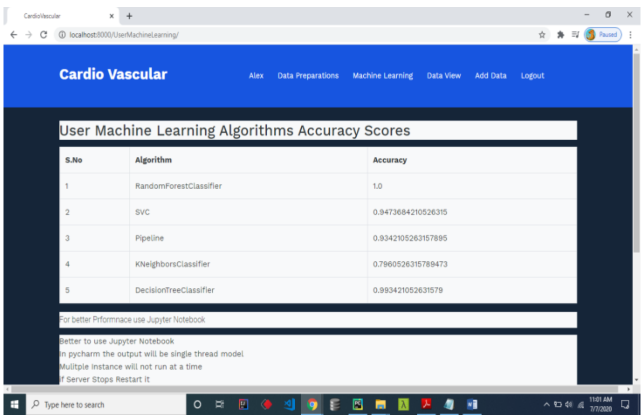
SVC Curve:

Algorithm Accuracy:

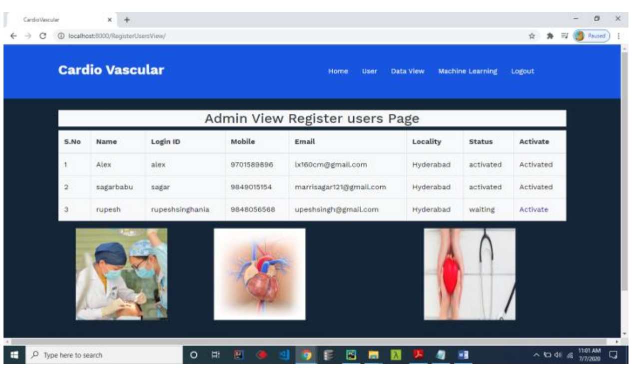
eISSN 1303-515



Accuracy on Client Side:



Admin Activate users:



CONCLUSION

In this work, machine learning classifiers such as Random Forest, Decision Tree, Logistic Regression, Support vector machine (SVM), K-nearest neighbors (KNN) were used in the prediction of Cardio Vascular Disease (CVD). The proposed method using a random forest machine learning classifier has achieved a greater accuracy of 85.71% with a ROC AUC score of 0.8675 which outperformed all the classifiers under analysis in classifying patients with Cardio Vascular Disease.

REFERANCES

[1] https://www.who.int/news-room/fact-sheets/detail/cardiovasculardiseases-(cvd)

[2] Kelly, B. B., & Fuster, V. (Eds.). (2010). Promoting cardiovascular health in the developing world: a critical challenge to achieve global health. National Academies Press.

[3] Poirier, Paul, et al. "Obesity and cardiovascular disease: pathophysiology, evaluation, and effect of weight loss: an update of the 1997 American Heart Association Scientific Statement on Obesity and Heart Disease from the Obesity Committee of the Council on Nutrition, Physical Activity, and Metabolism." Circulation 113.6 (2006): 898-918.

[4] Bhatnagar, Prachi, et al. "Trends in the epidemiology of cardiovascular disease in the UK." Heart 102.24 (2016): 1945-1952.

[5] Beunza, Juan-Jose, et al. "Comparison of machine learning algorithms for clinical event prediction (risk of coronary heart disease)." Journal of biomedical informatics 97 (2019): 103257.

[6] Zhao, Lina, et al. "Enhancing Detection Accuracy for Clinical Heart Failure Utilizing Pulse Transit Time Variability and Machine Learning." IEEE Access 7 (2019): 17716-17724.

[7] Borkar, Sneha, and M. N. Annadate. "Supervised Machine Learning Algorithm for Detection of Cardiac Disorders." 2018 Fourth International Conference on Computing Communication Control and Automation (ICCUBEA). IEEE, 2018.

[8] Omar Boursalie, Reza Samavi, Thomas E. Doyle. "M4CVD: Mobile Machine Learning Model for Monitoring Cardiovascular Disease." Procedia Computer Science, 63 (2015): 384-391.

[9] Chen, Rui, et al. "Using Machine Learning to Predict One-year Cardiovascular Events in Patients with Severe Dilated Cardiomyopathy." European Journal of Radiology (2019).

[10] Dhar, Sanchayita, et al. "A Hybrid Machine Learning Approach for Prediction of Heart Diseases." 2018 4th International Conference on Computing Communication and Automation (ICCCA). IEEE, 2018.

[11] Dinesh, Kumar G., et al. "Prediction of Cardiovascular Disease Using Machine Learning Algorithms." 2018 International Conference on Current Trends towards Converging Technologies (ICCTCT). IEEE, 2018.

[12] Mezzatesta, Sabrina, et al. "A machine learning-based approach for predicting the outbreak of cardiovascular diseases in patients on dialysis." Computer Methods and Programs in Biomedicine 177 (2019): 9-15.