

ANALYSIS AND PREDICTION OF CARDIO VASCULAR DISEASE

¹BADRAPU MANASA,²SAIKRISHNA.D

¹MCA Student, ²Assistant Professor

DEPARTMENT OF MCA

SREE CHAITANYA COLLEGE OF ENGINEERING, KARIMNAGAR

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.

I. 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.

II. SYSTEM ANALYSIS EXISTING SYSTEM:

Cosmology and machine learning for visualizing cardio vascular disease as a complex versatile clinical framework. Ontology and machine learning are the techniques that have been utilized in existing methodology. Therefore, it shows a viable cardiovascular choice to help instrument for taking care of mistakes in the clinical hazard appraisal of chest torment patients and helps clinicians adequately recognize intense angina/heart chest torment patients from those with different reasons for chest torments. Another machine learning procedure is Coronary Artery Disease method called N2 Genetic optimizer agent (another hereditary preparing) has been presented in this methodology. These outcomes are aggressive and practically identical to the best outcomes in the field.

DISADVANTAGES OF EXISTING SYSTEM:

- Machine learning-based coronary artery disease examined datasets, test sizes, highlights, areas of information accumulation, execution measurements, and applied ML are the basic methods that have been broken down in this methodology.
- Chronic Heart Failure Detection is an anticipated the Constant Cardiovascular breakdown identification from heart sounds utilizing a pile of machine learning classifiers. The strategies used to foresee comprises filtering segmentation and feature extraction to the model.

Algorithm: Cardiovascular Infection Expectation Framework, Genetic algorithm.

PROPOSED SYSTEM:

We worked on heart disease dataset obtained from UCI (University of California at Irvine) repository, the data set contained attributes such as age, sex, cp, trestbps, cho, fbs, restecg, thalach, ca, and target with 304 instances has taken. At first level, The dataset is first cleansed and processed using preprocessing techniques like Data Integration, Data transformation, Data reduction, and Data cleaning using pandas tool. The proposed framework a total of 304 patient records were visualized. Data visualization techniques helps the data scientist to understand the feasibility of the dataset. The box plot relationship between the sex and target attributes. The correlation matrix and histogram were represented.

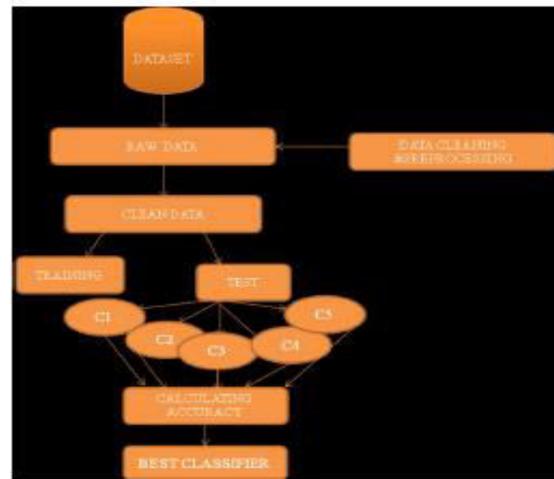
ADVANTAGES OF PROPOSED SYSTEM:

- 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.

Algorithm: Support Vector Machine (SVM), Decision Tree (DT), Random Forest (RF), K-Nearest Neighbors (KNN).

SYSTEM ARCHITECTURE:



III. IMPLEMENTATION

MODULES DESCRIPTION:

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.

IV. 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.

Further Enhancement

The area under the ROC curve measures how well a continuous variable predicts the outcome of interest: if the sensitivity increases steeply as the threshold for diagnosis is relaxed with only a relatively slow accumulation of false positives, the area under the ROC curve will be large; conversely, if the sensitivity increases slowly as the threshold for diagnosis is relaxed with a rapid accumulation of false positives, the area under the ROC curve will be correspondingly smaller. The differences in areas may be tested to determine whether they are statistically significant. We have used this approach to compare the fasting glucose value and the value 2 h after an oral glucose load with various multivariate models for predicting future (Cardio Vascular Disease) CVD.

REFERENCES

- [1] [https://www.who.int/news-room/fact-sheets/detail/cardiovascular-diseases-\(cvd\)](https://www.who.int/news-room/fact-sheets/detail/cardiovascular-diseases-(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.
- [13] Terrada, Oumaima, et al. "Classification and Prediction of atherosclerosis diseases using machine learning algorithms." 2019 5th International Conference on Optimization and Applications (ICOA). IEEE, 2019.
- [14] Alić, Berina, Lejla Gurbeta, and Almir Badnjević. "Machine learning techniques for classification of diabetes and cardiovascular diseases." 2017 6th Mediterranean Conference on Embedded Computing (MECO). IEEE, 2017.
- [15] Awan, Shahid Mehmood, Muhammad Usama Riaz, and Abdul Ghaffar Khan. "Prediction of heart disease using artificial neural network." *VFAST Transactions on Software Engineering* 13.3 (2018): 102-112.