FACULTY POSITION RECLASSIFICATION FOR SUCS

(DBM-CHED Joint Circular No. 3, series of 2022)

CERTIFICATION OF PERCENTAGE CONTRIBUTION

(Research Output with Multiple Authors)

Title of Research: Identification of Coronary Artery Disease Using Extra Tree Classification
Type of Research Output: Conference Paper (Scopus-Indexed, IEEE Paper)

Instruction: Supply ALL the names of the authors involved in the publication of Research output and indicate the contribution of each author in percentage. Each author shall sign the conforme column if he/she agrees with the distribution. The conforme should be signed by all the authors in order to be considered. Please prepare separate Certification for each output.

	Name of Authors	Current Affiliation	% Contribution	Conforme *(Sign if you agree with the % distribution)*
1	Nazia Wahid	Vishwakarma University	18%	
2	Abdelhamid Zaidi	Qassim University	18%	
3	Gaurav Dhiman	GBCC	18%	
4	Manika Manwal	Graphic Era Hill University	18%	
5	Dheresh Soni	VIT University	18%	
6	Renato R. Maaliw III	SLSU	10%	

** Should have a total of 100%* 100.00%

Prepared by:

Renato R. Maaliw III, DIT
(Name and Signature)
Faculty

Certified by:

Nicanor L. Guinto, Ph.D
(Name and Signature)
Director, Office of Research Services

Proceedings of the International Conference on Inventive Computation Technologies (ICICT 2023)
IEEE Xplore Part Number: CFP23F70-ART; ISBN: 979-8-3503-9849-6

Identification of Coronary Artery Disease using Extra Tree Classification

Nazia Wahid[1]
Assistant Professor, Department of Mathematics & Statistics, Faculty of Science & Technology, Vishwakarma University,
Pune (Maharashtra)-411048
wahidnazia010@gmail.com

Abdelhamid Zaidi[2]
Department of mathematics, College of Science, Qassim University,
P.O.Box 6644, Buraydah 51452, Saudi Arabia
A.ZAIDI@qu.edu.sa

Gaurav Dhiman[3]
Department of Computer Science, Government Bikram College of Commerce,
Patiala, India
gdhiman0001@gmail.com

Manika Manwal[4a,4b]
[4a]*Assistant Professor, CSE,Graphic Era Hill University,*
Dehradun, India
[4b]*Research Scholar, CSE,Graphic Era deemed to be University,*
Dehradun, India
mmanwal@gehu.ac.in

Dheresh Soni[5]
School of CSE, VIT BHOPAL University
dheresh.soni@vitbhopal.ac.in

Renato R. Maaliw III[6]
College of Engineering, Southern Luzon State University,
Lucban, Quezon, Philippines
rmaaliw@slsu.edu.ph

2023 International Conference on Inventive Computation Technologies (ICICT) | 979-8-3503-9849-6/23/$31.00 ©2023 IEEE | DOI: 10.1109/ICICT57646.2023.10134338

Abstract— **Caused by a shortage of oxygen-rich blood, cardiac artery disease occurs when coronary arteries get blocked. Arterial occlusion decreases the heart's blood flow and hence its oxygen supply. Plaque buildup is the leading cause of arterial obstruction. Both cholesterol and calcium can be found in plaque. Atherosclerosis refers to the development of plaque within the arteries. It is possible to have total ischemia due to an arterial clot. As arterial plaque ruptures, blood clots form. A diet rich in cholesterol can contribute to the development of plaque in a person's arteries. Cholesterol is transported in the bloodstream via lipoproteins, which are protein complexes. Lipoproteins come in two main types, high-density (HDL) and low-density (LDL). Plaque builds up in the arteries when levels of LDL are too high. The term "bad cholesterol" (LDL) is also commonly used. HDL is responsible for transporting cholesterol from the blood to the liver. HDL is the "good" or "healthy" cholesterol. Maintaining healthy levels of LDL and HDL is crucial for avoiding CAD. This research study has developed a ETC (Extra Tree Classification) approach for CAD diagnosis. The proposed approach employed the Z-Alizadeh Sani CAD dataset and employed SMOTE to address the issue of class imbalance. Classification was carried out utilizing classifiers such as XGBoost (extreme gradient boosting), KNN (K-nearest neighbour), SVM-Linear (support vector machine - linear), and SVM-RBF (support vector machine - radial basis function). After employing the GridSearch optimization approach, the hyper-parameters were fine-tuned to an accuracy of 95.16 percent.**

Keywords—Demographic Features, Machine Learning, Classifier, optimization, Feature Selection

I. INTRODUCTION

The most common sign of heart disease is chest pain. The cause of chest discomfort is a lack of oxygenated blood to the heart. In medical terms, this is called angina. It's possible the person will feel suffocated. Heart disease can also cause pain that radiates from the chest to other parts of the body. In addition to lightheadedness, nausea, and profuse perspiration, heart disease can cause dizziness. Heart disease can also cause swelling in the feet and general weakness [1]. All intelligent apps nowadays leverage machine learning capabilities to improve efficiency. Machine learning is useful because it can help extract information from raw data that would be difficult or impossible for traditional algorithms to extract [33]. In recent years, machine learning has expanded into every area of study.

Machine learning allows for automated learning by machines. It's utilized to give computers the capacity to reason as well as people [2]. The performance of a machine improves with time, and this training is facilitated by the machine's exposure to data [3].

When not caught in time, coronary artery disease (CAD) can cause cardiac arrest [23]. In terms of global mortality, it is by far the most important factor. Blockages in the arteries that provide blood to the heart muscles are the cause of coronary artery disease [4].

Cholesterol and calcium accumulate into plaque, which hardens and narrows arteries. That's medical jargon for what we call atherosclerosis. The endothelium is a thin layer of cells lining the inside of the arteries that helps keep blood flowing smoothly. Atherosclerosis occurs when plaque forms in the arterial wall due to endothelial degradation. The weakening of blood cells and subsequent disruption of the heart's blood flow are direct results of plaque buildup. If coronary artery disease is not recognized and treated in a timely manner, the patient's condition might worsen [23]. Plaque continues to build up and eventually hardens. As plaques burst, platelets can aggregate and cause clots to develop in the arteries. It can cause a total cutoff of blood flow, which can result in a heart attack [5].

Diagnosing CAD typically involves angiography. This technique is invasive and may have unwanted consequences. Thus, clinical data-based non-invasive techniques for CAD diagnosis are required. In this paper, we proposed using clinical data to diagnose CAD with the use of ETC (Extra Tree Classification). This approach was verified using the Z-Alizadeh Sani CAD dataset hosted by UCI (University of California, Irvine). The primary goal of this proposed technique is to minimize mortality by increasing CAD detection rates. Experiments are conducted using the Z-Alizadeh Sani CAD dataset. It's a clinical dataset including information on 303 people and 54 attributes. The dataset is unbalanced, including information from just 87 healthy people compared to 216 people with CAD. The dataset may be accessed through the UCI digital archive. There are primarily four kinds of characteristics in this dataset: ECG, and Demographic Factors Echocardiogram, Laboratory Tests, and Physical Examination [6].

II. RELATED WORK

The authors Choi et al. [7] trained a model to identify heart failure using RNN (recurrent neural network). Using the GRU of RNN, the scientists evaluated the correlation between EHR (electronic health records) events over time. Electronic Health Records were retrieved from Sutter PAMF (Palo Alto Medical Foundation). An assortment of one-hot vectors was used to characterize EHR events. N happenings were represented by an N-dimensional vector. Every n-dimensional vector has a single element set to 1 to denote the occurrence of the event and every other element set to 0. At time t, the GRU in the hidden layer h was fed the vector Xt as its input. Each each time stamp reveals a different secret condition. A scalar value, indicating the patient's risk score, was generated by applying logistic regression on the vector of the end state. AUC was calculated at 0.777, therefore the model did a good job.

One such approach, presented by Arabasadi et al. [8] and based on the use of a neural network, is one for the diagnosis of CAD. The neural network's weights were fine-tuned using GA. During teaching the ANN, the back propagation method was applied. Generalized Adaptiveness started with a population size of 100 chromosomes. The root mean square error (RMSE) of the untrained ANN was used to determine the fitness value of the chromosomes. In GA, the algorithm employed for selection was akin to a roulette wheel. The crossover probability between the two points was set at 1. In order to accomplish the mutation, a gaussian mutation was used. Each gene on a chromosome encoded one of the neural network's weights, thus each chromosome included all of the network's weights. SVM was utilised for feature selection. The Z-Alizadeh Sani dataset was used for the system's evaluation. The system was able to attain a 93.85% success rate.

The ANN model for estimating the likelihood of cardiac failure was created by Samuel et al. [9]. Fuzzy optimization was used to find the optimal weights for the network. The success rate was 91.10 percent. A method for estimating the likelihood of heart disease was created by Kim and Kang [10]. By doing a correlation analysis of traits, we were able to narrow down on the most important ones for making the diagnosis. They are strongly associated if a shift in the value of one feature affects the sensitivity of the other feature more than the average shift in the sensitivity of all characteristics. The hidden layer of the neural network was used to link the associated features. The neural network used to make illness predictions was only fed information that was actually useful. On the basis of the KNHANES-VI dataset, experiments were conducted (6th Korea National Health and Nutrition Examination Survey). The overall system reliability was 82.51%. In [11], Caliskan and Yusskel suggested a DNN (deep neural network model) for CAD diagnosis by fusing a softmax and two auto encoders. There were four datasets used to verify the model's accuracy by the authors. On the Cleveland dataset, the system obtained a success rate of 85.2%, on the Long Beach dataset, it reached 84%, on the Switzerland dataset it scored 92.2%, and on the Hungarian dataset it achieved 83.5%. Using a stacked ensemble model, Tama et al. [12] created a CAD diagnostic system. The stacked model was built with the help of the GB (gradient boosting), RF (random forest), and XGB (extreme gradient boosting) classifiers, with dimensionality reduction handled by PSO. The system has a precision of 91.18 percent on the Hungarian dataset, 86.49 percent on the Cleveland dataset, and 93.55% on the Statlog dataset.

Experimental work on three datasets was conducted by Terrada et al. [13] using DT, ANN, and AdaBoost classifiers for the diagnosis of heart disease. Using the Z-alizadeh Sani dataset, ANN achieved a best-in-class accuracy of 94%.

Using the J48, CART (classification and regression tree), and RF classifiers, Verma [14] created an ensemble model for CAD diagnosis. Z-Alizadeh data was used to verify the model. Overall, the method was accurate 84.82% of the time. In order to better anticipate the occurrence of cardiac disease, Javid et al. [15] created an ensemble model based on a voting process. The ensemble model was created using many classifiers, including KNN, RF, SVM, GRU (gated recurrent unit), and LSTM (long short term memory) [24]. The Cleveland data set served as the subject of the experiments. The dataset was divided into a training portion (80%) and a validation portion (20%). The method was accurate to the tune of 85.71%.

Models for the diagnosis of CAD were created by Joloudari et al. [16] utilising DT (decision tree), RT (random tree), CHAID (chi-squared automated interaction detection), and SVM (support vector machine). Using a feature ranking system, crucial characteristics were chosen. A 91.47 percent success rate was achieved by using a random tree. The dataset was split up into sections by Mienye et al. [17]. CART was used to create several models for the separated datasets (classification and regression tree). By pooling together distinct CART models, an ensemble model was created. For the Framingham dataset, the method attained 91% accuracy. Cleveland, Hungarian, Long Beach, and the Swiss datasets were all integrated into one by Spencer et al. [18]. After building the unified data set, we used the chi-square test to zero down on the most pertinent characteristics. The bayes net method was used to correctly categorise patients with heart disease 85% of the time. Using an SVM classifier, Gageloglu [19] was able to correctly categorise cardiac diseases with an accuracy of 84.81%.

Prediction of cardiovascular disease was performed using DT and KNN classifiers by Jothi et al. [20]. KNN was only able to reach 67% accuracy on the Cleveland dataset, but DT was able to obtain 81% accuracy.

Optimizing a random forest classifier with a combination of randomised search, grid search, and a genetic algorithm was accomplished by Valarmathi et al. [21,32]. SFS (sequential forward selection) was used to narrow down candidate traits to those most relevant to making a diagnosis. For the Z-Alizadeh dataset, the optimal random forest has an accuracies of 80.2% while diagnosing CAD. Using FCRLC, Bahani et al. [22] created a method for cardiac disease prediction (fuzzy rule based classification system with fuzzy clustering and linguistic modifiers). For the Cleveland dataset, the system was 83.17 percent accurate.

Shorewala [26] created a stacked model for cardiac disease diagnosis that employs KNN, SVM, and RF. The LASSO technique was used to narrow down candidate features. The model was tested on data from 70,000 people diagnosed with cardiovascular disease, and it was shown to be 75.1% accurate. L et al. [27] created a refined model for cardiac illness forecasting using a random forest classifier. Using a genetic algorithm, the collection of features was fine-tuned. For the Z-Alizadeh dataset, the model achieved an accuracy of 90.7%.

III. METHODOLOGY

Many trials utilizing four distinct classifiers on the CAD dataset led to the proposal of the ET-SVMRBF approach. To begin with, the data has been pre-processed so that it is in a usable condition for categorization [31]. As part of the preparation of the data, a one-hot encoding was applied to the dataset, and conventional scalar scaling was utilized. Using a conventional scalar transformation, the data is transformed so that every feature has a mean value of 0 and a standard deviation of 1 [28]. There are 216 records of people with CAD and 87 records of healthy people, a huge disparity. A model's ability to classify information accurately may be impaired by skewed data. With the help of SMOTE, we've achieved data parity (Synthetic Minority Oversampling Technique). There are 55 prediction attributes in the dataset.

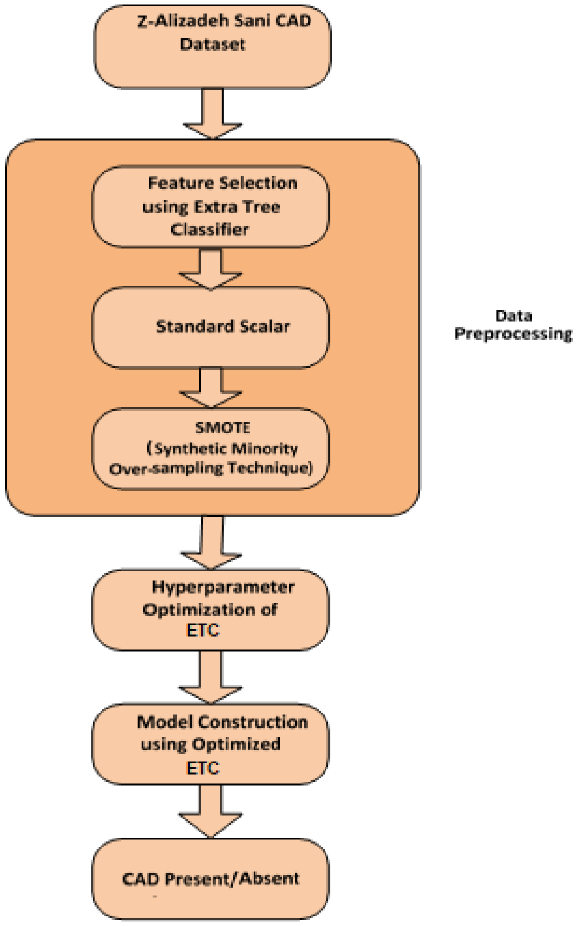

Fig.1. ETC Methodology for CAD Diagnosis

In order to narrow down the features, an additional tree classifier is used. 36 features were produced as a direct result. For the classification task, we used XGBoost (extreme gradient boosting), KNN (K- closest neighbour), SVM-Linear (support vector machine-linear), and SVM-RBF (support vector machine-random forest) (support vector machine-radial basis function) [25]. Because of its superior performance, the SVMRBF model's hyperparameters were adjusted using the GridSearch technique. At long last, this perfected model has been applied to the detection of CAD. Figure 1 demonstrates the ET-SVMRBF technique for detecting CAD.

The effectiveness of classification algorithms is affected by feature selection [29]. The value of features has been computed with the help of an additional tree classifier. This classifier does classification by combining the outcomes of many decision trees. To carry out feature selection, one generates several decision trees from the original training samples, feeding each tree a different, randomly generated set of k features. The Gini index of each feature is used by the decision tree to determine which characteristic is the most important. Finally, the Gini importance of each feature is used in conjunction with the aggregated results from these many de-correlated trees to yield feature importance. The relevance ranking of features is used to determine which K characteristics are the most important, and the top K features can have any value. A number of tests were conducted with varying K values in order to pinpoint the optimal value. By using K=36, the best outcomes were shown [30].

IV. Analysis of ETC Efficiency

ETC was tested on the Z-Alizadeh Sani CAD dataset and the results were analysed. Both continuous and categorical variables are present in the dataset. The ten-fold cross-validation procedure ensures the accuracy of the results. XGBoost, KNN, SVM-Linear, and SVM-RBF are just a few of the classifiers that have been used to develop various models on the CAD dataset. Performance measurements including as accuracy, sensitivity, specificity, precision, and F-Measure have been applied to the study of classifier efficacy. Other than these measures of efficacy, ROC curves have also been used to assess the quality of various models (receiver operating characteristic curve). A common scalar was utilized to normalize the numbers. SMOTE was used to achieve class balancing, which ensured that training data was representative of the population as a whole. First, CAD prediction using the entire dataset's characteristics was performed, and classifier performance was assessed. Figures 2, 3, 4, 5 and 6 illustrate the results of a comparative examination of the performance of the various classifiers. ETC has delivered the best performance.

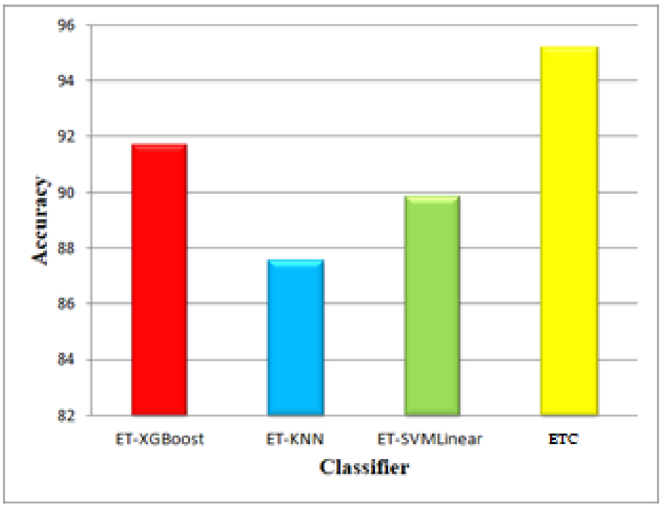

Fig.2. Accuracy comparison for different classifiers

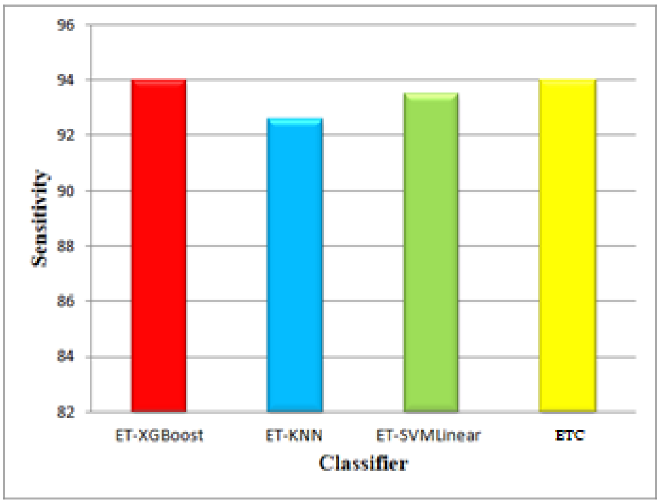

Fig.3. Sensitivity comparison for different classifiers

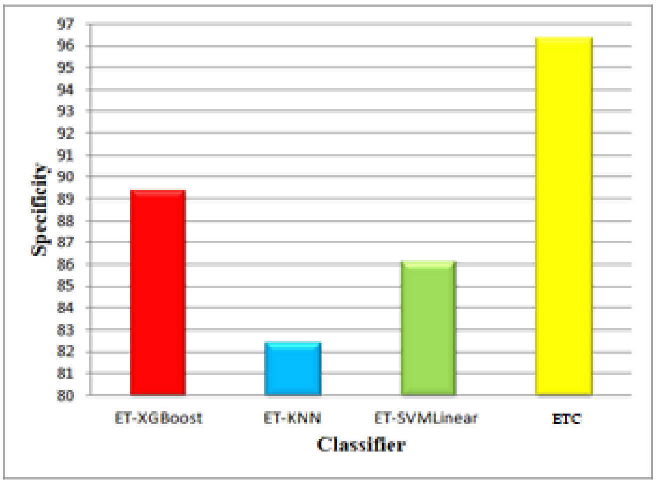

Fig.4. Specificity comparison for different classifiers

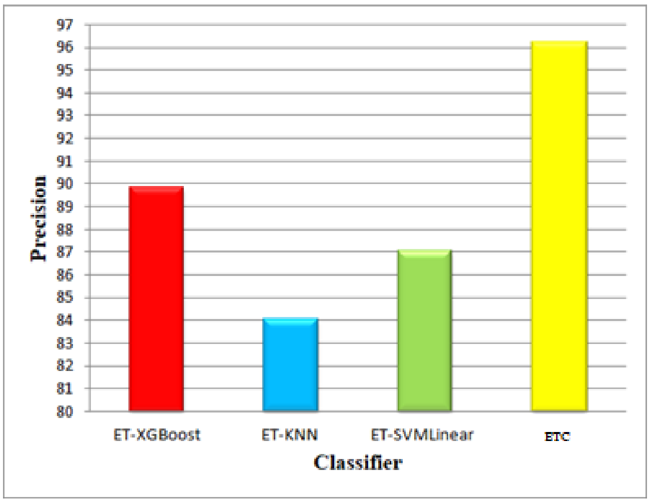

Fig.5. Precision comparison for different classifiers

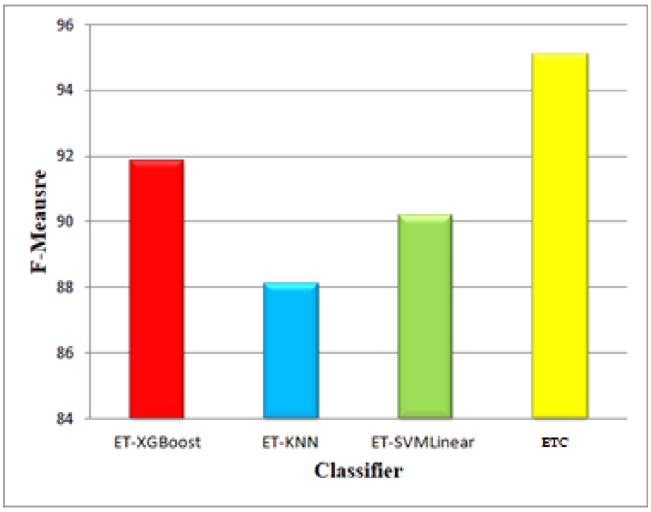

Fig.6. F-Measure comparison for all classifiers

V. CONCLUSION

As CAD is potentially fatal, an early diagnosis is of paramount importance. Using a combination of extra tree and SVM-RBF algorithms, ETC is proposed in this chapter as an approach for CAD diagnosis. The Z-Alizadeh Sani CAD dataset was used to assess the efficacy of four different classifiers: XGBoost, KNN, SVM-Linear, and SVM-RBF. The data has been preprocessed using the regular scalar, SMOTE, and additional tree feature selection techniques. Optimizing the model's performance with GridSearch, the accuracy of ETC is 95.79 percent. The results suggest that this strategy may be utilised to efficiently detect CAD using clinical data. The medical field is one where it has the potential to be of great use. In places with inadequate access to healthcare, such a device might be invaluable. It can help doctors improve the quality and precision of their diagnoses when used in conjunction with a second opinion.

REFERENCES

1. Quah, J. L., Yap, S., Cheah, S. O., Ng, Y. Y., Goh, E. S., Doctor, N., Leong, B. S., Tiah, L., Chia, M. Y., & Ong, M. E. (2014). Knowledge of signs and symptoms of heart attack and stroke among Singapore residents. BioMed Research International, 2014, 572425. https://doi.org/10.1155/2014/572425.
2. Ray, A., & Chaudhuri, A. K. (2021). Smart healthcare disease diagnosis and patient management: Innovation, improvement and skill development. Machine Learning with Applications, 3, 100011. https://doi.org/10.1016/j.mlwa.2020.100011.
3. L'heureux, A., Grolinger, K., Elyamany, H. F., & Capretz, M. A. (2017). Machine learning with big data: Challenges and approaches. IEEE Access, 5, 7776-7797. https://doi.org/10.1109/ACCESS.2017.2696365.
4. Wah, T. Y., Gopal Raj, R., & Iqbal, U. (2018). Automated diagnosis of coronary artery disease: A review and workflow. Cardiology Research and Practice, 2018, 2016282. https://doi.org/10.1155/2018/2016282.
5. Xu, H., Duan, Z., Miao, C., Geng, S., & Jin, Y. (2017). Development of a diagnosis model for coronary artery disease. Indian Heart Journal, 69(5), 634-639. https://dx.doi.org/10.1016%2Fj.ihj.2017.02.022.
6. Rajkumar, R., Anandakumar, K., & Bharathi, A. (2006). Coronary artery disease (CAD) prediction and classification-a survey. ARPN Journal of Engineering and Applied Sciences, 11(9), 5749-5754.
7. Choi, E., Schuetz, A., Stewart, W. F., & Sun, J. (2017). Using recurrent neural network models for early detection of heart failure onset. Journal of the American Medical Informatics Association, 24(2), 361-370. https://doi.org/10.1093/jamia/ocw112.
8. Arabasadi, Z., Alizadehsani, R., Roshanzamir, M., Moosaei, H., & Yarifard, A. A. (2017). Computer aided decision making for heart disease detection using hybrid neural network- genetic algorithm. Computer Methods and Programs in Biomedicine, 141, 19-26. https://doi.org/10.1016/j.cmpb.2017.01.004.
9. Samuel, O. W., Asogbon, G. M., Sangaiah, A. K., Fang, P., & Li, G. (2017). An integrated decision support system based on ANN and Fuzzy_AHP for heart failure risk prediction. Expert Systems with Applications, 68, 163-172.

https://doi.org/10.1016/j.eswa.2016.10.020.

10. Kim, J. K., & Kang, S. (2017). Neural network-based coronary heart disease risk prediction using feature correlation analysis. Journal of Healthcare Engineering, 2017, 1-13. https://doi.org/10.1155/2017/2780501.

11. Caliskan, A., & Yuksel, M. E. (2017). Classification of coronary artery disease data sets by using a deep neural network. EuroBiotech J, 1(4), 271-277. http://dx.doi.org/10.24190/ISSN2564-615X/2017/04.03.

12. Terrada, O., Hamida, S., Cherradi, B., Raihani, A., & Bouattane, O. (2020). Supervised machine learning based medical diagnosis support system for prediction of patients with heart disease. Advances in Science, Technology and Engineering Systems Journal, 5(5), 269-277. http://dx.doi.org/10.25046/aj050533.

13. Verma P. (2020). Ensemble models for classification of coronary artery disease using decision trees. International Journal of Recent Technology and Engineering, 8(6), 940-944. http://dx.doi.org/10.35940/ijrte.F7250.038620.

14. Javid, I., Alsaedi, A. K. Z., & Ghazali, R. (2020). Enhanced accuracy of heart disease prediction using machine learning and recurrent neural networks ensemble majority voting method, International Journal of Advanced Computer Science and Applications, 11(3), 540-551. http://dx.doi.org/10.14569/IJACSA.2020.0110369.

15. Srivastava, Deepak, et al. "Analysis of Protein Structure for Drug Repurposing Using Computational Intelligence and ML Algorithm." International Journal of Software Science and Computational Intelligence (IJSSCI) 14.1 (2022): 1-11.

16. Mienye, I. D., Sun, Y., & Wang, Z. (2020). An improved ensemble learning approach for the prediction of heart disease risk. Informatics in Medicine Unlocked, 20, 100402. https://doi.org/10.1016/j.imu.2020.100402.

17. Srivastava, D., Soni, D., Sharma, V., Kumar, P., & Singh, A. K. (2022). An Artificial Intelligence Based Recommender System to analyze Drug Target Indication for Drug Repurposing using Linear Machine Learning Algorithm. Journal of Algebraic Statistics, 13(3), 790-797.

18. Gazeloglu, C. (2020). Prediction of heart disease by classifying with feature selection and machine learning methods . Progress in Nutrition, 22(2), 660–670. https://doi.org/10.23751/pn.v22i2.9830.

19. Jothi, K. A., Subburam, S., Umadevi, V., & Hemavathy, K. (2021). Heart disease prediction system using machine learning. Materials Today: Proceedings. https://doi.org/10.1016/j.matpr.2020.12.901.

20. Valarmathi, R., & Sheela, T. (2021). Heart disease prediction using hyper parameter optimization (HPO) tuning. Biomedical Signal Processing and Control, 70, 103033. https://doi.org/10.1016/j.bspc.2021.103033.

21. Bahani, K., Moujabbir, M., & Ramdani, M. (2021). An accurate Fuzzy Rule- Based Classification Systems for heart disease diagnosis. Scientific African, e01019. https://doi.org/10.1016/j.sciaf.2021.e01019.

22. Shorewala, V. (2021). Early detection of coronary heart disease using ensemble techniques. Informatics in Medicine Unlocked, 100655. https://doi.org/10.1016/j.imu.2021.100655.

23. S. Gaba, A. Aggarwal, S. Nagpal, D. Kumar and P. Singh, "A Forecast of Coronary Heart Disease using Proficient Machine Learning Algorithms," 2021 Sixth International Conference on Image Information Processing (ICIIP), Shimla, India, 2021, pp. 517-522, doi: 10.1109/ICIIP53038.2021.9702640.

24. A. Aggarwal, S. Gaba and P. Singh, "Character Recognition using Approaches of Artificial Neural Network: A Review," CEUR Workshop Proceedings, vol. 3309, no. 1, pp. 186-193, 2022.

25. G. Bathla, P. Singh, R. K. Singh, E. Cambria and R. Tiwari, "Intelligent Fake reviews detection based on Aspect Extraction and Analysis using Deep Learning," Neural Computing and Applications, July 2022.

26. L, P.R., Jinny, S.V. & Mate, Y.V. Early prediction model for coronary heart disease using genetic algorithms, hyper-parameter optimization and machine learning techniques. Health and Technology. 11, 63–73 (2021). https://doi.org/10.1007/s12553-020-00508-4.

27. Tomar, D., & Agarwal, S. (2014). Feature selection based least square twin support vector machine for diagnosis of heart disease. International Journal of Bio-Science and Bio-Technology, 6(2), 69-82. http://dx.doi.org/10.14257/ijbsbt.2014.6.2.07.

28. Malakar, A. K., Choudhury, D., Halder, B., Paul, P., Uddin, A., & Chakraborty, S. (2019). A review on coronary artery disease, its risk factors, and therapeutics. Journal of Cellular Physiology, 234(10), 16812-16823. https://doi.org/10.1002/jcp.28350.

29. Zhenya, Q., & Zhang, Z. (2021). A hybrid cost-sensitive ensemble for heart disease prediction. BMC Medical Informatics and Decision Making, 21(1), 1-18.

30. Weinhaus, A. J., & Roberts, K. P. (2005). Anatomy of the human heart. In handbook of cardiac anatomy, physiology, and devices (pp. 51-79). Humana press. https://doi.org/10.1007/978-1-59259-835-9_4.

31. P. Borah et al., “Neurological Consequences of SARS-CoV-2 Infection and Concurrence of Treatment-Induced Neuropsychiatric Adverse Events in COVID-19 Patients: Navigating the Uncharted,” Front. Mol. Biosci., vol. 8, 2021, doi: 10.3389/fmolb.2021.627723.

32. A. Panwar, R. Yadav, K. Mishra, and S. Gupta, “Deep learning techniques for the real time detection of Covid19 and pneumonia using chest radiographs,” in EUROCON 2021 - 19th IEEE International Conference on Smart Technologies, Proceedings, 2021, pp. 250–253, doi: 10.1109/EUROCON52738.2021.9535604.

33. G. Chhabra, E. M. Onyema, S. Kumar, M. Goutham, S. Mandapati, and C. Iwendi, “Human emotions recognition, analysis and transformation by the Bioenergy Field in smart grid using image processing,” Electronics, vol. 11, no. 23, p. 4059, 2022.

6th International Conference on Inventive Computation Technologies, ICICT 2023 - Proceedings • Pages 787 - 792 • 2023 • 6th International Conference on Inventive Computation Technologies, ICICT 2023 • Lalitpur • 26 April 2023through 28 April 2023 • Code 189117

Document type
Conference Paper
Source type
Conference Proceedings
ISBN
979-835039849-6
DOI
10.1109/ICICT57646.2023.10134338
View more

Identification of Coronary Artery Disease using Extra Tree Classification

Wahid, Nazia[a] ; Zaidi, Abdelhamid[b] ; Dhiman, Gaurav[c] ; Manwal, Manika[d, e] ; Soni, Dheresh[f] ; Maaliw, Renato R.[g]
Save all to author list

[a] Department of Mathematics & Statistics, Faculty of Science & Technology, Vishwakarma University, Maharashtra, Pune, 411048, India
[b] Department of Mathematics, College of Science, Qassim University, P.O.Box 6644, Buraydah, 51452, Saudi Arabia
[c] Department of Computer Science, Government Bikram College of Commerce, Patiala, India
[d] CSE, Graphic Era Hill University, Dehradun, India
View additional affiliations

Abstract

Caused by a shortage of oxygen-rich blood, cardiac artery disease occurs when coronary arteries get blocked. Arterial occlusion decreases the heart's blood flow and hence its oxygen supply. Plaque buildup is the leading cause of arterial obstruction. Both cholesterol andcalcium can be found in plaque. Atherosclerosis refers to the development of plaque within the arteries. It is possible to have total ischemia due to an arterial clot. As arterial plaque ruptures, blood clots form. A diet rich in cholesterol can contribute to the development of plaque in a person's arteries. Cholesterol is transported in the bloodstream via lipoproteins, which are protein complexes. Lipoproteins come in two main types, high-density (HDL) and low-density (LDL). Plaque builds up in the arteries when levels of LDL are too high. The term "bad cholesterol" (LDL) is also commonly used. HDL is responsible for transporting cholesterol from the blood to the liver. HDL is the "good" or "healthy" cholesterol. Maintaining healthy levels of LDL and HDL is crucial for avoiding CAD. This research study has developed a ETC (Extra Tree Classification) approach for CAD diagnosis. The proposed approach employed the Z-Alizadeh Sani CAD dataset and employed SMOTE to address the issue of class imbalance. Classification was carried out utilizing classifiers such as XGBoost (extreme gradient boosting), KNN (K-nearest neighbour), SVM-Linear (support vector machine - linear), and SVM-RBF (support vector machine - radial basis function). After employing the GridSearch optimization approach, the hyperparameters were fine-tuned to an accuracy of 95.16 percent. © 2023 IEEE.

Sustainable Development Goals 2023
SciVal Topics

Author keywords

Classifier; Demographic Features; Feature Selection; Machine Learning; optimization

Indexed keywords

Sustainable Development Goals 2023 New

SciVal Topics

Sustainable Development Goals 2023
SciVal Topics

References (33)

1 Quah, J.L.J., Yap, S., Cheah, S.O., Ng, Y.Y., Goh, E.S., Doctor, N., Leong, B.S.-H., (...), Ong, M.E.H.
Knowledge of signs and symptoms of heart attack and stroke among Singapore residents

Identification of Coronary Artery Disease using Extra Tree Classification

Publisher: IEEE Cite This PDF

Nazia Wahid ; Abdelhamid Zaidi ; Gaurav Dhiman ; Manika Manwal ; Dheresh Soni ; Renato R. Maaliw **All Authors**

Abstract:
Caused by a shortage of oxygen-rich blood, cardiac artery disease occurs when coronary arteries get blocked. Arterial occlusion decreases the heart's blood flow and hence its oxygen supply. Plaque buildup is the leading cause of arterial obstruction. Both cholesterol and calcium can be found in plaque. Atherosclerosis refers to the development of plaque within the arteries. It is possible to have total ischemia due to an arterial clot. As arterial plaque ruptures, blood clots form. A diet rich in cholesterol can contribute to the development of plaque in a person's arteries. Cholesterol is transported in the bloodstream via lipoproteins, which are protein complexes. Lipoproteins come in two main types, high-density (HDL) and low-density (LDL). Plaque builds up in the arteries when levels of LDL are too high. The term "bad cholesterol" (LDL) is also commonly used. HDL is responsible for transporting cholesterol from the blood to the liver. HDL is the "good" or "healthy" cholesterol. Maintaining healthy levels of LDL and HDL is crucial for avoiding CAD. This research study has developed a ETC (Extra Tree Classification) approach for CAD diagnosis. The proposed approach employed the Z-Alizadeh Sani CAD dataset and employed SMOTE to address the issue of class imbalance. Classification was carried out utilizing classifiers such as XGBoost (extreme gradient boosting), KNN (K-nearest neighbour), SVM-Linear (support vector machine - linear), and SVM-RBF (support vector machine - radial basis function). After employing the GridSearch optimization approach, the hyper-parameters were fine-tuned to an accuracy of 95.16 percent.

Published in: 2023 International Conference on Inventive Computation Technologies (ICICT)

Date of Conference: 26-28 April 2023

Date Added to IEEE *Xplore*: 01 June 2023

ISBN Information:
Electronic ISBN:979-8-3503-9849-6
Print on Demand(PoD) ISBN:979-8-3503-9850-2

ISSN Information:
Electronic ISSN: 2767-7788
Print on Demand(PoD) ISSN: 2767-777X

INSPEC Accession Number: 23239495

DOI: 10.1109/ICICT57646.2023.10134338

Publisher: IEEE

Conference Location: Lalitpur, Nepal

I. Introduction

The most common sign of heart disease is chest pain. The cause of chest discomfort is a lack of oxygenated blood to the heart. In medical terms, this is called angina. It's possible the person will feel suffocated. Heart disease can also cause pain that radiates from the chest to other parts of the body. In addition to lightheadedness, nausea, and profuse perspiration, heart disease can cause dizziness. Heart disea[...]eet and general weakness [1]. All intelligent apps nowadays leverage machine learning capabilities to improve efficiency. Machine learning is useful because it can help extract information from raw data that would be difficult or impossible for traditional algorithms to extract [33]. In recent years,

Sign in to Continue Reading

Authors
Figures
References
Keywords
Metrics

Mail
Chat
Spaces
Meet
Compose
Inbox
Starred
Snoozed
Sent
Drafts
More
Labels
Acceptance Notifications
Book Chapter Publications
Certificates
Citations
Huawei
ISA
Journal Publications

ICICT 2023; Acceptance Letter

External

ICICT Conf <icict.conf@gmail.com>
to wahidnazia010, A.ZAIDI, gdhiman0001, mmanwal, dheresh.soni, me, Pardeep

Mar 25, 2023, 1:40 PM

Dear Author,
Hearty Congratulations!

Your paper is accepted and recommended for publication in IEEE ICICT 2023 Conference Proceedings.

Please refer the acceptance letter and review comments in the attachments.
Kindly ensure the following before uploading the final paper.

1. The format must be as per IEEE Template: Click here: Sample Paper Template
2. Minimum 15 references must be in the paper and all references must be cited in the text. Like [1], [2] ...
3. The article has few typographical errors which may be carefully looked at.
4. Complete the copyright form (Copyright Form (IEEE))
5. Review comments should be incorporated.

Mode of Presentation: Online

Reviewers Comments:

1. The paper briefly mentions hyperparameter optimization for the Extra Tree algorithm, but lacks in-depth discussions and justification for the chosen hyperparameter values. Proper hyperparameter tuning is crucial for achieving optimal model performance. The paper should have provided more details on the hyperparameter search space, optimization methods (e.g., grid search, random search, Bayesian optimization), and the rationale behind selecting specific values. Without a thorough exploration of hyperparameters, the reported results might not reflect the best possible performance of the Extra Tree model.

2. The paper lacks a comprehensive discussion on the choice of evaluation metrics. It merely mentions using "accuracy" for performance evaluation. While accuracy is a commonly used metric, it may not be appropriate for imbalanced datasets, which are common in medical diagnosis tasks. Precision, recall, F1-score, and area under the receiver operating characteristic curve (AUC-ROC) are essential metrics that should have been included. These metrics provide a more informative assessment of the model's performance, especially when dealing with imbalanced classes. Without a thorough evaluation using various metrics, the reliability of the model's reported accuracy is questionable.

3. The paper's conclusion does not effectively summarize the findings and limitations of the study. It should have discussed the strengths and weaknesses of the proposed Extra Tree approach compared to other classification methods commonly used for CAD detection. Additionally, the authors should have addressed the potential limitations of their study, such as data collection biases, data quality issues, and external validity concerns. A robust discussion of these limitations would provide readers with a better understanding of the reliability and generalizability of the proposed method.

Important Dates:
Last date for registration: 31.03.2023
Conference Date: 26-28, April 2023

Registration Details:

Refer web link: http://icicts.com/2023/registration.html

Payment Link:

INR/USD: https://www.townscript.com/v2/e/6th-international-conference-on-inventive-computation-technologies-110432/booking/tickets

Registration Procedure:

Kindly send the following documents to icict.conf@gmail.com

1. Final paper (source file (MS word/Latex File and PDF required),
2. Anyone author Signed IEEE copyright form
3. Snapshot of Successful transaction (payment proof)

Contact Details:

Mobile number: +918870489968
Mail id: icict.conf@gmail.com

Thanks & Regards
--
--
Conference Chair
ICICT 2023

One attachment • Scanned by Gmail

Acceptance Lette...

Reply | Reply all | Forward