Section A-Research paper
ISSN 2063-5346

Coronary Artery Disease Diagnosis using ET-SVMRBF

Shuchi Mandhanya*, Dr. N. Sateesh kumar
*Research Scholar; Department of Computer science and engineering, Sunrise University, Alwar
Research Supervisor, Associate professor, Department of Computer science and engineering, Sunrise University, Alwar
Email id: shuchi.indani20@gmail.com

Abstract: Angiography is the standard method for diagnosing CAD. This method is an invasive method with certain side effects. Therefore, non-invasive methods should be used to diagnose CAD using clinical data. In this section, an ET-SVMRBF (Extra Tree Support Vector Machine Radial Basis Function) method is proposed for CAD diagnosis using clinical data. The Z-Alizadeh Sani CAD dataset located at UCI (University of California, Irvine) was used to validate this model. The main goal of this strategy is to reduce mortality through timely diagnosis of CAD.

Key Words: Coronary Artery Disease, Plaque deposition, GridSearch optimization.

Introduction

CAD (Coronary Artery Disease) is a heart condition that can lead to a heart attack if not controlled in time. It is the leading cause of death worldwide. CAD is caused by a blockage in the arteries that supply blood to the heart muscle. 1-2

Plaque deposits of cholesterol and calcium can clog arteries. This is called atherosclerosis. The thin linings of blood vessels that allow blood to flow through blood vessels are called endothelial cells. Plaque forms on the artery wall when endothelial cells are damaged. This plaque blocks blood flow to the heart and the blood cells begin to weaken. CAD develops over time and can be fatal if left undiagnosed. Plaque continues to form over time and hardens. If a plaque rupture occurs, the platelets clump together to form a clot in the blood vessel. It can block all the blood causing a heart attack.[3-5]

Dataset [6-9]

The Z-Alizadeh Sani CAD dataset was used to perform the experiment. This is a medical record with 54 properties belonging to 303 people. This is a randomized database containing data on 216 CAD patients and 87 healthy individuals. The document is available online in the UCI repository. This information usually has four types of features: demographics, ECG, symptoms and diagnoses, tests, and echo [3]. The different features found in the data are shown in Table 1.

Table 1: Features in Z-Alizadeh Sani CAD dataset

Feature Category	Feature Name	Feature Range
Demographic Features	Weight	Between 48 and 120
	Age	Between 30 and 86
	Sex	Male, Female
	Diabetes Mellitus	Yes/No
	Ex-Smoker	
	Current Smoker	
	Hyper Tension	
	Family History	
	Body Mass Index	Between 18 and 41
	Dyslipidemia	Yes/No
	Airway Disease	
	Chronic Renal Failure	
	Cerebrovascular Accident	
	Congestive Heart Failure	
	Obesity	
	Thyroid	
Symptom and examination	Edema	Yes/No
	Systolic murmur	
	Chest Pain (Typical)	
	Atypical	
	Weak peripheral pulse	
	Exertional Chest Pain	
	Nonanginal Chest pain	
	Dyspnea	
	Lung Rales	
	Diastolic murmur	

	low Threshold angina	
	Blood Pressure	Between 90 and 190
	Function Class	1 to 4
	Pulse Rate	Between 50 and 110
ECG	ST Elevation	Yes/No
	Poor R Wave Progression	
	T inversion	
	Q Wave	
	Left Ventricular Hypertrophy	
	ST Depression	
	Rhythm	Sin, AF
Laboratory and echo	Lymphocyte	Between 7 and 60
	K (Potassium)	Between 3.0 and 6.6
	Valvular Heart Disease	Mild /Normal
	Blood Urea Nitrogen	Between 6 and 52
	Creatine	Between 0.5 and 2.2
	Low-density lipoprotein	Between 18 and 232
	Triglyceride	Between 37 and 1050
	Erythrocyte Sedimentation	Between 1 and 90
	Neutrophil	Between 32 and 89
	High-density lipoprotein	Between 15 and 111
	Hemoglobin	Between 8.9 and17.6
	Platelet	Between 25 and 742
	Fasting Blood Sugar	Between 62 and 400
	Sodium	Between 128 and 156
	Region with RWMA	0,1,2,3,4
	Ejection Fraction	Between 15 and 60
	Fasting Blood Sugar	Between 62 and 400

Proposed ET-SVMRBF Methodology [10-14]

The ET-SVMRBF method has been proposed by performing various experiments on the CAD dataset using four different methods. First, the data is preprocessed to obtain a form suitable for classification. One-bit encoding was applied to the dataset and scaled using standard scalars as part of data preprocessing. Standard scalar scale information with each feature having a mean of 0 and a standard deviation of 1 [4]. Data were not matched with the files of 216 CAD patients and 87 normal subjects.

Unequal information can affect the allocation of the model's resources. Data were synchronized using SMOTE (Synthetic Minority Oversampling Technique). Data included 55 predicted features.

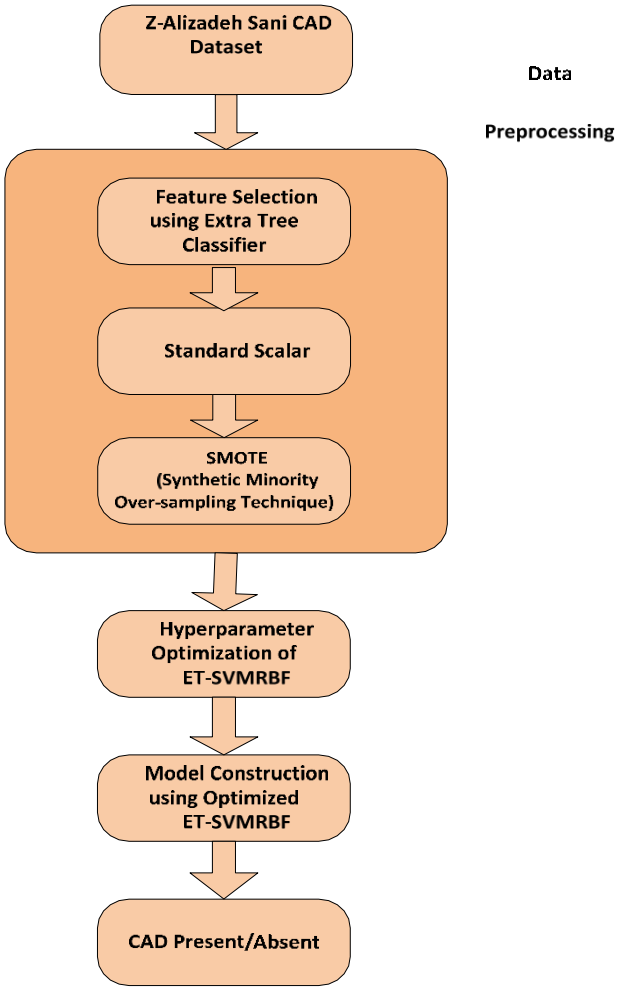

Figure 1: ET-SVMRBF methodology for CAD diagnosis

The specific process is reduced by the use of additional individual trees. It gives 36 properties. This classification reduction technique uses XGBoost (Extreme Gradient Boosting), KNN (K Nearest Neighbor), SVM-Linear (Support Vector Machine Linear), and SVM-RBF (Support Vector Machine-Radial Basis Function). SVMRBF gives the best results, so the hyperparameters of this model have been optimized using the GridSearch optimization method. Finally, the optimized model was used to diagnose CAD. The ET-SVMRBF method for CAD diagnosis is shown in Figure 1.

Feature Selection using Extra Tree [15-18]

Feature selection affects the performance of classification algorithms. Additional classification trees were used to calculate important features. This classifier provides the results of multiple decision trees for classification. To perform feature selection, multiple decision trees are created from the original training samples and each decision tree is given a different set of k features. Each decision tree selects the best features using the Gini index of each feature.

Then, the results of these multiple dissociation trees are combined to create baselines using the Gini values of each feature. Rank the features in order of importance and select the best K features for each value of K. Several experiments are done with different K values to obtain the best K. Best results are obtained with K = 36.

Performance Evaluation of ET-SVMRBF [19-23]

The performance of ET-SVMRBF was evaluated on the Z-Alizadeh Sani CAD dataset. The data contain a combination of continuous and categorical features. The distribution of constant behavior in the data is shown in Figure 2. Verify the results using the ten-fold cross-validation method.

different models were created using CAD datasets with different properties such as XGBoost, KNN, SVM-Linear, and SVM-RBF.

The performance measures of Accuracy, Sensitivity, Specificity, Precision, and F-Measure were used to evaluate the effectiveness of the study. In addition to these performances, the model was also evaluated using ROC (Receiver Operating Characteristic Curve).

The scalar method was used to bring the data to the same scale. The class assessment was performed using SMOTE to obtain equivalent data for training. First, all features of the data are used for CAD estimation and the performance of the class is evaluated. The performance of the product with the whole process is shown in Table 2.

Figure 2: Distribution of continuous attributes in the dataset

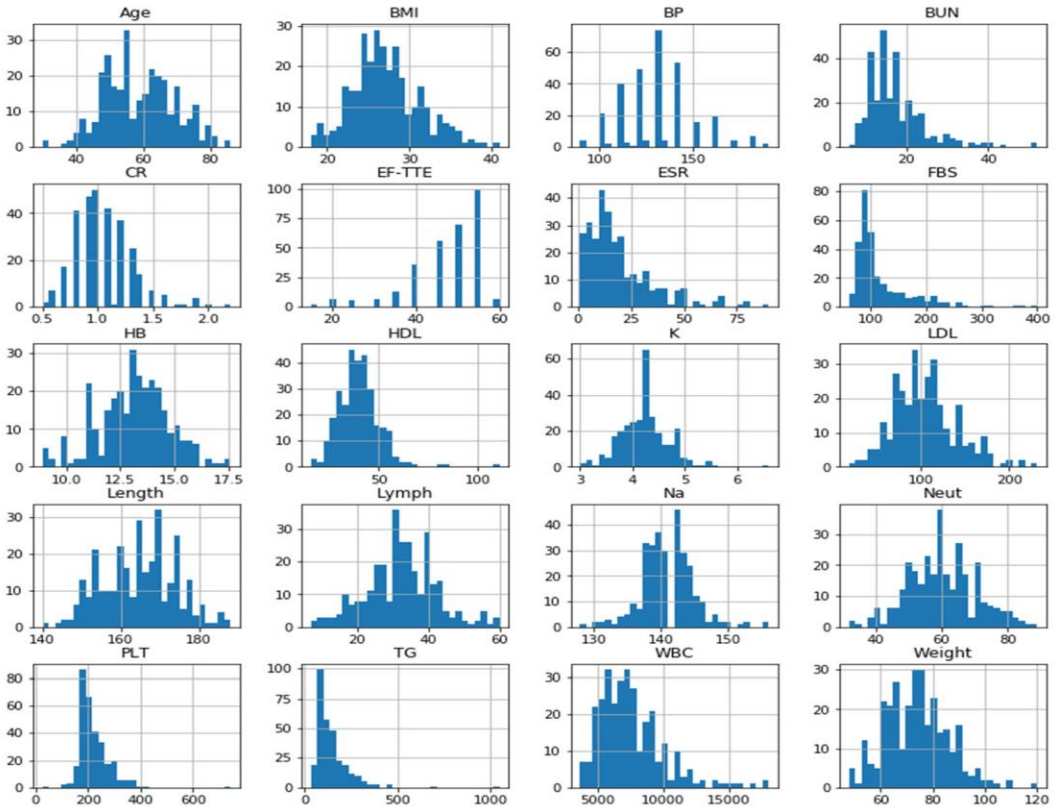

Table 2: Performance of classifiers with a complete feature set

Classifier	Accuracy	Sensitivity	Specificity	Precision	F-Measure
XGBoost	89.38	92.12	86.57	87.28	89.63
KNN	87.97	95.83	80.09	82.80	88.84
SVM-Linear	89.82	93.98	85.64	86.75	90.22
SVM-RBF	92.63	86.11	99.07	98.93	92.07

Optimization has been done to improve product performance. Model optimization is done in two steps: feature selection and hyperparameter tuning. Special selection is made with

using additional classification trees. 36 out of 55 features were selected with this method. Hyperparameter tuning was done using the GridSearch optimization method.

The performances of different products after optimization are shown in Table 3. After evaluating the performances of different products, SVM-RBF with C=1 and Gamma=0.1 is selected for the prepared method. [24-25]

Table 3: Performance of classifiers after model optimization

Classifier	Accuracy	Sensitivity	Specificity	Precision	F-Measure
ET-XGBoost	91.69	93.98	89.35	89.82	91.85
ET-KNN	87.52	92.59	82.40	84.03	88.10
ET-SVMLinear	89.82	93.51	86.11	87.06	90.17
ET-SVMRBF	95.16	93.98	96.29	96.20	95.08

The ROC of different products is shown in Figure 3. ET-SVMRBF gives the best value for AUROC (area under the ROC curve). A comparison of the performance of the different components shows that the ET-SVMRBF provides the best performance, as shown in Figures 4, 5, 6, 7, and 8.

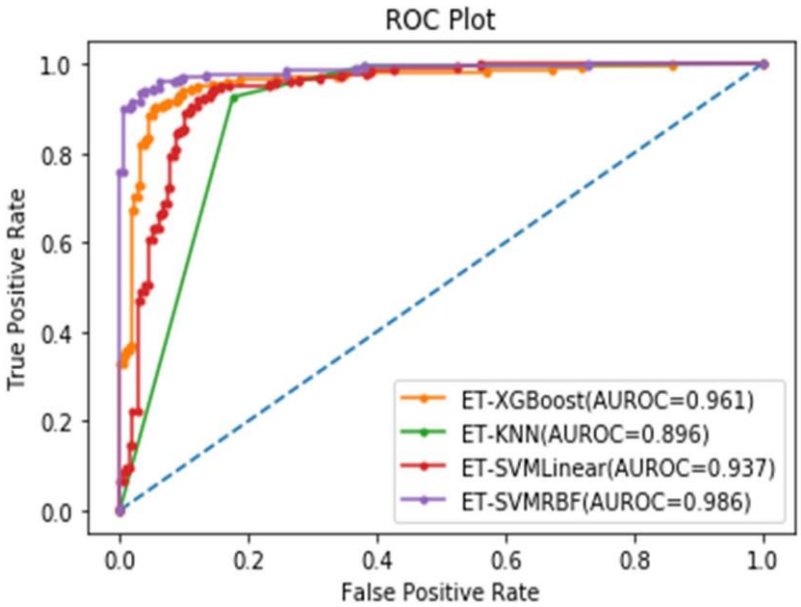

Figure 3: ROC Curve of different classifiers

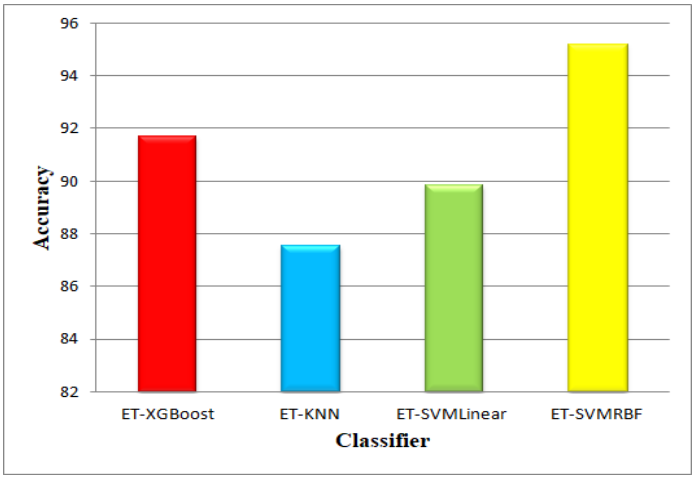

Figure 4: Comparison of accuracy of different classifiers

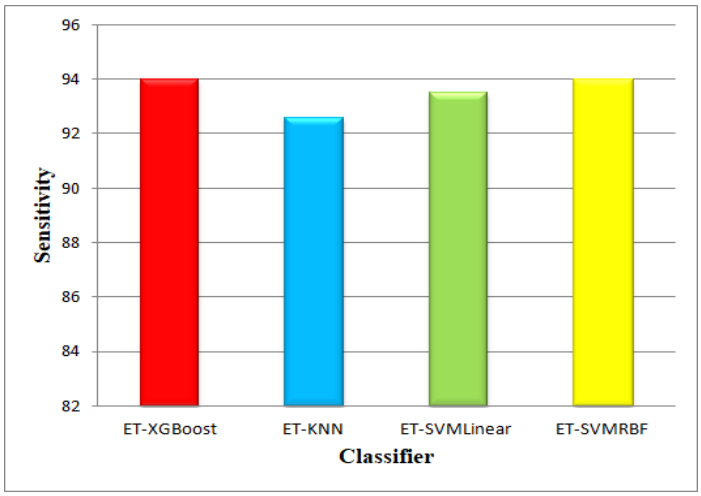

Figure 5: Comparison of the sensitivity of different classifiers

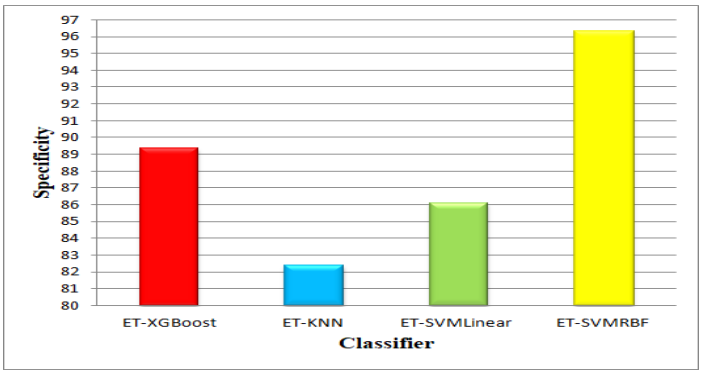

Figure 6: Comparison of specificity of different classifiers

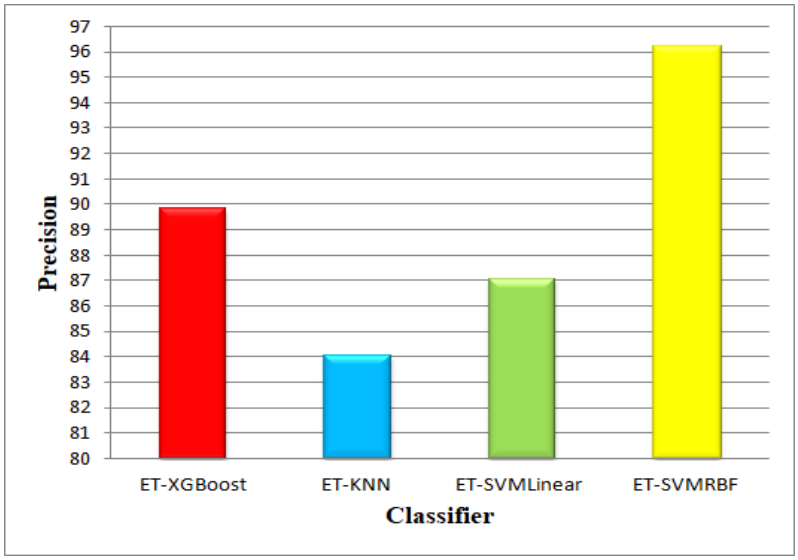

Figure 7: Comparison of precision of different classifiers

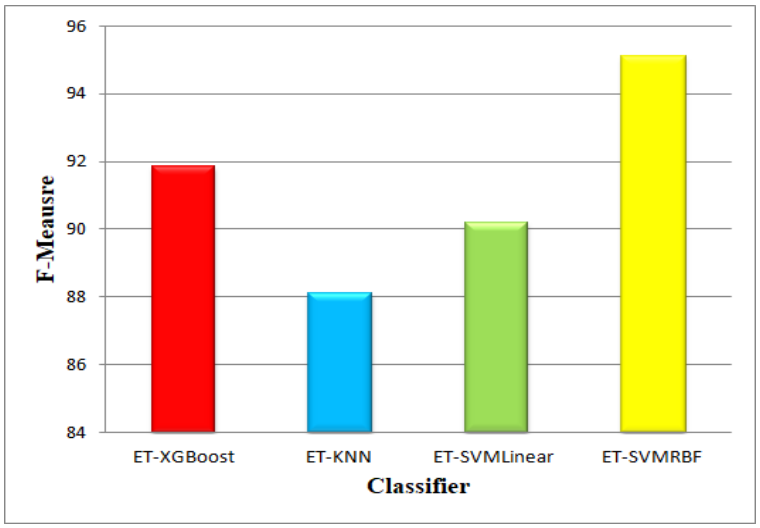

Figure 8: Comparison of F-Measure of different classifiers

Comparison of ET-SVMRBF with Existing Systems [26-32]

ET-SVMRBF has been tested in a simulation environment developed in Python. This method achieves an accuracy of 95.16%, surpassing some existing studies in the literature. In Table 4, the design results are compared with existing studies. Compared to existing systems, ET-SVMRBF improves the accuracy of CAD diagnosis. The accuracy comparison between the ET-SVMRBF and existing studies is shown in Figure 9.

Table 4: Comparison of ET-SVMRBF with the existing work

Study	Year	Dataset	Feature Selection	Classifier	Accuracy
Babic et al. [53]	2017	Z-Alizadeh Sani	Chi-square	Support vetor machine	86.67%
Qin et al. [57]	2017	Z-Alizadeh Sani	An ensemble algorithm based upon multiple feature selection	Adaptive boosting	93.70%
Arabasadi et al. [86]	2017	Z-Alizadeh Sani	Weight by SVM, Gini index, information gain, and principal component analysis.	GA optimized neural network	93.85%
Abdar et al. [67]	2019	Z-Alizadeh Sani	Genetic algorithm and particle swarm optimization	SVM	93.08%
Terrada et al. [72]	2020	Z-Alizadeh Sani	Not used	Artificial neural network	94%
Verma [73]	2020	Z-Alizadeh Sani	Not used	Stacked ensemble model	84.82%
Joloudari et al. [75]	2020	Z-Alizadeh Sani	Feature ranking	Random tree	91.47%
Valarmathi [80]	2021	Z-Alizadeh Sani	Sequential forward Selection	Random Forest	80.2%
L et al. [83]	2021	Z-Alizadeh Sani	Genetic algorithm	Random Forest	90.70%

Proposed ET-SVMRBF	2021	Z-Alizadeh Sani	Extra tree	SVM-RBF	95.16%

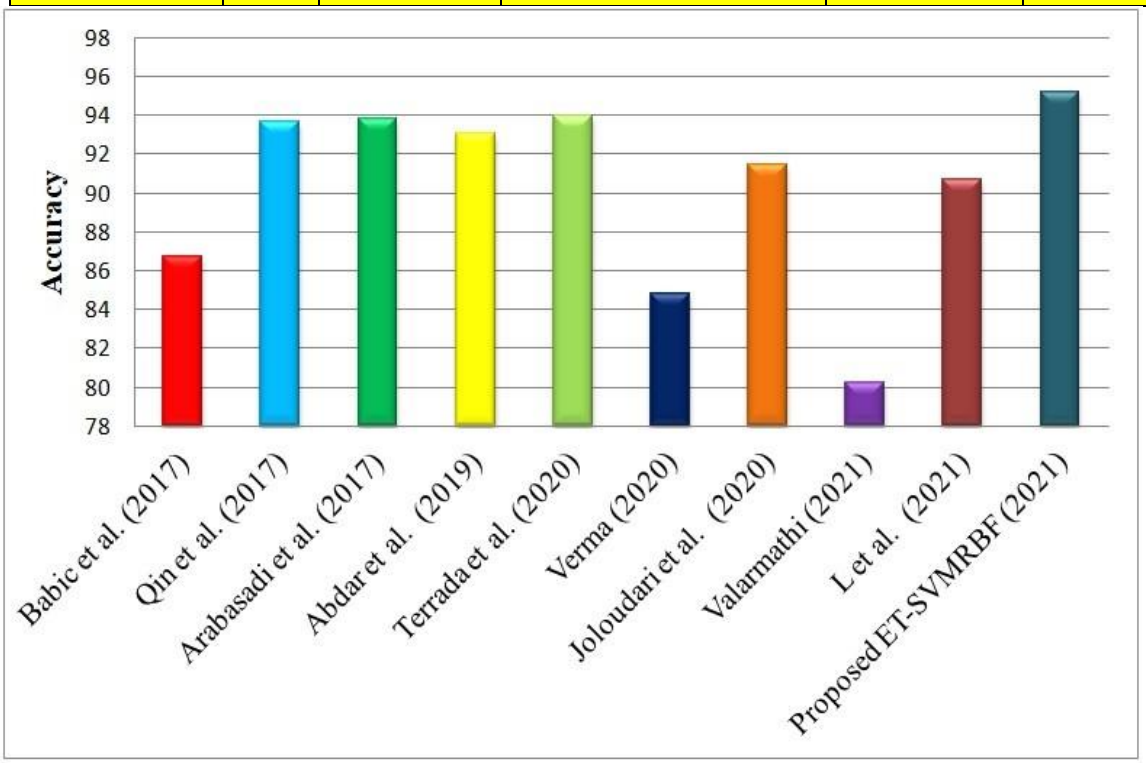

Figure 9: Comparison of ET-SVMRBF with existing work

Summary

CAD is a life-threatening condition and early diagnosis is important. In this section, ET-SVMRBF method is proposed for CAD diagnosis by combining complementary trees and SVM-RBF algorithm. The performance of the four classifiers XGBoost, KNN, SVM-Linear and SVM-RBF was evaluated on the Z-Alizadeh Sani CAD dataset and the best classifier was selected. Data preprocessing was done using scalar techniques, SMOTE and additive tree feature selection methods. Use GridSearch optimization to improve the performance of your model. The ET-SVMRBF gives an accuracy of 95.16%. The results show that this method can be used to effectively diagnose CAD based on clinical data. It is very useful in medicine. This is beneficial for countries with limited healthcare. It can be used by the doctor as a link to other techniques to increase the reliability and accuracy of the diagnosis.

References:

1. Babic, F., Olejar, J., Vantova, Z., & Paralic, J. (2017). Predictive and descriptive analysis for heart disease diagnosis. In 2017 Federated Conference on Computer Science and Information Systems (FEDCSIS) (pp. 155-163). IEEE. https://doi.org/10.15439/2017F219.
2. Dolatabadi, A. D., Khadem, S. E. Z., & Asl, B. M. (2017). Automated diagnosis of

coronary artery disease (CAD) patients using optimized SVM. Computer Methods and Programs in Biomedicine, 138, 117-126. https://doi.org/10.1016/j.cmpb.2016.10.011.

3. Kumar, S. U., & Inbarani, H. H. (2017). Neighborhood rough set-based ECG signal classification for diagnosis of cardiac diseases. Soft Computing, 21(16), 4721-4733. https://doi.org/10.1007/s00500-016-2080-7.
4. Shah, S. M. S., Batool, S., Khan, I., Ashraf, M. U., Abbas, S. H., & Hussain, S. A. (2017). Feature extraction through parallel probabilistic principal component analysis for heart disease diagnosis. Physica A: Statistical Mechanics and itsApplications, 482, 796-807. https://doi.org/10.1016/j.physa.2017.04.113.
5. Qin, C. J., Guan, Q., & Wang, X. P. (2017). Application of ensemble algorithm integrating multiple criteria feature selection in coronary heart disease detection. Biomedical Engineering: Applications, Basis and Communications, 29(06), 1750043. https://doi.org/10.4015/S1016237217500430.
6. Verma, L., Srivastava, S., & Negi, P. C. (2018). An intelligent noninvasive model for coronary artery disease detection. Complex & Intelligent Systems, 4(1), 11-18. https://doi.org/10.1007/s40747-017-0048-6.
7. Dhanaseelan, R., & Sutha, M. J. (2018). Diagnosis of coronary artery disease using an efficient hash table based closed frequent itemsets mining. Medical & Biological Engineering & Computing, 56(5), 749-759. https://doi.org/10.1007/s11517-017-1719-6.
8. David H. B. F., & Belcy S. A. (2018). Heart disease prediction using data mining techniques. ICTACT Journal on Soft Computing, 9, 1824-1830. https://doi.org/10.21917/ijsc.2018.0253.
9. Haq, A. U., Li, J. P., Memon, M. H., Nazir, S., & Sun, R. (2018). A hybrid intelligent system framework for the prediction of heart disease using machine learning algorithms. Mobile Information Systems, 2018, 3860146. https://doi.org/10.1155/2018/3860146.
10. Vijayashree, J., & Sultana, H. P. (2018). A machine learning framework for feature selection in heart disease classification using improved particle swarm optimization with support vector machine classifier. Programming and Computer Software, 44(6), 388-397. https://doi.org/10.1134/S0361768818060129.
11. Dwivedi, A. K. (2018). Performance evaluation of different machine learning techniques for prediction of heart disease. Neural Computing and Applications, 29(10), 685-693. https://doi.org/10.1007/s00521-016-2604-1.
12. Dogan, M. V., Grumbach, I. M., Michaelson, J. J., & Philibert, R. A. (2018). Integrated genetic and epigenetic prediction of coronary heart diseasein the Framingham Heart Study. PloS one, 13(1), e0190549. https://doi.org/10.1371/journal.pone.0190549.
13. Saqlain, S. M., Sher, M., Shah, F. A., Khan, I., Ashraf, M. U., Awais, M., & Ghani, A. (2019). Fisher score and Matthews correlation coefficient-based feature subset

selection for heart disease diagnosis using support vector machines. Knowledge and Information Systems, 58(1), 139-167. https://doi.org/10.1007/s10115-018-1185-y.

14. Abdar, M., Ksiazek, W., Acharya, U. R., Tan, R. S., Makarenkov, V., & Pławiak, P. (2019). A new machine learning technique for an accurate diagnosis of coronary artery disease. Computer Methods and Programs in Biomedicine, 179, 104992. https://doi.org/10.1016/j.cmpb.2019.104992.
15. Ayatollahi, H., Gholamhosseini, L., & Salehi, M. (2019). Predicting coronary artery disease: a comparison between two data mining algorithms. BMC Public Health, 19(1), 1-9. https://doi.org/10.1186/s12889-019-6721-5.
16. Latha, C. B. C., & Jeeva, S. C. (2019). Improving the accuracy of prediction of heart disease risk based on ensemble classification techniques. Informatics in Medicine Unlocked 16, 100203. https://doi.org/10.1016/j.imu.2019.100203.
17. Khennou, F., Fahim, C., Chaoui, H., & Chaoui, N. E. H. (2019). A machine learning approach: Using predictive analytics to identify and analyze high risks patients with heart disease. International Journal of Machine Learning and Computing, 9(6), 762-767. https://doi.org/10.18178/ijmlc.2019.9.6.870.
18. Almustafa, K. M. (2020). Prediction of heart disease and classifiers' sensitivity analysis. BMC Bioinformatics, 21(1), 1-18. https://doi.org/10.1186/s12859-020-03626-y.
19. Tama, B. A., Im, S., & Lee, S. (2020). Improving an Intelligent Detection System for Coronary Heart Disease Using a Two-Tier Classifier Ensemble. BioMed Research International, 2020, 1-10. https://doi.org/10.1155/2020/9816142.
20. 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.
21. 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.
22. 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 andApplications, 11(3), 540-551. http://dx.doi.org/10.14569/IJACSA.2020.0110369.
23. Joloudari, J. H., Joloudari, E. H., Saadatfar, H., GhasemiGol, M., Razavi, S. M., Mosavi, A., Nabipour, N. , S., & Nadai, L. (2020). Coronary artery disease diagnosis; ranking the significant features using a random trees model. International Journal of Environmental Research and Public Health, 17(3): 731. https://dx.doi.org/10.3390%2Fijerph17030731.
24. 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.

25. Spencer, R., Thabtah, F., Abdelhamid, N., & Thompson, M. (2020). Exploring feature selection and classification methods for predicting heart disease. Digital health, 6, 2055207620914777. https://doi.org/10.1177%2F2055207620914777.
26. Gazeloglu, C. (2020). Prediction of heart disease by classifying with featureselection and machine learning methods. Progress in Nutrition, 22(2), 660–670. https://doi.org/10.23751/pn.v22i2.9830.
27. 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.
28. 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.
29. 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.
30. 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.
31. 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.
32. 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.