IJFANS INTERNATIONAL JOURNAL OF FOOD AND NUTRITIONAL SCIENCES
ISSN PRINT 2319 1775 Online 2320 7876
Research paper UGC CARE Listed (Group -I) Journal Volume 10, Iss 1, 2021

A Hybrid Decision Support System for Heart Disease Prediction

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:

A cross-breed choice bolster framework is proposed that can help within the early discovery of heart maladies based on the clinical parameters of the persistent. Machine learning strategies have been utilized to develop this framework. Within the proposed framework, creators have not as it was taken care of the issue of lost values but too dealt with the include determination issue efficiently.

The major inspiration behind proposing this choice back framework is to create a framework, which can be utilized by any proficient individual within the nonattendance of a specialist to analyze heart infection at an early arrange. The framework depends totally on clinical information that does not require a heart pro specialist. The framework has been tried within the recreation environment created utilizing Python.

Key Words: Randomized Search, Hybrid Decision Support Systems, MICE Algorithm

Introduction

The heart is the foremost critical portion of the human body which is dependable for pumping oxygen-rich blood to other body parts through a organize of courses and veins. Any sort of clutter that influences our heart is heart infection. There are different sorts of heart maladies such as coronary supply route malady, innate heart illness, arrhythmia, etc. The understanding enduring from heart illness has different indications such as chest torment, woozy sensations, and profound sweating. Smoking, tall blood weight, diabetes, obesity, etc. are the most reasons behind heart disease.

Heart infection can be analyzed with the assistance of intrusive and non-invasive strategies.

ISSN PRINT 2319 1775 Online 2320 7876
Research paper UGC CARE Listed (Group -I) Journal Volume 10, Iss 1, 2021

Electrocardiograms, energetic electrocardiograms, phonocardiograms, and echocardiography are a few of the non-invasive strategies. Coronary angiography is the foremost broadly utilized intrusive strategy. Intrusive strategies of diagnosing the disease are more precise than non-invasive methods but these strategies are expensive and agonizing. Patients don't go for obtrusive strategies within the starting stages. Subsequently, there's a requirement for a method that can analyze heart maladies in a non-invasive way at less taking toll.[1-3]

Proposed Hybrid Heart Disease Prediction System

In this area, a cross breed heart malady forecast framework is proposed. The strategy of the proposed framework and calculations utilized to create this framework are examined.

Methodology

The half breed heart malady forecast framework comprises of three stages: information collection, information pre-processing, and show development. The Cleveland heart illness dataset gotten from the UCI (College of California, Irvine) store was utilized for performing the tests. This dataset is having 14 highlights out of which 8 are categorical highlights and 6 are numeric highlights. Target trait demonstrates whether a individual is enduring from heart illness or not. In this include, there are five conceivable values. The nonappearance of heart illness is indicated by the esteem and the diverse levels of illness are spoken to by values from 1 to 4. The dataset has 6 occasions having lost values. There are 4 lost values within the NMVCF highlight and 2 lost values within the THALM feature.4-5

In the pre-processing arrange, lost values are ascribed, highlight choice is done, highlight scaling is performed and lesson adjusting is done. Lost values are ascribed utilizing the MICE (multivariate ascription by chained equations) algorithm. After that Highlight determination is performed employing a cross breed GARFE (hereditary calculation recursive highlight end) calculation. The coefficient of all highlights is brought to the same esteem by employing a standard scalar guaranteeing that each highlight features a cruel of and a standard deviation of 1. Within the dataset, 164 occurrences are having a place to lesson 0, and 139 occurrences have a place to lesson 1. Course adjusting is performed utilizing Destroyed (manufactured minority oversampling method). It makes manufactured tests of minor classes coming about in an break even with number of tests of both classes. Classification is performed on chosen highlights utilizing NB (credulous Bayes), SVM (back vector machine), LR (calculated relapse), RF (irregular woodland), and AdaBoost (versatile boosting) classifiers. At long last, the classifier predicts whether a individual is having heart malady or not.6 The strategy of the proposed cross breed framework for heart illness expectation is appeared in Figure 1.

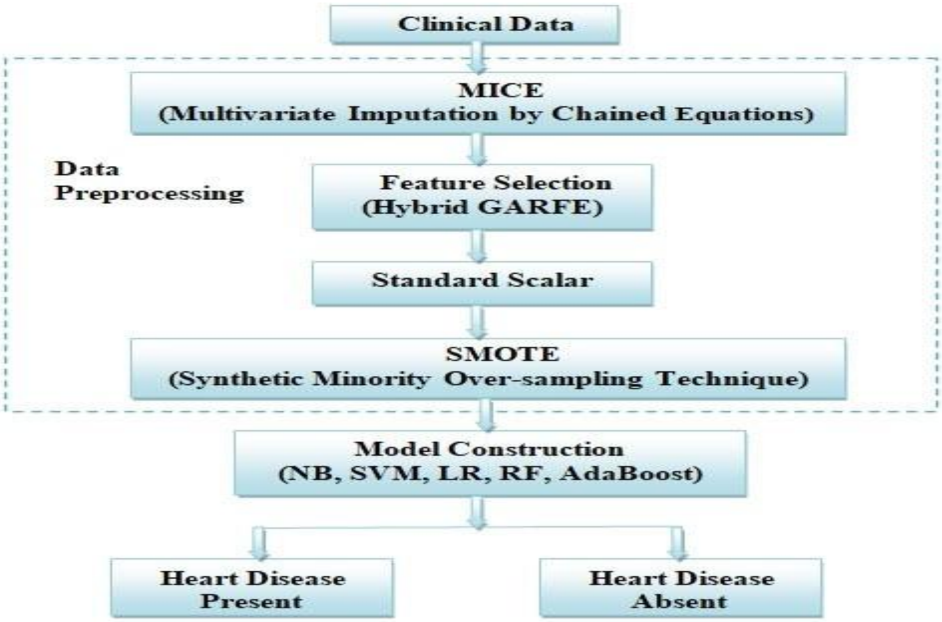
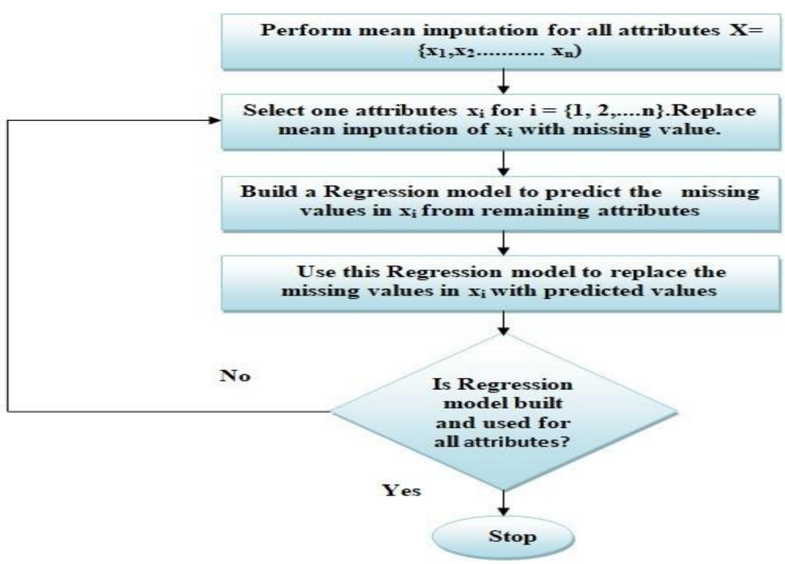


Figure 1 Proposed hybrid heart disease prediction system

Multivariate Imputation by Chained Equations Algorithm[7-9]

The MICE algorithm is used to fill in missing data in the proposed hybrid system. This algorithm uses a regression model to predict the value of the missing attribute from the remaining attributes in the dataset. Figure 2 depicts the steps of this algorithm.



ISSN PRINT 2319 1775 Online 2320 7876
 UGC CARE Listed (Group -I) Journal Volume 10, Iss 1, 2021

Figure 2: Multiple Imputation Chained Equations algorithm

ISSN PRINT 2319 1775 Online 2320 7876
 UGC CARE Listed (Group -I) Journal Volume 10, Iss 1, 2021

Performance Evaluation of the Hybrid System[10-13]

Various classification algorithms, including NB, SVM, LR, RF, and Adaboost, have been used to predict heart disease in patients. Cardiovascular disease was estimated using the various treatment options available in the Cleveland dataset. These parameters are used as a classification where class 1 means the person has the disease and class 0 means the person does not have the disease. Performance is measured in terms of accuracy, sensitivity, specificity, accuracy and F-measure. Verify results using the ten-fold cross-validation method.

Performance with All Features[14-16]

First, testing is done on all properties of the data, without using any progression or special selection. The classification performance of all methods is shown in Table 1. NB classifier gives the best performance among all methods, while SVM gives the best performance.

Table 1: Performance of classifiers on the full feature set

Classifier	Accuracy	Sensitivity	Specificity	Precision	F-Measure
NB	85.79	81.57	89.41	85.69	81.96
SVM	78.50	75.82	84.53	79.48	78.03
LR	84.80	78.13	88.80	84.81	81.68
RF	84.83	78.69	88.02	85.61	81.40
Adaboost	83.12	78.85	85.14	81.32	80.53

Performance Improvement using Scaling[17-18]

After applying the scaling preprocessing method, recheck the performance of the classification. The scalar method is used for data entry. Scaling causes changes in classifier performance. While the performance of some classifications has improved, the performance of some classifications has decreased. NB, RF and Adaboost do not cause any change in performance.

The accuracy of the SVM increased by 6.66%, while the accuracy of the LR decreased by 1.55%. Causes a 5.76% increase in sensitivity, 7. Specificity increased by 30%, precision increased by 9.10% and F-Measure increased by 7.36%. The results show that the index has a

ISSN PRINT 2319 1775 Online 2320 7876
 UGC CARE Listed (Group -I) Journal Volume 10, Iss 1, 2021

good effect on the SVM, but not on the performance of other classifications. The effect of the measurement on classification performance is shown in Table 2.

Table 2: Performance improvement using scaling

Classifier	Accuracy	Sensitivity	Specificity	Precision	F-Measure
NB	85.79	81.57	88.51	85.59	82.86
SVM	85.79	78.13	89.53	86.21	82.60
LR	83.50	79.41	85.87	82.37	80.54
RF	84.83	78.69	89.12	85.41	81.30
Adaboost	83.12	78.85	84.64	81.12	80.53

Performance Improvement using Class Balancing[19]

After scaling is applied, the performance of the classifier is further improved by scaling the classes. The SMOTE algorithm is used to evaluate the class. It produced 164 samples from class 0 and class 1. Table 5.3 shows the consequences of class balance on the performance of classes. Figures 3, 4, 5, 6, and 7 show the evolution of accuracy, sensitivity, specificity, precision, and F-Measure for classifiers with equal classes, respectively.

Table 3: Performance improvement using class balancing

Classifier	Accuracy	Sensitivity	Specificity	Precision	F-Measure
NB	84.07	83.92	88.19	87.62	85.72
SVM	85.16	82.09	88.19	87.36	85.64
LR	84.24	81.48	86.97	86.16	83.75
RF	84.85	82.70	86.97	86.35	84.48
Adaboost	83.34	78.26	84.36	85.41	82.76

ISSN PRINT 2319 1775 Online 2320 7876
  UGC CARE Listed (Group -I) Journal Volume 10, Iss 1, 2021

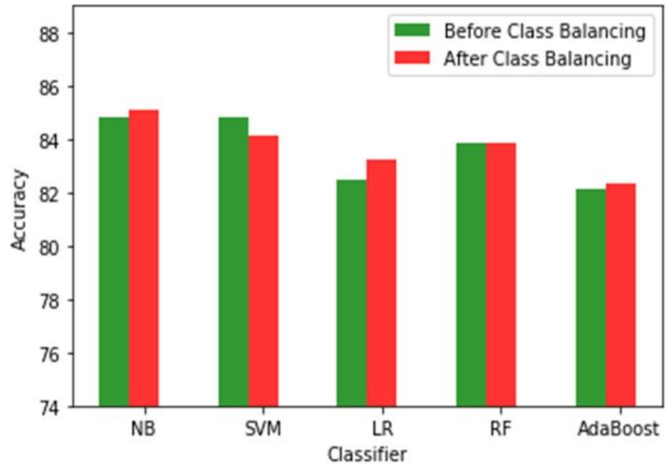


Figure 3: Increase in accuracy of classifiers using class balancing

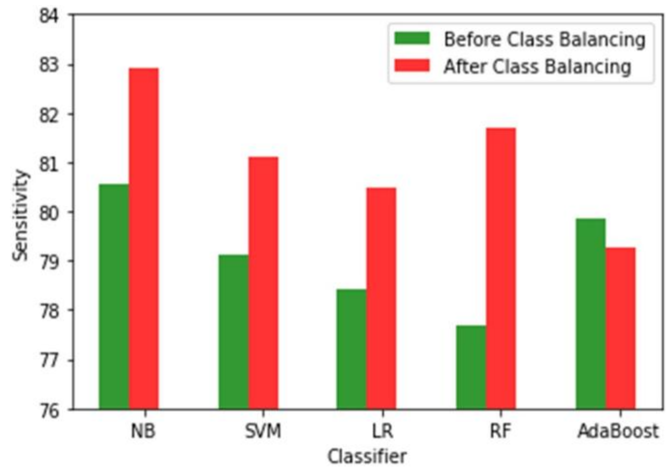


Figure 4: Increase in sensitivity of classifiers using class balancing

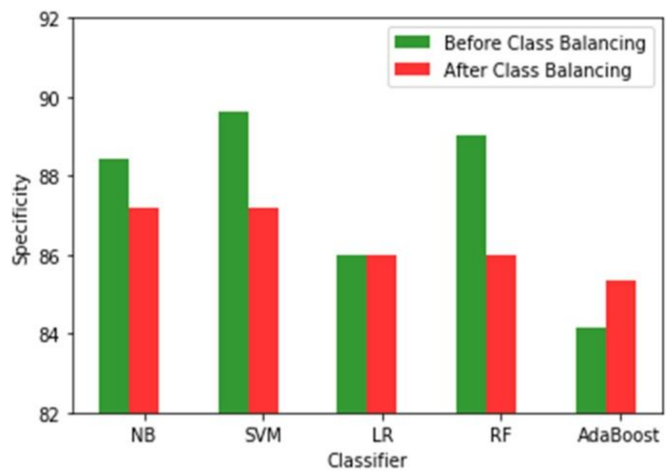


Figure 5: Increase in specificity of classifiers using class balancing

IJFANS INTERNATIONAL JOURNAL OF FOOD AND NUTRITIONAL SCIENCES
ISSN PRINT 2319 1775 Online 2320 7876
 UGC CARE Listed (Group -I) Journal Volume 10, Iss 1, 2021

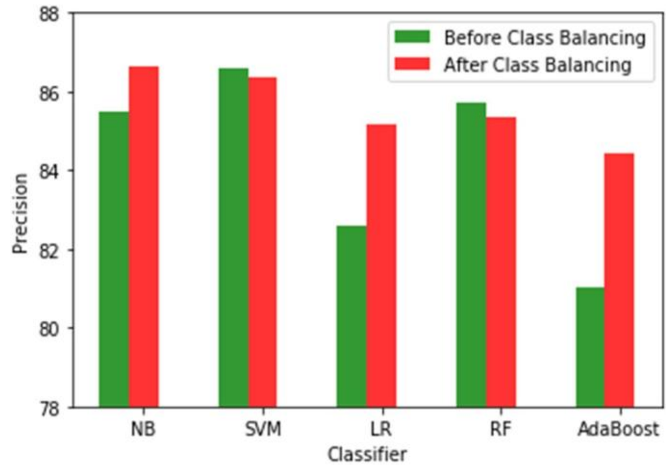


Figure 6: Increase in precision of classifiers using class balancing

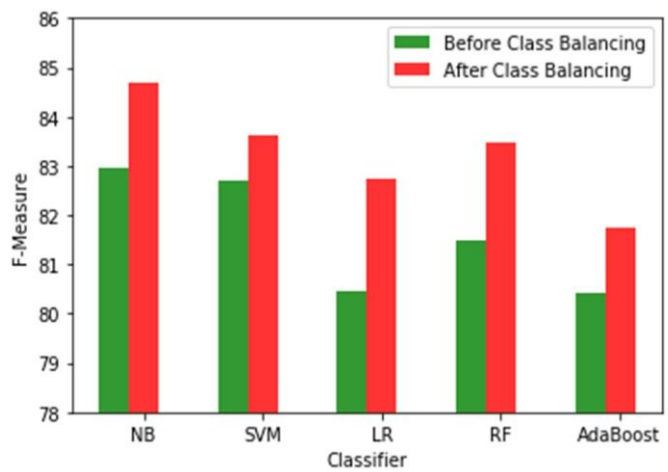


Figure 7: Increase in F-Measure of classifiers using class balancing

Performance Improvement Using Feature Selection[20-22]

The accuracy of the classifier is further enhanced by feature selection. The hybrid GARFE algorithm is used for special selection. This hybrid algorithm selects 8 features out of 13 features. As shown in Table 4, feature selection improves the performance of all classifications except NB. SVM accuracy only increased by 0. 33% LR improved by 2.19%. RF accuracy increased by 3.27%. Adaboost increases accuracy by up to 5.16%. The best results were obtained using Adaboost and RF. Adaboost increases sensitivity by 5.23%, specificity by 5%, precision by 4%, and F-Measure by 5%. 38% RF increases sensitivity from to 2.98%, specificity by 3.54%, precision by 3.64%, and F-Measure by 3. 31% Figures 8, 9, 10, 11 and 12 show the evolution of the classification system by special selection in

ISSN PRINT 2319 1775 Online 2320 7876
 UGC CARE Listed (Group -I) Journal Volume 10, Iss 1, 2021

terms of accuracy, sensitivity, specificity, precision and F-Measure, respectively.

Table 4: Performance improvement using feature selection

Classifier	Accuracy	Sensitivity	Specificity	Precision	F-Measure
NB	83.65	83.31	85.75	85.37	84.33
SVM	83.46	82.09	88.80	85.92	84.91
LR	84.07	83.92	88.19	87.62	85.73
RF	87.60	85.14	87.02	89.46	87.25
Adaboost	87.59	84.53	88.63	89.96	87.16

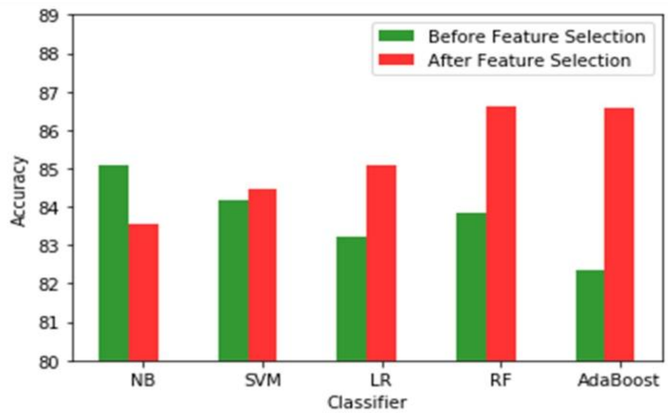


Figure 8: Increase in accuracy of classifiers using feature selection

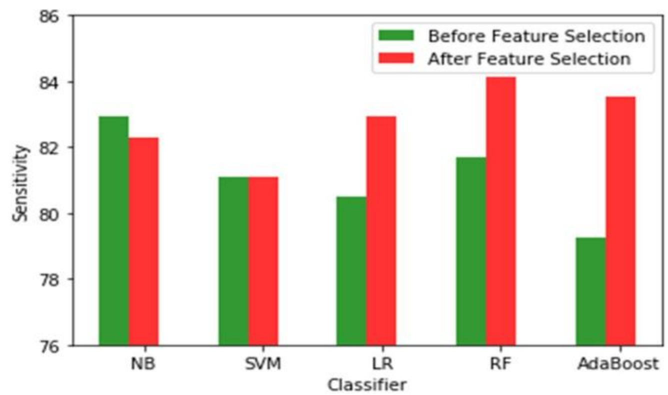


Figure 9: Increase in sensitivity of classifiers using feature selection

ISSN PRINT 2319 1775 Online 2320 7876
 UGC CARE Listed (Group -I) Journal Volume 10, Iss 1, 2021

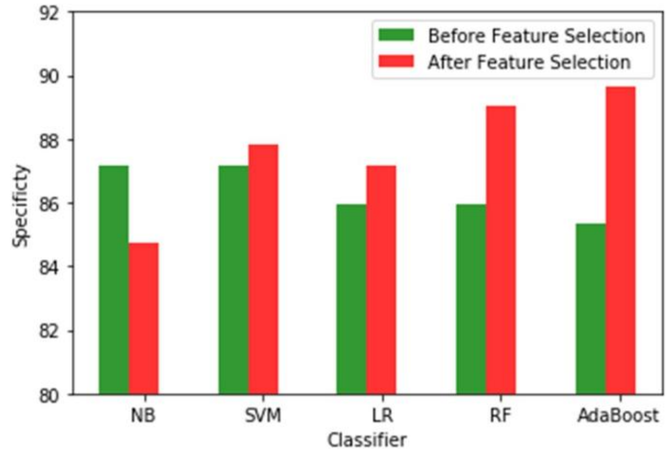


Figure 10: Increase in the specificity of classifiers using feature selection

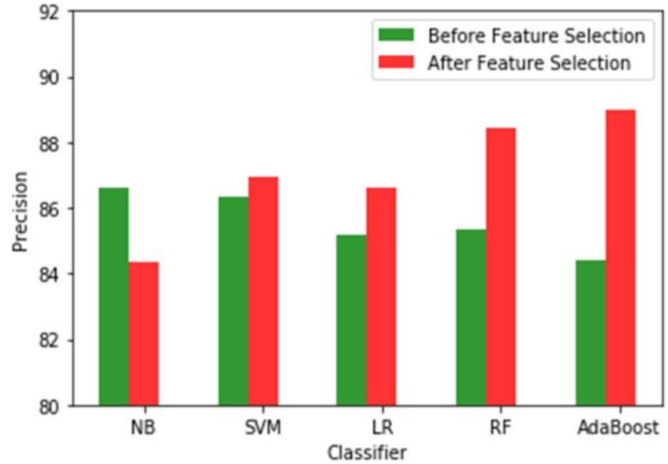


Figure 11: Increase in the precision of classifiers using feature selection

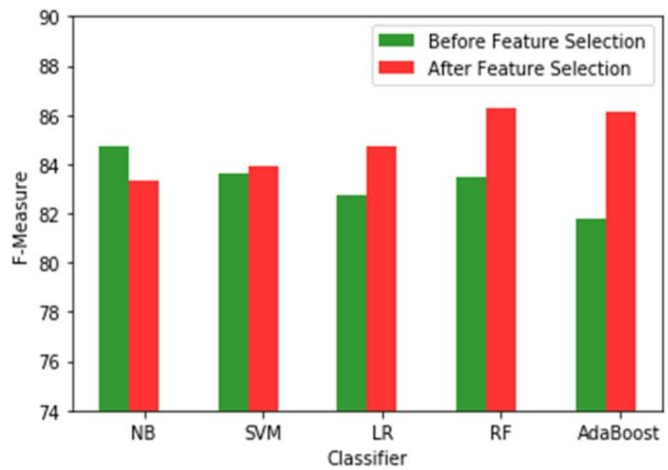


Figure 12: Increase in F-Measure of classifiers using feature selection

ISSN PRINT 2319 1775 Online 2320 7876
 UGC CARE Listed (Group -I) Journal Volume 10, Iss 1, 2021

The results show that Random Forest combined with MICE, GARFE, Scaling and SMOTE gives the best results. Size reduction using custom options helps greatly improve RF performance.

Comparison of Hybrid System with Existing Systems[23-25]

The proposed system achieved 86.60% accuracy. Table 5 shows the comparison of the proposed method with existing ones. Compared to the current system, the improved accuracy of the hybrid system is shown in Figure 13.

Table 5 Comparison of hybrid heart disease prediction system with existing systems

Study	Year	Dataset Used	Handling of Missing Values	Feature Selection Method	Classifiers	Accuracy
Jabbar et al. [49]	2016	Cleveland	Rows having the missing values were deleted	Chi-square method	Random forest classifier	83.60%
Shah et al. [56]	2017	Cleveland	Rows having the missing values were deleted	Probabilistic principal component analysis	Support vector machine	82.18%
Vijayashree and Sultana [62]	2018	Cleveland	Rows having the missing values were deleted	Improved particle swarm optimization	Random forest, naive Bayes, multilayer perceptron, support vector machine	84.36%
Saqlain et al. [65]	2018	Cleveland	Rows having missing values were deleted	Fisher score, forward and reverse feature selection	Support vector machine	81.19%

ISSN PRINT 2319 1775 Online 2320 7876
 UGC CARE Listed (Group -I) Journal Volume 10, Iss 1, 2021

Latha and Jeeva [68]	2019	Cleveland	Rows having the missing values were deleted	Features were selected by creating different feature subsets randomly	Results of naive bayes, bayes network, multilayer perceptron and random forest classifiers were combined using voting mechanism	85.48%
Gageloglu et al. [78]	2020	Cleveland	Rows having the missing values were deleted	Not applied	Support vector machine	85.14%
Verma and Mathur [73]	2020	Cleveland	Rows having the missing values were deleted	A hybrid correlation and cuckoo search method	Multilayer perceptron	85.48%
Tama et al. [71]	2020	Cleveland	Rows having the missing values were deleted	Particle swarm optimization	Two-tier ensemble method using random forest, gradient boosting, and extreme gradient boosting classifiers	85.71%
Javid et al. [74]	2020	Cleveland	Rows having the missing values were deleted	Not Applied	Voting based ensemble method	85.71

Jothi et al. [79]	2021	Cleveland	Mean and Mode Imputation	Not Mentioned	Decision tree	81%
Bahani et al. [81]	2021	Cleveland	Rows having the missing values were deleted	Not applied	Fuzzy rule - based classification system with fuzzy clustering and linguistic modifiers	83.17%
Proposed Hybrid System	2021	Cleveland	MICE	Hybrid GARFE	Random forest classifier	86.60%

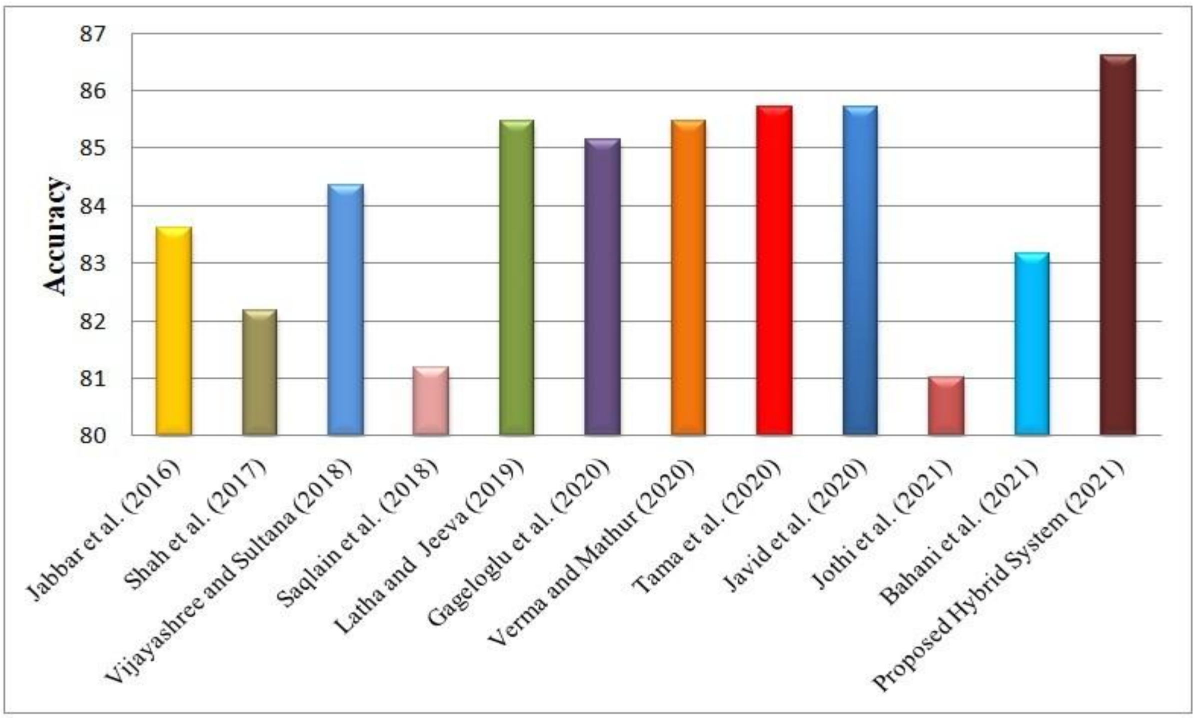


Figure 13 Comparison of accuracy of the proposed hybrid system with existing systems

Summary

The leading cause of cardiac death is late diagnosis. A hybrid heart disease prediction has been proposed. The system can help doctors detect diseases early. It can help provide patients with appropriate treatment in a timely manner. The beauty of the system is that it will rely on all medical knowledge that does not require a cardiologist.

Whether it's a worthless, feature-level option, or a separate option-level; When developing a hybrid decision-making process, the best algorithm is determined and implemented. The Cleveland dataset was used in a simulated environment created using Python to test and compare the methods. The optimization of the system is done using the CV () random search function. It performs better than other hybrid decision support methods in the literature. Random Forest gives the highest accuracy at 86.60%

References:

1. Sarker, I. H. (2021). Machine learning: Algorithms, real-world applications and research directions. SN Computer Science, 2(3), 1-21. https://doi.org/10.1007/s42979-021-00592-x.
2. Patel, S., & Patel, A. (2016). A big data revolution in health care sector: Opportunities, challenges and technological advancements. International Journal of Information Sciences and Techniques, 6(1/2), 155-162. 10.5121/ijist.2016.6216.
3. 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.
4. 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.
5. Masetic, Z., & Subasi, A. (2016). Congestive heart failure detection using random forest classifier. Computer Methods and Programs in Biomedicine, 130, 54-64. https://doi.org/10.1016/j.cmpb.2016.03.020.
6. Zhenya, Q., & Zhang, Z. (2021). A hybrid cost-sensitive ensemble for heart disease prediction. BMC Medical Informatics and Decision Making, 21(1), 1-18. https://doi.org/10.1186/s12911-021-01436-7.

7. 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.
8. Alizadehsani, R., Habibi, J., Hosseini, M. J., Mashayekhi, H., Boghrati, R., Ghandeharioun, A., Bahadorian, B. & Sani, Z. A. (2013). A data mining approachfor diagnosis of coronary artery disease. Computer Methods and Programs in Biomedicine, 111(1), 52-61. https://doi.org/10.1016/j.cmpb.2013.03.004.
9. Karunathilake, S. P., & Ganegoda, G. U. (2018). Secondary prevention of cardiovascular diseases and application of technology for early diagnosis. BioMed research international, 2018. https://doi.org/10.1155/2018/5767864.
10. Olin, J. W., White, C. J., Armstrong, E. J., Kadian-Dodov, D., & Hiatt, W. R. (2016). Peripheral artery disease: evolving role of exercise, medical therapy, and endovascular options. Journal of the American College of Cardiology, 67(11), 1338-1357. https://doi.org/10.1016/j.jacc.2015.12.049.
11. Yeboah, K., Puplampu, P., Ainuson, J., Akpalu, J., Gyan, B., & Amoah, A. G. (2016). Peripheral artery disease and exertional leg symptoms in diabetes patients in Ghana. BMC Cardiovascular Disorders, 16(1), 1-9. https://doi.org/10.1186/s12872-016-0247-x.
12. Setty, H. S. S. N., Patil, S. S. G., Ramegowda, R. T., Vijaykumar, V. I., & Manjunath, C. (2017). Comprehensive approach to congenital heart defects. Journal of Cardiovascular Disease Research, 8(1), 1-5.10.5530/jcdr.2017.1.1.
13. Inamdar, A. A., & Inamdar, A. C. (2016). Heart failure: diagnosis, management and utilization. Journal of Clinical Medicine, 5(7), 62. https://dx.doi.org/10.3390%2Fjcm5070062.
14. Yancy, C. W., Jessup, M., Bozkurt, B., Butler, J., Casey, D. E., Drazner, M. H., Fonarow, G.C., Geraci, S. A., Horwich T., Januzzi J. L., Johnson M. R., Kasper E.K., Levy W. C., Masoudi F. A., McBride P. E. , McMurray J. J., Mitchell J. E., Peterson P. N., Riegel B., Sam F., Stevenson L. W., Tang W. H., Tsai E. J., & Wilkoff, B. L. (2013). 2013 ACCF/AHA guideline for the management of heart failure: a report of the American College of Cardiology Foundation/American Heart Association Task Force on Practice Guidelines. Journal of the American College of Cardiology, 62(16), e147-e239. https://doi.org/10.1016/j.jacc.2013.05.019.

ISSN PRINT 2319 1775 Online 2320 7876
 UGC CARE Listed (Group -I) Journal Volume 10, Iss 1, 2021

15. Dhinakaran S., Tamilanban T., & Chitra V. (2019). Arrhythmia's: types, pathophysiology and therapy: a review. International Research Journal of Pharmacy, 10(2), 23-30. http://dx.doi.org/10.7897/2230-8407.100236.
16. 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.
17. Amadi, C. E., Lawal, F. O., Mbakwem, A. C., Ajuluchukwu, J. N., & Oke, D. A. (2018). Knowledge of cardiovascular disease risk factors and practice of primary prevention of cardiovascular disease by Community Pharmacists in Nigeria: a cross-sectional study. International Journal of Clinical Pharmacy, 40(6), 1587- 1595. https://doi.org/10.1007/s11096-018-0744-3.
18. Hasan, M. S., Ebrahim, Z., Mahmood, W. H. W., & Ab Rahman, M. N. (2017). Decision support system classification and its application in the manufacturing sector: A review. Jurnal Teknologi, 79(1). https://doi.org/10.11113/jt.v79.7689.
19. Bonczek, R. H., Holsapple, C. W., & Whinston, A. B. (2014). Foundations of decision support systems. Academic Press, 17-23.
20. Power, D. J. (2002). Decision support systems: concepts and resources for managers. Greenwood Publishing Group, 9-10.
21. Jiang, Y., Qiu, B., Xu, C., & Li, C. (2017). The research of clinical decision support system based on three-layer knowledge base model. Journal of Healthcare Engineering, 2017. https://doi.org/10.1155/2017/6535286.
22. Sebaa, A., Nouicer, A., Tari, A., Tarik, R., & Abdellah, O. (2017). Decision support system for health care resources allocation. Electronic physician, 9(6), 4661. https://dx.doi.org/10.19082%2F4661.
23. Kumar, N., Narayan Das, N., Gupta, D., Gupta, K., & Bindra, J. (2021). Efficient automated disease diagnosis using machine learning models. Journal of Healthcare Engineering, 2021. https://doi.org/10.1155/2021/9983652.
24. Jayatilake, S. M. D. A. C., & Ganegoda, G. U. (2021). Involvement of machine learning tools in healthcare decision making. Journal of Healthcare Engineering, 2021. https://doi.org/10.1155/2021/6679512.
25. 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.