

Classification and Prediction of Heart Disease using Novel Random Forest Algorithm by Comparing Logistic Regression for Obtaining Better Accuracy

T Poojitha, Mahaveerakannan R

Department of Computer Science and Engineering, Saveetha School of Engineering, Saveetha Institute of Medical and Technical Sciences, Saveetha University, Chennai, Tamil Nadu, India, Pincode:602105.

ABSTRACT

Aim: Heart attacks are usually caused due to blockages, partially or completely, of the heart's veins or arteries that constrict the flow of blood from or to the heart. The primary objective of this review aims to be seen as the most appropriate algorithm to give us the ideal prediction. We will be comparing the novel Random forest with Logistic regression to find out which of these can give us the best accuracy. **Material and Methods:** The study used 143 samples with novel Random Forest and Logistic Regression is executed with varying training and testing splits for foreseeing the accuracy of coronary disease prediction with the 80% of G-power value and heart disease data were gathered from multiple web sources, including latest The study's findings and criterion were 0.05%, with a 95% probability value, average, and confidence interval. The performance accuracy rate of the classifiers is used to evaluate the coronary disease dataset. There was a statistically significant value test between the novel Random Forest and Logistic Regression is 0.046 ($p < 0.05$). **Results and Discussion:** The accuracy of predicting coronary disease in the novel Random Forest 90.16 % and Logistic Regression 85.25 % is obtained. **Conclusion:** This study concludes that the Prediction of Coronary disease using the novel Random Forest (RF) algorithm looks to be fundamentally superior to the Logistic Regression (LR) with increased precision.

Keywords

Coronary Disease, Novel Random Forest Algorithm, Logistic Regression, Machine Learning, Classification, Prediction.

Imprint

T Poojitha, Mahaveerakannan R. Classification and Prediction of Heart Disease using Novel Random Forest Algorithm by Comparing

Logistic Regression for Obtaining Better Accuracy. *Cardiometry*; Special issue No. 25; December 2022; p. 1538-1545; DOI: 10.18137/cardiometry.2022.25.15381545; Available from: <http://www.cardiometry.net/issues/no25-december-2022/classification-prediction-heart-disease>

INTRODUCTION:

The prediction of Coronary Disease and to increase the precision of predicting the disease using novel Random Forest and comparing it with Logistic Regression. In the course of the last 10 years, coronary disease or cardiovascular sickness remains the essential premise of death around the world (Steiner et al. 2021). An estimation by means of the World wellbeing association, that over 17.9 million passouts happen each overall year due to cardiovascular disorders, mainly 80% of the passouts are only in view of coronary vein infection also, cerebral stroke (Woodruff et al. 2021). Estimating an individual's chances of adequate blood flow to the heart is one of the significant issues of cardiovascular disease. The prediction technique might be built by applying a number of longitudinal review study autonomous-regression analyses (Sharma 2021). The main motivation of this paper is to focus on diagnostic tests and see some of the daily health habits that contribute to cardiovascular disease (CVD). Most importantly, the different machine learning algorithms are compared using intelligent optimization algorithms. Manual classification data is used for healthy or unhealthy (Tithi et al. 2019). Data science plays a fundamental role in handling a gigantic information measurement in the sphere of medical services. Because cardiovascular disease is a perplexing task, there is a need to automate the schedule phase in minimizing any problems and to advise the patient far ahead of time (Rajdhan et al. 2020a). Finding and predicting Cardiovascular infection (CVD) requires more accuracy, flawlessness, and rightness on the grounds that a little mix-up can cause weariness issues or the death of the individual. The significance of this exploratory study is to bring coronary illness forecasts utilizing AI that can assist with recognizing the clinical information, for example, hypertension, diabetes, cigarettes smoked each day, etc is taken as information and afterward, these elements are modeled for prediction (Maranhao Neto et al. 2021). At the same time, we can apply this research in clinical field correctness and in some complicated

tasks like fatigue problems. Our team has extensive knowledge and research experience that has translated into high quality publications (Chellapa et al. 2020; Lavanya, Kannan, and Arivalagan 2021; Raj R, D, and S 2020; Shilpa-Jain et al. 2021; Sudha S, R, and P 2021; Ramadoss, Padmanaban, and Subramanian 2022; Wu et al. 2020; Kalidoss, Umapathy, and Rani Thirunavukkarasu 2021; Kaja et al. 2020; Antink et al. 2020; Paul et al. 2020; Malaikolundhan et al. 2020)

A ton of examinations has been finished based on the classification of Hearts utilizing Machine Learning Algorithms. More than 120 articles were published in Science Direct and nearly 140 articles were found in google scholar. Related to work on coronary disease prediction using machine learning the authors obtained better and more accurate results with the application of KNN accuracy of 97.3% (Nambi 2018). Coronary illness prediction and classification utilizing AI calculations enhanced by molecule swarm streamlining and subterranean insect province improvement, the authors achieved better and more accurate results having 99.65% (J. Singh 2017). Comparing different algorithms of novel Random Forest arrangement looking for better execution in coronary illness determination, WEKA (R. Singh and Rajesh 2019), classification system using Logistic model tree algorithm and Novel Random Forest Algorithm, the authors gained accurate results using different techniques (Wang et al. 2021). A half and half savvy framework system for the forecast of coronary illness utilizing AI calculations (Westerlund et al. 2021), the intrusive based methods to the diagnosing of the coronary illness depend on concerning symptoms to address these intricacies painless clinical choice emotionally supportive network comprises of the fake neural organization (ANN), Logistic relapse (LR), AdaBoost (AB), support vector machine (SVM), Fuzzy Logic (FL), Naive Bayes (NB), and unpleasant set [9,10] has been created has been created by the different researches and broadly utilized for coronary illness conclusion (Baig 2020). The experimental results indicate that the novel Random Forest yields 89.14% in terms of the accuracy for Resting ECG and CP (chest pain) of the datasets. This scientific study has been presented and published in over 55 indexed journals. (Rajdhan et al. 2020b).

The existing research study predicts the classification using probability but it does not group the attributes (Qi and Zhang, n.d.). Another research paper has developed the performance of the order and its dataset

will yield a limit of 54.30% of precision utilizing AI calculations and this will make an exploration hole to foresee the reliable and sensible ways of managing making an early finding to achieve brief organization of the infection (Chatterjee and Chandran 2019). The aim of this research is to utilize the original Random Forest (RF) algorithm to group the best arrangement of value administrations and compelling determination of anticipating coronary illness at an early stage by achieving better accuracy.

MATERIALS AND METHODS:

This exploration study was done in the Artificial Intelligence Research center at the Department of Computer Science and Engineering, Saveetha School of Engineering, Saveetha Institute of Medical and Technical Sciences, Chennai. The innovative Random Forest (RF) method and the Logistic Regression (LR) technique are used in this research project. The G-power programme version 3.1.10 was used to forecast each scale factor, yielding 286 total number of observations and 143 number of respondents per group with 80% G power values, a minimum significant value of 0.05, and a probability value of 95%. (Rajdhan et al. 2020b).

This research makes use of a dataset obtained from the kaggle open-access dataset. One of the most popular online networks for information researchers and artificial intelligence professionals. It allows clients to search for and discover the datasets they need, as well as give a customisable Jupyter notebook environment with free GPU and 6 hours of runtime. The dataset consists of 76 attributes (Westerlund et al. 2021). The attributes are age, sex, chest torment type (4 sorts), resting circulatory strain, serum cholesterol in mg/dl, fasting glucose > 120 mg/dl, most extreme pulses accomplished, and 6 others. The dataset has 14 columns and 304 rows containing data on symptoms associated with coronary disease, including duplicate, null, and missing values. Pre-processing of the datasets is done to remove the duplicate and null values with the help of functions that are provided by Microsoft excel. The dataset that is collected consists of all numerical forms; there is no need to convert the data. The unwanted attribute such as age, exacting attributes that are not necessary for the prediction is removed (Saranya S et al. 2020).

RF falls within the umbrella of adaptive boosting. RFs may be utilized for effectiveness as well as foresight. For implementing an RF, the random state function is used to create forest-like structures and com-

bine them. The overfitting of data is done with help of calculating lowered, this is accomplished by dividing the dataset and assessing all splits by creating the rf. Predict.

The RF method uses the following pseudocode to apply a coronary disease dataset. The datasets will be fed into the pseudocode, and the final result will be passed through the accuracy and classification parameters.

Pseudocode of the Novel Random Forest Algorithm

```
from sklearn.ensemble import RandomForestClassifier

max_accuracy = 0

for x in range(2000):
    rf = RandomForestClassifier(random_state=x)
    rf.fit(X_train,Y_train)
    Y_pred_rf = rf.predict(X_test)
    current_accuracy = round(accuracy_score(Y_pred_rf,Y_test)*100,2)
    if(current_accuracy>max_accuracy):
        max_accuracy = current_accuracy
        best_x = x

print(max_accuracy)
print(best_x)

rf = RandomForestClassifier(random_state=best_x)
rf.fit(X_train,Y_train)
Y_pred_rf = rf.predict(X_test)
```

Innovative Random Forest Algorithm is extensively used for both recurrence and identification issues that must be addressed. This algorithm can be used in various places such as banking, prediction works, health, stock markets, artificial intelligence, etc. To get more accurate and stable estimates the novel random forest will create a forest-like structure and combine them. Subsets from both datasets and attributes are selected arbitrarily and get trained. using this method overfitting of data can be lowered. This algorithm takes lower training time than many other algorithms on large datasets with maintaining the precision of the accuracy when a huge part of data is not present.

Pseudocode of the Logistic Regression algorithm

```
from sklearn.linear_model import LogisticRegression

lr = LogisticRegression()

lr.fit(X_train,Y_train)

Y_pred_lr = lr.predict(X_test)
```

Logistic regression is essentially an administered grouping computation. In a representation problem, the objective variable (or yield), y, might have essen-

tially distinct features for a particular situation approach of features(or inputs), X. Contrary to standard reasoning, vital backslide is a backslide model. The model constructs a backslide technique to forecast the possibility that a given data segment will have a place in the class labelled «1.» When a decision edge is introduced, it converts into a request methodology. The edge's positioning regard is an indispensable piece of Logistic backslide and is dependent upon the course of action issue itself. The decision for the value of the edge regard is fundamentally affected by the potential gains of precision and survey.

Statistical Analysis

The statistical analysis is applied using For each group, descriptive statistical analysis such as mean, standard deviation, and standard error were performed using the Statistical Package for Social Sciences (SPSS). The test was performed to analyze factors across the review gatherings. Hence the independent variable in this study is mean, entropy, variance, and contrast (Woodruff et al. 2021). The dependent variables are fever, heart rate, bp, cholesterol. The independent sample t-test is performed to compare the accuracy performance of both Group-1 and Group-2 using the RF and the LR algorithms.

RESULTS

The RF and LR learning algorithms are compared with 10 examples using various 70% training and 30% evaluation samples with varied amounts of entries in the dataset, and the conclusions are displayed in Table 1. The dataset contains 304 rows, and the accuracy of the RF and LR algorithms is determined for 10 samples (iterations). It is observed that the RF algorithm is significantly better than the LR algorithm. The preprocessing is used to clean the data to apply for the machine learning algorithm, due to the original dataset containing more errors and null values, it has been preprocessed. The dataset has been resized and made ready for the next stage, as the network allows. The algorithm for preprocessing the dataset can be implemented using Spyder functions.

From Table 1, the statistical analysis of the Novel RF algorithm and LR algorithm. For 10 iterations, the mean accuracy value, standard deviation, and accuracy for Novel RF and LR algorithms are obtained. The RF algorithm performed 90.16% better than the LR method (85.25%).

Table 1

Correlation of the RF and LR techniques with N=10 examples of the dataset and the greatest performance of 90.16% and 85.25% in the samples (when N=1) using dataset size=304 and 70% training and 30% testing data.

Sample (N)	Dataset size / rows	RF Accuracy in %	LR Accuracy in %
1	304	90.16	85.25
2	285	89.79	84.10
3	260	89.50	83.85
4	235	89.15	83.35
5	206	88.68	82.90
6	189	88.50	82.12
7	150	87.60	81.79
8	115	86.45	81.55
9	100	85.25	81.12
10	86	84.10	78.36

Table 2

Statistical outcomes of the RF and LR techniques. For RF and LR algorithms, the mean accuracy value, standard deviation, and standard error mean are calculated after 10 iterations. The mean of the RF 89.14% algorithm fared better than the LR 84.12% method..

Algorithms (Accuracy)	Sample (N)	Mean	Std Deviation	Std error mean
Random Forest	10	89.14	2.9654	0.93775
Logistic Regression	10	84.12	3.7039	0.17128

In Table 2 we observed after performing statistical analysis of the Novel RF algorithm and the LR algorithm. Mean accuracy value, the standard deviation is observed that the RF algorithm performed better than the LR algorithm. The Novel RF algorithm obtained 2.9654 standard deviations with 0.93775 standard error while the LR algorithm obtained 3.7039 standard deviations with 0.17128 standard error. After comparing the 10 accuracies of both the algorithms, a random forest algorithm was obtained with the mean of 89.14% and logistic regression obtained with the mean

of 84.12. Also, the independent sample t-test was utilized to look at the exactness of two calculations and a genuinely tremendous distinction was seen with the RF model having 90.16% precision and LR has 85.25% of exactness. When contrasted with the other calculation's exhibition, the proposed RF classifier is fundamentally better compared to the SVM algorithm.

Table 3 shows the separate sample t-test with relevance values and standard error. The «p» value was less than 0.05, which is deemed clinically meaningful, and 95% confidence intervals were produced. The independent sample t-test was utilized to think about the precision of two calculations and a genuinely massive distinction was taken, note is 0.046 ($p < 0.05$) The Novel RF model obtained 90.16% accuracy. When compared to the performance of the other strategies, the suggested Novel RF outperforms LR substantially.

The symptom shown here is the remainder of the ECG comparison as seen in Fig. The red individuals are at danger because of this one attribute, whereas the blue people are not at risk because of this one attribute; the values range from 0 to 300.

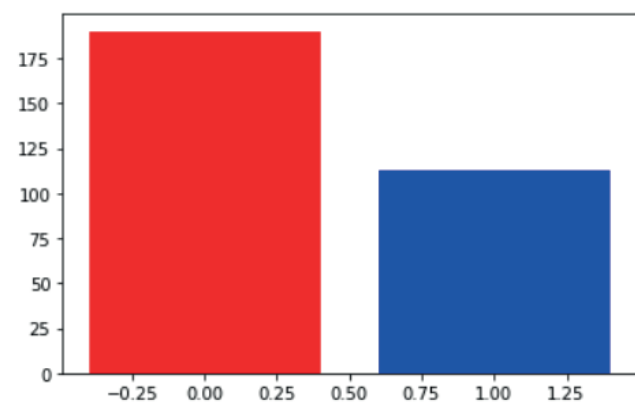


Fig. 1. Based on input values ranging from 0 to 300, the red individuals are at risk of cardiovascular breakdown, while the blue people are not at risk.

From Figure 2 Comparison of the chest pain between people of different age groups having the pain

Table 3

The significance level Novel RF and LR algorithms are significant in the Independent sample t-test ($p = 0.046$). As a result, with a 95% confidence interval, both the Novel RF and the LR algorithms have a significance level smaller than 0.05.

Accuracy	Levene's test for equality of variances		T-test for Equality of means					95% Confidence Interval of the Difference	
	F	Sig	t	df	Sig (2-tailed)	Mean difference	Std Error difference	Lower	Upper
Equal Variances Assumed	.748	.046	3.348	18	0.004	5.0230	1.5004	1.8707	8.1752
Equal Variances Not assumed	-	-	3.348	17.178	0.004	5.0230	1.5004	1.8598	8.1861

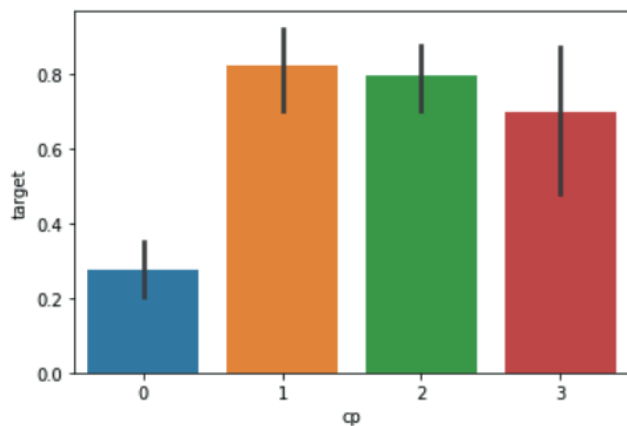


Fig. 2. Comparison of the chest pain between people of different age groups having the pain in the heart using target attribute, the value ranges from 0.0 to 0.8.

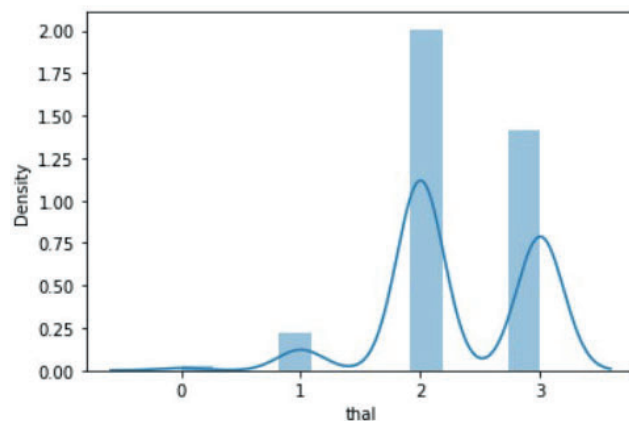


Fig.3. Comparison of attributes between thal and density to determine the chest pain intensity the values range between 0.00 to 2.00.

in the heart using target attribute, the values range from 0.0 to 0.8.

From Figure 3 Comparison of attributes between thal and density to determine the chest pain intensity the values range between 0.00 to 2.00.

From Figure 4 the mean accuracy of the two algorithms has been shown and compared the mean error and the graph shows that Novel RF is better than the LR algorithm.

The Novel RF model obtained 90.16% and the LR algorithm already predicted 85.25%. performance value. Finally, as shown in Fig. 4, we compared the mean error of the Novel RF and LR algorithms. As a result, the Novel RF method outperforms the LR algorithm substantially.

DISCUSSION

Novel random forest as convincingly appears better than Logistic Regression with improved accuracy. The novel Random Forest classifier shows some When compared to Logistic Regression, There is a distinction in terms of effectiveness, scoring rate, and overall efficiency. Data mining is crucial in the health-care business for new illness identification. A variety of tests must be performed on the patient in order to diagnose an ailment. On the other hand, using data mining techniques can minimize the number of tests necessary.

To Support this work, a Comparison between each algorithm's performance in terms of accuracy, and loss proved that Random Forest is the best overall. The

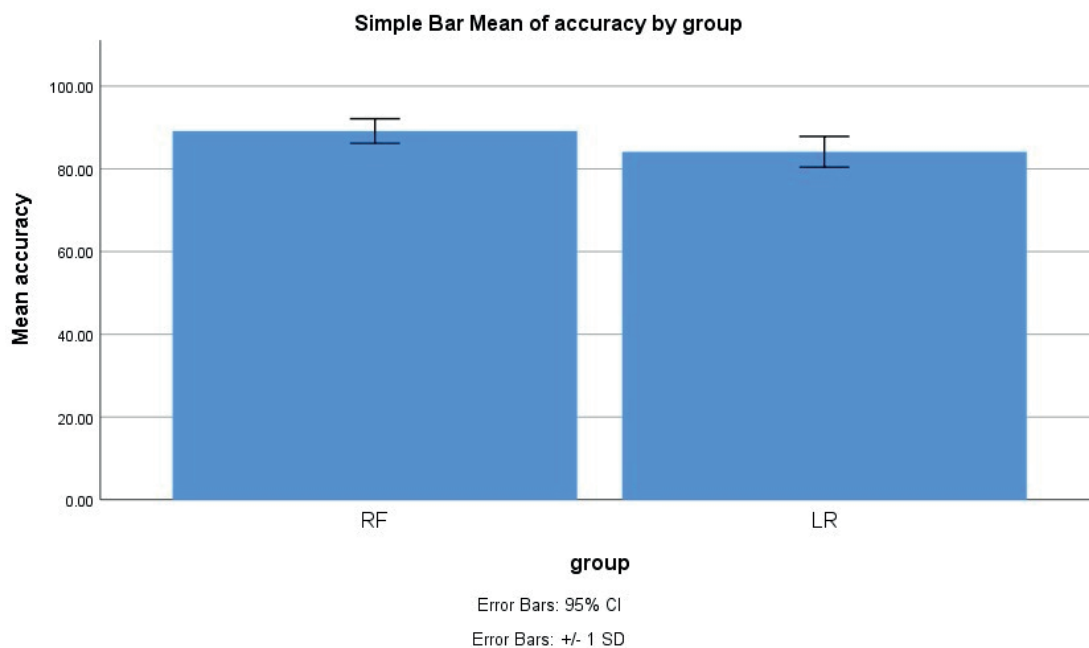


Fig. 4. Comparison of RF and LR algorithms with regards to precision. The mean precision of the RF algorithm is better than LR and the standard deviation of RF is slightly better than the LR algorithm. X-Axis: RF vs LR algorithm Y-Axis: Mean accuracy of detection \pm 1 SD.

outcome of this paper is similar to (Shah, Patel, and Bharti 2020) where they found that the Novel Random Forest Algorithm shows better performance with 89% accuracy than Logistic Regression. Hence the novel Random Forest can be used to predict heart disease efficiently (Zhang et al. 2021). The Positive other finding backup, the experiment was carried out using regressive testing of all possible variations that are there for the given models (*Risk Prediction in Cardiovascular Disease-Current Status and Future Challenges* 2009; Kaur, Sharma, and Sharma 2019). Other study findings, in contrast to this study, the novel Random Forest method involved the use of various tree depth levels along with the use of confidence vote and Gini index criteria. It resulted in an accuracy of 89.14% (*Risk Prediction in Cardiovascular Disease-Current Status and Future Challenges* 2009; Kaur, Sharma, and Sharma 2019)

Some other research findings employed a low sample size to counteract this study number of algorithms for classification but we are using as many as possible machine learning algorithms for analysis. In other existing studies, they only applied for specific calculations advanced by molecule multitude and subterranean insect state improvement (Saranya S et al. 2020). But In this proposed research work, it can be found that the staging framework has an enormous number of hyperparameters that are to be calibrated. The classification models can be enhanced in order to improve the features of classification (Tithi et al. 2019).

Most of the previous work done is based on the data from the specific dataset. The limitation of this research study having a lack of accuracy in predicting heart disease is improved by machine learning algorithms. In the future, more attributes in the dataset can be included and can also be used as datasets for better accuracy. Artificial intelligence can also be utilized in the future to anticipate the seriousness of coronary illness.

CONCLUSION

The main aim of the study is to quantify the precision in the classification of heart disease using medical data. This research study applied RF algorithm for the detection of heart disease from a dataset with parameters achieved has been compared with the LR algorithm. The results obtained show that the RF algorithm has found 90.16% of accuracy on predicting heart disease than the 85.25% of the LR algorithm.

DECLARATION

Conflict of Interests

No conflict of interest in this manuscript.

Authors' Contributions

Author PT was involved in data collection, data analysis and manuscript writing. Author MR was involved in the conceptualization, data validation and critical review of the manuscript.

Acknowledgments

The authors would like to express their gratitude towards Saveetha School of Engineering, Saveetha Institute of Medical and Technical Sciences (Formerly known as Saveetha University) for providing the necessary infrastructure to carry out this work successfully.

Funding

We thank the following organizations for providing financial support that enabled us to complete the study.

1. Verk Technologies, Rangareddy, Telangana
2. Saveetha University
3. Saveetha Institute of Medical And Technical Sciences
4. Saveetha School of Engineering

REFERENCES

1. Antink, Christoph Hoog, Joana Carlos Mesquita Ferreira, Michael Paul, Simon Lyra, Konrad Heimann, Srinivasa Karthik, Jayaraj Joseph, et al. 2020. "Fast Body Part Segmentation and Tracking of Neonatal Video Data Using Deep Learning." *Medical & Biological Engineering & Computing* 58 (12): 3049–61.
2. Baig, Daniyal. 2020. *Making Heart Diseases Detectable. The Invention of an Algorithm for Systematically Predictions*. GRIN Verlag.
3. Chatterjee, Debmalya, and Saravanan Chandran. 2019. "Prediction and Classification of Heart Disease Using AML and Power BI." *Proceedings of the 11th International Joint Conference on Computational Intelligence*. <https://doi.org/10.5220/0008381505080515>.
4. Chellapa, L. R., S. Rajeshkumar, M. I. Arumugham, and S. R. Samuel. 2020. "Biogenic Nanoselenium Synthesis and Evaluation of Its Antimicrobial, Antioxidant Activity and Toxicity." *Bioinspired Biomimetic and Nanobiomaterials*, July, 1–6.

5. Kaja, Rekha, Anandh Vaiyapuri, Mohamed Sherif Sirajudeen, Hariraja Muthusamy, Radhakrishnan Unnikrishnan, Mohamed Waly, Samuel Sundar Doss Devaraj, Mohamed Kotb Seyam, and Gopal Nambi S. 2020. "Biofeedback Flutter Device for Managing the Symptoms of Patients with COPD." *Technology and Health Care: Official Journal of the European Society for Engineering and Medicine* 28 (5): 477–85.
6. Kalidoss, Ramji, Snehalatha Umapathy, and Usha Rani Thirunavukkarasu. 2021. "A Breathalyzer for the Assessment of Chronic Kidney Disease Patients' Breathprint: Breath Flow Dynamic Simulation on the Measurement Chamber and Experimental Investigation." *Biomedical Signal Processing and Control* 70 (September): 103060.
7. Kaur, Gagandeep, Anshu Sharma, and Anurag Sharma. 2019. "Heart Disease Prediction Using KNN Classification Approach." *International Journal of Computer Sciences and Engineering*. <https://doi.org/10.26438/ijcse/v7i5.416420>.
8. Lavanya, M., P. Muthu Kannan, and M. Arivalagan. 2021. "Lung Cancer Diagnosis and Staging Using Firefly Algorithm Fuzzy C-Means Segmentation and Support Vector Machine Classification of Lung Nodules." *International Journal of Biomedical Engineering and Technology* 37 (2): 185.
9. Malaikolundhan, Harikrishna, Gowsik Mookkan, Gunasekaran Krishnamoorthi, Nirosha Matheswaran, Murad Alsawalha, Vishnu Priya Veeraraghavan, Surapaneni Krishna Mohan, and Aiting Di. 2020. "Anticarcinogenic Effect of Gold Nanoparticles Synthesized from Albizia Lebbeck on HCT-116 Colon Cancer Cell Lines." *Artificial Cells, Nanomedicine, and Biotechnology* 48 (1): 1206–13.
10. Maranhao Neto, Geraldo A., Iuliia Pavlovska, Anna Polcrova, Jeffrey I. Mechanick, Maria M. Infante-Garcia, Jose Pantaleón Hernandez, Miguel A. Araujo, Ramfis Nieto-Martinez, and Juan P. Gonzalez-Rivas. 2021. "Prediction of Cardiorespiratory Fitness in Czech Adults: Normative Values and Association with Cardiometabolic Health." *International Journal of Environmental Research and Public Health* 18 (19). <https://doi.org/10.3390/ijerph181910251>.
11. Nambi, Vijay. 2018. *Biomarkers in Cardiovascular Disease*. Elsevier Health Sciences.
12. Paul, M., S. Karthik, J. Joseph, M. Sivaprakasam, J. Kumutha, S. Leonhardt, and C. Hoog Antink. 2020. "Non-Contact Sensing of Neonatal Pulse Rate Using Camera-Based Imaging: A Clinical Feasibility Study." *Physiological Measurement* 41 (2): 024001.
13. Qi, Zhenya, and Zuoru Zhang. n.d. "A Hybrid Cost-Sensitive Ensemble for Heart Disease Prediction." <https://doi.org/10.21203/rs.2.22946/v5>.
14. Rajdhan, Apurb, Avi Agarwal, Milan Sai, Dundigalla Ravi, Poonam Ghuli, and R V College of Engineering. 2020a. "Heart Disease Prediction Using Machine Learning." *International Journal of Engineering Research and*. <https://doi.org/10.17577/ijert-v9is040614>.
15. – – – . 2020b. "Heart Disease Prediction Using Machine Learning." *International Journal of Engineering Research and*. <https://doi.org/10.17577/ijert-v9is040614>.
16. Raj R, Kathiswar, Ezhilarasan D, and Rajeshkumar S. 2020. "β-Sitosterol-Assisted Silver Nanoparticles Activates Nrf2 and Triggers Mitochondrial Apoptosis via Oxidative Stress in Human Hepatocellular Cancer Cell Line." *Journal of Biomedical Materials Research. Part A* 108 (9): 1899–1908.
17. Ramadoss, Ramya, Rajashree Padmanaban, and Balakumar Subramanian. 2022. "Role of Bioglass in Enamel Remineralization: Existing Strategies and Future Prospects-A Narrative Review." *Journal of Biomedical Materials Research. Part B, Applied Biomaterials* 110 (1): 45–66.
18. Risk Prediction in Cardiovascular Disease-Current Status and Future Challenges. 2009.
19. Shah, Devansh, Samir Patel, and Santosh Kumar Bharti. 2020. "Heart Disease Prediction Using Machine Learning Techniques." *SN Computer Science*. <https://doi.org/10.1007/s42979-020-00365-y>.
20. Sharma, Prateek. 2021. "Review on Heart Disease Prediction Using Machine Learning." *International Journal for Research in Applied Science and Engineering Technology*. <https://doi.org/10.22214/ijraset.2021.33848>.
21. Shilpa-Jain, D. P., Jogikalmat Krithikadatta, Dinesh Kowsky, and Velmurugan Natanasabapathy. 2021. "Effect of Cervical Lesion Centered Access Cavity Restored with Short Glass Fibre Reinforced Resin Composites on Fracture Resistance in Human Mandibular Premolars- an in Vitro Study." *Journal of the Mechanical Behavior of Biomedical Materials* 122 (October): 104654.
22. Singh, Jagdeep. 2017. *Hybrid Technique for Associative Classification of Heart Diseases: Prediction of Heart Disease Using Data Mining*.
23. Singh, Reetu, and E. Rajesh. 2019. "Prediction of Heart Disease by Clustering and Classification Tech-

- niques.” *International Journal of Computer Sciences and Engineering*. <https://doi.org/10.26438/ijcse/v7i5.861866>.
24. S, Saranya, S. Saranya, P. Deepika, and S. Sasi-kala. 2020. “COMPREHENSIVE REVIEW ON HEART DISEASE PREDICTION USING OPTIMI-ZATION TECHNIQUES.” *EPRA International Journal of Research & Development (IJRD)*. <https://doi.org/10.36713/epra3887>.
 25. S, Sudha, Kalpana R, and Soundararajan P. 2021. “Quantification of Sweat Urea in Diabetes Using Elec-tro-Optical Technique.” *Physiological Measurement* 42 (9). <https://doi.org/10.1088/1361-6579/ac1d3a>.
 26. Steiner, Jill M., Erica Lokken, Elizabeth Bayley, Jai-mie Pechan, Anna Curtin, Jonathan Buber, and Cath-erine Albright. 2021. “Cardiac and Pregnancy Out-comes of Pregnant Patients With Congenital Heart Disease According to Risk Classification System.” *The American Journal of Cardiology*, October. <https://doi.org/10.1016/j.amjcard.2021.08.037>.
 27. Tithi, Sushmita Roy, Afifa Aktar, Fahimul Aleem, and Amitabha Chakrabarty. 2019. “ECG Data Anal-ysis and Heart Disease Prediction Using Machine Learning Algorithms.” 2019 IEEE Region 10 Sym-posium (TENSYP). <https://doi.org/10.1109/ten-symp46218.2019.8971374>.
 28. Wang, Ke, Jing Tian, Chu Zheng, Hong Yang, Jia Ren, Yanling Liu, Qinghua Han, and Yanbo Zhang. 2021. “Interpretable Prediction of 3-Year All-Cause Mortality in Patients with Heart Failure Caused by Coronary Heart Disease Based on Machine Learning and SHAP.” *Computers in Biology and Medicine* 137 (October): 104813.
 29. Westerlund, Annie M., Johann S. Hawe, Matthias Heinig, and Heribert Schunkert. 2021. “Risk Predic-tion of Cardiovascular Events by Exploration of Mo-lecular Data with Explainable Artificial Intelligence.” *International Journal of Molecular Sciences* 22 (19). <https://doi.org/10.3390/ijms221910291>.
 30. Woodruff, Rebecca C., Ipuniuese Eliapo-Unutoa, Howard Chiou, Maria Gayapa, Sara Noonan, Pradeep S. B. Podila, Victoria Rayle, et al. 2021. “Period Prev-alence of Rheumatic Heart Disease and the Need for a Centralized Patient Registry in American Samoa, 2016 to 2018.” *Journal of the American Heart Associa-tion*, October, e020424.
 31. Wu, Shuang, Shanmugam Rajeshkumar, Malini Madasamy, and Vanaja Mahendran. 2020. “Green Syn-thesis of Copper Nanoparticles Using Cissus Vitiginea and Its Antioxidant and Antibacterial Activity against Urinary Tract Infection Pathogens.” *Artificial Cells, Nanomedicine, and Biotechnology* 48 (1): 1153–58.
 32. Zhang, Dengqing, Yunyi Chen, Yuxuan Chen, Shengyi Ye, Wenyu Cai, Junxue Jiang, Yechuan Xu, Gongfeng Zheng, and Ming Chen. 2021. “Heart Dis-ease Prediction Based on the Embedded Feature Se-lection Method and Deep Neural Network.” *Journal of Healthcare Engineering* 2021 (September): 6260022.