



International Journal of Innovative Research in Science Engineering and Technology (IJIRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)



Impact Factor: 8.699

Volume 14, Issue 3, March 2025

Cardiovascular Disease Prediction Using Machine Learning

Ms. P. Priya B.Tech., M.E, J.Vincent Herpert, A.Aropio Sham, K.Santhana Karuppu

Assistant Professor, Department of Computer Science and Design, Sethu Institute of Technology, Virudhunagar
District, Tamil Nadu, India

Department of Computer Science and Design, Sethu Institute of Technology, Virudhunagar District, Tamil Nadu, India

Department of Computer Science and Design, Sethu Institute of Technology, Virudhunagar District, Tamil Nadu, India

Department of Computer Science and Design, Sethu Institute of Technology, Virudhunagar District, Tamil Nadu, India

ABSTRACT: Heart Attack is a term that assigns a large number of medical conditions related to heart. The key to Heart (Cardiovascular) diseases to evaluate large scores of data sets, compare information that can be used to predict, Prevent, Manage such as Heart attacks. Heart Disease is mainly because of stress, family backgrounds, High blood Pressure, etc... Data analytics is used to incorporate world for its valuable use to controlling, contravasting and Manage a large data sets. It can be applied with an much success to predict, prevent, Managing a Cardiovascular Diseases. To solve this we aims to implement the Data Analytics based on SVM and Genetic Algorithm to diagnosis of heart diseases. This result reveal the Genetic Algorithm as best optimized Prediction Models.

KEYWORDS: Heart Attack, Cardiovascular Diseases, Heart Disease, Data Analytics, Prediction, Prevention, Management, High Blood Pressure, Stress, Family History, Large Data Sets, Machine Learning, SVM (Support Vector Machine), Genetic Algorithm, Diagnosis, Optimization, Prediction Models, Medical Conditions, Risk Factors, Data Processing, Health Monitoring.

I. INTRODUCTION

Identifying heart disease is challenging due to multiple risk factors, including diabetes, high blood pressure, high cholesterol, and abnormal pulse rates. These factors contribute to the complexity of diagnosing and predicting heart-related conditions. To address this, various data mining techniques and neural networks have been applied to assess the severity of heart disease in individuals. Classification methods such as K-Nearest Neighbor (KNN), Decision Trees (DT), Logistic Regression (LR), and Support Vector Machine (SVM) are commonly used to analyze medical data and categorize patients based on their risk levels.

Given the intricate nature of heart disease, a careful and precise approach is necessary to ensure accurate predictions and effective medical intervention. Failure to manage the condition properly may lead to severe complications, including heart failure or premature death. From a medical and data science perspective, heart disease prediction benefits significantly from advanced data mining techniques. These methods help uncover patterns related to metabolic syndromes, which are crucial for early diagnosis and treatment.

Classification algorithms play a vital role in identifying at-risk individuals by analyzing vast medical datasets and detecting subtle trends that may indicate the onset of heart disease. With continuous advancements in machine learning and predictive analytics, healthcare professionals can leverage data-driven insights to improve early detection, prevention strategies, and personalized treatment plans. Ultimately, integrating medical knowledge with computational techniques enhances the accuracy of heart disease predictions, leading to better patient outcomes and reduced mortality rates.

II. LITERATURE SURVEY

According to Ordonez [1] the heart disease can be predicted with some basic attributes taken from the patient and in their work have introduced a system that includes the characteristics of an individual human being based on totally 13 basic attributes like sex, blood pressure, cholesterol and others to predict the likelihood of a patient getting affected by heart disease. They have added two more attributes i.e. fat and smoking behaviour and extended the research dataset. The data mining classification algorithms such as Decision Tree, Naive Bayes, and Neural Network are utilized to make predictions and the results are analysed on Heart disease database.

Yilmaz, [2] have proposed a method that uses least squares support vector machine (LS-SVM) utilizing a binary decision tree for classification of cardiogram to find out the patient condition. Duff, et al. [3] have done a research work involving five hundred and thirty-three patients who had suffered from cardiac arrest and they were integrated in the analysis of heart disease probabilities. They performed classical statistical analysis and data mining analysis using mostly Bayesian networks.

Frawley, et al. [4] have performed a work on prediction of survival of Coronary heart disease (CHD) which is a challenging research problem for medical society. They also used 10-fold cross-validation methods to determine the impartial estimate of the three prediction models for performance comparison purposes. Lee, et al. proposed a novel methodology to expand and study the multi-parametric feature along with linear and nonlinear features of Heart Rate Variability diagnosing cardiovascular disease.

They have carried out various experiments on linear and non-linear features to estimate several classifiers, e.g., Bayesian classifiers, CMAR, C4.5 and SVM. Based on their experiments, SVM outperformed the other classifiers. Noh, et al. suggested a classification method which is an associative classifier that is constructed based on the efficient FPgrowth method. Because the volume of patterns can be diverse and huge, they offered a rule to measure the cohesion and in turn allow a tough choice of pruning patterns in the pattern-generating process.

Parthiban, et al. [7] have proposed a new work in which the heart disease is identified and predicted using the proposed Coactive Neuro-Fuzzy Inference System (CANFIS). Their model works based on the collective nature of neural network adaptive capabilities and based on the genetic algorithm along with fuzzy logic in order to diagnose the occurrence of the disease. The performance of the proposed CANFIS model was evaluated in terms of training performances and classification accuracies. Finally, their results show that the proposed CANFIS model has great prospective in predicting the heart disease.

Singh, et al. [8] have done a work using, one partition clustering algorithm (K-Means) and one hierarchical clustering algorithm (agglomerative). K-means algorithm has higher effectiveness and scalability and converges fast when production with large data sets. Hierarchical clustering constructs a hierarchy of clusters by either frequently merging two smaller clusters into a larger one or splitting a larger cluster into smaller ones. Using WEKA data mining tool, they have calculated the performance of k-means and hierarchical clustering algorithm on the basis of accuracy and running time.

Guru, et al. [9] have proposed the computational model based on a multilayer perceptron with three layers is employed to enlarge a decision support system for the finding of five major heart diseases. The proposed decision support system is trained using a back propagation algorithm amplified with the momentum term, the adaptive learning rate and the forgetting mechanics.

Palaniappan, et al. [10] have carried out a research work and have built a model known as Intelligent Heart Disease Prediction System (IHDPS) by using several data mining techniques such as Decision Trees, Naïve Bayes and Neural Network. Shantakumar, et al. [11] have done a research work in which the intelligent and effective heart attack prediction system is developed using Multi-Layer Perceptron with Back-Propagation. Accordingly, the frequency patterns of the heart disease are mined with the MAFIA algorithm based on the data extracted.

III. PROPOSED METHODOLOGY

The proposed methodology focuses on developing a machine learning-based model for predicting cardiovascular diseases with high accuracy. The system utilizes patient medical records, including risk factors such as age, blood pressure, cholesterol levels, pulse rate, and diabetes history, to make predictions. By leveraging advanced classification algorithms, the model aims to detect the likelihood of heart disease in individuals, allowing for early diagnosis and timely intervention. Machine learning techniques, such as Support Vector Machine (SVM), Decision Trees (DT), Logistic Regression (LR), and Neural Networks, are employed to improve predictive performance and enhance the reliability of the system.

One of the primary goals of this approach is to assist in reducing healthcare costs by facilitating early detection and targeted treatment plans. By predicting the risk of heart disease before severe complications arise, doctors can recommend preventive measures and cost-effective treatments, ultimately minimizing the financial burden on patients. Early intervention not only improves patient outcomes but also reduces hospitalizations and expensive surgical procedures, making healthcare more accessible and efficient.

The model evaluates key performance parameters such as accuracy, processing time, and energy efficiency to ensure optimal functionality. Accuracy is a critical factor in determining the reliability of predictions, while elapsed time measures the efficiency of the model in generating real-time results. Additionally, energy consumption is assessed to optimize computational resources, ensuring that the system remains lightweight and can be deployed in real-time healthcare applications without excessive power usage.

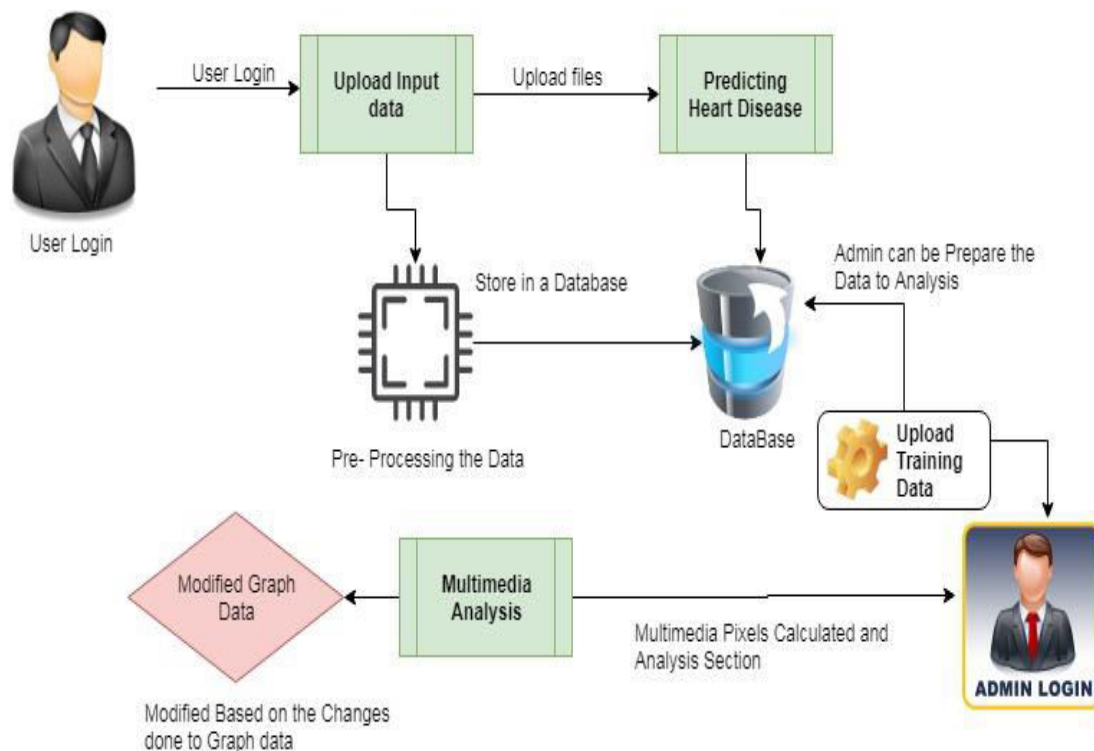


Fig1: Architecture Diagram

The To develop a robust and precise prediction system, feature selection and data preprocessing are implemented as essential steps. The dataset undergoes normalization, handling of missing values, and outlier detection to improve data quality. Feature selection techniques such as Principal Component Analysis (PCA) and Recursive Feature Elimination (RFE) are applied to extract the most relevant attributes, reducing computational complexity while maintaining high predictive performance. This step enhances the efficiency of machine learning algorithms and prevents overfitting.

The proposed system also incorporates real-time monitoring and cloud-based integration to facilitate continuous tracking of patient health. By connecting to electronic health records (EHRs) and wearable devices, the model can collect and analyze real-time patient data, enabling doctors to receive timely alerts about potential heart disease risks. Cloud integration allows for secure storage, remote accessibility, and scalability, ensuring that the system can be deployed in various healthcare settings for widespread use.

Finally, the evaluation and validation phase ensures that the proposed model performs effectively across diverse datasets. Cross-validation techniques such as k-fold validation are applied to assess model performance on different subsets of data. Additionally, comparative analysis with traditional diagnostic methods is conducted to highlight the advantages of machine learning-based predictions. By continuously refining and optimizing the model, this methodology aims to provide a reliable, cost-effective, and efficient solution for cardiovascular disease prediction, ultimately improving patient care and reducing mortality rates.

IV. RESULT AND DISCUSSION

1. UPLOAD TRAINING DATA:

The rule generation process is carried out in two distinct stages. In the first stage, an SVM model is constructed using the training dataset, ensuring that the model learns patterns and relationships within the data. During each fold of cross-validation, the trained SVM model is applied to predict class labels, allowing for a systematic evaluation of its predictive capability. This iterative process ensures that the model is effectively trained and tested across different subsets of data, enhancing its robustness and generalizability.

In the second stage, the generated rules are assessed based on their accuracy, precision, recall, and F-measure using the remaining 10% of test data in each fold. These evaluation metrics help determine the reliability and effectiveness of the classification model. Additionally, ruleset size and mean rule length are measured for each fold of cross-validation, providing insights into the complexity and interpretability of the generated rules. This dual-stage approach ensures that the rule-based classification model is both accurate and efficient while maintaining a balance between model performance and rule complexity.

2. DATA PRE- PROCESSING:

After collecting various patient records, the heart disease dataset undergoes a pre-processing phase to ensure data quality and consistency. The dataset consists of 303 patient records, among which 6 records contain missing values. To maintain the integrity of the analysis, these incomplete records are removed, leaving 297 complete patient records for further processing. This step ensures that only high-quality data is utilized, reducing inconsistencies that could impact the accuracy of predictive modeling. In the pre-processing stage, the dataset is structured to support both multiclass classification and binary classification for different attributes. Multiclass classification helps in categorizing the severity of heart disease into different levels, while binary classification differentiates between the presence and absence of the disease. This approach enhances the flexibility of the predictive model, allowing it to provide more detailed insights into heart disease risk assessment.

3. PREDICTING HEART DISEASE:

The training set and test set are kept distinct to ensure an unbiased evaluation of the model's performance. In this study, we employ this approach to validate the universal applicability of the proposed methods, ensuring that the results are generalizable across different datasets. By maintaining this separation, the model's ability to accurately classify new, unseen data is thoroughly assessed, strengthening its reliability for heart disease prediction. To further enhance the robustness of the model, we utilize k-fold cross-validation, where the entire dataset is systematically divided into multiple subsets for both training and testing. This iterative approach ensures that every data point is used for both training and validation at different stages, leading to a more comprehensive and accurate assessment of the classifier's ability to predict heart stroke risk. This method effectively reduces overfitting and improves the generalization of the model across diverse patient data.

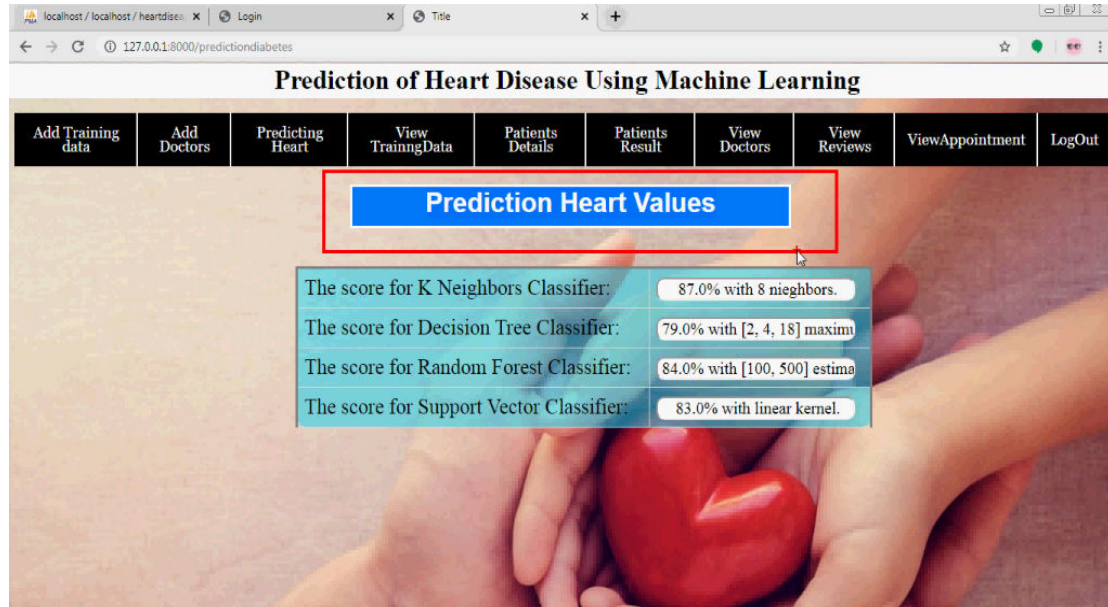
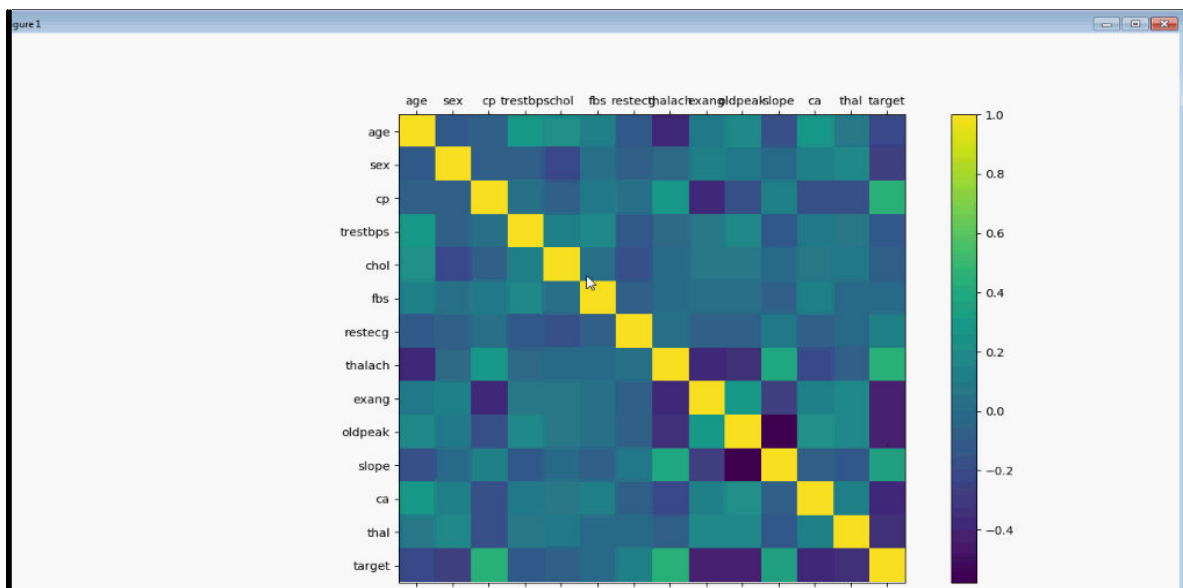


Fig2: Predicting Heart Disease

4. GRAPHICAL REPRESENTATIONS:

The evaluation of the proposed system is conducted by analyzing both approvals and disapprovals, providing a comprehensive assessment of its effectiveness. These outcomes are visually represented using graphical notations, including pie charts, bar charts, and line charts, which help in illustrating trends, distributions, and comparisons in a clear and interpretable manner. This visual approach enhances the understanding of the system's performance and aids in identifying areas for improvement.

Furthermore, the data used for analysis is dynamically updated, allowing for real-time insights and adaptability to new information. By continuously processing and integrating new data, the system ensures that its predictions and evaluations remain relevant and accurate. This dynamic approach facilitates ongoing monitoring and refinement, leading to a more efficient and data-driven decision-making process in heart disease analysis.



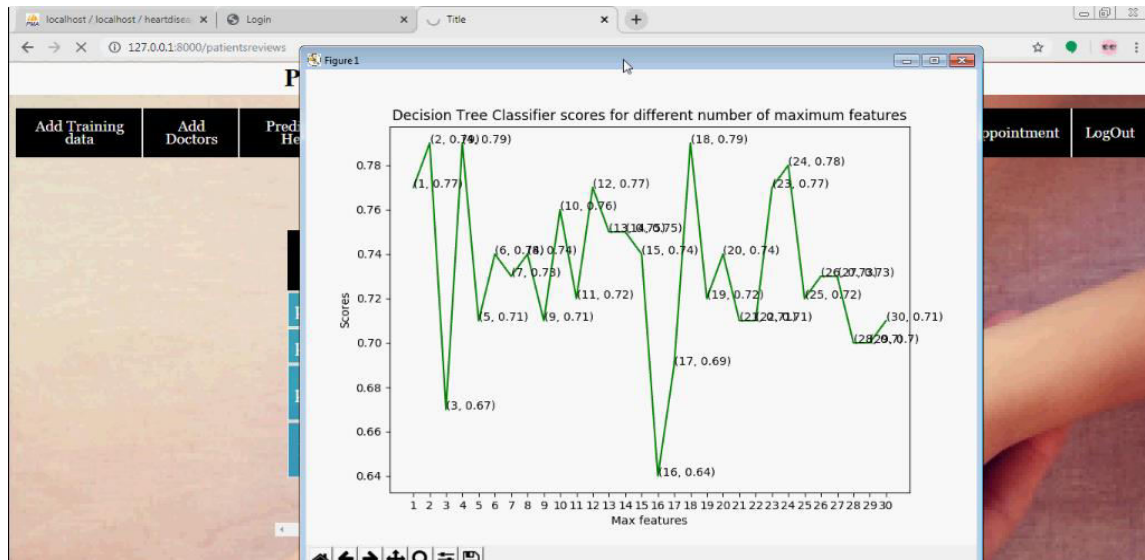


Fig3: Graphical Representation

V.CONCLUSION

Identifying the processing of raw healthcare data of heart information will help in the long term saving of human lives and early detection of abnormalities in heart conditions. Machine learning techniques were used in this work to process raw data and provide a new and novel discernment towards heart disease. Heart disease prediction is challenging and very important in the medical. However, the mortality rate can be drastically controlled if the disease is detected at the early stages and preventative measures are adopted as soon as possible. Further extension of this study is highly desirable to direct the investigations to real-world datasets instead of just theoretical approaches and simulations. The proposed hybrid HRFLM approach is used combining the characteristics of Random Forest (RF) and Linear Method (LM). HRFLM proved to be quite accurate in the prediction of heart disease. The future course of this research can be performed with diverse mixtures of machine learning techniques to better prediction techniques. Furthermore, new feature selection methods can be developed to get a broader perception of the significant features to increase the performance of heart disease prediction.

REFERENCES

1. Al-Maqaleh, B. M., & Al-Wesabi, S. A. (2023). Prediction of heart disease based on machine learning using Apache Spark. *Journal of Big Data*, 10(1), 1-20. <https://doi.org/10.1186/s40537-023-00712-7>
2. Amin, M. S., Chiam, Y. K., & Varathan, K. D. (2019). Identification of significant features and data mining techniques in predicting heart disease. *Tehnički vjesnik*, 26(1), 149-155. <https://doi.org/10.17559/TV-20170611124840>
3. Anwar, S. M., Gul, M., Majid, M., Alnowami, M., & Khan, M. K. (2018). Medical image analysis using convolutional neural networks: A review. *Journal of Medical Systems*, 42(11), 226.
4. Aslam, M. W., Zhu, Z., & Nandi, A. K. (2013). Feature generation using genetic programming with comparative partner selection for diabetes classification. *Expert Systems with Applications*, 40(13), 5402-5412.
5. Bashir, S., Qamar, U., & Khan, F. H. (2019). Heterogeneous classifiers fusion for dynamic selection of heart disease diagnosis. *International Journal of Medical Informatics*, 128, 123-138.
6. Chaurasia, V., & Pal, S. (2014). Data mining approach to detect heart diseases. *International Journal of Advanced Computer Science and Information Technology (IJACSIT)*, 2(4), 56-66.
7. Chen, X., & Liu, C. (2020). Prediction of cardiovascular diseases using hybrid machine learning model based on random forest and particle swarm optimization. *Journal of Medical Systems*, 44(2), 1-11.
8. Choudhury, A., & Gupta, D. (2019). A survey on medical diagnosis of heart disease using data mining techniques. *Global Journal of Computer Science and Technology*.

9. Dey, A. K., & Chakraborty, S. (2018). Machine learning techniques for heart disease prediction: A comparative study. *Proceedings of the 2018 International Conference on Computer, Electrical & Communication Engineering (ICCECE)*, 1-4. <https://doi.org/10.1109/ICCECE.2018.8534703>
10. Dutta, S., & Dutta, S. (2016). Prediction of heart disease using hybrid ensemble classifier. *International Journal of Computer Applications*, 139(2), 6-11.
11. Gudadhe, M., Wankhade, K., & Dongre, S. (2010). Decision support system for heart disease based on support vector machine and artificial neural network. *2010 International Conference on Computer and Communication Technology (ICCCT)*, 741-745. <https://doi.org/10.1109/ICCCT.2010.5640373>
12. Gupta, S., & Kumar, D. (2018). Heart disease prediction using data mining techniques. *2018 International Conference on Intelligent Sustainable Systems (ICISS)*, 104-109. <https://doi.org/10.1109/ISS1.2018.8418734>
13. Ha, J., & Park, C. (2020). Prediction of heart disease using machine learning algorithms. *Journal of Information and Communication Convergence Engineering*, 18(3), 153-157. <https://doi.org/10.6109/jicce.2020.18.3.153>
14. 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, 1-21.
15. Jabbar, M. A., Deekshatulu, B. L., & Chandra, P. (2013). Heart disease prediction using lazy associative classification. *2013 International Mutli-Conference on Automation, Computing, Communication, Control and Compressed Sensing (iMac4s)*, 40-46. <https://doi.org/10.1109/iMac4s.2013.6526390>
16. Carlos Ordonez, "Improving Heart Disease Prediction using Constrained Association Rules", Technical Seminar Presentation, University of Tokyo, 2004.
17. Franck Le Duff, CristianMunteanb, Marc Cuggiaa and Philippe Mabob, "Predicting Survival Causes After Out of Hospital Cardiac Arrest using Data Mining Method", *Studies in Health Technology and Informatics*, Vol. 107, No.2, pp. 1256-1259, 2004.
18. W.J. Frawley and G. Piatetsky-Shapiro, "Knowledge Discovery in Databases: An Overview", *AI Magazine*, Vol. 13, No. 3, pp. 57-70, 1996.
19. Heon Gyu Lee, Ki Yong Noh and Keun Ho Ryu, "Mining Bio Signal Data: Coronary Artery Disease Diagnosis using Linear and Nonlinear Features of HRV", *Proceedings of International Conference on Emerging Technologies in Knowledge Discovery and Data Mining*, pp. 56-66, 2007.
20. Kiyong Noh, HeonGyu Lee, Ho-Sun Shon, Bum Ju Lee and Keun Ho Ryu, "Associative Classification Approach for Diagnosing Cardiovascular Disease", *Intelligent Computing in Signal Processing and Pattern Recognition*, Vol. 345, pp. 721-727, 2006.
21. Latha Parthiban and R. Subramanian, "Intelligent Heart Disease Prediction System using CANFIS and Genetic Algorithm", *International Journal of Biological, Biomedical and Medical Sciences*, Vol. 3, No. 3, pp. 1-8, 2008.
22. Niti Guru, Anil Dahiya and Navin Rajpal, "Decision Support System for Heart Disease Diagnosis using Neural Network", *Delhi Business Review*, Vol. 8, No. 1, pp. 1-6, 2007.
23. Sellappan Palaniappan and Rafiah Awang, "Intelligent Heart Disease Prediction System using Data Mining Techniques", *International Journal of Computer Science and Network Security*, Vol. 8, No. 8, pp. 1-6, 2008.



INTERNATIONAL
STANDARD
SERIAL
NUMBER
INDIA



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

 9940 572 462  6381 907 438  ijirset@gmail.com



www.ijirset.com

Scan to save the contact details