

Research Article

Metaheuristic Methods for Efficiently Predicting and Classifying Real Life Heart Disease Data Using Machine Learning

Elia Ramirez-Asis ¹, Magna Guzman-Avalos ², Bireshwar Dass Mazumdar ³,
D Lakshmi Padmaja ⁴, Manmohan Mishra ⁵, Deepali S Hirolikar ⁶,
and Karthikeyan Kaliyaperumal ⁷

¹Department of Education Administration, Universidad Cesar Vallejo, Huaraz, Peru

²Department of Management and Educational Sciences, Universidad Nacional Santiago Antúnez de Mayolo, Huaraz, Peru

³Department of Computer Science and Engineering, United University Prayagra, Prayagraj, India

⁴Department of Information Technology, Anurag University, Hyderabad, Telengana State, India

⁵United Institute of Management, Prayagraj, India

⁶Department of Information Technology, PDEA'S College of Engineering Manjari Bk, Savitribai Phule Pune University, Pune, India

⁷IT and IoT-HH Campus, Ambo University, Ambo, Ethiopia

Correspondence should be addressed to Karthikeyan Kaliyaperumal; karthikeyan@ambou.edu.et

Received 9 March 2022; Accepted 23 April 2022; Published 25 May 2022

Academic Editor: Vijay Kumar

Copyright © 2022 Elia Ramirez-Asis et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The heart attack happens if the flow of blood leads to blocks in any of the blood veins and vessels liable for delivering blood into internal parts of the heart. In the modern life activities and habits, the males and females hold the same responsibility and burden of risk. The absence of understanding frequently leads to a postponement in dealing with the heart attack issues, which could worsen the injury and in most of the situations shown to be dead. Several researchers have applied data mining techniques to diagnose illnesses, and the results have been encouraging. Some methods forecast a specific illness, whereas others predict a wide spectrum of illnesses. In addition, the accuracy of sickness predictions can be improved. This post went into great length on the many approaches of data classification that are currently available. Algorithms primarily represent themselves through representations. Data classification is a typical but computationally intensive task in the area of information technology. A huge amount of data must be analysed in order to come up with an effective plan for fighting disease. Metaheuristics are frequently employed to tackle optimization issues. The accuracy of computing models can be improved by using metaheuristic techniques. Early disease diagnosis, severity evaluation, and prediction are all popular uses for artificial intelligence. For the sake of patients, health care costs, and slowed course of disease, this is a good idea. Machine learning approaches have been used to achieve this. Using machine learning and metaheuristics, this study attempts to classify and forecast human heart disease.

1. Introduction

Heart-related diseases [1, 2] consume around a million lives of peoples every year, creating this as the primary reason. In the year 2016, around 920,00 people had heart attacks and nearly half of them occurred suddenly without prior symptoms. Sudden death is the only symptom for heart disease. One death among five is due to heart problems in

India. Heart disease has considerably become more when compared with past decades and has turned out to be the primary reason of death. It is highly challenging for healthcare professionals to identify quickly and precisely.

So, it is essential to implement computer expertise in this analysis to help healthcare professionals to detect in the early stages with enriched accuracy. The objective of this research is to precisely and proficiently evaluate

heart-related hospitalizations based on the offered medical account of the patient. The approaches of this research are innovative for this domain. This encourages the research to forecast heart-related hospitalization. To save the irreversible lives of human beings, early diagnosis and prediction of the disease is mandatory than any other business and profit. So, this research focuses on early and efficient detection of heart disease at higher accuracy levels using data mining algorithms and history of past patient records. The feature of classification algorithms in data mining is analysed in this research for effective prediction. The datasets used by the researcher and algorithms are naturally impure and contain missing, irrelevant, and outdated values with the system and human errors. Since, the efficiency of prediction algorithms completely depends on the input dataset, it is necessary to focus on data cleaning (preprocessing) process prior to the actual mining process. This will result in improved accuracy levels of prediction. Hence, this system focused on preprocessing, mining, and prediction of heart disease in earlier stages using the history of databases [3].

There are many different types of cardiac illness, each of which affects a different organ within the heart [4]. Cardiovascular illnesses include any form of cardiac disease, and the diseases associated with the heart are discussed in greater detail further down in this section. CAD, which is another term for coronary heart disease, is the most common type of heart disease in the world and is also known as coronary artery disease (CAD). Inflammation of the blood vessels and arteries, caused by fat accumulation, is a source of complaint.

The data confirm that the heart diseases are double that of the average ratio of neighboring countries of India in the world. Regardless of actuality a developing concern, many Indians are not conscious about the heart disease and its related preliminary indications. The disease history related to generations of the family is considered as the general and strong threat factors; most of heart-related diseases are because of well-known manageable reasons such as sugar level, cholesterol, abnormal blood pressure, unhealthy diet, habit of smoking, style of inactive life, stress, and abnormal weights. In current situation, the daily life style and habits have become the primary reason and risk factor for getting heart disease. Heart attack is the major reason for deaths, particularly amongst youngsters and teens in India. As per the Indian Heart Association, mostly 50 percent of heart issues are in men under the age of fifty years and 26 percent of every attack is below the age of 40 [5, 6].

The heart attack happens if the flow of blood leads to blocks in any of the blood veins and vessels liable for delivering blood into internal parts of the heart. In the modern life activities and habits, the males and females hold the same responsibility and burden of risk. The absence of understanding frequently leads to a postponement in dealing with the heart attack issues, which could worsen the injury and in most of the situations shown to be dead. September 29th of every year is renowned to be the World's Heart Day (WHD) to bring alertness of the heart-related issues. Different themes were introduced every year to tackle the many causes of heart-related diseases and to bring awareness among the public [7–10].

Machine learning models can solve highly critical issues by automatically detecting the characteristics of the input data, and deep learning models can adapt to changes in the problem that they are attempting to answer. Using the inferred data, machine learning models will be able to uncover and analyse characteristics in data patterns that have not yet been presented to the user. Because even low-computing models will be able to accomplish this, a significant amount of time will be saved [10].

Heart disease has considerably become more when compared with past decades and has turned out to be the primary reason of death. It is highly challenging for healthcare professionals to identify quickly and precisely. Using machine learning and metaheuristics, this study attempts to classify and forecast human heart disease.

2. Literature Survey

Cardiac disease and ECG datasets are back-eliminated from the classifier model built by the authors in [11]. The feature choices have improved categorization techniques and reduced the number of inputs, according to experience. Seventy eight percent increase in performance was achieved with just a 19% reduction in the size of the arrhythmia dataset that was utilized in this investigation. In comparison to the prior data, this new set is 85% better and has just four distinct features. In a previous research, redundant features were shown to improve the performance of classifiers.

A surge development approach-based fuzzy master system is outlined [12]. To handle UCI machinery cardiac datasets, this method was developed particularly for them. Decision tree algorithms are used to identify the most important qualities for optimal diagnosis and therapy. Fuzzy rules are used to generate the output data. Fuzzy approximation is used to get the outcome. An expert system based on the particle swarm optimization approach has a 93.27% accuracy rate. There is a huge advantage to this system when compared to other classification methods, which are difficult to understand the output model given by fuzzy expert systems.

A firefly-based method based on rough sets was proposed by authors in [13] as a foundation for an accurate prediction system. The high complexity and uncertainty associated with heart disease datasets may be reduced by including both fuzzy and rough theoretical notions. With the roughest-based fuzzy learning approach, it is feasible to find optimum answers while consuming minimal computer resources. Support vector machines and artificial neural networks cannot match these results when it comes to heart disease prediction and medication prescribing.

Scientists have devised a novel method for forecasting ventricular arrhythmia. An ECG signal processor that is completely integrated into the system is used in this work to construct a pain prediction system. A certain set of ECG parameters may be used to predict whether or not a person would have ventricular arrhythmia. To evaluate and monitor the sites, an ECG waveform is recognized and noted (PQRST). This procedure is carried out using real-time and flexible methods. Controlling the ECG signal fluctuations efficiently and accurately is the goal of these techniques. The

American Heart Association's collection of cardiovascular signals is utilized to assess the system's performance. It seems that the previous methods' accuracy metrics are comparable to the new ones, based on simulation findings. The system is simulated using an ASIC (application-specific integrated circuit) (ASIC). ESP-based ventricular arrhythmia forecasting is implemented for the first time on an ASIC.

A Naïve Bayes classification approach for heart disease detection was highlighted by researchers [14]. In today's society, heart illness has serious ramifications, as shown in this book. The Naive Bayes classifier is used in conjunction with statistical methods to accurately predict and diagnose cardiac disorders. It uses data preprocessing methods to handle the massive and complex gathering of medical data. Cardiac disease is classified using a discretization algorithm. In this case, directed variation with equal frequencies was used as the discretization approach. The stat log heart database contains datasets on heart disease that are used in this investigation. The findings show that this approach provides more accurate measurements than earlier techniques.

As mentioned in [15], the linear SVM classifier model works as follows. After isolating a hyperplane from a given dataset using categorized instructing tests, this difference classifier outputs an optimal hyperplane. Thus, the newly created instances of the input data model may be further classified in this manner. A hyperplane is a line that, in a two-dimensional space, splits the hyperplane into two parts. Both sides of the dividers include each class. In a nutshell, SVM separates classes.

Using deep learning and a linear SVM classification, the authors of [16] performed an original investigation. The softmax layer is replaced with a linear combinational machine in deep convolutional networks. Margin-based loss may be used to replace cross-entropy loss since it is more efficient. According to published studies, the SVM may be used to return various layers of a deep complex network. On the second layer, a deep convolution network replaces SVM. Ultimately, the goal of this study is to aid in the development of facial recognition software for use by humans.

Using logistic regression data analysis, the authors of [17] give an in-depth look at the statistical technique. Logistic degenerate is a statistical process for analysing dual dependent variables. An efficient method for regression analysis is logistic regression. The logistic model's parameters are estimated using logistic regression techniques. Logistic models use independent elements to calculate the probability of an event occurring.

Models of logistic regression provide the highest accurate classification results and are broadly used in a broad range of fields. Predictive models for heart disease are frequently evaluated using this technique. Overfitting is avoided with this method, which yields more precise findings. Nonlinear connections, on the contrary, make things more complicated and time intensive. In addition, this technique performs effectively as an evaluation tool for healthcare firms rather than a categorization model.

It was found that combining K-means and Apriori yielded the best results [18]. The dataset is first gathered using k-means gathering. The Apriori method is then used to

determine the most often recurring item sets. A "bottom-up" approach is used by the Boolean association rule to get better results. Real-world scenarios in the heart disease prediction system provide a succession of challenging questions to patterns. Predictive analysis relies on categorization to ensure that the input data are accurately identified and mapped. There are two categories of data: those that have been tagged and those that have not. In addition to the single target quality, the labelled data include many predictive characteristics. Using all the target characteristics, the class label is denoted. Only those features that have been labelled have predictive properties that are not unlabeled. The basic goal of the classification process is to appropriately classify not labelled data using categorization models derived from labelled examples (historical data). The first step is to create a training model that already has the proper class (or goal values) to build it.

3. Methodology

Figure 1 illustrates an example of how heart disease can be predicted using genetics. This method is based on a database of heart disease cases. The extraction of features is accomplished through the use of a procedure known as PCA. For the purpose of categorising the data, a number of machine learning algorithms are used. The results of the testing can be used to improve disease prediction. Using machine learning in conjunction with metaheuristics is a lethal combination when it comes to accurate forecasting.

First ever developing conclusion tree-based approach is J. Ross Quinlan's ID-3 method. This strategy makes use of entropy and information gain measurements. The entropy of the functional characteristics is computed iteratively, starting with a nodule. Split attributes refer to the subsets of a dataset that have been divided based on the feature with the lowest error rate (entropy) and the greatest information gain. In the absence of an accurate categorization of its target classes, the algorithm repeats itself for each subset of data. The terminal nodes of a decision tree are defined as the final subset of the branch's nonterminal nodes. The split attribute specifies nonterminal nodes, whereas class labels are denoted by terminal nodes. An ID-3-based conclusion tree method developed in [19] allows for the early diagnosis of cardiac problems.

According to Peterson et al. [20], using a K-nearest neighbor strategy for pattern detection and data categorization is the most resilient method in the field. Distance functions or similarity measures are used in K-nearest neighbor algorithms. In order to classify freshly defined instances, a similarity measure is employed, and all instances are kept. In order to efficiently classify, it makes use of the instance-based learning approach. Based on the votes of the classes immediately adjacent to it, each new instance of the dataset is allocated a category. Both the training and testing datasets are used to calculate the distance metric. As soon as k has been chosen, the algorithm estimates how far apart the two instances are.

As a nonprobabilistic binary linear classification approach, PSO-SVM (particle swarm optimization) support vector machine is the best choice [21]. Using this method,

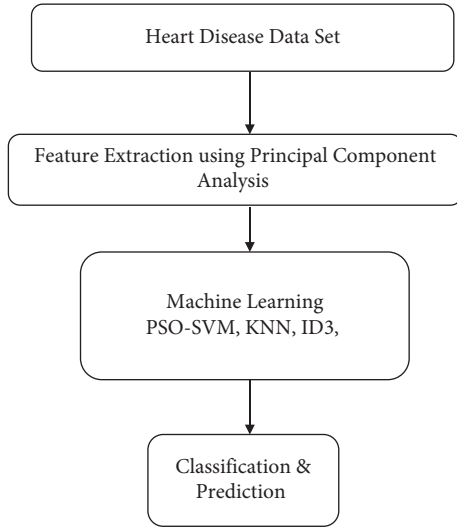


FIGURE 1: Machine learning and metaheuristics methods for classification and prediction of heart disease data.

samples may be divided into a single or several target classes. A single point is used to represent each piece of data. With each new group, it widens because of the clear divisions. The goal classes of the new occasions are remapped depending on which side of the space they fall. Nonlinear classification is possible if the input datasets are not tagged. The support vector machine employs an apart learning approach to classify the data because the instances cannot be allocated to target classes. More instances are added when the clusters based on functions have been formed. There is proof of a nonlinear support vector machine recommendation system. Nonlinear support vector machine approaches are the most often utilized way for dealing with unlabeled data.

4. Results and Analysis

Using data from the UCI machinery cardiac illness dataset [22], we conducted an analytical investigation. ID-3, C4.5, Random Forest, KNN, and SVM algorithms employ 303 entries from the Cleveland database as input. Weka is used to preprocess the incoming data set. The correctness of the data has been improved as a result of this preprocessing. Out of 303 instances, 240 instances were used for the training of machine learning predictors and remaining 63 instances were used for the testing of the machine learning predictors. Figures 2–4 show performance of machine learning predictors on the basis of different performance comparison parameters:

$$\begin{aligned} \text{Accuracy} &= \frac{TP + TN}{TP + TN + FP + FN}, \\ \text{Sensitivity} &= \frac{TP}{TP + FN}, \\ \text{Specificity} &= \frac{TN}{TN + FP}, \end{aligned} \quad (1)$$

where TP = True Positive, TN = True Negative, FP = False Positive, and FN = False Negative.

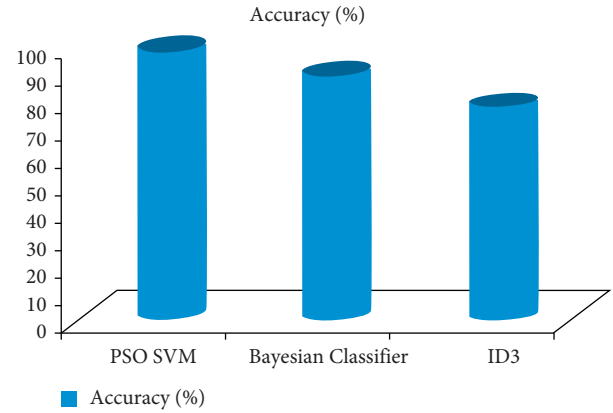


FIGURE 2: Accuracy of classifiers for heart disease dataset.

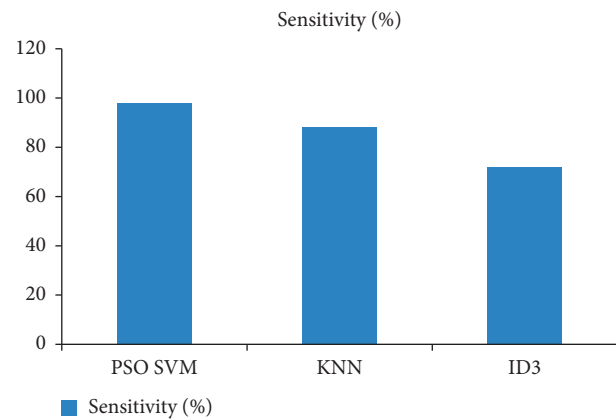


FIGURE 3: Sensitivity of classifiers for heart disease dataset.

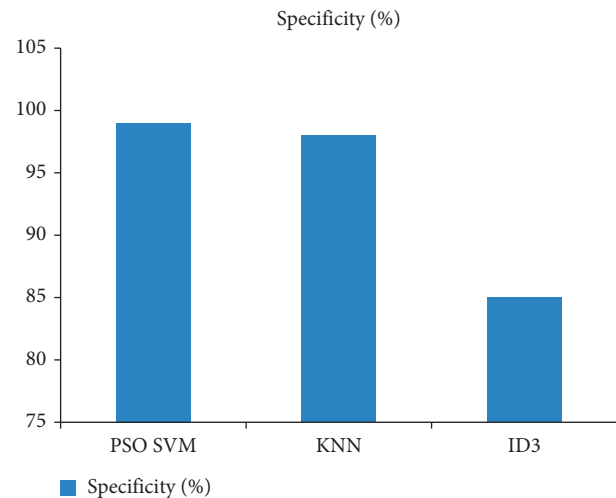


FIGURE 4: Specificity of classifiers for heart disease dataset.

5. Conclusion

The heart attack happens if the flow of blood leads to blocks in any of the blood veins and vessels liable for delivering blood into internal parts of the heart. In the modern life activities and habits, the males and females hold the same responsibility and burden of risk. The absence

of understanding frequently leads to a postponement in dealing with the heart attack issues, which could worsen the injury and in most of the situations shown to be dead. Metaheuristics are frequently employed to tackle optimization issues. The accuracy of computing models can be improved by using metaheuristic techniques. Early disease diagnosis, severity evaluation, and prediction are all popular uses for artificial intelligence. This study presented machine learning and metaheuristics methods for early and accurate detection of cardiac illness. Cleveland dataset is used for the innovative analysis. PSO-SVM algorithm is performing better than other machine learning predictors.

Data Availability

The data used to support the findings of the study can be obtained from the corresponding author upon request.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

References

- [1] A. Singh and R. Kumar, "Heart disease prediction using machine learning algorithms," in *Proceedings of the 2020 International Conference on Electrical and Electronics Engineering (ICE3)*, pp. 452–457, Gorakhpur, India, February 2020.
- [2] V. Sharma, S. Yadav, and M. Gupta, "Heart disease prediction using machine learning techniques," in *Proceedings of the 2020 2nd International Conference on Advances in Computing, Communication Control and Networking (ICACCCN)*, pp. 177–181, Greater Noida, India, December 2020.
- [3] P. Motarwar, A. Duraphe, G. Suganya, and M. Premalatha, "Cognitive approach for heart disease prediction using machine learning," in *Proceedings of the 2020 International Conference on Emerging Trends in Information Technology and Engineering (Ic-ETITE)*, pp. 1–5, Vellore, India, February 2020.
- [4] R. Manne and S. C. Kantheti, "Application of artificial intelligence in healthcare: chances and challenges," *Current Journal of Applied Science and Technology*, vol. 40, no. 6, pp. 78–89, 2021.
- [5] R. J. P. Princy, S. Parthasarathy, P. S. Hency Jose, A. Raj Lakshminarayanan, and S. Jeganathan, "Prediction of cardiac disease using supervised machine learning algorithms," in *Proceedings of the 2020 4th International Conference on Intelligent Computing and Control Systems (ICICCS)*, pp. 570–575, Madurai, India, May 2020.
- [6] J. I. Zong Chen and P. Hengjinda, "Early prediction of coronary artery disease (CAD) by machine learning method - a comparative study," *Journal of Artificial Intelligence and Capsule Networks*, vol. 3, no. 1, pp. 17–33, 2021.
- [7] A. Mehmood, M. Iqbal, Z. Mehmood et al., "Prediction of heart disease using deep convolutional neural networks," *Arabian Journal for Science and Engineering*, vol. 46, no. 4, pp. 3409–3422, 2021.
- [8] P. Karunakaran, Y. Babiker Hamdan, and P. Sathish, "Early prediction of autism spectrum disorder by computational approaches to fMRI analysis with early learning technique," *Journal of Artificial Intelligence and Capsule Networks*, vol. 2, no. 4, pp. 207–216, 2020.
- [9] V. Hemamalini, S. Rajarajeswari, S. Nachiyappan et al., "Food quality inspection and grading using efficient image segmentation and machine learning-based system," *Journal of Food Quality*, vol. 2022, Article ID 5262294, 6 pages, 2022.
- [10] A. Raghuvanshi, U. K. Singh, and C. Joshi, "A review of various security and privacy innovations for IoT applications in healthcare," in *Advanced Healthcare Systems*, pp. 43–58, Wiley, Hoboken, NJ, USA, 2022.
- [11] S. Shilaskar and A. Ghatol, "Feature selection for medical diagnosis :evaluation for cardiovascular diseases," *Expert Systems with Applications*, vol. 40, no. 10, pp. 4146–4153, 2013.
- [12] S. Muthukaruppan and M. J. Er, "A hybrid particle swarm optimization based fuzzy expert system for the diagnosis of coronary artery disease," *Expert Systems with Applications*, vol. 39, no. 14, Article ID 11657, 2012.
- [13] N. C. Long, P. Meesad, and H. Unger, "A highly accurate firefly based algorithm for heart disease prediction," *Expert Systems with Applications*, vol. 42, no. 21, pp. 8221–8231, 2015.
- [14] S. Aydin, M. Ahanpanjeh, and S. Mohabbatiyan, "Comparison and evaluation data mining techniques in the diagnosis of heart disease," *International Journal on Computational Science & Applications*, vol. 6, no. 1, pp. 1–15, 2016.
- [15] G. B. Berikol, O. Yildiz, and İ. T. Özcan, "Diagnosis of acute coronary syndrome with a support vector machine," *Journal of Medical Systems*, vol. 40, no. 4, pp. 84–18, 2016.
- [16] A. Gupta, L. Kapoor, and M. Wattal, "C2C (cloud-to-cloud): an ecosystem of cloud service providers for dynamic resource provisioning," in *Proceedings of the advances in Computing and Communications*, Kochi, India, July 2011.
- [17] B. Hssina, A. Merbouha, H. Ezzikouri, and M. Erritali, "A comparative study of decision tree ID3 and C4.5," *International Journal of Advanced Computer Science and Applications*, vol. 4, no. 2, pp. 13–19, 2014.
- [18] C. Sharma, A. Bagga, R. Sobti, M. Shabaz, and R. Amin, "A robust image encrypted watermarking technique for neuro-degenerative disorder diagnosis and its applications," *Computational and Mathematical Methods in Medicine*, vol. 2021, Article ID 8081276, 14 pages, 2021.
- [19] A. Gupta, "Performance insight 360: a cloud-based quality management framework for educational institutions in India," in *Proceedings of the 2013 IEEE 15th Conference on Business Informatics*, Vienna, Austria, July 2013.
- [20] H. Z. Almarzouki, H. Alsulami, A. Rizwan, M. S. Basingab, H. Bukhari, and M. Shabaz, "An internet of medical things-based model for real-time monitoring and averting stroke sensors," *Journal of Healthcare Engineering*, vol. 2021, Article ID 1233166, 9 pages, 2021.
- [21] S. Zhang, L. Yao, A. Sun, and Y. Tay, "Deep learning based recommender system," *ACM Computing Surveys*, vol. 52, no. 1, pp. 1–38, 2020.
- [22] Heart Disease, "Heart disease data set," <https://archive.ics.uci.edu/ml/datasets/heart+disease>.