International Journal of
INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING
ISSN:2147-6799 www.ijisae.org Original Research Paper

A Study on Functional Behavior of Machine Learning Model for Cardiac Disease Classification

[1]Ritika, [2]Rajender Singh, [3]Sandeep Dalal

Submitted: 09/11/2022 **Revised**: 20/01/2023 **Accepted**: 11/02/2023

Abstract: Cardiac diseases are most prevalent these days with high mortality ratio. The causes and symptoms of the heart diseases also vary according to the type of heart disease. In recent years, Heart disease diagnosis has attracted researchers to provide some automated and online solutions to detect heart disease at an early stage. AI and machine learning algorithms have already contributed a lot in this field and have been proved to be reliable and most efficient. Various feature weight identification and optimization heuristics have also been integrated with machine learning models for detecting heart disorders accurately. This review encapsulates the study on the research works to optimize various stages of machine learning models. It also intends to discuss the significance of pre-processing specially feature selection for machine learning algorithm in detail. The recent advancements in machine learning algorithms, methodologies and the performance gain are also provided in this article.

Keywords: *Cardiac Disease; Machine Learning; Feature Selection; Medical domain*

1. Introduction

Cardiovascular disease is the most critical disease with highest death rate. The cardiac attack is the cause of sudden death. In such attack situation, the chances of patient survival are very less and its treatment costs are very high. The chances of patient survival can be increased if the cardiovascular disease can be predicted at the earlier stage. The main cause of heart disease is the heavy work load in daily routine, stress, tension, depression[1][2]. A healthy individual also affected by tension and depression that causes heart disease and cardiac attack. The depression now days very common due to busy lifestyle. There are many prognostic factors that increase the severity of cause's cardiac problems because of depression. Depression can cause hyper activity, variable heart rate and platelet responsiveness [3].

There are different types of cardiac problems and diseases based on the damaged part of heart. Each of the category or type has its own risks and severity. The foremost heart disease type is Coronary Artery Disease (CAD). It is most common heart disease that causes because of fat within blood vessels and veins. CAD affects the blood and oxygen flow into the heart. CAD changes the structure of heart that can cause a serious heart problem. It is identified as a congenital issue that can exist in babies during birth. It can cause the heart pain that happens with non-continuous blood supply. This kind of heart pain is called Angina Pectoris. Pain in angina pectoris is an early symptom of heart attack. The severity of the pain depends on the duration of pain. The irregular flow of blood can cause congestive heart failure. Cardiomyopathy is another heart disease that occurs with the change in the shape of muscle. The change in appearance or infection to heart pump is known as Cardiomyopathy. Hypertension, alcohol consumption and viral infection are the major causes of Cardiomyopathy. Arrhythmia is another heart disease that occurs periodically with varied heartbeat. This abnormal heart rate can disturb the electrical connections of heart. Myocarditis heart disease is a virus infected disease that changes the shape and structure of muscles. The symptoms of this rare disease are fever, joint pain and legs bulging [4][9][20].

With the advancement in the health sciences, the symptom analysis can be analyzed at the earlier

1 Department of Computer Science and Applications, MDU, Rohtak, Haryana, India
ORCID ID : 0000-0003-2385-487X
2 Department of Computer Science and Applications, MDU, Rohtak, Haryana, India
ORCID ID : 0000-0003-3602-1965
3 Department of Computer Science and Applications, MDU, Rohtak, Haryana, India
ORCID ID : 0000-0002-2924-3474
** ritikarathee@gmail.com*

stage by the machine itself to predict the heart disease. Machine learning is one of phenomenon approach to predict the cardiovascular disease accurately. The health information and symptom are collected from patient and a series of data mining operations are applied to predict the heart disease. The health information collected in real-time can have large number of features including patient health history and symptoms. There are many data mining methods to evaluate the significance of available features. These data mining filters are used to reduce the dimension of dataset and to remove various impurities that exist in the dataset [6][7][13].

This paper explored the characteristics, severities, issues, symptoms and preventive methods associated with cardiac diseases. In this section, the severity, symptoms and causes of cardiac diseases are discussed and presented. The section also discussed different kinds of cardiac diseases, their causes and symptoms. The need of cardiac disease prediction is also described in this section. In section II, the usage of machine learning algorithm in medical disease prediction is defined. The standard model with associated stages is described in this section. The machine learning methods used by the researchers are also discussed. In section III, the machine learning methods proposed by the earlier researchers are discussed. The research methods and performance improvement gained by the researchers are discussed in this section. In section IV, the conclusion of the work is provided.

2. Machine Learning Methods for Medical Disease Prediction

Healthcare system is a critical application area that requires expert concern and rules to make the decisions. Data mining and machine learning is gaining the attention of researchers to optimize the performance of healthcare system. Various data mining and machine learning techniques were proposed by the researchers to predict the existence of different diseases. Machine learning systems can predict and classify medical disease accurately. The system must be capable to handle the real-time challenging. The potential risk of patients and conditions should be addressed by the model. The risk analysis based accurate disease prediction is the key challenge for the machine learning models. The involvement of machine learning and AI models within healthcare domain ensures the better treatment and healthy life of patients. Lot of researches has contributed their efforts to improve the strengths of different machine learning and data processing algorithm to predict and classify various diseases. A standard machine learning model for predicting or classifying any medical disease is shown in Figure 1.

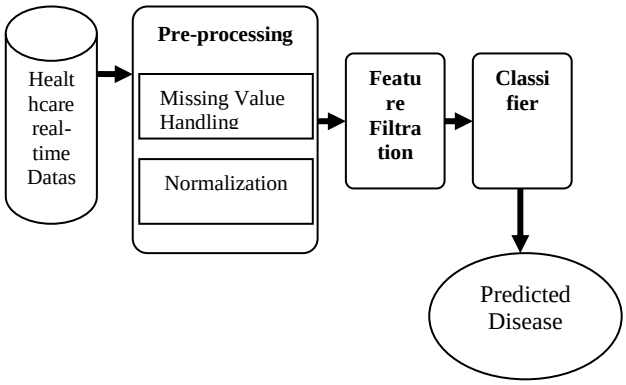

Figure 1: Standard Model for Medical Disease Prediction

The effectiveness of any of machine learning model depends on the authenticity and validity of the training set. The training set must contains the labeled data that can be used for training. The data should be collected from real-time environment. The data should be independent from location, gender and age. For a specific disease analysis, the dataset should include patient basic details, medical history, symptoms and environmental factors. Once the data is collected and available, it can have various real-time issues such as missing data or the representational problem for different features. The processing methods are applied to resolve these issues and to transform the raw data to a standard form. This normalized dataset can have number of features that are less significant to predict a specific disease. In such case, the feature analysis methods are applied to evaluate the significance of each feature. Info Gain, Gain Ratio, PCA, Chi Square analysis etc. are the most and most significant feature selection methods proven in the earlier researches. The feature extraction methods improve the performance of the machine learning model by reducing the dimension and dataset volume. The reliability of the machine learning can be improved by removing the insignificant and non-relevant features from a dataset. The dataset with selected and effective features is finally processed by the classification model to predict or classify the disease. The machine learning models are applied by the researchers in healthcare system to predict different kinds of diseases. Table 1 shows some of

contributions of researches with research methodology for different heart diseases.

Table 1: Machine Learning Models for different Medical Diseases

Author	Disease	Machine Learning Model	Methodology/ Significance
Mohan et al.[24]	Cardiovascular Disease	Hybrid random forest with a linear model (HRFLM)	Used different combinations of features with HRFLM and achieved 88.7% accuracy
Chaubeyet al.[25]	Thyroid Disease	Logistic regression, KNN and Decision Tree	Achieved 96.87% accuracy for KNN based Model
Oriol et al.[26]	Alzheimer Disease	Genetic variation data with machine learning algorithms	95% AUC
Jadhav et al.[27]	Regional Illness	CNN based Uni-model Ailment threat prediction Model	Selection tree based hybrid model achieved 94.8% accuracy
Krishnamoorthi et al.[30]	Diabetes	Unique intelligent diabetes mellitus prediction framework (IDMPF)	Achieved 83% accuracy with lesser error rate

3. Machine Learning Models for Cardiac Disease Prediction

In the recent years, various machine learning, hybrid and optimized learning methods and models were designed and proposed by the researchers for predicting the cardiac disease in the early stage. Various feature selection methods with expert decisions and machine learning [21] methods were proposed to optimize the performance of heart disease prediction. In this section, the studies are presented to explore various methods proposed by the researchers. The functional flow of cardiac disease prediction under expert observation is shown in Figure 2.

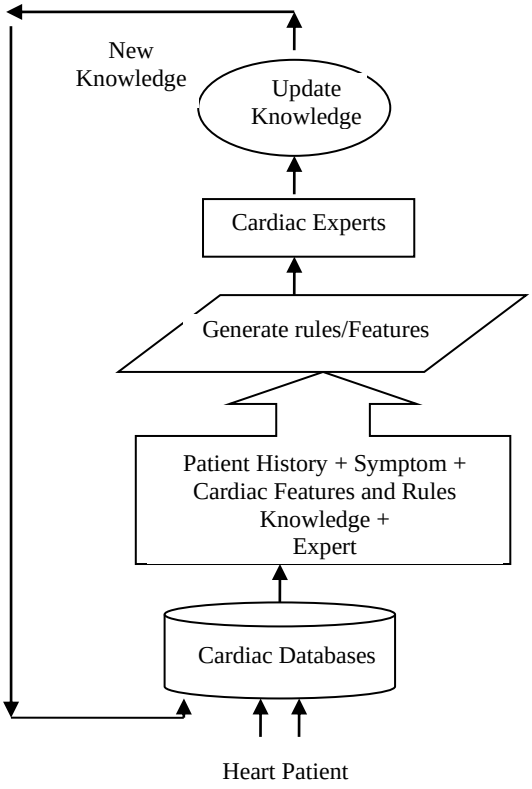

Figure 2: Functional Flow for Cardiac Disease Prediction/Classification

Figure 2 shows the functional flow to predict the cardiac disease at the earlier stage in a real and automated environment. Such automated system can be implemented by adapting any machine learning model. This model is intelligent and learns from a new patient history and medical information is available. This heart patient detail is included in the cardiac database. This updated information is

further used to generate the features for the machine learning model. These features include patient's symptoms, history, cardiac disease features and the associated rules defined by the expert. All this information is used collectively to generate the rules to predict and classify the cardiac disease. The cardiac experts are involved in the system to generate the rules and features and to update the dataset. The more the rules and information will be involved in the system; the maximum accuracy will be achieved. The machine learning and predictive models proposed and optimized by the earlier researchers are discussed in this section.

Mohan et al. [5] proposed hybrid random forest with a linear model (HRFLM) and achieved the accuracy of cardiac disease prediction up to 88.7%. In this model, a less error classifier was integrated as intermediate stage to extract the effective features. This selected feature space was processed by the hybrid classifier for predicting the heart disease. The proposed model achieved the high accuracy in comparison with conventional machine learning algorithms. Li et al. [8] defined an effective feature filtration and classification based model for optimizing the heart disease prediction. The feature selection methods used in this work are Relief, Least absolute shrinkage and minimum redundancy maximal relevance. These methods improved the database reliability by removing the redundant and irrelevant features. Author also proposed a fast conditional mutual information feature selection method to gain high reliability. The filtered dataset was processed by the SVM classifier to predict and classify the heart disease. The proposed model achieved the high accuracy of 92.37%. Ali et al.[16] provided a study on supervised machine learning methods with experimentation and analysis against accuracy and performance measures. The feature ranking method was applied to identify the effective features. This work identified that the KNN, random forest and decision tree methods achieved the maximum accuracy.

The ensemble learning techniques were used by the researchers to combine the features of multiple classifiers and to improve the capability of the classification method. The hybrid classification models were also built by the researchers to improve the accuracy and performance of classification method. Latha et al. [12] provided a comparative and analytical study to identify the performance gain achieved by the ensemble learning algorithms. Author identified that the ensemble learning method can improve the performance of weak classifier by including the feature of other classification method into it. The study was performed on boosting and bagging methods and achieved maximum of 7% increase in accuracy of a weak classifier. Ali et al.[14] proposed a smart health care system by using the ensemble deep learning approach with feature fusion method. The conditional probability and weighted information gain based method is defined for generating the feature weighs. The ensemble deep learning model is applied on selected features and achieved an accuracy of 98.5%. Bharti et al.[15] combined the machine learning and deep learning methods and achieved 94.2% accuracy. The isolation forest feature filter is applied to remove the irrelevant features.

Table 2: Machine Learning Models for Heart Disease Prediction

Author	Machine Learning Model	Methodology	Datasets	Performance
Fitriyaniet al.[10]	Heart Disease Prediction Model	Combined DBSCAN, SMOTE-ENN with XGBoost Classifier	Statlog and Cleveland	95.90% and 98.40% Accuracy
Haqet al.[11]	Hybrid Intelligent System Framework	Applied multiple feature selection methods with different classifiers	Cleveland Heart Dataset	89% Accuracy for Relief Feature Selection +SVM (Linear)

Gokulnathet al.[17]	GA-SVM Model	Genetics based feature selection stage is followed by SVM classifier	Statlog Dataset	88.34% Accuracy
Maji et al.[18]	Hyrbid Model	Combined Decision Tree with Neural Network	Statlog Dataset	78.14% Accuracy
Manikandan et al.[19]	Artificial Fish Swarm Optimization Algorithm	Fish behaviour analysis model was integrated to optimize performance	Statlog Dataset	97.68% Accuracy
Gárate-Escamilaet al.[22]	Hybrid Feature selection CHI-PCA with Random Forest	Used Chi Square and PCA methods for Feature Selection and Random Forest for Classification	Cleveland, Hungarian and Cleveland Hangarian Datasets	98.7% Accuracy 99% Accuracy 99.4% Accuracy
Sharma et al.[23]	Optimized Model with MAPO based Feature Selection Method	Modified Artificial Plant Optimization is combined with different machine learning algorithms	Statlog Dataset. Framingham Heart Study Dataset	87.25% dimension reduction and 90% accuracy
Singh et al.[28]	Genetic with Naïve Bayes classification	Genetic algorithm is integrated within probabilsistic estimation to optimize performance of Naïve Bayes Classifier	Statlog Dataset	97.14% Accuracy
Arabasadiet al.[29]	Genetic with Artificial Neural Network	Used genetic for feature selction and ANN for classification	Cleveland Dataset	93.85% Accuracy

4. Conclusion

Cardiac disease is a critical disease with a high treatment cost and death rate. The causes of occurrence of this disease are also varied and high in the busy work environment of an individual. Early prediction of cardiac disorder symptoms is the only way to save a life. The researchers have proposed various machine learning models to predict cardiac disease at an early stage. This paper provided a detailed exploration of cardiac disease, its symptoms, risks, and behavior. The paper described the stages of a functional machine learning model. The contribution of the researchers

in optimizing each of the stages is also described in this paper. The paper setup the base to design a new cardiac disease prediction model by applying the appropriate feature selection and disease prediction methods. The hybrid model can be designed by combining highly effective feature selection and prediction stages.

References

[1] Celano, Christopher M., and Jeff C. Huffman. "Depression and cardiac disease: a review." Cardiology in review 19, no. 3 (2011): 130-142.

[2] Lespérance, François, and Nancy Frasure-Smith. "Depression in patients with cardiac disease: a practical review." Journal of psychosomatic research 48, no. 4-5 (2000): 379-391.

[3] Huffman, Jeff C., Christopher M. Celano, Scott R. Beach, Shweta R. Motiwala, and James L. Januzzi. "Depression and cardiac disease: epidemiology, mechanisms, and diagnosis." Cardiovascular psychiatry and neurology 2013 (2013).

[4] Del Re, Dominic P., DulguunAmgalan, Andreas Linkermann, Qinghang Liu, and Richard N. Kitsis. "Fundamental mechanisms of regulated cell death and implications for heart disease." Physiological reviews 99, no. 4 (2019): 1765-1817.

[5] Mohan, Senthilkumar, ChandrasegarThirumalai, and Gautam Srivastava. "Effective heart disease prediction using hybrid machine learning techniques." IEEE access 7 (2019): 81542-81554.

[6] Amin, Mohammad Shafenoor, Yin Kia Chiam, and Kasturi DewiVarathan. "Identification of significant features and data mining techniques in predicting heart disease." Telematics and Informatics 36 (2019): 82-93.

[7] Kavitha, R., & Kannan, E., 2016. An efficient framework for heart disease classification using feature extraction and feature selection technique in data mining. International Conference on Emerging Trends in Engineering, Technology and Science (ICETETS), Pudukkottai, pp. 1-5.

[8] Li, Jian Ping, Amin UlHaq, Salah Ud Din, Jalaluddin Khan, Asif Khan, and Abdus Saboor. "Heart disease identification method using machine learning classification in e-healthcare." IEEE Access 8 (2020): 107562-107582.

[9] De Hert, Marc, Johan Detraux, and Davy Vancampfort. "The intriguing relationship between coronary heart disease and mental disorders." Dialogues in clinical neuroscience 20, no. 1 (2018): 31.

[10] Fitriyani, Norma Latif, Muhammad Syafrudin, GanjarAlfian, and Jongtae Rhee. "HDPM: an effective heart disease prediction model for a clinical decision support system." IEEE Access 8 (2020): 133034-133050.

[11] Haq, Amin Ul, Jian Ping Li, Muhammad Hammad Memon, Shah Nazir, and Ruinan Sun. "A hybrid intelligent system framework for the prediction of heart disease using machine learning algorithms." Mobile Information Systems 2018 (2018).

[12] Latha, C. Beulah Christalin, and S. Carolin Jeeva. "Improving the accuracy of prediction of heart disease risk based on ensemble classification techniques." Informatics in Medicine Unlocked 16 (2019): 100203.

[13] Wang, Heru, Jinlong Wei, Qingshuang Zheng, Lingbin Meng, Ying Xin, Xia Yin, and Xin Jiang. "Radiation-induced heart disease: a review of classification, mechanism and prevention." International journal of biological sciences 15, no. 10 (2019): 2128.

[14] Ali, Farman, Shaker El-Sappagh, SM Riazul Islam, Daehan Kwak, Amjad Ali, Muhammad Imran, and Kyung-Sup Kwak. "A smart healthcare monitoring system for heart disease prediction based on ensemble deep learning and feature fusion." Information Fusion 63 (2020): 208-222.

[15] Bharti, Rohit, Aditya Khamparia, Mohammad Shabaz, Gaurav Dhiman, Sagar Pande, and Parneet Singh. "Prediction of heart disease using a combination of machine learning and deep learning." Computational intelligence and neuroscience 2021 (2021).

[16] Ali, Md Mamun, Bikash Kumar Paul, Kawsar Ahmed, Francis M. Bui, Julian MW Quinn, and Mohammad Ali Moni. "Heart disease prediction using supervised machine learning algorithms: performance analysis and comparison." Computers in Biology and Medicine 136 (2021): 104672.

[17] Gokulnath, Chandra Babu, and S. P. Shantharajah. "An optimized feature selection based on genetic approach and support vector

machine for heart disease." Cluster Computing 22, no. 6 (2019): 14777-14787.

[18] Maji, Srabanti, and Srishti Arora. "Decision tree algorithms for prediction of heart disease." In Information and communication technology for competitive strategies, pp. 447-454. Springer, Singapore, 2019.

[19] Barani, A. M., R. Latha, and R. Manikandan. "Implementation of Artificial Fish Swarm Optimization for Cardiovascular Heart Disease." International Journal of Recent Technology and Engineering (IJRTE) 8, no. 4S5 (2019): 134-136.

[20] Moholdt, Trine, Carl J. Lavie, and Javaid Nauman. "Sustained physical activity, not weight loss, associated with improved survival in coronary heart disease." Journal of the American College of Cardiology 71, no. 10 (2018): 1094- 1101.

[21] Kannan, R., and V. Vasanthi. "Machine learning algorithms with ROC curve for predicting and diagnosing the heart disease." In Soft Computing and Medical Bioinformatics, pp. 63-72. Springer, Singapore, 2019.

[22] Gárate-Escamila, Anna Karen, Amir Hajjam El Hassani, and Emmanuel Andrès. "Classification models for heart disease prediction using feature selection and PCA." Informatics in Medicine Unlocked 19 (2020): 100330.

[23] Sharma, Prerna, Krishna Choudhary, Kshitij Gupta, Rahul Chawla, Deepak Gupta, and Arun Sharma. "Artificial plant optimization algorithm to detect heart rate & presence of heart disease using machine learning." Artificial intelligence in medicine 102 (2020): 101752.

[24] S. Mohan, C. Thirumalai and G. Srivastava, "Effective Heart Disease Prediction Using Hybrid Machine Learning Techniques," in IEEE Access, vol. 7, pp. 81542-81554, 2019, doi: 10.1109/ACCESS.2019.2923707.

[25] Chaubey, Gyanendra, Dhananjay Bisen, Siddharth Arjaria, and Vibhash Yadav. "Thyroid disease prediction using machine learning approaches." National Academy Science Letters 44, no. 3 (2021): 233-238.

[26] De Velasco Oriol, J., Vallejo, E.E., Estrada, K., Taméz Peña, J.G. and Disease Neuroimaging Initiative, 2019. Benchmarking machine learning models for late-onsetalzheimer's disease prediction from genomic data. BMC bioinformatics, 20(1), pp.1-17.

[27] Jadhav, Saiesh, Rohan Kasar, Nagraj Lade, Megha Patil, and Shital Kolte. "Disease prediction by machine learning from healthcare communities." International Journal of Scientific Research in Science and Technology (2019): 29-35.

[28] Navdeep Singh and Sonika Jindal, " Heart Disease Prediction System using Hybrid Technique of Data Mining Algorithms", International Journal of Advance Research, Ideas and Innovations in Technology, Vol.4, Issue 2, 2018.

[29] Zeinab Arabasadiet al. , " Computer aided decision making for heart disease detection using hybrid neuralnetwork- Genetic algorithm" , Computer Methods and Programs in Biomedicine-ELSEVIER, Vol. 141, pp.19-26, April- 2017.

[30] Krishnamoorthi, Raja, Shubham Joshi, Hatim Z. Almarzouki, Piyush Kumar Shukla, Ali Rizwan, C. Kalpana, and Basant Tiwari. "A Novel Diabetes Healthcare Disease Prediction Framework Using Machine Learning Techniques." Journal of Healthcare Engineering 2022 (2022).