



e-ISSN: 2319-8753 | p-ISSN: 2347-6710

# IJIRSET

International Journal of Innovative Research in  
**SCIENCE | ENGINEERING | TECHNOLOGY**



# INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

Volume 12, Issue 2, February 2023

**ISSN** INTERNATIONAL  
STANDARD  
SERIAL  
NUMBER  
INDIA

Impact Factor: 8.118



9940 572 462



6381 907 438



ijirset@gmail.com



www.ijirset.com

# Survey on Heart Disease Prediction System using Machine Learning

Mansi Dayanand Barbade, Aditi Sidu Gadade, Payoshni Sanjay Gade, Diksha Ramesh Sancheti,  
Prof.Swati B Dhadake

Department of Computer Engineering, Smt Kashibai Navale College of Engineering (SKNCOE), India

**ABSTRACT:** The main cause of death in the world is heart disease. Based on clinical data, a machine learning (ML) system can identify cardiac disease in its early stages and reduce mortality rates. However, ML has consistently struggled with class imbalance and high dimensionality problems, which limit accurate predictive data analysis in numerous real-world applications, including the identification of heart disease. This study suggests a novel approach to resolve these problems and enhance the ability to predict the presence of heart disease and patients' survival. It uses supervised infinite feature selection (Inf-FSSs) to identify the most important features, Improved Weighted Random Forest (IWRF) to forecast heart disease, and Bayesian optimization to fine-tune the new hyperparameters for IWRF. Performance measurements like accuracy and f-measure are used to compare the proposed model to other hybrid models and demonstrate its superiority. According to the results, the suggested Inf-FSS-IWRF model outperformed other models in terms of accuracy and F-measure for both datasets. The suggested model outperformed the others by an accuracy improvement of 2.4% and 4.6% on both datasets, respectively, in a comparison study to compare with earlier investigations.

**KEYWORDS:** CVD detection, heart disease classification, feature selection, random forest, imbalance, Bayesian optimization.

## I. INTRODUCTION

The most common cause of death in the world is cardiovascular disease (CVD), also referred to as heart disease. The World Heart Federation has conducted research that found that cardiovascular disease accounts for one out of every three fatalities [1]. The World Health Organization (WHO) predicts that by 2030, over 23.6 million individuals would pass away from cardiovascular disease (CVD), largely from heart failure and strokes [2]. The only way to halt this form of mortality and lower the overall death toll is to prevent CVD at an early stage. Due to a number of contributing factors, including high cholesterol, high blood pressure, smoking, diabetes, being overweight and obese, and many more factors, diagnosing CVD can be challenging.

To get effective results in the medical field, clinical decision-making based on hybrid machine learning (ML) models is applied. The clinical datasets do, however, present certain difficulties, primarily because of their high dimensionality and class imbalance. Applying ML without addressing these problems consequently has an impact on the techniques' accuracy [4]. The literature has therefore presented a variety of ML-based prediction techniques for CVD detection and survival. Earlier studies focused on feature selection and employed several ML algorithms to identify CVD (FS). To predict cardiac disease, using rough sets (RS) to identify the most important features and feeding them to the chaos firefly algorithm [5] and backpropagation neural network (BPNN) [6] (HD).

## II. RELATED WORK

**R. ThangaSelvi, I. Muthulakshmi-** This paper provides an extensive health application system based on optimum artificial neural networks (OANN) for the diagnosis of heart disease, which is regarded as the world's worst disease. The teaching and learning-based optimization (TLBO) algorithm for ANN and the distance-based misclassified instance removal (DBMIR) process are the two primary components of the proposed OANN (TLBO-ANN). A Big Data framework like Apache Spark is used to create the suggested model. The two modes of operation for the presented OANN model are offline prediction and online prediction.

**C. Sowmiya, P. Sumitra-** The most significant topic in the analysis of clinical data is purposefully the prediction of heart disease from the clinical data. Data size is enormous, particularly in the healthcare industry. Data mining (DM)

uses unprocessed medical data to help in prediction and decision-making. The vast collection is transformed into informative data through DM. The data mining techniques are currently used in several research to predict cardiac disease. There hasn't been much research done on identifying the key characteristics that are significant in predicting heart disease. The purpose of this work is to develop an improved method for predicting mortality in patients with congestive heart failure using unique feature selection and classification techniques.

**Davide Chicco and Giuseppe Jurman**- In this study, we examine data from 299 heart failure patients who were gathered in 2015. We use a number of machine learning classifiers to rank the features that correlate to the most significant risk variables as well as forecast the patients' survival. By using conventional biostatistical tests, we also conduct a different feature ranking analysis and compare the outcomes to those produced by the machine learning algorithms. Serum creatinine and ejection fraction are categorically identified as the two most important characteristics by both feature ranking algorithms, hence we base our machine learning survival prediction models only on these two variables.

**Bayu Adhi Tama and Sunghoon Lim**-The choice to use classifiers is frequently arbitrary because it mostly depends on the type of data and classification task. In this work, a wide-array classifier from six different families—tree, ensemble, neural, probability, discriminant, and rule-based classifiers—is evaluated in comparison to one another. To show the generalizability of the classifiers in multiple illness prediction, a variety of real-world publically available datasets from various diseases are considered in the experiment.

**Bayu Adhi Tama, Sun Im and Seungchul Lee**-This article discusses a novel CHD detection method based on machine learning methods, such as classifier ensembles. The result is the construction of a two-tier ensemble in which some ensemble classifiers are used as base classifiers for another ensemble. The class label prediction of three ensemble learners, namely random forest, gradient boosting machine, and extreme gradient boosting, is blended using a stacked architecture. Multiple heart disease datasets, including Z-Alizadeh Sani, Statlog, Cleveland, and Hungarian, are used to evaluate the detection model, supporting its generalizability.

**A. Michael Morey, Frédéric Noo**-For cancer imaging, positron emission tomography (PET) images are commonly reconstructed with an in-plane pixel size of about 4 mm. This study set out to determine how employing smaller pixels affected overall oncologic lesion detection. Using experimental phantom data from the Utah PET Lesion Detection Database, which simulated whole-body FDG PET cancer imaging for a 92 kg patient, a number of observer studies were carried out. On a BiographmCT time-of-flight (TOF) PET/CT scanner, 24 scans were collected over the course of 4 days, with up to 23 lesions (diam. 6–16 mm) scattered throughout the phantom on each scan. Using ordered-subsets expectation-maximization (OSEM) with and without point spread function (PSF) modelling and TOF, images with 2.036 mm and 4.073 mm pixels were reconstructed.

**NAN ZHAO, AIFENG REN, MASOOD UR REHMAN, ZHIYA ZHANG, XIAODONG YANG, AND FANGMING HU**-This paper investigates the implementation of a physical layer security algorithm as an alternative. The algorithm is based on the behavior fingerprint developed using the wireless channel characteristics. The usability of the algorithm is established through experimental results, which show that this authentication method is not only effective, but also very suitable for the energy-, resource-, and interface-limited WBAN medical applications.

**ABID ISHAQ, SAIMA SADIQ, MUHAMMAD UMER 1, SALEEM ULLAH, SEYEDALI MIRJALILI, VAIBHAV RUPAPARA, AND MICHELE NAPPI**-This study examines the 299 hospitalised patients with heart failure who survived. The goal is to identify key characteristics and efficient data mining approaches that can improve the precision of the survivor prediction for cardiovascular patients. This study uses nine classification models to predict the patient's survival: the Decision Tree (DT), Adaptive Boosting Classifier (AdaBoost), Logistic Regression (LR), Stochastic Gradient Classifier (SGD), Random Forest (RF), Gradient Boosting Classifier (GBM), Extra Tree Classifier (ETC), Gaussian Naive Bayes Classifier (G-NB), and Support Vector Machine (SVM).

**NORMA LATIF FITRIYANI, MUHAMMAD SYAFRUDIN, GANJAR ALFIAN, AND JONGTAE RHEE**-This study suggests an efficient heart disease prediction model (HDPM) for a CDSS, which consists of XGBoost to predict heart disease, a hybrid Synthetic Minority Over-sampling Technique-Edited Nearest Neighbor (SMOTE-ENN) to balance the training data distribution, and Density-Based Spatial Clustering of Applications with Noise (DBSCAN) to detect and eliminate outliers static data.



**ZhenhaoJiang ,Tingting Pan , Chao Zhang and Jie Yang**—The number of synthetic samples produced by SMOTE for each positive sample depends on the categorization contribution degree. OS-CCD eliminates oversampling from noisy spots and adheres to the spatial distribution characteristics of the original samples on the class border. AUC, ROC, F1-score, and accuracy tests on twelve benchmark datasets show that OS-CCD beats six traditional oversampling approaches.

### III. EXISTING SYSTEM/OPEN ISSUES

The EHDPS predicts the likelihood of patients getting heart disease. It enables significant knowledge, eg, relationships between medical factors related to heart disease and patterns, to be established. We have employed the multilayer perceptron neural network with backpropagation as the training algorithm.

Negative comments

here are instruments available which can predict heart disease but either they are expensive or are not efficient to calculate chance of heart disease in human. Early detection of cardiac diseases can decrease the mortality rate and overall complications.

### IV. CONCLUSION

In order to predict the existence of CVD and cardiac patient survival, this paper attempts to offer an accurate and effective machine learning ensemble model. The suggested model incorporates BO, IWRF, and Inf-FSs. With the help of these three techniques, the most important features are chosen, the imbalanced data classification problem that arises in medical datasets is addressed, and the weighting factor is adjusted. Two open datasets are used to test the proposed model and to compare it to earlier research.

### REFERENCES

1. R. T. Selvi and I. Muthulakshmi, “An optimal artificial neural networkbased big data application for heart disease diagnosis and classification model,” J. Ambient Intell. Humanized Comput., vol. 12, no. 6, pp. 6129–6139, Jun. 2021.
2. C. Sowmiya and P. Sumitra, “A hybrid approach for mortality predictionfor heart patients using ACO-HKNN,” J. Ambient Intell. Humanized Comput., vol. 12, no. 5, pp. 1–8, 2020.
3. D. Chicco and G. Jurman, “Machine learning can predict survival ofpatients with heart failure from serum creatinine and ejection fractionalone,” BMC Med. Informat. Decis. Making, vol. 20, no. 1, pp. 1–16, Dec. 2020.
4. B. A. Tama and S. Lim, “A comparative performance evaluation of classification algorithms for clinical decision support systems,” Mathematics, vol. 8, no. 10, p. 1814, Oct. 2020.
5. B. A. Tama, S. Im, and S. Lee, “Improving an intelligent detection systemfor coronary heart disease using a two-tier classifier ensemble,” BioMedRes. Int., vol. 2020, Apr. 2020, Art. no. 9816142.
6. Michael Morey, FrédéricNoo, “Effect of Using 2 mm Voxels on Observer Performance for PET Lesion Detection” Member, IEEE, and Dan J. Kadrmas, Member, IEEE
7. Michael Morey, FrédéricNoo, “Biometric Behavior Authentication Exploiting Propagation Characteristics of Wireless Channel”
8. ABID ISHAQ, SAIMA SADIQ, MUHAMMAD UMER , SALEEM ULLAH ,SEYEDALI MIRJALILI, VAIBHAV RUPAPARA, AND MICHELE NAPPI, “Improving the Prediction of Heart Failure Patients’ Survival Using SMOTE and Effective Data Mining Techniques”
9. N. L. Fitriyani, M. Syafrudin, G. Alfian, and J. Rhee, “HDPM: An effectiveheart disease prediction model for a clinical decision support system,” IEEE Access, vol. 8, pp. 133034–133050, 2020.
10. Z. Jiang, T. Pan, C. Zhang, and J. Yang, “A new oversampling methodbased on the classification contribution degree,” Symmetry, vol. 13, no. 2, p. 194, Jan. 2021.



**INNO SPACE**  
SJIF Scientific Journal Impact Factor  
**Impact Factor: 8.118**



**ISSN** 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