International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 11 Issue: 07 | July 2024 www.irjet.net p-ISSN: 2395-0072

Heart Disease Identification Methods using Machine Learning and efficient data balancing techniques

Ranu Masram[1], Prof. Sudeep Kishore Sharma[2], Mr.Nagendra Kumar[3]

[1]Reseacrh Scholar, Department of CSE, Shri ram Institute of Science and Technology, Jabalpur, M.P.
[2]Professor, Departmet of CSE, Shri Ram Institute of Science and Technology, Jabalpur, M.P.
[3]Professor, Departmet of CSE, Shri Ram Institute of Science and Technology, Jabalpur, M.P.

Abstract –A heart attack is a life-threatening event that is very difficult to predict. Early diagnosis and prompt treatment can reduce mortality. According to the British health foundation (b.h.f.), 1 In 14 people worldwide have heart disease or heart disease. It is also estimated that approximately 200 million people suffer from heart disease. Medical records consist of many different sections collected from different sources, each offering a different perspective on the patient's condition. Machine learning has proven to be an excellent method for predicting non-standard data. Algorithms such as svc, k nearest neighbor, decision tree, random forest classifier, XGBoost and cnn can be used for early detection of viruses. Data mining techniques are used to collect data from clinical sites and use it to determine initial diagnoses of diseases without the need for intervention from doctors. Application form. Machine learning algorithms are used especially to predict the risk of heart attack. This article is about using machine learning to predict a patient's risk of heart attack. This research paper focuses on classification techniques in machine learning in healthcare with the aim of finding similar models that can make important predictions and help in early detection of diseases. Data on various characteristics such as age, gender and cholesterol levels were used to develop the prediction models. The model was trained and tested on various data sets to determine its accuracy and ability to predict heart disease. The results of this study can be used to develop more accurate methods to predict heart disease risk and reduce the number of deaths from heart disease.

Keywords: Heart attack prediction, Data balancing, Data Pre-processing, Machine learnings, SMOTE, DBSMOTE, CNN.

1. INTRODUCTION

Cardiovascular disease, also known as cardiovascular disease (cvd), is the leading cause of death worldwide. In a recent study, the world heart federation found that one-third of deaths was caused by heart disease [1]. According to statistics from the world health organization (who), more than 23.6 million people may die from cvd, mainly stroke and heart failure, by 2030 [2]. Cvd can be caused by many factors, including stress, alcohol, smoking, poor diet, poor lifestyle, and other health problems such as high blood pressure or diabetes. However, most cvd diseases are completely treatable when diagnosed early [3]. In this case, doctors need to approach the diagnosis and prognosis of heart failure more specifically. New techniques for data analysis can lead to early diagnosis of cvd through evaluation of patients' medical records [2]. Forest and finally diagnosis. The testing process will take longer than expected depending on people's health and well-being, natural and genetic factors in lifestyle. However, now healthcare can be done by estimating and measuring risk for more serious prevention of infection, thus achieving results resulting in better health (4). This has led to the development of advanced techniques for analyzing heart disease (hd) clinical data to identify advanced failures. Many studies have used machine learning (ml) algorithms to predict cvd from clinical data. However, medical records are still highly problematic due to class inequality and size. Therefore, using machine learning without solving these problems reduces the efficiency and accuracy of the method. Previous researchers focused on specific selection (fs) and used various machine learning methods to predict cvd. Uzun et al. [5] Developed a heart disease decision system based on rough system (rs) and chaos firefly algorithm (cfars-ar) to select the best features, and then obtained a type 2 fuzzy logic system for HD detection. The design achieved an accuracy rate of 88.3%. Authors in [6] combined rs with Backpropagation neural network (bpnn) to predict cvd. Additionally, Dwivedi performed the comparison of hd prediction using different ml models such as artificial neural network (ann), logistic regression (lr), classification tree and naive bayes (nb). The authors concluded that lr outperformed other methods in detecting cvd [7]. Additionally, huck et al. A comparative study was conducted using different ml models (e.g. Ann, rf and lr) with different fs (e.g. Click). The authors report that removing features affects the performance of the model. This study concluded that lr with click fs achieved an accuracy of 89% compared to other methods used in the same study [8]. Amin et al. Comparative analysis, analyzing the main features and using seven ml models including ann, lr, decision tree (dt), support vector machine (svm), nearest neighbor (knn), nb and hybrid model (voting with lr and voting). Was carried out. Note) [9]. The results showed that the hybrid model achieved the best accuracy (87.41%). In another study, a hybrid model consisting of rf and linear model (hrflm) was developed to improve the accuracy of cvd prediction. This work has been applied to different variants of the traits; later, vijayashree et al. A recent project introduced a weighting decision based on population

diversity and optimization process. They also showed how the pso system can be improved and reduced the number of features with the help of svm. Therefore, their method achieved an accuracy of 84.36% [11]. In a successful attempt to improve accuracy, ali et al. The hd detection method was improved by adding two sets of svms. The first svm removes irrelevant features, while the second svm is used to predict presence and absence. Therefore, the model developed by stacked achieved an accuracy of 92.22% [12]. Additionally, gupta et al. Develop a machine learning algorithm for hd diagnosis based on famd and rf. Famd selects baseline features and rf predicts the absence and presence of hd. The results showed that the proposed model achieved an accuracy of 93.44% [13]. The author of [14] conducted a comparative study. In most cases, authors use different types of classification. Conditional inference tree forest (cforest) outperforms other classifiers. Tamar et al. [15] Introduced a 2-layer model for feature selection based on particle swarm optimization (pso) for cardiovascular disease prediction. The accuracy of the model is 93.55%. However, the above studies have some limitations in detecting hd using the proposed method due to insufficient clinical data. Several studies have been conducted on the creation of decision support and balanced strategies to solve the above problems. Fitriani et al. nearby neighbors heart determining predictions are made. Additionally, the xgboost classifier can predict the patient. The established method achieved an accuracy of 95.9% [16]. Waqar et al. Proposed smote-based deep learning to predict heart attack. The authors use the smote method to balance the data without making any special selections. Parallel data was trained and tested with a deep neural network to predict the presence of a heart attack, achieving 96% accuracy [17]. Recently, Ishaq et al. Smote was used to balance the profile distribution and extreme tree (et) of options to predict patient survival using rf importance ranking [18]. Some researchers also use mixed methods to examine disease severity, Salary et al. A hybrid approach is introduced, comprising a genetic algorithm (ga) for feature selection, modified kNN, and a recurrent network for weight distribution. The results showed that this method achieved an accuracy of 62.1% [19]. In the same context, Wiharto et al. A hybrid method based on binary tree (BT) and svm is proposed to predict HD disease. This study achieved an accuracy of 61.86% [20].

Unlike Khateeb et al. Perform inverse filtering on the data and then use kNN (ibk) to estimate the significance level. The accuracy of this model is as high as 79.20% [21]. Magesh and Swarnalatha developed the optimal fs method using cluster-based dt learning (cdtl) to find important features and then using rf to estimate the weights. The developed model achieved a prediction accuracy of 89.3% [22]. Recently, authors in [23] developed a decision for hd weight prediction based on ml fusion approach. The results showed that the proposed model achieved a prediction accuracy of 75%.

Most studies focus on the importance of hd features in selecting the best feature and removing artifacts to improve the performance of machine learning models. However, in the field of machine learning, data inconsistencies and hyperparameters changes can occur and affect the performance of predictive models, as many machine learning methods learn no differences between the best ones to achieve optimal performance across various tasks and datasets. Most focus on complex machine learning with many hyperparameters and large datasets. However, none of the previous studies have combined the hyperparameters optimization (hpo) method with parallel data and machine learning techniques in a way that affects the performance of the model. Previous studies have reported improved classification model performance compared to bayesian optimization (bo) and random search (rs) by integrating hyperband (hb) into ml hpo [24] , [25] . However, to our knowledge, there are no studies combining hb to improve smote and ml to predict the presence and severity of cvd. This article is believed to be the first to study cvd detection and severity classification and use hpo to predict heart disease. We use smote, Undersampling, and oversampling to help resolve the inconsistency in both estimation problems. We are working on ml algorithm variants to predict heart failure and its severity: random forest, logistic regression, knn, decision tree (dt), XGBoost and cnn. The selected classifiers were optimized using hb together with smote to find the hyperparameters with the best performance, since there are many good ml hyperparameters for different data.

2. Related Work

Poor functioning of the heart for this reason is called heart disease. There are many types of heart disease. The most common types are heart failure (hf) and coronary artery disease (cad). The main cause of heart failure (hf) is blockage or narrowing of the arteries. Coronary arteries also supply blood to the heart. Cad is a type of heart disease and is a well-known cardiovascular disease worldwide. Heart failure is a pressing health problem in the world today, with 26 million adults worldwide reported to suffer from heart disease. According to reports, approximately 31% of deaths worldwide are caused by heart failure. In 2005, heart failure was the leading cause of death worldwide, causing 17.5 million deaths; more than 80% of these occurred in low- and middle-income countries such as china. Additionally, heart failure is expected to account for more than 80% of future deaths in developing countries. Heart disease-related events and deaths have also increased among women; heart failure has become the leading cause of death in china. Different factors have been identified through clinical and clinical studies as risk factors for heart failure and cad. Risk factors fall into two broad categories. The first category includes non-modifiable risk factors such as age, gender, family history. The second category includes modifiable risk factors such as smoking, healthy diet, and high cholesterol.

Therefore, the risks in the second category can be eliminated or controlled with lifestyle changes and drug therapy.

To diagnose or predict heart failure, doctors use a variety of methods. Currently, angiography is the most commonly used method by doctors to diagnose cad and is considered the best method to diagnose cad or hf caused by cad. However, the high cost and side effects of angiography are significant limitations in diagnosing heart disease or cardiac disease. In addition, the diagnosis of cad requires detailed examination of many factors, which makes the doctor's job difficult. These questions have inspired the development of noninvasive heart failure detection mechanisms. Additionally, the heart failure diagnostic process focuses on reviewing the patient's medical history, the cardiologist's assessment of various symptoms, and physical examination reports. Therefore, traditional methods used to diagnose heart failure are time-consuming and can lead to misdiagnoses due to human error. Therefore, to avoid these problems, we need to develop an automatic learning system that can effectively and quickly diagnose heart failure. As can be seen from the literature, many automated diagnostic methods have been proposed in the past to diagnose heart failure using various machine learning techniques, for example, naive bayes (nb), support vector machine (svm), stacked and optimized. Integration of svm, fuzzy logic, artificial neural network (ann), deep neural network (dnn) and ann. Canfila et al. A classification method based on multilayer perceptron (mlp), Backpropagation learning algorithm and biomedical test values is proposed to diagnose heart failure with feature selection algorithm.

Thanks to the feature selection process, the total number of 13 features is reduced to 8. The accuracy of the training data is 89.56%, and the accuracy of the validation data is 80.99%. Paul et al. Fuzzy decision support system (fdss) was introduced for the diagnosis of heart disease. The proposed system achieved 80% accuracy. Verma et al. A new hybrid method has been proposed for the diagnosis of coronary artery disease (cad). The accuracy of this method is 88.4%. The proposed model increased the performance of the classification algorithm on the Cleveland dataset by 11.4%. Shah et al. proposed an elimination-based method to reduce its value. The plan uses principal component analysis (ppca). The proposed method achieved an accuracy of 82.18% on the Cleveland dataset. Dwivedi et al. Tested the performance of different learning systems in predicting heart failure [26]. The maximum distribution based on logistic regression is 85%. Amin et al. Evaluated different data mining methods and identified important features for predicting heart failure [27]. According to experimental results, the most effective data mining technology in terms of classification accuracy in predicting heart failure is 87.4%. Recently, Resul et al. With the improvement of neural network model integration, the accuracy of detecting heart failure reaches 89.01%, the sensitivity reaches 80.95%, and the specificity reaches 95.91%. Samuel et al. By combining ann and fuzzy analysis hierarchy (fuzzy-ahp), a new hybrid decision support system is proposed, which increases the heart failure accuracy to 91.10%. Additionally, ali et al. With the development of a new diagnostic method combining two svm, the accuracy of predicting heart failure increased to 92.22%. The first svm is used to select important features, while the second svm is used for prediction. Both models have been optimized using new search methods. Recently, Paul et al. With the recommendation of the modified weight fuzzy system, the accuracy of predicting heart failure was increased to 92.31%. However, the accuracy of the prediction still needs a lot of improvement. The same process is repeated until the number of values (generated from random locations) reaches the size of the eigenvectors. It should be noted that whenever the feature subset is used for the rf algorithm for classification and the grid search algorithm is used to search for the best hyperparameters of the rf. The heart is the hardest working organ of our body. Important factors for future heart disease are high cholesterol, high blood pressure, diabetes, and people with a family history of this disease. Heart diseases are often caused by lifestyle. The cad is the artery that supplies oxygen and blood to the heart. It is the cause of death almost everywhere in the world. Coronary artery disease usually occurs when cholesterol accumulates on the artery walls and forms plaque. The arteries become compressed, making it difficult for blood to travel to the heart. It can also cause diamond-shaped blood clots to block all blood flow, which when the blockage occurs is called coronary occlusion. Therefore, the intervention is blockage of the heart (myocardial infarction). Contraction of the heart muscle suddenly stopped. Blood viscosity plays an important role in maintaining vascular homeostasis. Hematocrit is the cell volume (pcv) of blood. Various machine learning algorithms are used to check large amounts of data and accuracy. The proposed model [28] uses extreme gradient boost, DT, kNN, svm, naive bayes, random forest, Ann and hyperparameters-tuned random forest algorithm for cvd prediction. . They are general and useful methods in machine learning. Random forest identifies the most important features of a particular problem. Heart disease is a disease of the heart and many diseases that affect the heart. When you have heart disease, heart problems can occur due to disease in the heart and damage to the blood vessels. Coronary heart disease is a heart disease that reduces blood flow to the heart. Heart disease occurs when blood flow is low. It is necessary to evaluate and predict a heart attack based on symptoms. This article took and analyzed a sample of existing data on patients with heart failure symptoms [29]. Use python language to predict the truth. As we all know, most of the diseases that cause death in the world are heart diseases and then cancer, which is a very long and dangerous disease that occurs in the same people all over the world.

Diseases and problems like this do not appear immediately. Scientists and doctors have shown that this is an ongoing process that occurs from time to time, with some symptoms appearing suddenly. Finally, when the heart stops, the heart

cannot pump the blood it needs to all parts of the body, and due to the blood vessels in the arteries, the heart cannot receive enough blood as itself. To heart failure and death. This article provides an overview of the data science and algorithms to create a hybrid model that can predict heart disease in patients over a sufficiently long period of time. Additionally, the system must inform patients in advance about the benefits and preventive measures in line with internationally accepted standards. The hybrid model aims to create real-world solutions with the help of data science algorithms, a program of naive bayes, artificial neural networks, support vector machines and hybrid naive bayes, support vector machines and artificial neural networks to predict and report heart disease patients. The accuracy, specificity, and sensitivity of naive bayes, artificial neural networks, support vector machines, and hybrid naive bayes, support vector machines, and artificial neural networks were evaluated. The combined model increased the accuracy of predicting heart disease by about 2 percent. The composite model also showed high specificity and sensitivity of 82.11% and 91.47%, respectively. This article also looks at different traits and shows a strong link to heart disease, including genetics, physical activity, total fat intake, stress, and work. This study [31] suggests new directions in research, and these ideas can be used for smart devices and research data to improve the diagnosis and treatment of heart disease.

3. Proposed Work

Figure 3.1 represents the architecture of the proposed prediction system. First step is to collect heart disease dataset. After this data pre-processing will be performed. After pre-processing the dataset, it is split into training set and testing set. Training set is used to train the algorithm and testing set is used for testing purpose. Proposed algorithm takes training dataset as the input to train on various samples and produces a trained model based on proposed algorithm. Testing data is than applied on the model for performance evaluation. For predicting the result, best performed model will be applied. The proposed system uses multiple Undersampling techniques and Multiple Oversampling techniques as well as DBSMOTE for balancing the dataset for improving the performance of the models.

This is different from oversampling that involves adding example to the minority class to reduce the skew in the class distribution. In case of Undersampling techniques, randomly data is selected from majority class and delete it from training dataset. This is called Random Undersampling.

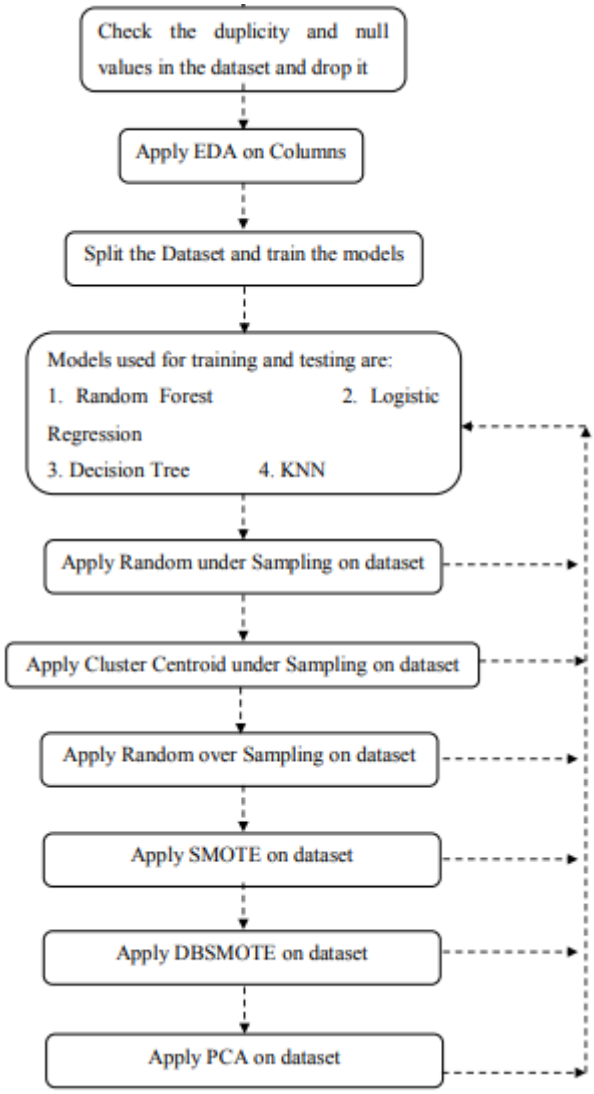

Figure 3.1: Proposed model architecture.

SMOTE is also an oversampling techniques. SMOTE is applied to balance the imbalanced classifiers that have many challenges. With imbalanced classifiers, many algorithms have poor performance on the minority class.

Perhaps the most widely used approach to synthesizing new examples is called the **Synthetic Minority Oversampling TEchnique**, or SMOTE for short. The approach is effective because new synthetic examples from the minority class are created that are plausible that is, are relatively close in feature space to existing examples from the minority class.

DBSMOTE generates synthetic instances along a shortest path from each positive instance to a pseudo-centroid of a minority-class cluster. Consequently, these synthetic instances are dense near this centroid and are sparse far from this centroid.

In data cleaning & pre-processing, system will first check the name of duplicate data and removes it. Now, system will

check missing entries in the dataset column wise. After this, there are no missing entries in the dataset. Next we will move towards exploring the dataset by performing detailed EDA. In EDA, firstly we will check the shape of the dataset with important columns one by one. A check for summary statistics of numerical columns, distribution of heart disease, gender, age wise distribution chest pain type, rest ECG and ST slope will be performed.

The next step is to perform data balancing and removal of unbalanced data. The system uses Random Undersampling, Oversampling, SMOTE and DBSMOTE for data balancing and testing score method for this. The dataset is divided into train and test set. Than in feature normalization all the numeric features in the range of 0 to 1. In cross validation, we will build different baseline models and perform cross validation to filter top performing baseline models to be used in level 0. After this model building and evaluation has been performed on the various parameters. On the basis of result prediction has been performed.

4.4 Proposed Algorithm

Step 1: Read the dataset.

Step 2: Normalize the numerical values of data.

Step 3: Delete the null values and duplicate records.

Step 4: Train the six classifiers one by one.

Step 5: Test the classifiers and predict the results.

Step 6: Apply random Under Sampling techniques on datasets.

Step 7: Repeat step 4 and 5.

Step 8: Apply Cluster Centroid Sampling techniques on datasets.

Step 9: Repeat step 4 and 5.

Step 10: Apply Random over sampling techniques on datasets.

Step 11: Repeat step 4 and 5.

Step 12: Apply SMOTE techniques on datasets.

Step 13: Repeat step 4 and 5.

Step 14: Apply DBSMOTE techniques on datasets.

Step 15: Repeat step 4 and 5.

Step 16: Apply PCA techniques on datasets.

Step 17: Repeat step 4 and 5.

Step 18: End of algorithm.

Basically, we felt the need to improve the current studies in this field and analyzed previous models to determine what might be lacking, after which we took the initiative to devise a solution after applying data balancing techniques that might reshape the current datasets and provide a better of results that makes the system suitable for practical implementation.

4. Result Analysis

The classification performance can be evaluated in terms of accuracy. Accuracy explains correctly classified instances of the symptoms with respect to heart disease. The table below shows the Accuracies comparisons of all six classifiers before data balancing and after data balancing techniques.

Table 6.1: Accuracy Comparison before data balancing techniques.

Classifiers Used	Accuracy
Random Forest	89%
Logistic Regression	89.82%
KNN	88.62%
Decision Tree	83.47%
XGBoost	89.83%
CNN	90%

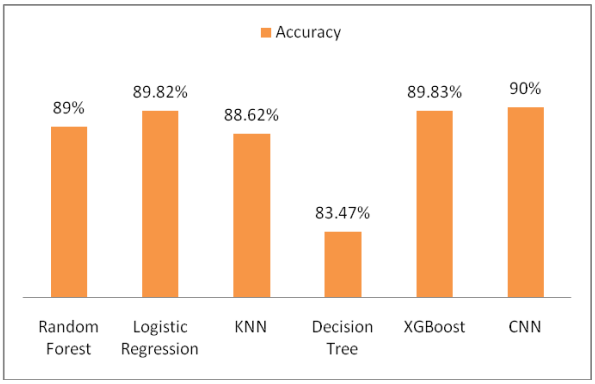

Figure 6.61: Accuracy Comparison Chart.

From above chart it is viewed that the CNN Classifier has better performance as compared to other classifiers i terms of accuracy.

5. Conclusion

The main motivation of this paper is to provide an insight about detecting and curing heart disease using machine learning technique. For this thesis, data were collected from Kaggle Data Sets. All attributes are numeric-valued. These attributes are fed into Random Forest, Logistic Regression, KNN, Decision Tree, XGBoost and CNN, in which CNN gave the better result with the highest accuracy most of times after data balancing techniques applied. Valid performance is achieved using CNN algorithm in diagnosing heart diseases

and can be further improved by increasing the number of attributes.

Thus, in an environment similar to that of the used dataset, if all the features are preprocessed such that they acquire normal distribution, all the data balancing techniques are applied, CNN is a good selection to obtain a robust prediction model. And, such models provide a valuable assistant to the society for health care management domain.

6. References

[1] R. T. Selvi and I. Muthulakshmi, "An optimal artificial neural network based big data application for heart disease diagnosis and classification model," J. Ambient Intell. Humanized Comput. vol. 12, no. 6, pp. 6129–6139, 2021.

[2] G. Bazoukis, S. Stavrakis, J. Zhou, S. C. Bollepalli, G. Tse, Q. Zhang, J. P. Singh, and A. A. Armoundas, ''Machine learning versus conventional clinical methods in guiding management of heart failure patients—A systematic review,'' Heart Failure Rev., vol. 26, no. 1, pp. 23–34, Jan. 2021.

[3] A. Makhlouf, I. Boudouane, N. Saadia, and A. R. Cherif, ''Ambient assistance service for fall and heart problem detection,'' J. Ambient Intell. Humanized Comput. vol. 10, no. 4, pp. 1527–1546, 2018.

[4] M. Chen, S. Gonzalez, V. Leung, Q. Zhang, and M. Li, ''A 2G-RFIDbased e-healthcare system,'' IEEE Wireless Commun. Mag., vol. 17, no. 1, pp. 37–43, Feb. 2010.

[5] N. C. Long, P. Meesad, and H. Unger, ''A highly accurate firefly based algorithm for heart disease prediction,'' Expert Syst. Appl., vol. 42, no. 21, pp. 8221–8231, 2015.

[6] K. B. Nahato, K. N. Harichandran, and K. Arputharaj, ''Knowledge mining from clinical datasets using rough sets and Backpropagation neural network,'' Comput. Math. Methods Med., vol. 2015, pp. 1–13, Mar. 2015.

[7] A. K. Dwivedi, ''Performance evaluation of different machine learning techniques for prediction of heart disease,'' Neural Comput. Appl., vol. 29, no. 10, pp. 685–693, 2018.

[8] A. U. Haq, J. P. Li, M. H. Memon, S. Nazir, and R. Sun, ''A hybrid intelligent system framework for the prediction of heart disease using machine learning algorithms,'' Mobile Inf. Syst., vol. 2018, pp. 1–21, Dec. 2018.

[9] M. S. Amin, Y. K. Chiam, and K. D. Varathan, ''Identification of significant features and data mining techniques in predicting heart disease,'' Telematics Inform., vol. 36, pp. 82–93, Mar. 2019.

[10] S. Mohan, C. Thirumalai, and G. Srivastava, ''Effective heart disease prediction using hybrid machine learning techniques,'' IEEE Access, vol. 7, pp. 81542–81554, 2019.

[11] J. Vijayashree and H. P. Sultana, ''A machine learning framework for feature selection in heart disease classification using improved particle swarm optimization with support vector machine classifier,'' Program. Comput. Softw. vol. 44, no. 6, pp. 388–397, Nov. 2018.

[12] L. Ali, A. Niamat, J. A. Khan, N. A. Golilarz, X. Xingzhong, A. Noor, R. Nour, and S. A. C. Bukhari, ''An optimized stacked support vector machines based expert system for the effective prediction of heart failure,'' IEEE Access, vol. 7, pp. 54007–54014, 2019.

[13] A. Gupta, R. Kumar, H. S. Arora, and B. Raman, ''MIFH: A machine intelligence framework for heart disease diagnosis,'' IEEE Access, vol. 8, pp. 14659–14674, 2020.

[14] 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.

[15] B. A. Tama, S. Im, and S. Lee, ''Improving an intelligent detection system for coronary heart disease using a two-tier classifier ensemble,'' BioMed Res. Int., vol. 2020, pp. 1–10, Apr. 2020.

[16] N. L. Fitriyani, M. Syafrudin, G. Alfian, and J. Rhee, ''HDPM: An effective heart disease prediction model for a clinical decision support system,'' IEEE Access, vol. 8, pp. 133034–133050, 2020.

[17] M. Waqar, H. Dawood, H. Dawood, N. Majeed, A. Banjar, and R. Alharbey, ''An efficient SMOTE-based deep learning model for heart attack prediction,'' Sci. Program., vol. 2021, pp. 1–12, Mar. 2021.

[18] A. Ishaq, S. Sadiq, M. Umer, S. Ullah, S. Mirjalili, V. Rupapara, and M. Nappi, "Improving the prediction of heart failure patients' survival using SMOTE and effective data mining techniques,'' IEEE Access, vol. 9, pp. 39707–39716, 2021.

[19] N. Salari, S. Shohaimi, F. Najafi, M. Nallappan, and I. Karishnarajah, "A novel hybrid classification model of genetic algorithms, modified Knearest neighbor and developed Backpropagation neural network,'' PLoS ONE, vol. 9, no. 11, Nov. 2014, Art. no. e112987.

[20] W. Wiharto, H. Kusnanto, and H. Herianto, "Performance analysis of multiclass support vector machine classification for diagnosis of coronary heart diseases,'' 2015, arXiv: 1511.02352.

[21] N. Khateeb and M. Usman, ‘‘Efficient heart disease prediction system using K-nearest neighbor classification technique,’’ in Proc. Int. Conf. Big Data Internet Thing (BDIOT), 2017, pp. 21–26.

[22] G. Magesh and P. Swarnalatha, “Optimal feature selection through a cluster-based DT learning (CDTL) in heart disease prediction,’’ Evol. Intell. vol. 14, no. 2, pp. 583–593, Jun. 2021.

[23] H. B. Kibria and A. Matin, “The severity prediction of the binary and multi-class cardiovascular disease—A machine learning-based fusion approach,’’ Comput. Biol. Chem., vol. 98, Jun. 2022, Art. No. 107672.

[24] S. Shin, B. Ko, and H. So, ‘‘Noncontact thermal mapping method based on local temperature data using deep neural network regression,’’ Int. J. Heat Mass Transf., vol. 183, Feb. 2022, Art. no. 122236.

[25] P. Lacerda, B. Barros, C. Albuquerque, and A. Conci, ‘‘Hyperparameter optimization for COVID-19 pneumonia diagnosis based on chest CT,’’ Sensors, vol. 21, no. 6, p. 2174, Mar. 2021.

[26] A. K. Dwivedi, ‘‘Performance evaluation of different machine learning techniques for prediction of heart disease,’’ Neural Comput. Appl., vol. 29, no. 10, pp. 685–693, 2018.

[27] M. S. Amin, Y. K. Chiam, K. D. Varathan, ‘‘Identification of significant features and data mining techniques in predicting heart disease,’’ Telematics Inform., vol. 36, pp. 82–93, Mar. 2019.

[28] D. R. Krithika and K. Rohini, "Ensemble Based Prediction of Cardiovascular Disease Using Bigdata analytics," 2021 International Conference on Computing Sciences (ICCS), 2021, pp. 42-46, doi: 10.1109/ICCS54944.2021.00017.

[29] N. M. Lutimath, N. Sharma and B. K. Byregowda, "Prediction of Heart Disease using Random Forest," 2021 Emerging Trends in Industry 4.0 (ETI 4.0), 2021, pp. 1-4, doi: 10.1109/ETI4.051663.2021.9619208.

[30] M. J. A. Junaid and R. Kumar, "Data Science And Its Application In Heart Disease Prediction," 2020 International Conference on Intelligent Engineering and Management (ICIEM), 2020, pp. 396-400, doi: 10.1109/ICIEM48762.2020.9160056.

[31] T. Obasi and M. Omair Shafiq, "Towards comparing and using Machine Learning techniques for detecting and predicting Heart Attack and Diseases," 2019 IEEE International Conference on Big Data (Big Data), *2019,* pp. 2393-2402, doi: 10.1109/BigData47090.2019.9005488.