JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

A Comprehensive Survey of Heart Disease Prediction using Machine Learning Algorithms

[1]Koduri Naga Rajani, [2]M. Aruna Safali

[1]Mtech Student, [2]Professor
[12]Department of CSE,
[12]Dhanekula Institute of Engineering and Technology, Vijayawada, India

Abstract: Modern medicine says that heart disease is the most common health problem in the world. It is also one of the main causes of death. Heart disease is more dangerous, and if it isn't found early on, it can even have bad outcomes. To diagnose patients, methods such as electronic health records, continuous body monitoring, and figuring out their health conditions by putting medical sensor projections on their bodies and using wearable tech are used. Data mining techniques are used to efficiently sort the health data that is collected because the human body creates a huge amount of data all the time. The classification of health data is also the most important step because it needs to be done correctly to find heart disease early. Most medical experts and doctors agree that one of the main reasons diseases that can't be cured don't work is because they are hard to spot early on. Saving lives is the hardest thing for doctors to do, so it's very important to figure out what's wrong with patients right away. The goal of this study is to lower the risk of heart disease by using a good feature selection and classification-based prediction system. So, one of the main problems with the way things are done now is that heart disease can't be predicted earlier and more accurately. For that reason, this study tries to create a classifier that works well and accurately predicts the early stages of heart disease. There are two stages of research that are done in this project: feature selection and extraction. The chosen features from the combined Cleveland and Statlog heart dataset were taken out and picked using the Correlation-based Feature Selection (CFS) method. After that, the datasets were fed into both single classifiers and ensemble classifiers to see which hyper parameters were best at predicting what would happen. The results of the experiments show that optimizing hyperparameters makes the model more accurate. Compared to similar works, the proposed work had an Area under curve (AUC) of 0.997 and an accuracy of 97.92%, which was better than those works. To do this, parameter optimizations for the Rotation Forest ensemble classifier were used to get certain features from the CFS method. It was found that the results of this study were much better than those of earlier studies that focused on predicting heart disease.

***Index Terms* – Cardiovascular Disease (CVD),Correlation Feature Selection (CFS),Area Under Curve(AUC).**

I. Introduction

Heart disease is a substantial contributor to the global high mortality rate, as per a report by the World Health Organization (WHO). In general, cardiovascular disease (CVD) is the leading cause of death worldwide and accounts for approximately 30% of all deaths. According to recent research, the total number of deaths is expected to rise by approximately 22 million by 2030. The American Heart Disease Association has published a report indicating that most American adults are affected by cardiovascular disease, which affects 121.5 million adults. In 2018, Korea accounted for 45 percent of the total number of deaths and ranked third in terms of the most common causes of death. Cardiovascular disease is a condition in which the flow of blood is obstructed by plaque in the arterial walls, resulting in a heart attack or stroke [1-9].

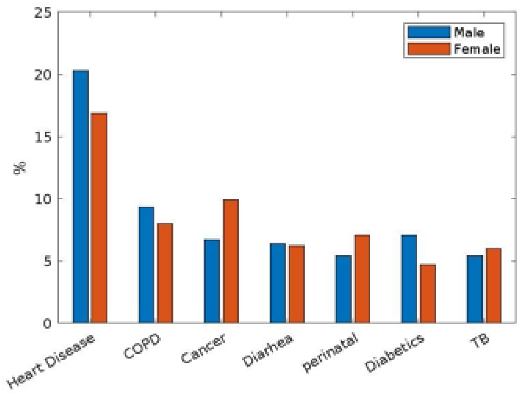

Figure1 Percentage of CVD death rates in males and females across globe

Figure 2 Risk factor of heart disease

In the event of a heart attack, it is imperative to receive prompt medical attention to prevent heart damage and preserve the patient's life. The medical professional employs emerging technology to continuously monitor the data of heart patients and provide them with advice to facilitate their recovery from heart disease. Consequently, the utilization of computer technology in the field of medicine has been on the rise. In the health care center, a variety of techniques are employed to predict and detect heart disease. The medical care industries contain a vast number of datasets that contain conceptual knowledge, which is extremely beneficial for the generation of appropriate decisions. The data provided can be used to make the necessary decisions, necessitating the extensive use of data mining techniques. The disease is predicted using data mining techniques, which have been developed with a variety of characteristics to ascertain whether the individual is experiencing a heart attack. In the same vein, the reduced time required for heart disease prediction enhances the diagnosis of the disease with the highest accuracy rate and decreases the incidence of heart attacks.

The data mining techniques [10-21] utilize data analysis to explain the past and predict the future. These data mining techniques are a fusion of a variety of fields, including artificial intelligence, database technology, and machine learning. Data mining techniques are employed in a variety of applications, and they are particularly effective in the early detection of diseases. The general process of predicting heart disease involves the extraction of significant data from a vast amount of data through the feature extraction process. Subsequently, the extracted data will be trained using the selected dataset, and the data will be subjected to the testing process. Diverse classification methodologies are implemented to accomplish this. This method is also referred to as KDD, or Knowledge Discovery of Data. A variety of methods are employed to predict the development of heart disease. (1) Decision tree, (2) SVM-Support Vector Machine, (3) Neural network, and (4) KNN-K-Nearest Neighbours are the four primary types of data mining techniques. In addition to these methods, a variety of other methods are employed to predict heart disease. Heart disease and cardiovascular disease have been predicted using a variety of machine learning techniques. The typical architecture for predicting heart disease involves preprocessing the data, extracting features, and either predicting the disease or evaluating the performance obtained during the prediction process. The following is a partial list of the various types of predicting techniques. In general, three distinct procedures are implemented to predict heart disease: feature selection, preprocessing the data, and classification methods. The performance analysis is conducted after the prediction/classification method.

II. LITERATURE SURVEY

Analysis of different machine learning algorithms [22-29] applied in CVD prediction on different datasets. Prediction of chronic heart disease using data mining techniques.

Sravani N. et al. [28] developed a predictive model for chronic heart disease using data mining techniques. During the pre-processing of the Framingham dataset, any missing values in the attributes were replaced with the mean of the other values within the same attribute. Next, divide the dataset into training and testing sets and apply prediction models using XGBoost and LR techniques. Attained a peak accuracy of 86.86% and 87.46% using LR and XGBoost, respectively. It has been determined that LR is appropriate for small and medium-sized datasets, while XGBoost is more favorable for large datasets because of its rapid response.

Investigating feature selection and classification techniques for the prediction of heart disease , Spencer et al. [29] evaluated the effectiveness of machine learning models using different feature-selection techniques. Four heart disease datasets, namely Cleveland, Long-Beach-VA, Hungarian, and Switzerland, were combined to create a dataset with 720 instances and 14 features. Principal Component Analysis (PCA), Chi-square, Relief, and Symmetry Uncertainty (SU) feature-selection methods were used to select the most relevant features. The dataset was then evaluated using eight supervised machine learning algorithms: Bayes Net, Logistic Regression, Stochastic Gradient Descent (SGD), K-Nearest Neighbours (KNN) with k=21, Adaboost M1 with Decision Stump, Adaboost M1 with Logistic Regression, JRip, and Random Forest (RF). Attained an accuracy of 85.1%, precision of 84.83%, and recall of 85.59% by employing Chi squared feature selection in conjunction with the Bayes Net classifier.

Accurate Forecasting of Cardiovascular Disease Applying Machine Learning Algorithms with Relief and LASSO Feature Selection Techniques, Ghosh et al. [30] devised hybrid techniques, including Decision Tree Bagging Method (DTBM), Random Forest Bagging Method (RFBM), K-NN Bagging Method (KNNBM), AdaBoost Boosting Method (ABBM), and Gradient Boosting Method (GBBM), which integrate conventional classifiers with bagging and boosting techniques. Combined five datasets, namely Cleveland, Long Beach VA, Switzerland, Hungarian, and Statlog, which had identical attributes. Conducted pre-processing techniques to address the issue of missing values. Next, I utilised the Relief and LASSO feature selection (FS) techniques to identify the most optimal features. Assessed the models' performance by calculating accuracy, sensitivity, error rate, precision, and F1 score, as well as NPV, FPR, and FNR. Attained a peak accuracy of 99.15% by employing the RFBM algorithm in conjunction with the Relief FS technique.

Accurate Forecasting of Cardiovascular Disease Applying Machine Learning Algorithms Utilising Relief and LASSO feature selection techniques, Ghosh, P. et al., introduced a highly effective model for predicting heart disease. Aggregated five datasets from Cleveland, Long Beach V, Switzerland, Hungarian, and Statlog and conducted data preprocessing to address missing values. Perform feature selection using Relief and LASSO methods. Train hybrid classifiers, including DT Bagging, RF Bagging, KNN Bagging, AdaBoost Boosting, and GB Boosting, by combining traditional classifiers with bagging and boosting techniques. Assessed the effectiveness of these models by calculating accuracy, sensitivity, error rate, precision, and F1 score, as well as NPV, FPR, and FNR. Attained a precision rate of 99.05% using the RFBM and Relief FS techniques. The following features were selected using Relief FS: age, chol, fbs, restecg, thalach, exang, ca (10). The features selected using LASSO FS are age, cp, trestbps, chol, thalach, old peak, slope, and thal. There are a total of 11 features.

Classification of heart disease utilising data mining tools and machine learning techniques, IliasTougui et al[22] conducted a comparative analysis of the performance of six machine learning algorithms (LR, SVM, KNN, ANN, NB, and RF) in classifying heart disease. The analysis was carried out using six commonly used data mining tools: Orange, Weka, RapidMiner, Knime, MATLAB, and Scikit-Learn. The UCI Cleveland dataset was selected, which consists of 303 instances, 13 features, and one target variable. Prior to analysis, the dataset underwent pre-processing to eliminate any missing values. Next, the six machine learning techniques were implemented on the dataset using 10-fold cross validation. The confusion matrices were then extracted to compute the accuracy, sensitivity, and specificity, which are three performance metrics. The MATLAB ANN technique achieved the highest accuracy of 85.76% and sensitivity of 83.74%, while RapidMiner's SVM method achieved the highest specificity of 94.28%.

Formation of a hierarchical random forest using a nonlinear regression model for the prediction of cardiovascular diseases, M. Aslam et al [27] introduced a new approach by employing a Hierarchical Random Forest Formation combined with a Nonlinear Regression Model to forecast CVD using the Cleveland dataset, which consists of 14 distinct characteristics. Conducted data preprocessing and partitioned the dataset instances, allocating 70% for training and 30% for testing. The bottom-up approach is employed with Artificial Neural Networks (ANN) to extract notable features, while the combination of Support Vector Machines (SVM) and the Apriori algorithm is utilised for classification. The HRFFNRM method achieved an accuracy of 90.13%, precision of 91.14%, and a misclassification rate of 8.84%.

Heart disease prediction through the application of feature selection and machine learning techniques, C. Gazeloglu et al[18], utilised 18 machine learning models and 3 feature selection techniques (Correlation-based FS, Chi-Square, and Fuzzy Rough Set) to determine the optimal combination for predicting heart disease diagnosis on the Cleveland dataset. The dataset consisted of 303 patients and 14 variables. The SVM (PolyKernel) algorithm achieved an accuracy of 85.248% without feature selection. The Naïve Bayes algorithm achieved an accuracy of 84.718% with feature selection using CFS, selecting 6 features. The radial basis function (RBF) network achieved an accuracy of 81.288% with feature selection using Fuzzy Rough Set and Chi-Square FS, selecting 7 features.

Heart disease prediction is accomplished through the utilisation of a hybrid machine learning model, M. Kavitha et al[19] developed an innovative hybrid model to forecast heart disease by utilising two machine learning algorithms: decision tree and random forest. The Cleveland heart dataset was obtained from the USI repository and divided into a 70:30 ratio for training and testing purposes. Utilised Decision Tree (DT), Random Forest (RF), and a hybrid model combining DT and RF. At last, I have developed an application using TkInter in Python to evaluate the hybrid model for a single user input. Attained an accuracy rate of 88.1% using the hybrid model.

An enhanced ensemble learning methodology for forecasting the risk of heart disease, I. D. Mienye et al. [23] developed an ensemble method by combining CART models with an accuracy-based weighted ageing classifier ensemble (AB-WAE) on the Cleveland dataset (303 instances) and the Framingham dataset (4238 instances). Implemented and assessed various machine learning algorithms including K-Nearest Neighbours (KNN), Logistic Regression (LR), Linear Discriminant Analysis (LDA), Support Vector Machines (SVM), Classification and Regression Tree (CART), Gradient Boosting (GB), and Random Forest (RF). Performance evaluation was conducted using accuracy, sensitivity, precision, and F1 score metrics. The proposed approach yielded an accuracy of 93.1% and 91.2% with the Cleveland and Framingham datasets, respectively.

Heart Disease Prediction Applying a fusion of machine learning and deep learning techniques, Rohit Bharti[28]and his colleagues examined three techniques for predicting heart disease using machine learning and deep learning (DL) on the UCI heart dataset. The first method does not involve feature selection or outlier detection. The second method includes feature selection but does not involve outlier detection. The third method includes both feature selection and outlier detection. Asserted that the third methodology yielded superior outcomes compared to the initial two approaches. In this study, various classifiers including LR, KNN, SVM, RF, DT, and DL were trained. Additionally, the LASSO feature selection technique was employed to identify the most significant features for heart disease. Obtained an accuracy of 84.68%, specificity of 77.67%, and sensitivity of 85.0% using the K-Nearest Neighbours (KNN) algorithm. In contrast, achieved accuracy, specificity, and sensitivity of 94.12%, 82.13%, and 83.21%, respectively, using deep learning.

III. CLASSIFICATION OF ML ALGORITHMS BASED THEIR PERFORMANCE FOR DIFFERENT DATASETS.

This section presents the latest methods for identifying cardiovascular disease (CVD) using machine learning (ML) algorithms, as evidenced by multiple successful research studies. Table 1 contains a summary of the works that are related to the topic.

S. No	Author	Technique	Pros	Cons
1	Javeed [24]	FS: Random Search Algorithm Classifier: RF Holdout: (70:30) validation	Accuracy, sensitivity, specificity, MCC, and ROC-AUC SW:Python programming	Only one classifier has been trained and the dataset samples become fewer when using hold-out validation
2	Alam et al. [25]	FS: Info Gain, Gain Ratio, Correlation, One R, and Relief F Classifiers: RF 10-fold cross-validation (CV)	Accuracy and AUC SW: Weka	One classifier has been utilized and the accuracy achieved was a moderate value
3	Mohamed et al. [27]	FS:Parasitism-predation algorithm Classifiers: KNN with 10-fold CV	Accuracy and execution time SW: Matlab 2017a	Only one classifier has been trained& achieved moderate accuracy.
4	Bharti et al. [28]	FS: LASSO Classifiers: LR, KNN, SVM, RF, DT, and DL with 10-fold cross-validation	Accuracy, specificity, and sensitivity SW: Python programming	Achieved good accuracy of 94.2%. However, the specificity and sensitivity are poor.
5	Gupta et al. [26],	Classifiers: LR, SVM, NB, DT, KNN, and RF 70:30 train-test split	Accuracy, sensitivity, specificity, precision, and F1 score SW: Python programming	No feature selection method was employed. Lesser number of observations for validating the model

Table 1. Summary of related works

IV. PROPOSED METHODOLOGY

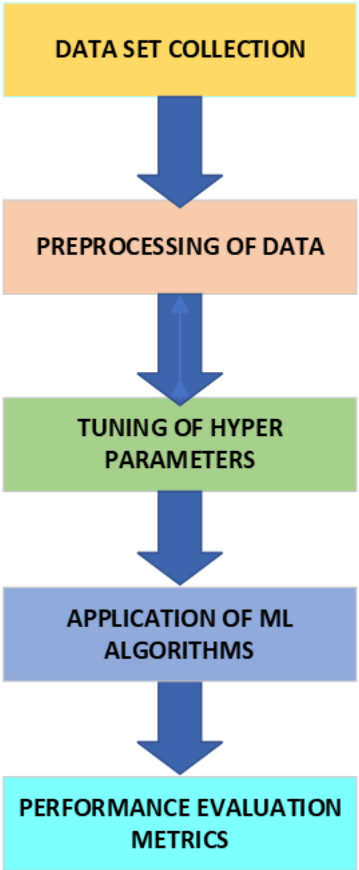

Figure 3 Proposed Methodology

4.1ML Algorithms for Heart Disease Prediction

To predict the accuracy of CVD, the following machine learning (ML) classifier algorithms are implemented: Support Vector Machines (SVM), K-Nearest Neighbours (KNN), Naïve Bayes, Decision Tree, AdaBoost M1, Bagging, Artificial Neural Networks (ANN), and Rotation Forest.

Support Vector Machines (SVM):

SVM based on the theory of statistics, SVM is an algorithm for supervised machine learning that addresses classification problems. SVM typically employs a hyper plane to partition the data into two classes by the greatest possible margin. SVMs employ the nonlinear kernel, a mathematical operation utilized to transform the data, to execute nonlinear classification. To enhance the efficacy of SVM, a variety of kernels are implemented on multiple data sets. SVMs find extensive applicability in a wide range of domains, including pattern recognition and the medical field.

K-Nearest Neighbors (KNN)

KNN is an example of a classification function-determining ML algorithm. In the process of regression, weights are assigned to each neighboring element. This algorithm is implemented to classify instances in a more efficient fashion. It is utilized to solve classification, regression, and nonlinear classification problems. It is a straightforward process to implement. The KNN classifier requires optimization of the distance weighting and the number of nearest neighbors.

Naïve Bayes:

A genus of algorithms known as Naive Bayes classifiers is founded upon Bayes' Theorem. Notwithstanding the "naive" presumption of feature independence, these classifiers are extensively employed in machine learning due to their straightforwardness and effectiveness. The Naïve Bayes algorithm is employed to address classification challenges. It operates rapidly and facilitates prediction for data with a high dimension.

Decision Tree (DT)

DT is a classifier that generates decisions primarily using a graphical model resembling a tree structure. Its primary application is in medical diagnosis; therefore, it is an effective predictive model. From root to leaf, the algorithm is represented as a tree model. It is straightforward to interpret, comprehend, and implement, requires minimal data preparation, can process a wide variety of data types, and generates robust classifiers. It operates by generating a set of principles and requires only minimal computation. It aims to accomplish two goals: firstly, to produce an ideal classifier; and secondly, to comprehend the predictive structure of the problem.

AdaBoost M1 (AB):

AdaBoost generates a robust classifier through the combination of multiple weak classifiers. The fundamental concept entails attributing greater significance to data that has been inadequately classified while reducing the weight assigned to data that has been adequately classified. Thus, as each weak classifier concentrates more on cases that were incorrectly classified, the outcomes are enhanced. In this instance, decision trees served as the foundational approach. There are three steps in the algorithmic process. Initially, the weights are initialized by assigning an equal weight to each of the N samples. During the second step, a weak classifier is trained with the condition that its weight is increased to assign greater significance to the data if it is accurately classified, and decreased if it is inaccurately classified. Finally, a strong classifier is constructed by combining the weak classifiers acquired in each training.

Bagging:

Ensemble learning is a technique that is frequently used for the purpose of minimizing variance in chaotic data sets. Bagging, which is also known as bootstrap aggregation, is another name for this technique. The process of bagging involves the replacement of a random sample of data from a training set. This means that each data point may be selected multiple times during the bagging process. A prediction model's variance is reduced because of its implementation, which is necessary to address trade-offs between bias and variance. To avoid overfitting of data, the classification process typically involves the use of bagging.

Artificial Neural Networks (ANN):

An ANN-based approach to decision-making is represented graphically. The architecture comprised three distinct layers: an input layer, concealed layers, and an output layer. Deter mining the activation function and the number of layers and nodes are intricate aspects of ANN design. The optimization or customization of the structure and design of ANN is necessary for each individual application. An increasing number of cancer prognosis and diagnosis models can benefit from the implementation of ANN. It finds utility in an extensive array of classification applications.

Rotation Forest (RF)

The rotation forest classifier is utilized in a wide range of data mining applications since it is an effective classification method. Furthermore, it is a robust ensemble algorithm for tree-based feature extraction that enhances the diversity of the base classifier. Rotation Forest is an ensemble classification technique that, like Random Forest, mitigates one of the major shortcomings of Random Forest. This is the ability of its component models, decision trees, to partition the feature space only orthogonally, which means that they are perpendicular to the feature axes. Rotation Forest is like Random Forest.

$$\text{Accuracy} = TN + TP/ TN + TP + FN + FP \quad (1)$$

$$\text{Sensitivity } S = TP/ TP + FN \quad (2)$$

$$\text{Specificity} = TN/ TN + FP \quad (3)$$

$$\text{Precision } P= TP/ TP + FP \quad (4)$$

$$\text{F1 Score} = 2*P*S/P+S \quad (5)$$

V. RESULTS AND DISCUSSION

The results after simulation of datasets using different ML algorithms in terms of different performance metrics.

CLASSIFIER	Accuracy	AUC	Specificity	Sensitivity(S)	Precision(P)	F1 Score
SVM	93.2%	0.928	92.3%	93.2%	93.24%	93.20%
Decision Tree	95.42%	0.965	94. 2%	95.42%	95.42%	95.42%
Naïve Bayes	95.2%	0.961	94.1%	95.2%	95.2%	95.2%
KNN	97.03%	0.985	96.3%	97.03%	97.02%	97.01%
AdaBoost	97.32%	0.988	96.2%	97.32%	97.30%	97.31%
ANN	96.92%	0.976	95.62%	96.92%	96.92%	96.92%
Bagged Tree	94.86%	0.984	93.66%	94.86%	94.84%	94.85%
RF	97.92%	0.995	97.61%	97.92%	97.91%	97.93%

Table 2. Performance Measure Analysis of Classifiers in presence of optimization

The illustration of performance metrics is graphically presented as show in the figure 4

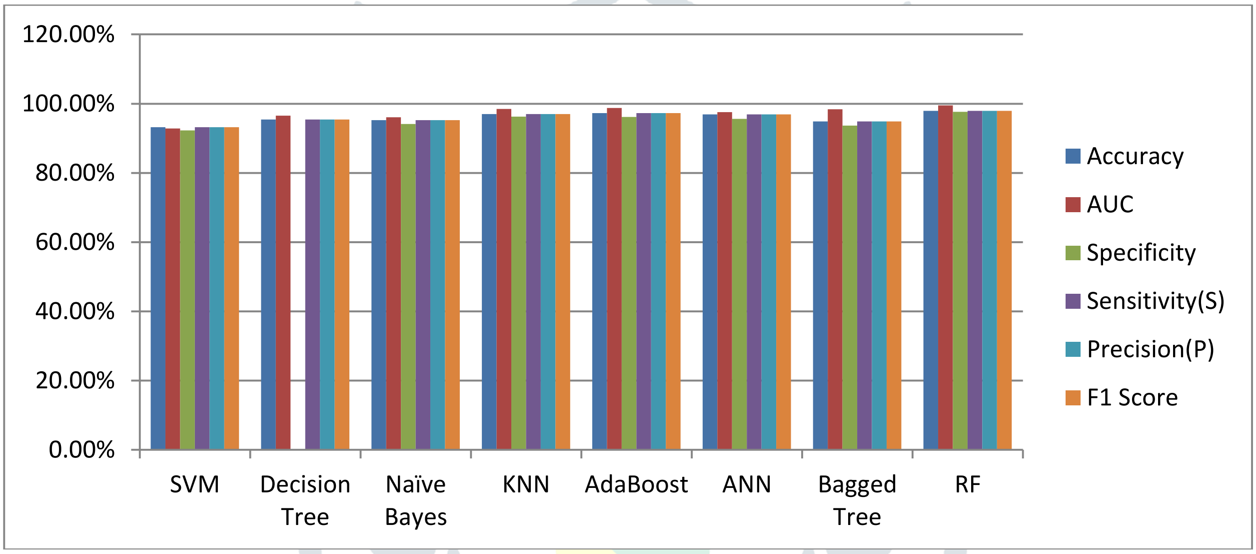

Figure 4. Performance Measure Analysis of Classifiers in presence of optimization

A significant number of fatalities occur annually because of cardiovascular diseases. Early detection of heart disease has the potential to preserve numerous lives. Through the implementation of correlation-based feature selection and hyper parameter optimization on the combined heart dataset (Cleveland and Statlog), ML models have demonstrated improved capability in forecasting the likelihood of coronary heart disease. The Rotation Forest classifier demonstrated the highest accuracy of 97.9% and AUC of 0.995 for 14 out of 566 instances. Because of the labor-intensive process of hyper parameter optimization, the training period for the machine learning models on the specified components was prolonged. Potential future research directions include the integration of innovative feature selection methods and automatic hyper parameter optimization strategies into machine learning models, with the aim of augmenting their predictive capacities. With the selected hyper parameters, the proposed algorithms demonstrate the highest accuracy in comparison to random forest for the Cleveland and Statlog datasets.

Acknowledgment

We would like to thank the management of Dhanekula Institute of Engineering and Technology for providing computing facilities for completion of research work.

REFERENCES

[1] Ali, A. 2001.Macroeconomic variables as common pervasive risk factors and the empirical content of the Arbitrage Pricing Theory. Journal of Empirical finance, 5(3): 221–240.

[2] Basu, S. 1997. The Investment Performance of Common Stocks in Relation to their Price to Earnings Ratio: A Test of the Efficient Markets Hypothesis. Journal of Finance, 33(3): 663-682.

[3] Bhatti, U. and Hanif. M. 2010. Validity of Capital Assets Pricing Model.Evidence from KSE-Pakistan.European Journal of Economics, Finance and Administrative Science, 3 (20).

[1] Nalluri, S., Vijaya Saraswathi, R., Ramasubbareddy, S., Govinda, K., Swetha, E. (2020). Chronic Heart Disease Prediction Using Data Mining Techniques. In: Raju, K., Senkerik, R., Lanka, S., Rajagopal, V. (eds) Data Engineering and Communication Technology. Advances in Intelligent Systems and Computing, 3(1079):672-680.

[2] Rohit Bharti, Aditya Khamparia, Mohammad Shabaz, Gaurav Dhiman, Sagar Pande, Parneet Singh.2021.Prediction of Heart

Disease Using a Combination of Machine Learning and Deep Learning,Computational Intelligence and Neuroscience, 11(3):987-1010.

[3] Mienye, Ibomoiye Domor, Yanxia Sun, and Zenghui Wang. 2020.An improved ensemble learning approach for the prediction of heart disease risk. Informatics in Medicine Unlocked 20 (2020): 100402.

[4] M.Kavitha, G. Gnaneswar, R. Dinesh, Y. R. Sai and R. S. Suraj.2021. Heart Disease Prediction using Hybrid machine Learning Model.2021 6th International Conference on Inventive Computation Technologies (ICICT), Coimbatore, India: 1329-1333, doi: 10.1109/ICICT50816.2021.935859.

[5] E.K. Hashi and M.S.U. Zaman. 2020.Developing a Hyperparameter Tuning Based Machine Learning Approach of Heart Disease Prediction", Journal of Applied Science & Process Engineering, 7(2): 631-647.

[6] Jonnavithula .2020. Role of machine learning algorithms over heart biseases prediction. AIP Conference Proceeding, vol. 2292(1): 040013.

[7] T. Porter and B. Green.2009.Identifying diabetic patients: A data mining approach. in Proc.

[8] Mediacentre.2017. From World Health Organization. Accessed.

[9] A. K. Sen, S. B. Patel, and D. P. Shukla.2013.A data mining technique for prediction of coronary heart disease using neuro-fuzzy integrated approach two level,'' Int. J. Eng. Comput. Sci., 2(9): 1663–1671.

[10] T. Helmy, S. M. Rahman, M. I. Hossain, and A. Abdelraheem.2013.Non- linear heterogeneous ensemble model for permeability prediction of oil reservoirs.Arabian J. Sci. Eng., 38(6): 1379–1395.

[11] S. Zhao, M. Jiang, B. Qin, T. Liu, C. Zhai, and F. Wang.2020.Structural and textual information fusion for symptom and disease representation learning. IEEE Trans. Knowl. Data Eng., early access.

[12] J. P. Li, A. U. Haq, S. U. Din, J. Khan, A. Khan, and A. Saboor.2020.Heart disease identification method using machine learning classification in E- healthcare. IEEE Access, 8(1): 107562–107582.

[13] GAZELOĞLU, Cengiz. 2020.Prediction of heart disease by classifying with feature selection and machine learning methods.Progress in Nutrition 2(2):1017-1030.

[14] Aslam, Mohamed, and K. Jaisharma. 2021.Hierarchical Random Forest Formation with Nonlinear Regression Model for Cardiovascular Diseases Prediction. 2021 International Conference on Computer Communication and Informatics (ICCCI).

[15] Tougui, Ilias, Abdelilah Jilbab, and Jamal El Mhamdi. 2020. Heart disease classification using data mining tools and machine learning techniques. Health and Technology 10 (2): 1137-1144.

[16] Ghosh, Pronab, 2021.Efficient prediction of cardiovascular disease using machine learning algorithms with relief and LASSO feature selection techniques. IEEE Access 9 (2021): 19304-19326.

[17] Reddy, Karna Vishnu Vardhana, Irraivan Elamvazuthi, Azrina Abd Aziz, Sivajothi Paramasivam, Hui Na Chua, and Satyamurthy Pranavanand. 2023. An Efficient Prediction System for Coronary Heart Disease Risk Using Selected Principal Components and Hyper parameter Optimization.Applied Sciences 13(1): 118.

[18] Valarmathi, R., and T. Sheela.2021.Heart disease prediction using hyper parameter optimization (HPO) tuning." Biomedical Signal Processing and Control 70 (2021): 103033.

[19] García-Ordás, M.T., Bayón-Gutiérrez, M., Benavides, C.2023.Heart disease risk prediction using deep learning techniques with feature augmentation", Multimedia Tools Appl 8(2): 31759–31773 .

[20] Kondababu A, Siddhartha V, Kumar BB, Penumutchi B .2021.A comparative study on machine learning based heart disease prediction", Mater Today.

[21] Das, Surajit, Samaleswari P. Nayak, Biswajit Sahoo, and Sarat Chandra Nayak.2024.Machine Learning in Healthcare Analytics: A State-of-the-Art Review.Archives of Computational Methods in Engineering (2024): 1-40,

[22] Nimmagadda, Satyanarayana Murthy, Sowmya Sree Agasthi, Abbas Shai, Dimple Kavitha Raj Khandavalli, and Janaki Ram Vatti.2023.Kidney failure detection and predictive analytics for CKD using machine learning procedures." Archives of Computational Methods in Engineering 4 (2023): 2341-2354.

[23] Javeed, Ashir, Sanam Shahla Rizvi, Shijie Zhou, Rabia Riaz, Shafqat Ullah Khan, and Se Jin Kwon. 2020.Heart risk Failure prediction using a novel feature selection method for feature refinement and neural network for classification. Mobile Information Systems 2020 (2020): 1-11.

[24] Pasha, Syed Javeed, and E. Syed Mohamed.2020.Novel Feature Reduction (NFR) model with machine learning and data mining algorithms for effective disease risk prediction.IEEE Access 8 (2020): 184087-184108.

[25] Gupta, Ankur, Rahul Kumar, Harkirat Singh Arora, and Balasubramanian Raman.2019.MIFH: A machine intelligence framework for heart disease diagnosis. IEEE access 8 (2019): 14659-14674.

[26] Mohamed, Al-Attar A., S. A. Hassan, A. M. Hemeida, Salem Alkhalaf, M. M. M. Mahmoud, and Ayman M. Baha Eldin. 2020. Parasitism–Predation algorithm (PPA): A novel approach for feature selection. Ain Shams Engineering Journal , 2 (2020): 293-308.

[27] Srinivas, Rayudu, Ravi kiran Bagadi, T. Rama Reddy, Neti Praveen, and G. Aparanjini. 2024.An efficient hybrid optimization algorithm for detecting heart disease using adaptive stacked residual convolutional neural networks. Biomedical Signal Processing and Control 87 (2024): 105522.

[28] Peteti, Bala Srinivas, and Durgesh Nandan. 2023.Heart Disease Classification/Prediction: A Review." Revue d'Intelligence Artificielle 37(2):1015-1023.

[29] Pasha,2021.Novel Feature Reduction (NFR) model with machine learning and data mining algorithms for effective disease risk prediction.IEEE Access 8 (2020): 184087-184108