Asian Journal of Cardiology Research

Volume 7, Issue 1, Page 348-359, 2024; Article no.AJCR.128065

Optimizing Heart Disease Prediction: A Comparative Study of Machine Learning Techniques

Mohammad Raquibul Hossain [a*]

[a] *Department of Applied Mathematics, Noakhali Science and Technology University, Noakhali, Bangladesh.*

Author's contribution

The sole author designed, analysed, interpreted and prepared the manuscript.

Article Information

DOI: https://doi.org/10.9734/ajcr/2024/v7i1238

Open Peer Review History:
This journal follows the Advanced Open Peer Review policy. Identity of the Reviewers, Editor(s) and additional Reviewers, peer review comments, different versions of the manuscript, comments of the editors, etc are available here: https://www.sdiarticle5.com/review-history/128065

Original Research Article

Received: 14/10/2024
Accepted: 17/12/2024
Published: 20/12/2024

ABSTRACT

Healthcare is an important sector which needs continued modernization with technology-based state-of-the-art system. Main focus of this paper is applying artificial intelligence, especially machine learning techniques to build heart disease prediction system. In this sector, diagnosis of diseases like heart disease by thorough test reports investigation of doctors is crucial, challenging and time-consuming. In addition to doctor's investigation, artificial intelligence techniques can assist and alleviate healthcare hassles and challenges. This paper presents prediction system for heart disease using nine machine learning methods. The experimented heart disease dataset has 918 sample and 11 features (key features include Age, and Fasting Blood Sugar, correlation between Maximum Heart Rate, Gender, Chest Pain Type, Exercise Angina, peak slope of exercise ST segment). Among the methods experimented on heart disease data, Categorical Boosting (CatBoost) and Light Gradient Boosting Machine (LGBM) were outperformers among the methods with highest accuracy scores and Recall scores. Higher Recall scores contribute to correctly predict the true positive values or correctly predicting true heart disease patients which is very important in

**Corresponding author: Email: raquib.math@gmail.com;*

Cite as: *Hossain, Mohammad Raquibul. 2024. "Optimizing Heart Disease Prediction: A Comparative Study of Machine Learning Techniques". Asian Journal of Cardiology Research 7 (1):348-59. https://doi.org/10.9734/ajcr/2024/v7i1238.*

sectors like healthcare. Also, CatBoost and Random Forest were best performing methods in 10-fold cross-validation test. All these experimented results reflect that artificial intelligence or specifically machine learning algorithm-based prediction system for heart disease can be a very assisting tool for physicians and overall healthcare system. This research encourages further investigation with more and large datasets. Continued research and availability of more large datasets can improve prediction accuracy to higher satisfaction level of doctors which can help develop predictive ecosystem.

Keywords: Disease prediction; heart failure prediction; machine learning; artificial intelligence; prediction system; LGBM; random forest; CatBoost.

1. INTRODUCTION

Cardio-vascular disease (CVD), commonly known as heart disease, is currently a threat to the global population, and is a major challenge to health care. The mortality rates are high if diagnosis and treatment is not done early. Hence it is very important for patients to be checked and diagnosed early enough. However, conventional diagnostic techniques are usually slow, costly, and may at times, be affected by human errors. Artificial Intelligence especially machine learning (ML) technique is therefore a potential automated tool in healthcare sector since it is capable of analyzing data in a manner that human cannot observe. Using machine learning approaches, researchers and healthcare experts can discover methods and algorithms that would enable proper anticipation of heart diseases with better performance. Through continual investigation and assess of these methods and algorithms on large datasets, experts can develop smart and efficient disease prediction systems for different diseases like heart disease based-on relevant input data. These systems rely on complex mathematical models to analyse patient information including clinical history, healthy habits, and the results of clinical tests among others with a view of reviewing them and give timely recommendations about possible risk issues. A prediction system for heart disease by implementing techniques of ML can accelerate healthcare service and ease the works of doctors.

There are many studies on disease prediction including heart disease. Comparative study of Wang [1] employed machine learning for heart failure and found importance of z-score and oversampling in data preprocessing. Applicability of machine learning was focused in the work of Olsen et al. [2] for heart failure prediction, its diagnosis and classification. For hospitalization and mortality prediction [3] used machine learning where random forest was best performing model. The work of Sciomer et al. [4] focused role BMI, age and gender in heart failure prognosis. To predict survival of heart failure patients [5] used dataset collected in 2015 from 299 patients and employed machine learning. A focused review on the employment of artificial intelligence in heart failure prediction was done by Khan et al.[6] which include effective machine learning algorithms like random forest, support vector machine, principal component analysis, artificial neural networks etc. A method involving motion-based analysis was developed by Guo et al. [7] to predict survival of patients with heart disease. Machine learning algorithms were explored in the study of Singhal [8] for the prediction of congestive heart failure. Deep learning was employed by Li et al.[9] for a prediction system development of heart failure mortality. Convolutional neural network (abbreviated as CNN), a deep learning model, was applied in the work of Akbilgic et al.[10] to predict heart failure using ECG. The review study of Wang et al. [11] embroils risk prediction of heart failure in adult patients who inherited heart disease. To predict heart disease [12] used hybrid machine learning techniques and obtained 88.7% accuracy from hybrid random forest with a linear model. The work of Shah, Patel, & Bharti, [13] used heart disease data containing 303 samples and 14 features and found highest accuracy from k-nearest neighbor. In predicting heart disease, logistic Regression and k-NN were found to perform comparatively better in the study of Jindal et al. [14]. Random forest performed best for heart disease prediction in the study of Sharma, Yadav, & Gupta [15].To predict heart disease [16] reviewed and applied machine learning techniques where artificial neural network (with 86.91% accuracy) was best performing model.

The main focus of this study is the use of ML methods in heart disease prediction. An ML-based heart disease prediction system can be used in early diagnosis and personalized

treatment through development and deployment of the models. By incorporating AI into computations for heart disease, we can get closer to developing effective, cheap and effective solutions to healthcare problems facing the world today.

2. MATERIALS AND METHODS

This section contains description of datasets, experimented machine learning models, the disease prediction system, evaluation metrics and cross-validation method.

2.1 Dataset Description

This heart disease dataset was collected from Kaggle which was derived by Fedesoriano [17] from 5 databases (i.e., i. Cleveland, ii. Hungarian, iii. Switzerland, iv. Long Beach VA and v. Stalog) [18]. It has 11 features, 1 target variable and a total of 918 samples. The features are i. Age, ii. Sex, iii. ChestPainType, iv. RestingBP, v. Cholesterol, vi. FastingBS, vii. RestingECG, viii. MaxHR, ix. ExerciseAngina, x. Oldpeak and xi. ST_Slope. The features are briefly described below in Table 1.

The dataset has five numeric features (i.e., Age, RestingBP, Cholesterol and MaxHR) and seven categorical variables including the target/class variable (i.e., Sex, ChestPainType, FastingBS, RestingECG, ExerciseAngina, ST_Slope and HeartDisease). Statistical summary of the numeric features are presented in Table 2.

Correlations among variables are presented in the following correlation matrix in Fig. 1. It shows that high correlations are correlation between Age and HeartDisease is 0.28, correlation between FastingBS and HeartDisease is 0.27, correlation between MaxHR and HeartDisease is -0.40, correlation between Sex_M and HeartDisease is 0.31, correlation between ChestPainType_ATA and HeartDisease is -0.40, correlation between ExerciseAngina_Y and HeartDisease is 0.49, correlation between ST_Slope_Flat and HeartDisease is 0.55, correlation between ST_Slope_Up and HeartDisease is -0.62.

2.1.1 Train-test split

Two train-test splits were experimented in this research study for 918 samples. These are 70:30 split (643 samples in in-sample training set and 273 samples in hold-out out-sample test set) and 90:10 split (826 samples in in-sample training set and 92 samples in hold-out out-sample test set). In addition, 10-fold cross-validation is performed to evaluate average prediction accuracy throughout the dataset.

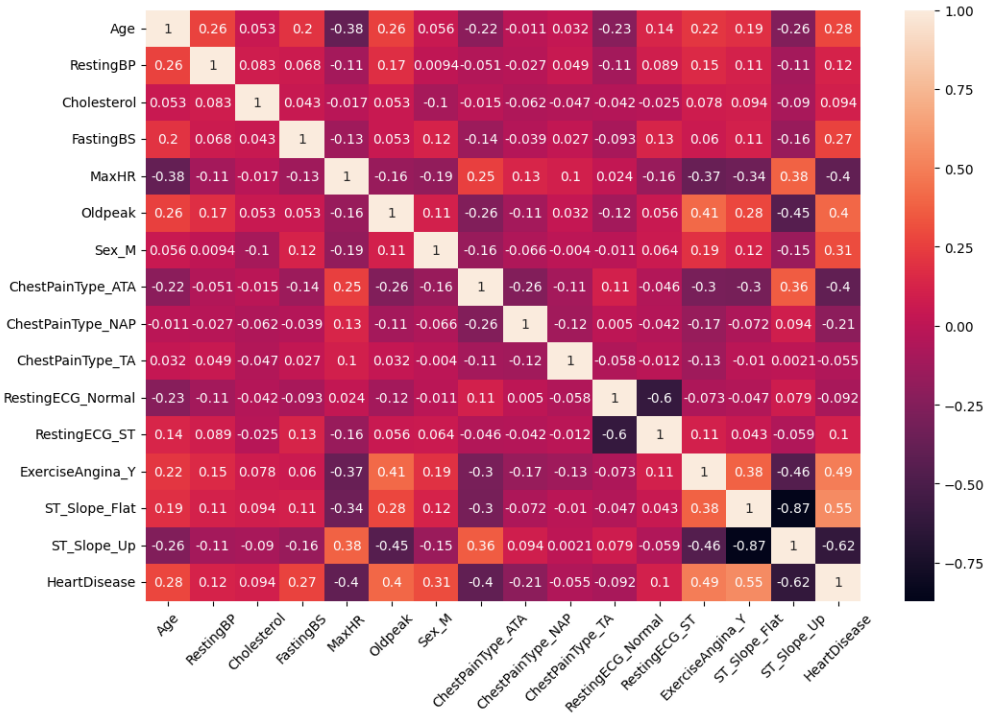

Fig. 1. Correlation matrix of numeric features and target feature/class of heart disease

Table 1. Brief description of the feature variables of heart disease dataset

Attribute/feature	Description	Value
Age	Patient's age	Numeric [years]
Sex	Patient's sex	M (for Male), F (for Female)
ChestPainType	Chest pain type	TA (for Typical Angina), ATA (for Atypical Angina), NAP (for Non-Anginal Pain), ASY (for Asymptomatic)
RestingBP	Resting blood pressure	Numeric [mm Hg]
Cholesterol	Serum cholesterol	Numeric [mm/dl]
FastingBS	Fasting blood sugar	1: if FastingBS > 120 mg/dl, 0: otherwise
RestingECG	Resting electrocardiogram results	Normal: Normal, ST: having ST-T wave abnormality (T wave inversions and/or ST elevation or depression of > 0.05 mV), LVH: showing probable or definite left ventricular hypertrophy by Estes' criteria.
MaxHR	Maximum heart rate achieved	Numeric [value between 60 and 202]
ExerciseAngina	Exercise-induced angina	Y: Yes, N: No
Oldpeak	ST value	Numeric [value measured in depression]
ST_Slope	The peak slope of exercise ST segment	Up (for upsloping), Flat (for flat), Down (for downsloping)
HeartDisease	Output class	1: heart disease, 0: Normal

Table 2. Statistical summary of numeric features of heart disease dataset

Feature	Sample Size	Mean	Standard Deviation	Minimum	25%	50%	75%	Maximum
Age	918	53.51	9.43	28	47	54	60	77
RestingBP	918	132.54	17.99	80	120	130	140	200
Cholesterol	918	244.64	53.32	85	214	244.6	267	603
MaxHR	918	136.81	25.46	60	120	138	156	202
Oldpeak	918	0.89	1.07	-2.6	0	0.6	1.5	6.2

2.2 Machine Learning Models

For heart disease prediction, 09 machine learning models were experimented in this research study. These are Logistic Regression (LR), Support Vector Machine Classifier (SVC), k-Nearest Neighbors (kNN), Gaussian Naïve Bayes (GNB), Decision Tree (DT), Random Forest (RF), Extreme Gradient Boosting (XGBoost), Light Gradient Boosting Machine (LGBM) and Categorical Boosting (CatBoost). These models are briefly described in the following sections.

2.2.1 Logistic regression

Logistic Regression (abbreviated as LR) is a simple and widely-used model for the prediction of two-scale dependent variables, such as yes/no or 0/1. LR algorithm takes a probability value to predict the possibility of an element of input data being classified in a particular class or into the category of the second class. The model has a specific function, i.e., sigmoid, which gives its predictions as the probability values between 0 and 1. During training, the algorithm acquires the best values that the architecture of the model requires for the prediction to be more precise. It is popular because it is easy, quick to train, and the results are probabilities that can be very useful in a decision-making process. The only drawback is that it supposes a linear relationship of the features with the result or outcome data; actually, it often may not be so.

2.2.2 Support vector classifier

Support Vector Classifier (abbreviated as SVC) is a robust algorithm that classifies data into different classes, searching a hyper-plane that gives maximum margin from the class means so that given model can accurately classify. Support vector machines are concentrated on the data points that are at a shortest distance from this boundary these points known as support vectors are vital in determining the decision boundary. If the data is not separable with ease, there is a trick that SVM uses: the kernel method, where it transforms the data into higher dimensions such that it may be easily separated. The Support Vector Machine is a proper way to solve simple problems and also complex ones. When there are many features or dimensions in the given problem, SVC serves the purpose reasonably well. Its major drawback, however, is that it is relatively tedious to run when datasets are very big. Besides, SVC is somewhat sensitive to additional tuning if one needs to get the best of the tool.

2.2.3 k-nearest neighbors

The k-Nearest Neighbors (abbreviated as kNN) classifier is one of the basic ML algorithms used for classification problems. In the kNN process, k number of similar points around some new unseen point is taken and the new point is classified according to the greatest number of neighbors. Consequently, similar data come close together in the data space. When it is required to classify a new data, kNN looks at the other data which are in the neighborhood of that dataset, and it would give the new data the same label as most of the neighbor data. k is one of the important parameters which indicates how much neighbors to check. When k is smaller, the model would tend to be sensitive to noise and is more difficult to predict. When k gets higher, the model would become more stable and would be hard to predict fine details. In other words, it is a "vote" by its nearest neighbors, where the object is classified as the majority voting class of them.

2.2.4 Gaussian naïve Bayes

The GNB (i.e., Gaussian Naïve Bayes in elaboration) classifier is a type of probability-based classification prediction model in machine learning. It assumes that the features or attributes of the data usually follow a normal, Gaussian distribution. GNB model calculates the probability of each class given features and for a new data point, it gives the label of the class having the highest probability. This simplifies the calculations, which is why the attribution as “naïve”. Although this may not hold all of the time, the algorithm often works well in practice. GNB takes the distribution of each feature and estimates how likely a data point is a member of any individual class, and chooses the class that has the highest likelihood. It is fast, easy to implement, and works well with all kinds of data.

2.2.5 Decision tree

Decision Tree (abbreviated as DT) classifier is a simple and basic ML algorithm that asks a series of simple questions about features of the data and based on the answer learns to make decisions. Further, it splits the data at every step depending on the most important feature to develop a tree-like structure. Each branch is a decision rule, and each leaf would be the final class or another decision rule, a branch. Usually, it is possible to see how it is making predictions, and the decisions are based on if-then rules. Hence, DT easily interpretable. Decision trees are useful in both classification and regression

tasks. However, they can overfit if not tuned properly.

2.2.6 Random forest

The Random Forest (abbreviated as RF) Classifier is an ensemble approach in which predictions are combined using a collection of Decision Trees. The model's name Random Forest is adopted because instead of training a single tree, RF trains a forest of trees, each using a random sample of the data. For classification, RF makes a prediction by considering all of its constituent trees' predictions, taking the most common (or, if it is a regression, its mean). RFs are very powerful in the process of increasing accuracy while minimizing errors. Therefore, if the classification or regression problem involves complex data, it increases the accuracy and reduces the error sufficiently. Moreover, the risk of overfitting is also generally lower than that for a single Decision Tree.

2.2.7 LGBM

LGBM (or Light Gradient Boosting Machine in elaboration) is a strong classifier model in machine learning. This algorithm is based on decision trees aimed to improve the accuracy of the predictions. An LGBM builds trees one after the other such that each new tree consists of the newly corrected errors found in the previous one. In short, LGBM is designed to be fast and efficient, especially on large datasets. The underlying process of LGBM involves "gradient boosting" which helps improve by finding errors of previous trees and fixing the error with the next tree. Because LGBM is very fast, it is often one of choice in many classification tasks even in competitions or real-world applications.

2.2.8 Extreme gradient boosting (XGBoost)

XGBoost classifier is an extremely efficient and powerful machine learning model considering both high accuracy and speed. It builds decision trees one by one, with each correcting the errors from previous trees. It works in a gradient boosting manner. It is designed to operate on large-size data sets efficiently and perform well through fine tuning of the model and reducing errors. In competitions and real-world applications, XGBoost is in widespread use because it does well and allows to customize it for different tasks.

2.2.9 CatBoost

The CatBoost classifier (or Categorical Boosting in elaboration) is a type of machine learning model. It is particularly good at handling categorical data, labels or categories. CatBoost is a gradient boosting algorithm reimplemented atop the gradient boosting framework in C++, along with XGBoost and LightGBM, focusing on how categorical features are treated. CatBoost is faster and more accurate than other approaches. CatBoost automatically converts categorical variables without extra preprocessing. It uses decision trees in a way that lowers overfitting and works with both small and large datasets. CatBoost's simplicity, high performance, and its ability to handle real-world messy data make it known overall.

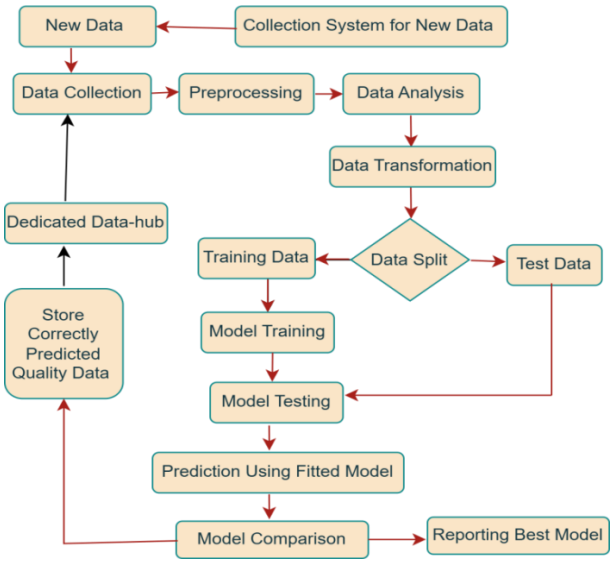

Fig. 2. Flowchart of prediction system for heart disease using machine learning techniques

2.3 Prediction System for Heart Disease

The prediction system for heart disease is presented in the Fig. 2. The system uses new data as well as previously stored data stored in a dedicated data-hub which can be for a single organization or a large community of well-connected medical services and diagnostic centers. Stored data will generally be clean and useful which need not be preprocessed as previously done while the new data require preprocessing and analysis. After data collection from single or both sources, preprocessing, data analysis and data transformation (i.e., MinMax scaler which is defined as $X_{sc} = \frac{X - X_{min}}{X_{max} - X_{min}}$) is applied on feature variables. Then, train-test data split is done as per split ratio. Using training data, different models are trained (namely LR, SVC, kNN, GNB, DT, RF, LGBM, XGBoost and CatBoost) are trained or fitted. Trained models are then used to predict outcome of hold-out test data. Based on model accuracy for test data, models are ranked, stored and made ready for future unknown data. Also, correctly predicted quality data are stored in the centralized data-hub for future use, i.e., training or data-fitting and testing purpose.

2.4 Performance Metrics

In binary classification problems, a prediction can be correct or incorrect in four different ways. Correct in 2 ways and incorrect in 2 other ways. The correct predictions are True Positive (TP) and True Negative (TN). The incorrect predictions are False Positive (FP) and False Negative (FN). TP is a positive value which is predicted correctly; while FP is actually a negative value which is incorrectly predicted as positive. Also, FN is actually a positive value which is incorrectly predicted as a negative value; while TN is a negative value which is predicted correctly. The frequently used evaluation metrics of machine learning classification problems are i. accuracy, ii. precision, iii. recall and iv. F1-score. These metrics (used in this paper) are defined as follows:

$$\text{Accuracy} = \frac{\text{number of all true values}}{\text{number of all values}} = \frac{\text{n(TP)} + \text{n(TN)}}{\text{n(TP)} + \text{n(TN)} + \text{n(FP)} + \text{n(FN)}}$$

$$\text{Precision} = \left(\frac{\text{n(TP)}}{\text{n(TP)} + \text{n(FP)}}\right)$$

$$\text{Recall} = \left(\frac{\text{n(TP)}}{\text{n(TP)} + \text{n(FN)}}\right)$$

$$\text{F}_1 - \text{score} = \text{Harmonic mean of Precision and Recall} = \left(\frac{2 * \text{Precision} * \text{Recall}}{\text{Precision} + \text{Recall}}\right)$$

The most frequently used performance metric of classification problem is accuracy. For high class imbalanced data, precision is more reliable than accuracy.

Also, ROC curve is widely used visual performance tool for binary classification. A receiver operating characteristic (ROC) curve is the graph of true positive rate ($\text{TPR} = \frac{\text{n(TP)}}{\text{n(TP)+n(FN)}}$) against false positive rate ($\text{FPR} = \frac{\text{n(FP)}}{\text{n(FP)+n(TN)}}$) at each threshold.

2.5 k-fold Cross-validation

The k-fold cross-validation is a method by which performance of a machine learning model is validated against potential data bias though various sub-setting of the dataset for training (with known data) and testing (with unknown data). It partitions the dataset into equally partitioned k-subsets (or folds). The model is trained on (k-1) sets and tested on the remaining fold. It used k iterations and during iteration, each fold is the test set only once. Then, all iterations are averaged to give final results indicating model's performance. It is an approach to guarantee that every data point is used for training and testing such that it is a good way to evaluate models. However, it is time consuming since the model is retrained k times. Commonly, 10-fold or 5-fold cross-validation is in practice.

3. RESULTS AND DISCUSSION

The results of experimented two types of train-test splits are presented below in tables and graphs. Table 3. (includes Precision, Recall, F1-score and accuracy), Fig. 3. (graphical accuracy comparison) and Fig. 5. (ROC curve) present model performances scores on hold-out test data for 09 models for 70:30 train-test split. These results reflect that the outperforming model was catBoost with accuracy 90.22%, F1-score 90.08% and recall 89.98%. Also, DT was least performing among all models in the same train-test split ratio with accuracy 83.33%, F1-score 83.08% and recall 82.97%.

Table 4., Fig. 4. and Fig. 6. present model performances scores on out-sample test data of the 09 models for 90:10 train-test split which

include scores of Precision, Recall, F1-score and accuracy, graph of accuracy comparison and ROC curve. These results reflect that the outperforming model was LGBM with accuracy 96.74%, F1-score 96.71% and recall 96.82%. Also, DT was least performing among all models in the same train-test split ratio with accuracy 89.13%, F1-score 89% and recall 89%.

Table 5. presents 10-fold cross-validation results (which include average prediction accuracy along with standard deviation) of the 09 ML models for the heart disease dataset. Based on average accuracy and standard deviation, CatBoost (with accuracy 92.04%) was outperforming model and decision tree (with accuracy 86.7%) was least performing method.

Table 3. Prediction results of 09 machine learning models for heart disease test data with 70:30-split.

Model	Precision	Recall	F1	Accuracy
LR	87.19	87.13	87.16	87.32
SVC	87.98	87.78	87.87	88.04
kNN	85.72	85.66	85.69	85.87
GNB	89	89	89	89.13
DT	83.21	82.97	83.08	83.33
RF	86.99	86.56	86.73	86.96
LGBM	89.12	88.84	88.96	89.13
XGBoost	87.53	87.53	87.53	87.68
CatBoost	90.19	89.98	90.08	90.22

Table 4. Prediction results of 09 machine learning models for heart disease test data with 90:10-split

Model	Precision	Recall	F1	Accuracy
LR	92.74	91.94	92.24	92.39
SVC	93.66	93.16	93.37	93.48
kNN	89.21	88.76	88.94	89.13
GNB	93.4	93.4	93.4	93.48
DT	89	89	89	89.13
RF	95.6	95.6	95.6	95.65
LGBM	96.62	96.82	96.71	96.74
XGBoost	94.43	94.62	94.51	94.57
CatBoost	95.6	95.6	95.6	95.65

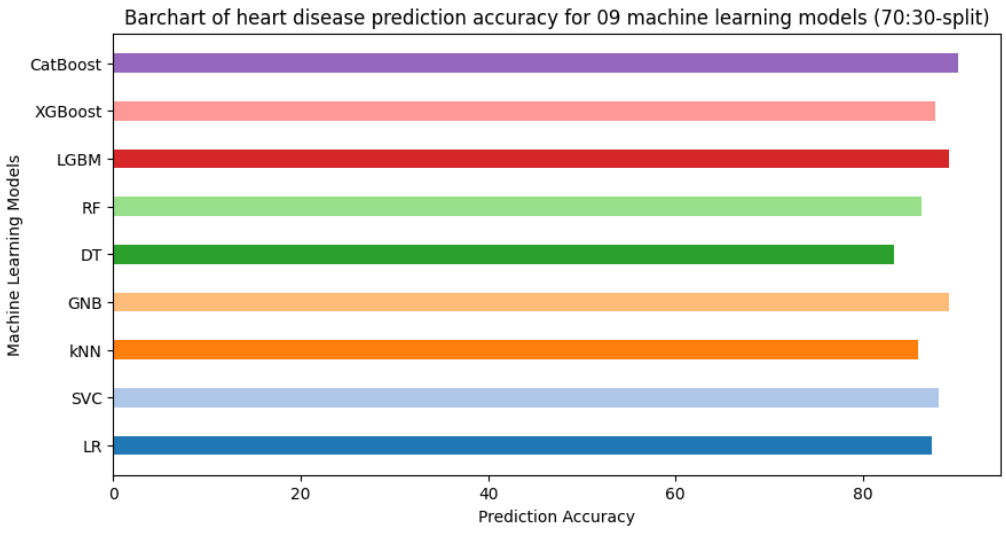

Fig. 3. Barchart of test accuracy comparison of 09 models for heart disease prediction with 70:30-split

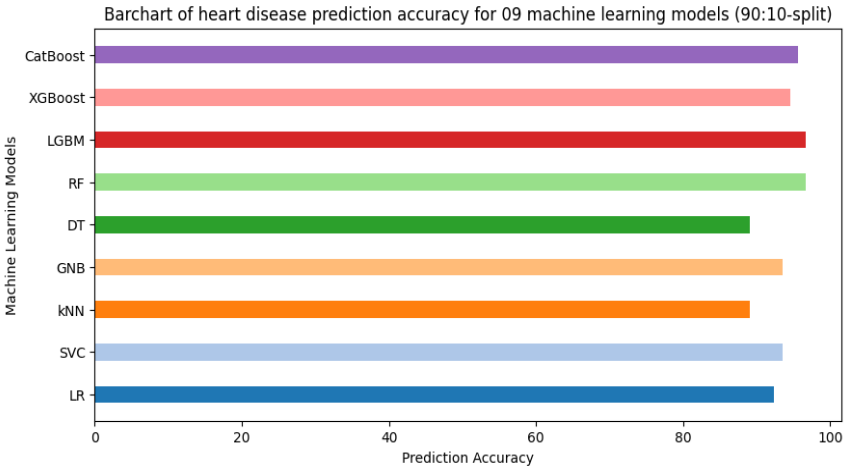

Fig. 4. Barchart of test accuracy comparison of 09 models for heart disease prediction with 90:10-split

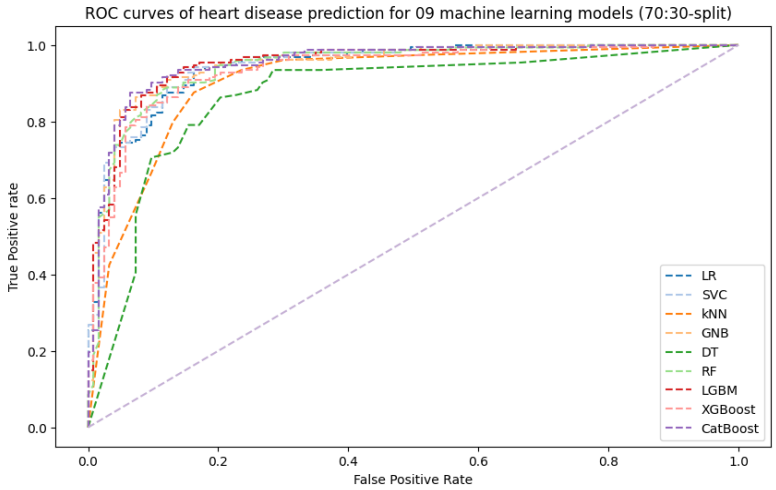

Fig. 5. ROC curves of 09 machine learning models for heart disease prediction with 70:30-split

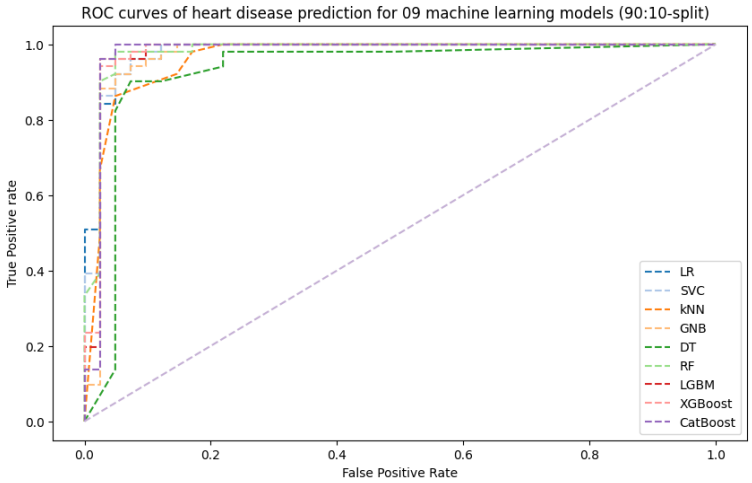

Fig. 6. ROC curves of 09 machine learning models for heart disease prediction with 90:10-split

Table 5.10-fold cross-validation of heart disease test data prediction accuracy

Model	Average Accuracy	Standard Deviation
LR	90.68	3.48
SVC	91.17	2.78
kNN	90.62	3.71
GNB	90.54	3.51
DT	86.7	3.74
RF	91.39	3.19
LGBM	90.93	3.5
XGBoost	90.75	3.93
CatBoost	92.04	2.85

As per the results found from the experimented models on heart disease dataset, varied prediction performance of the different models is clear. The train-test split ratio 90:10 produced comparatively better result than 70:30 split. It reflects that more quality data sample can produce better prediction on out-sample test data. In this scenario, almost all models produced improved predictions, e.g., CatBoost's accuracy improved from 90.22% to 95.65%. Also, LGBM's accuracy improved from 89.13% to 96.74%. Even the least performing model DT's accuracy improved from 83.33% to 89.13%. CatBoost being ensemble method produced better accuracy balancing between underfitting and overfitting. On the other hand, due to its intrinsic characteristic of overfitting tendency, DT performed very poorly. Essentially, all models are neither same nor similar. Data features jointly contribute to proper model training as well as testing for correct predictions. Again, false negative cases are very sensitive and import in some fields especially in medical fields and therefore, Recall score is crucial. For 70:30 test-train split, CatBoost's Recall was 89.98% while it improved to 95.6% in 90:10 split ratio. Also, LGBM's Recall improved from 88.84% to 96.82%. Even, least performing DT's Recall improved from 82.97% to 89%. Also, in the 10-fold cross-validation results (Table 5.) CatBoost was best performing ML model with highest average prediction accuracy of 92.04 and standard deviation of these prediction accuracies is comparatively low, i.e., 2.85%. Thus, comparatively outperforming models like CatBoost or LGBM can be used through continual assessment through disease prediction system. If a model performed better, it will be ranked high; if performed worse, it will be ranked low. Machine learning approaches mainly being data-driven, the best models are required to best performers on the unseen test data. Thus, in each field of application or even in each data category like heart disease data, specific model can do better than others due to the properties and shapes of the input features. Even, practitioners may experience that particular model frequently can outperform other models.

Limitation of this study is that it experimented a medium-sized dataset (with 918 samples and 11 features) combined from other similar single datasets available and only 09 models were experimented. Therefore, more data and various other potential ML models can give different results and insights. Hence, further research with more datasets and ML models are required to expect less biased prediction through adequate quality data and robust models.

Primarily, in addition to doctor's and practitioner's investigation, such machine learning based prediction can give them support and confidence on currant patient data through past available data. Also, increasing quality data through further research, continued investigation and assessment, more accuracy and Recall score might be achieved where high quality data and best performing model will provide highly improved performance. Such eco-system of disease out-come prediction along with data analytics can increase doctor's confidence in efficient diagnosis. If usefulness is proven and disease prediction technology is user friendly, it can tremendously impact as well as accelerate healthcare service.

Future plan is to work on development and implementation of machine learning as well as deep learning for other healthcare or disease data and any other data-driven decision-making fields.

4. CONCLUSION

This study focused on a heart disease prediction system where 09 machine learning techniques were experimented on heart disease datasets

with 11 feature variables for binary classification. Among the models, CatBoost and LGBM were found to be outperforming among others with accuracy scores of 90.22% and 89.13% respectively in 70:30 train-test split and 95.65% and 96.74% respectively for 90:10 train-test split. Also, both models Recall score are satisfactorily high (i.e., 89.98% and 88.84% respectively in 70:30 train-test split and 95.6% and 96.82% respectively in 90:10 train-test ratio. In addition, CatBoost achieved highest average accuracy of 92.04% in 10-fold cross-validation test. Therefore, such heart disease prediction system can be of significant assistance to healthcare and diagnostic services. Also, incremental health data in future can boost usefulness of machine learning-based disease prediction system.

CONSENT AND ETHICAL APPROVAL

It is not applicable.

DISCLAIMER (ARTIFICIAL INTELLIGENCE)

Author(s) hereby declare that NO generative AI technologies such as Large Language Models (ChatGPT, COPILOT, etc) and text-to-image generators have been used during writing or editing of this manuscript.

ACKNOWLEDGEMENTS

The author would like to thank Andras Janosi, M.D. (Hungarian Institute of Cardiology. Budapest), William Steinbrunn, M.D. (University Hospital, Zurich, Switzerland), Matthias Pfisterer, M.D. (University Hospital, Basel, Switzerland) and Robert Detrano, M.D., Ph.D (V.A. Medical Center, Long Beach and Cleveland Clinic Foundation). The authors is also thankful to the Donor David W. Aha (aha '@' ics.uci.edu).

COMPETING INTERESTS

Author has declared that no competing interests exist.

REFERENCES

1. Wang J. Heart failure prediction with machine learning: A comparative study. In Journal of Physics: Conference Series. 2021;2031:012068. IOP Publishing.
2. Olsen CR, Mentz RJ. Anstrom KJ, Page D, Patel PA. Clinical applications of machine learning in the diagnosis, classification, and prediction of heart failure. Am. Heart J. 2020;229:1–17.
3. Angraal S, Mortazavi BJ, Gupta A, Khera R, Ahmad T, Desai NR, Jacoby DL, Masoudi FA, Spertus JA, Krumholz HM. Machine learning prediction of mortality and hospitalization in heart failure with preserved ejection fraction. JACC Heart Fail. 2020;8:12–21.
4. Sciomer S, Moscucci F, Salvioni E, Marchese G, Bussotti M, Corrà U, Piepoli MF. Role of gender, age and BMI in prognosis of heart failure. Eur. J. Prev. Cardiol. 2020;27:46–51.
5. Chicco D, Jurman G. Machine learning can predict survival of patients with heart failure from serum creatinine and ejection fraction alone. BMC Med. Inform. Decis. Mak. 2020;20:1–16.
6. Khan MS, Arshad MS, Greene SJ, Van Spall HG, Pandey A, Vemulapalli S, Perakslis E, Butler J. Artificial intelligence and heart failure: A state-of-the-art review. Eur. J. Heart Fail. 2023;25: 1507–1525.
7. Guo S, Zhang H, Gao Y, Wang H, Xu L, Gao Z, Guzzo A, Fortino G. Survival prediction of heart failure patients using motion-based analysis method. Comput. Methods Programs Biomed. 2023;236:107547.
8. Singhal S. Predicting congestive heart failure using predictive analytics in AI. Int. J. Creat. Res. Comput. Technol. Des. 2023;5:1–10.
9. Li D, Fu J, Zhao J, Qin J, Zhang L. A deep learning system for heart failure mortality prediction. Plos One. 2023;18:e0276835.
10. Akbilgic O, Butler L, Karabayir I, Chang PP, Kitzman DW, Alonso A, Chen LY, Soliman EZ. ECG-AI: electrocardiographic artificial intelligence model for prediction of heart failure. Eur. Heart J.-Digit. Health. 2021;2:626–634.
11. Wang F, Harel-Sterling L, Cohen S, Liu A, Brophy JM, Paradis G, Marelli AJ. Heart failure risk predictions in adult patients with congenital heart disease: A systematic review. Heart . 2019;105:1661–1669.
12. Mohan S, Thirumalai C, Srivastava G. Effective heart disease prediction using hybrid machine learning techniques. IEEE Access. 2019;7:81542–81554.
13. Shah D, Patel S, Bharti SK. Heart disease prediction using machine learning techniques. SN Comput. Sci.2020;1:345.
14. Jindal H, Agrawal S, Khera R, Jain R, Nagrath P. Heart disease prediction using

machine learning algorithms. In IOP conference series: materials science and engineering. 2021;1022: 012072. IOP Publishing.
15. Sharma V, Yadav S, Gupta M. Heart disease prediction using machine learning techniques. In 2020 2nd international conference on advances in computing, communication control and networking (ICACCCN). 2020:177–181. IEEE.
16. Riyaz L, Butt MA, Zaman M, Ayob O. Heart disease prediction using machine learning techniques: a quantitative review. In International Conference on Innovative Computing and Communications: Proceedings of ICICC2021.2022;3:1–94. Springer.
17. Heart Disease Data. (n.d.). Avaliable:https://www.kaggle.com/datasets/fedesoriano/heart-failure-prediction,(Accessed: 2024-03-19)
18. Heart Failure Prediction Dataset. (n.d.). Avaliable:https://archive.ics.uci.edu/dataset/45/heart+disease, (Accessed: 2024-03-18)

Peer-review history:
The peer review history for this paper can be accessed here:
https://www.sdiarticle5.com/review-history/128065