

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/343012772>

Data-Driven Diagnosis of Heart Disease

Article in *International Journal of Computer Applications* · July 2020

DOI: 10.5120/ijca2020920549

CITATIONS

12

READS

544

2 authors, including:



[M. Rubaiyat Hossain Mondal](#)

Bangladesh University of Engineering and Technology

92 PUBLICATIONS 1,494 CITATIONS

SEE PROFILE

Data-Driven Diagnosis of Heart Disease

Md. Istiaq Habib Khan

Institute of Information and Communication
Technology, Bangladesh University of Engineering
and Technology, Dhaka-1205, Bangladesh

M. Rubaiyat Hossain Mondal

Institute of Information and Communication
Technology, Bangladesh University of Engineering
and Technology, Dhaka-1205, Bangladesh

ABSTRACT

This paper focuses on the data-driven diagnosis of heart disease using three freely available datasets. The first dataset has 303 instances with 14 attributes, the second dataset has 462 instances with 10 attributes and the third dataset has 70000 instances with 12 attributes. Scikit-learn library of Python programming language is used for data analysis purpose. Univariate feature selection algorithm is applied in order to find the most valuable attributes and risk factors associated with heart disease. Experimental results show that the most important attribute of the first dataset is the maximum heart rate achieved by a patient, while that of the second and third dataset is the patient age. Next, the heart disease is predicted using several machine learning algorithms including support vector machine (SVM), decision tree, k-nearest neighbors (kNN), logistic regression, naïve Bayes, random forest and majority voting. The training and testing portion of each dataset is separated using holdout and cross-validation methods. The performance of different algorithms for three datasets are evaluated in terms of testing accuracy, precision, recall and F1-score. It is shown here that majority voting as a combination of logistic regression, SVM and naïve Bayes exhibits the best accuracy of 88.89% when applied to the first dataset.

Keywords

feature selection, heart disease, SVM, logistic regression, recall, machine learning, disease prediction.

1. INTRODUCTION

Heart attack or myocardial infarction is one of the deadliest diseases in the world at present as it is the major cause of death and disability in many developed and developing countries [1]. Most heart attacks occur due to coronary artery disease. The patient suffering from a heart attack needs treatment within a very short time. So, it is very important to find out if a person is at risk of having a heart attack considering the risk factors associated with it. Machine learning algorithms [2-4] are considered in different application areas including disease prediction. A number of studies have also been reported for coronary artery disease or heart disease [5-10]. However, the findings of these works vary and one of the reasons behind this is the consideration of different attributes and collection of different datasets by different authors. Accuracy of the results is reduced when the medical data is incomplete. On the other hand, accurate analysis of heart datasets enables early heart disease detection, and optimizes patient care and treatment facilities. Therefore, research is still needed to find the most important attributes and how selection of the attributes influences the prediction of heart disease. This paper focuses on multiple datasets for finding the important attributes associated with heart disease. The paper then focuses on predicting heart disease by applying different machine learning algorithms to different datasets. The contributions of this paper can be summarized as

- To apply feature selection method to find the most influential attributes of three different datasets for predicting heart disease.

- To apply different machine learning algorithms to the three datasets for the prediction of heart disease.

- To compare the performance of the classifiers for the case of three datasets, and to compare with the results reported in the literature.

The findings may be used as automatic prediction of heart disease assisting practitioners in the clinical treatment of cardiovascular abnormality. The rest of the paper is organized as follows. Section 2 provides a brief description of the relevant work of heart disease. The overall methodology of this research is presented in Section 3. Results and discussion are provided in Section 4. Section 5 provides the concluding remarks.

2. RELATED WORK

A number of research papers report the use of machine learning algorithms to predict heart disease [5-10]. Different data mining techniques, performance tools and methods have been implemented which provide different perspective on the prediction of heart disease.

In [5], linguistic fuzzy-rule based classification system is used to predict heart disease [5]. In this study [5], a new methodology is proposed where fuzzy-rule based classification system is combined with interval-valued fuzzy sets. An evidence based fuzzy decision support system is applied for the diagnosis of coronary artery disease [6]. Three processing stages are used for the implementation of decision support system development. Rough set theory is used in [6] for rule generation from training dataset. Rule selection is done by transforming the rules into a decision table based on unseen dataset. Finally, selected rules are transformed into fuzzy rules where fuzzy rules weighing is also applied. Different types of decision tree algorithms are used to find out which performs better in predicting heart disease [7]. This study in [7] uses a model that combines discretization, decision tree type and voting to find out a more accurate method for heart disease prediction. The sensitivity, specificity and accuracy are calculated in order to compare the performance of different types of decision trees. A computer-based noninvasive coronary artery disease diagnosis system is used in [8]. The target of the study in [8] is to design a clinically interpretable fuzzy rule-based system. Discretization is done for the interval-scale variables and then the fuzzy rule-based system is formulated based on a neuro-fuzzy classifier. Multiple logistic regression and sequential feature selection methods are used. The combination of multiple logistic regression and neuro-fuzzy classifier method has exhibited the best performance [8]. A web-based fuzzy logic expert system is developed for the diagnosis of heart disease in [9]. The system consists of fuzzification module, knowledge-based

interface engine and defuzzification module. Fuzzification module operates on every input based on appropriate membership function. Then, the interface engine triggers the appropriate rule from knowledge base to find out the output value using appropriate defuzzification method. The study in [10] uses a comparative analytical approach to show that ensemble techniques can be applied to improve accuracy in heart disease prediction. Ensemble techniques like bagging, boosting etc. are proved to be effective in improving the prediction accuracy of weak classifiers. Feature selection technique is applied furthermore and it has enhanced the prediction accuracy [10]. However, none of the above research work show the variation in the results using multiple datasets.

3. METHODOLOGY

In this research, experiments were performed to classify normal and abnormal hearts from the samples of suspected heart patients available in the dataset. Using the concepts of disease diagnosis reported in [11], the following processes were followed in this work.

Three different datasets were used in this research. The first dataset [12] hereinafter termed as “dataset 1” was collected from UCI machine learning repository containing 303 instances (139 abnormal hearts) and 14 attributes including the target attribute. This means there are 13 attributes for predicting the target attribute. However, six rows had a large number of null values, and thus these instances were removed. The remaining 297 instances (137 abnormal hearts) were considered in this research. The second dataset [13] hereinafter termed as “dataset 2” was collected from Kaggle containing 10 attributes including the target attribute and 462 instances (160 abnormal hearts). The third dataset [14] hereinafter termed as “dataset 3” was collected from Kaggle containing 12 attributes including the target attribute and 70000 instances (34979 abnormal hearts). The data analysis of this work was carried out using Python programming language. For the deployment of python, Anaconda distribution [15] was used which is a free and open-source distribution that simplify package management. It includes a graphical user interface known as Anaconda navigator. For the implementation of this research, Jupyter notebook was used to run the codes for data analysis.

After that feature selection was performed to find the ranking of the features. The features were then used to classify the data using classification algorithms. Different supervised machine learning algorithms including support vector machine (SVM), decision tree, k-nearest neighbors (kNN), naïve Bayes, random forest, logistic regression and majority voting classifiers were used [16-24]. To implement these algorithms, scikit-learn library was used which is a free software machine learning library which is included in the Anaconda distribution package. The accuracy score for predicting heart disease was calculated for three different datasets. This was done by holdout method and by cross-validation method. For the case of holdout method, the percentage of training and testing data was set to four different values. These were testing size of 10%, 15%, 20% and 25% of the total data samples. The *train_test_split()* class from scikit-learn library was used to split the datasets into train and testing portions. For cross-validation, the total dataset was divided into k equal groups. Different parameters associated with the different learning algorithms were altered and applied to compare between the results to find out the desired condition that gives the highest accuracy score. In SVM algorithm, three different kernels were applied: linear,

RBF and sigmoid. For RBF and sigmoid kernels, the C value was altered from 1 to 5, and the gamma value was altered between ‘auto’ and ‘scale’. In case of kNN classifier, the value of k was changed from 1 to 50 to find out the best possible accuracy score. In logistic regression algorithm, two different solvers were applied: lbfgs and liblinear. The performance of the classifiers was compared in terms of testing accuracy, precision, recall and F1-score.

First feature selection is performed to find the best attributes of a dataset that lead to the diagnosis of heart disease. In this research, the feature selection method is *univariate feature selection* method where each feature is scored individually on certain specified criteria and the features are then selected based on the higher scores or higher ranks. The *SelectKBest()* class was applied using chi-square distribution function, *chi2()* to find the k number of best features from each dataset [11]. Next, all the features and a subset of most important features were then taken into the classifiers for heart disease prediction. The classifiers used in this research are described in the following. Support-vector machine (SVM) has several kernels, these are linear, sigmoid, radial basic functions (RBF), and polynomials kernels [19-20]. These SVM kernels have a regularization parameter C to control the trade-off between training and testing accuracy values. Moreover, the RBF, sigmoid and polynomial kernels have an addition parameter known as gamma, γ to control the impact of training data samples.

Similar to the work of diagnosis of diseases reported in the literature [11], this work considers a number of metrics for appropriate diagnosis of the patients with heart disease. The metrics are testing accuracy, precision, recall, and F1-measure. These metrics can be defined using several terms such as true positive (TP), true negative (TN), false negative (FN) and false positive (FP). In the context of this work, TP refers to the suspected heart patient that are correctly classified as abnormal which mean the patients have heart disease. The terms TN is the number of suspected patients having normal condition of the heart. The term FN refers to the people who actually have heart disease but remains undetected by the system. Furthermore, FP refers to the number of samples who are wrongly detected to have heart disease. The accuracy is the percentage of all normal and abnormal vectors that are correctly classified. Training accuracy and testing accuracy are defined as the accuracy obtained for training and testing samples, respectively. Precision, recall and F1 score can be mathematically expressed using TP , TN , FP and FN as shown in [11]. The recall value is important in the context of any disease including heart disease prediction, as it measures the correctness of classifying an abnormal heart case as abnormal. A very high value of recall indicates that there are only a few cases where a heart patient is wrongly classified as normal. Confusion matrix determines the performance of a classification algorithm by showing the predicted instances and the actual instances. In addition, area under the receiver operating characteristics (ROC) curve known as AUC is an important metric of classifiers. The AUC indicates how well a classifier can separate the positive and negative cases of a data sample.

4. RESULTS AND DISCUSSION

Currently, the diagnosis of heart disease is more important than ever, because patients of heart disease are more vulnerable to coronavirus [24-25] disease 2019 (COVID-19). This section provides results and associated discussion on data-driven diagnosis of heart disease.

4.1 Feature Ranking

Table 1 shows the ranking of features of dataset 1 based on the univariate feature selection algorithm. It can be seen that thalach is the most prominent feature in dataset 1. Table 2 shows the ranking of features of dataset 2. It can be seen that age is the most prominent feature in dataset 2. Table 3 shows the ranking of features of dataset 3. It can be seen that similar to dataset 2, age is the most prominent feature in dataset 3.

Table 1: Feature ranking for dataset 1

Feature	Description	Rank
thalach	Maximum heart rate achieved	1
ca	Number of major vessels (0-3) colored by fluoroscopy	2
oldpeak	ST depression induced by exercise relative to rest	3
thal	3 = Normal; 6 = Fixed defect; 7 = Reversible defect	4
exang	Exercise induced angina	5
age	Age in years	6
chol	Serum cholesterol in mg/dl	7
trestbps	Resting blood pressure (in mmHg on admission to the hospital)	8
cp	Chest pain type	9
restecg	Resting electrocardiographic results	10
slope	The slope of the peak exercise ST segment	11
sex	1 = Male; 0 = Female	12
fbs	Fasting blood sugar > 120 mg/dl: 1 = true; 0 = false	13

Table 2: Feature ranking for dataset 2

Feature	Description	Rank
age	Age in years	1
tobacco	Cumulative tobacco	2
adiposity	A numeric vector	3
alcohol	Current alcohol consumption	4
sbp	Systolic blood pressure	5
ldl	Low density lipoprotein cholesterol	6
famhist	Family history of heart disease	7
typea	Type-A behavior	8
obesity	A numeric vector	9

Table 3: Feature ranking for dataset 3

Feature	Description	Rank
age	Age in days	1
ap_lo	Diastolic blood pressure	2
ap_hi	Systolic blood pressure	3

Feature	Description	Rank
weight	Weight in kg	4
cholesterol	1 = Normal; 2 = Above normal; 3 = Well above normal	5
gluc	1 = Normal; 2 = Above normal; 3 = Well above normal	6
active	Physical activity	7
smoke	Whether patient smokes or not	8
alco	Alcohol intake	9
height	Height in cm	10
gender	1 = Female; 2 = Male	11

4.2 Classification Results for Dataset 1

This section provides the results and associated discussion on the performance of heart disease prediction using dataset 1. Table 4 shows the testing accuracy of dataset 1 for different classifiers. In this case, the parameters of the classifiers are varied. For example, for the case of SVM, three different kernels are considered and the values of C and gamma are varied. Table 4 indicates that among all the classifiers, majority voting exhibits the best accuracy of 88.89% when the test size is 15% of the data samples. Moreover, it can be seen that each classifier has its own best accuracy score for a set of conditions. These are SVM (kernel=linear, C=2; test size 10%): 85.67%, logistic regression (solver=liblinear; test size 15%): 85.11%, naïve Bayes classifier (test size 15%): 84.22%, random forest classifier (test size 20%): 83.83%, KNN classifier (k=5; test size 10%): 76.67%, and decision tree classifier (test size 10%): 76.33%. Next confusion matrix and classification reports are calculated for these conditions and the results are shown in Table 5. It can be seen from Table 5 that among all the classifiers, majority voting has the best precision of 90%, the best recall of 86% and the best F1 score of 88%. The recall value of 86% mean that the classifier can correctly classify the abnormal heart case as abnormal 86 times out of 100. In other words, there is a probability of 14% that a patient with abnormal heart is undetected or treated as normal. Note that the AUC value is calculated for majority voting classifier and it is found to be 86.61%. This means the classifier has a probability of 86.61% of separating the cases of abnormal and normal hearts.

The testing accuracy of the dataset 1 with cross-validation is discussed next. Table 6 shows the testing accuracy of dataset 1 with cross-validation for different classifiers. For dataset 1 using cross-validation method, the best accuracy of 84.83% is obtained by majority voting classifier with logistic regression, SVM and naïve Bayes where k=10. This result is for the case when all the 13 features are used for prediction. This means that the classification of normal and abnormal hearts can be done with 84.83% accuracy when majority voting is used on all the attributes of dataset 1. The highest accuracy values of the individual classifiers for dataset 1 is illustrated in Figure 1. It can be seen from Figure 1 that SVM, logistic regression, naïve Bayes, random forest, and majority voting exhibit accuracy values above 80%.

Next, univariate feature selection method is applied to see the impact of reducing the number of features in predicting heart patients. From the experiments, the best result is obtained when the highest 11 ranked features shown in Table 1 are

used for classification. Using those 11 features, an 85.15% accuracy is obtained with majority voting classifier in

classifying normal and abnormal cases.

Table 4: Accuracy score for dataset 1

Algorithms			Test=25%	Test=20%	Test=15%	Test=10%
Decision Tree Classifier			74.27%	73.83%	75.33%	76.33%
SVM	Kernel = linear	C = 1	83.73%	83.83%	85.11%	84.67%
		C = 2	82.27%	83.16%	83.55%	85.67%
		C = 3	81.73%	82.33%	80.89%	83.33%
		C = 4	81.87%	81.66%	83.78%	84.33%
		C = 5	82.8%	83.33%	81.56%	84%
	Kernel = rbf; Gamma = auto	C = 1	53.87%	53.5%	57.55%	57.67%
		C = 2	52.66%	53.33%	55.3%	56.33%
		C = 3	56.13%	53.83%	59.11%	57.3%
		C = 4	56.53%	56.67%	52.22%	55.66%
		C = 5	56.27%	54.16%	56.89%	53.33%
	Kernel = rbf; Gamma = scale	C = 1	67.86%	66.33%	64.67%	63.66%
		C = 2	67.6%	66.83%	68%	68.33%
		C = 3	69.2%	66.83%	67.11%	65.67%
		C = 4	69.07%	69.16%	68.89%	66.33%
		C = 5	68.13%	70.5%	70%	67.33%
	Kernel = sigmoid; Gamma = auto	C = 1	54.66%	51.33%	53.78%	55%
		C = 2	53.07%	56.16%	54.44%	54.33%
		C = 3	54.53%	57.83%	51.56%	54%
		C = 4	51.6%	53.67%	55.78%	57.33%
		C = 5	54.27%	51%	55.56%	54.3%
	Kernel = sigmoid; Gamma = scale	C = 1	53.2%	54.17%	54.66%	59.33%
		C = 2	53.33%	60.16%	58.22%	60.67%
		C = 3	58.8%	54.67%	61.11%	57%
		C = 4	59.07%	58.66%	56.44%	57.67%
		C = 5	55.2%	55.17%	52.89%	57.66%
kNN Classifier			69.33% [k=14]	73.33%	68.89% [k=5]	76.67%
Naïve Bayes Classifier			83.33%	81.3%	84.22%	83%
Random Forest Classifier			80.8%	83.83%	82.89%	80.33%
Logistic Regression	Solver = lbfgs		84.13%	82.17%	81.78%	83.33%
	Solver = liblinear		84.53%	81.33%	85.11%	84.3%
Majority Voting Classifier			85.33%	88.3%	88.89%	86.67%

Table 5: Confusion matrix and classification report for dataset 1

Algorithms	Confusion Matrix				Classification Report		
	TN	FP	FN	TP	Precision	Recall	F1-score
Decision Tree Classifier, Test size: 10%	12	4	3	11	73%	79%	76%
SVM (Kernel = linear; C=2), Test size: 10%	14	2	2	12	86%	86%	86%
kNN Classifier (k = 5), Test size: 10%	10	3	5	12	80%	71%	75%
Naïve Bayes, Test size: 15%	17	3	4	21	88%	84%	86%
Random Forest Classifier, Test size: 20%	29	3	7	21	88%	75%	81%
Logistic Regression (Solver = liblinear), Test size: 15%	20	3	4	18	86%	82%	84%
Majority Voting Classifier, Test size: 15%	22	2	3	18	90%	86%	88%

Table 6: Accuracy score with cross-validation for dataset 1

Algorithms		k=10	k=20	k=30
Decision Tree Classifier		77.13%	75.07%	74.3%
SVM	Kernel = linear	83.15%	82.81%	82.23%
	Kernel = rbf; Gamma = auto	54.89%	54.9%	54.83%
	Kernel = rbf; Gamma = scale	66.34%	65.98%	67.51%
	Kernel = sigmoid; Gamma = auto	53.89%	53.9%	53.92%
	Kernel = sigmoid; Gamma = scale	53.55%	53.9%	53.92%
KNN Classifier		68.71% [k=23]	69.76% [k=7]	69.8% [k=30]
Naïve Bayes Classifier		84.5%	83.81%	83.05%
Random Forest Classifier		83.15%	83.45%	82.13%

Algorithms		k=10	k=20	k=30
Logistic Regression	Solver = lbfgs	83.16%	82.78%	82.26%
	Solver = liblinear	83.5%	83.48%	82.6%
Majority Voting Classifier		84.83%	83.81%	82.9%



Figure 1: Accuracy score with cross-validation for dataset 1

4.3 Classification Results for Dataset 2

This section provides the results and associated discussion on the performance of heart disease prediction using dataset 2. Experimental results show that among all the classifiers, majority voting exhibits the best accuracy of 76.60% when the test size is 10% of the data samples. Moreover, it can be seen that each classifier has its own best accuracy score for a set of conditions. These are SVM (kernel=linear; C=1; test size 25%): 73.88%, logistic regression (solver=lbfgs; test size 10%): 73.83%, naïve Bayes (test size 15%): 72.28%, KNN classifier (k=11; test size 10%): 70.2%, random forest (test size 25%): 70.17% and decision tree (test size 25%): 61.55%. Next confusion matrix and classification reports are calculated for these conditions and the results are shown in Table 7. It can be seen from Table 7 that among all the classifiers, majority voting and logistic regression have the best precision of 69%, naïve Bayes has the best recall of 65% and logistic regression has the best F1 score of 65%. Note that the AUC value is calculated for majority voting classifier and it is found to be 74.70%. The testing accuracy of the dataset 2 with cross-validation is discussed next. Table 8 shows the testing accuracy of dataset 2 with cross-validation for different classifiers. For dataset 2 using cross-validation method, the best accuracy of 72.72% is obtained by logistic regression (solver = lbfgs) where k=30. This result is for the case when all the 9 features are used for prediction. Next, univariate feature selection method is applied to see the impact of reducing the number of features. From the experiments, the best result is obtained when the highest 8 ranked features shown in Table 2 are used for classification. Using those 8 features, a 73.62% accuracy is obtained with logistic regression classifier.

4.4 Classification Results for Dataset 3

This section provides the results and associated discussion on the performance of heart disease prediction using dataset 3. Experimental results show that among all the classifiers,

majority voting exhibits the best accuracy of 74.80% when the test size is 10% of the data samples. Moreover, it can be seen that each classifier has its own best accuracy score for a set of conditions. These are SVM (kernel=linear; C=3; test size 20%): 72.98%, random forest classifier (test size 15%): 72.06%, KNN classifier (k=47; test size 10%): 71.88%, logistic regression (solver=liblinear; test size 20%): 71.64%, decision tree classifier (test size 20%): 63.55%, and naïve Bayes classifier (test size 10%): 59.4%. Next confusion matrix and classification reports are calculated for these conditions and the results are shown in Table 9. It can be seen from Table 9 that among all the classifiers, majority voting has the best precision of 76%, random forest has the best recall of 69%, random forest and SVM have the best F1 score of 71%. Note that the AUC value is calculated for majority voting classifier and it is found to be 73.02%. The testing accuracy of the dataset 3 with cross-validation is discussed next. Table 10 shows the testing accuracy of dataset 3 with cross-validation for different classifiers. For dataset 3 using cross-validation method, the best accuracy of 72.22% is obtained by SVM (linear kernel) where k=10. This result is for the case when all the 11 features are used for prediction. Next, univariate feature selection method is applied to see the impact of reducing the number of features. From the experiments, there is no improvement in the accuracy score rather the accuracy decreases with the reduction in the number of features.

4.5 Comparative Results

From sections 4.2-4.4 it can be seen that the results are different for different classifiers. Results also vary when the parameters of the classifiers are varied, holdout or cross-validation methods are used. Table 11 presents the conditions at which the best accuracy values are obtained for three datasets. It can be seen that among all the cases shown in Table 11, the highest accuracy value of 88.89% is obtained for dataset 1 using majority voting algorithm. Figure 2 presents

the best accuracy values that are obtained for holdout/cross-validation methods and for three different datasets used in this work. From Figure 2 it can be seen that the accuracy value is always higher for the case of dataset 1 compared to other two datasets.

Among the three individual datasets used in this paper, the dataset 1 from UCI repository is also used in [7-8, 10] for the prediction of heart disease. A comparison is done among the best results obtained from [7-8, 10] with the best result obtained in this work for this particular dataset. This comparison is shown in Table 12. It can be seen from Table

12 that the majority voting used in this work gives an accuracy of 88.89% which is better than other accuracy values of 84.10%, 84% and 85.48% reported in [7], [8] and [10], respectively. Hence, the proposed majority voting algorithm as a combination of logistic regression, SVM and naïve Bayes can predict heart disease more reliably than the classifiers mentioned in the literature. This proposed machine learning algorithm along with sensors and wireless communication techniques [26-28] can contribute in the concept of Internet of medical things (IoMT) that will identify and monitor heart disease in real time.

Table 7: Confusion matrix and classification report for dataset 2

Algorithms	Confusion Matrix				Classification Report		
	TN	FP	FN	TP	Precision	Recall	F1-score
Decision Tree Classifier, Test size: 25%	54	23	16	23	50%	59%	54%
SVM (Kernel = linear; C=1), Test size: 25%	63	16	19	18	53%	49%	51%
KNN Classifier (k = 11), Test size: 10%	23	7	9	8	53%	47%	50%
Naïve Bayes, Test size: 15%	32	12	9	17	59%	65%	62%
Random Forest Classifier, Test size: 25%	56	21	16	23	52%	59%	55%
Logistic Regression (Solver = lbfgs), Test size: 10%	24	5	7	11	69%	61%	65%
Majority Voting Classifier, Test size: 10%	26	4	8	9	69%	53%	60%

Table 8: Accuracy score with cross-validation for dataset 2

Algorithms		k=10	k=20	k=30
Decision Tree Classifier		62.55%	62.12%	63.32%
SVM	Kernel = linear	70.78%	71.19%	71.17%
	Kernel = rbf; Gamma = auto	65.37%	65.36%	65.42%
	Kernel = rbf; Gamma = scale	66.45%	65.35%	65.41%
	Kernel = sigmoid; Gamma = auto	65.37%	65.36%	65.42%
	Kernel = sigmoid; Gamma = scale	60.41%	59.34%	58.54%
KNN Classifier		67.94% [k=16]	68.36% [k=16]	67.75% [k=16]
Naïve Bayes Classifier		71.003%	70.34%	70.83%
Random Forest Classifier		69.69%	68.6%	70.14%
Logistic Regression	Solver = lbfgs	72.07%	72.27%	72.72%
	Solver = liblinear	71.86%	71.63%	72.1%
Majority Voting Classifier		72.08%	72.27%	72.49%

Table 9: Confusion matrix and classification report for dataset 3

Algorithms	Confusion Matrix				Classification Report		
	TN	FP	FN	TP	Precision	Recall	F1-score
Decision Tree Classifier, Test size: 20%	4533	2560	2526	4381	63%	63%	63%
SVM (Kernel = linear; C=3), Test size: 20%	5330	1603	2283	4784	75%	68%	71%
KNN Classifier (k = 47), Test size: 10%	2760	798	1224	2218	74%	64%	69%
Naïve Bayes, Test size: 10%	3183	391	2401	1025	72%	30%	42%
Random Forest Classifier, Test size: 15%	3775	1411	1623	3691	72%	69%	71%
Logistic Regression (Solver = liblinear), Test size: 20%	5307	1783	2315	4595	72%	66%	69%
Majority Voting Classifier, Test size: 10%	2717	715	1297	2271	76%	64%	69%

Table 10: Accuracy score with cross-validation for dataset 3

Algorithms		k=10	k=20	k=30
Decision Tree Classifier		63.4%	63.13%	63.4%
SVM	Kernel = linear	72.22%	72.1%	72.2%
	Kernel = rbf; Gamma = auto	52.84%	53.24%	53.28%
	Kernel = rbf; Gamma = scale	59.78%	59.72%	59.72%
	Kernel = sigmoid; Gamma = auto	50.04%	50.04%	50.03%
	Kernel = sigmoid; Gamma = scale	40.2%	41.12%	40.18%

Algorithms		k=10	k=20	k=30
KNN Classifier		71.16% [k=47]	71.22% [k=43]	71.23% [k=30]
Naïve Bayes Classifier		59.02%	59.02%	59.03%
Random Forest Classifier		71.56%	71.43%	71.43%
Logistic Regression	Solver = lbfgs	69.8%	69.87%	69.87%
	Solver = liblinear	70.73%	70.77%	70.7%
Majority Voting Classifier		71.58%	71.58%	71.48%

Table 11: Comparative results for three datasets

Dataset	Accuracy (Holdout, All features)	Accuracy (Cross- validation, All features)	Accuracy (Cross-validation, Selected Features)
Dataset 1	All 13 features, Majority voting, test size=15%: 88.89%	All 13 features, Majority voting, k=10: 84.83%	Best 11 features, Majority voting, k=10: 85.15%
Dataset 2	All 9 features, Majority voting, test size=10%: 76.60%	All 9 features, Logistic regression, k=30: 72.72%	Best 8 features, Logistic regression, k=30: 73.62%
Dataset 3	All 11 features, Majority voting, test size=10%: 74.80%	All 11 features, SVM (linear kernel), k=10: 72.22%	Best/All 11 features, SVM (linear kernel), k=10: 72.22%

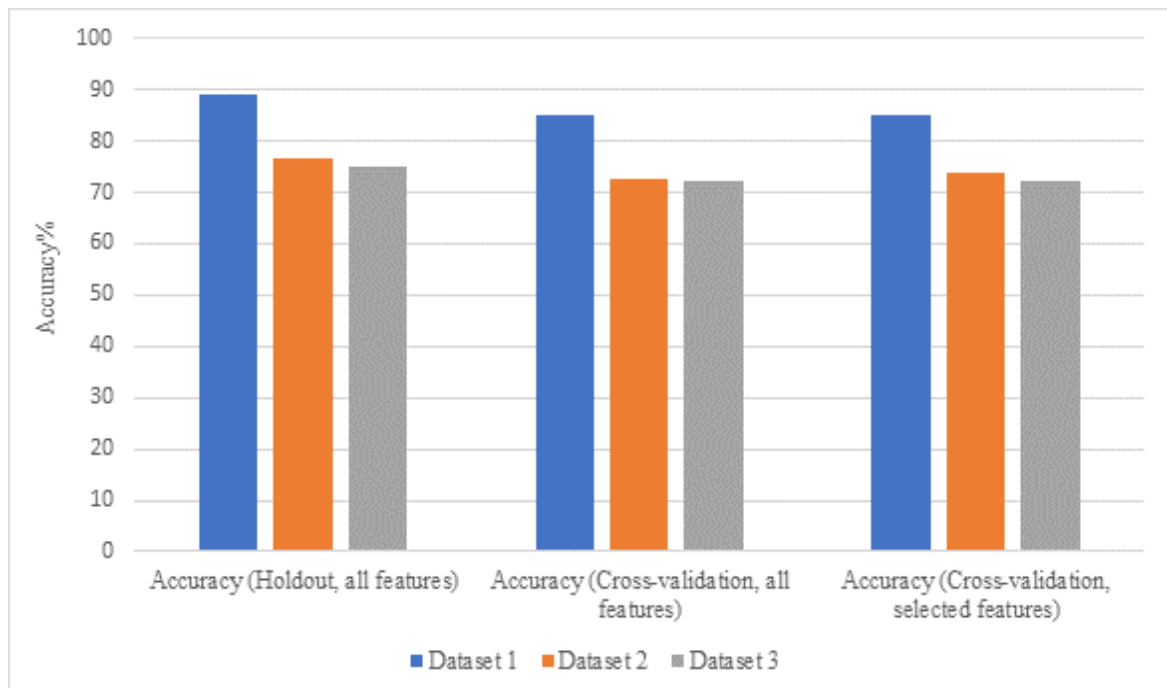


Figure 2: Comparative results for three datasets

Table 12: Comparative results of this work with the literature

Source	Approach	Accuracy
Shouman et al. (2011) [7]	Nine voting equal frequency discretization gain ratio decision tree	84.1%
Marateb et al. (2015) [8]	Multiple logistic regression + Neuro-fuzzy classifier	84%
Latha et al. (2019) [10]	Majority vote with naïve Bayes, Bayes net, random forest and multilayer perceptron	85.48%
This work (for dataset 1)	Majority vote with logistic regression, SVM and naïve Bayes	88.89%

5. CONCLUSION

The goal of this paper is to predict heart disease using different datasets. In this paper, seven different machine learning algorithms are applied to three individual datasets. Different parameters associated with each particular machine learning algorithms are altered and applied to find out which possible case predicts the heart disease accurately. Experimental results show that the most important attribute of

dataset 1 is 'thalach' which means the maximum heart rate achieved by a patient, while that of datasets 2 and 3 are patient age. For dataset 1 using holdout method, the best accuracy of 88.89% is obtained when majority voting classifier is used with logistic regression, SVM and naïve Bayes. This means majority voting algorithm can classify normal patients as normal and abnormal patients as abnormal with an accuracy of 88.89%. For the case of cross-validation method, the best result of 84.83% accuracy for dataset 1 is

obtained by the same classifier. For dataset 2 using holdout method, the best accuracy of 76.60% is obtained when majority voting classifier is used with logistic regression, SVM and naïve Bayes. For the case of cross-validation method, the best result of 72.72% accuracy for dataset 2 is obtained by logistic regression. For dataset 3 using holdout method, the best accuracy of 74.80% is obtained when majority voting classifier is used with logistic regression, SVM and naïve Bayes. For the case of cross-validation method, the best result of 72.22% accuracy for dataset 3 is obtained by SVM. Furthermore, it has been shown here that for dataset 1 and dataset 2, the accuracy values slightly improve when selected features rather than all the features are used for classification of heart disease. On the other hand, for dataset 3, there is no improvement in accuracy values with feature selection. Results also show that among all the datasets and classifiers used in this work, the best accuracy is obtained by majority voting when used as a combination of logistic regression, SVM and naïve Bayes and when used on dataset 1. At this particular case, the majority voting gives an accuracy of 88.89% which is better than other accuracy values reported in the literature using the same dataset. Furthermore, majority voting classifies the heart samples of dataset 1 with a recall, a precision, a F1-score and an AUC value of 86%, 90%, 88% and 86.61%, respectively.

It is observed from this research that the results of holdout method vary with the variation in testing size, and the variation occurs even when the testing size is the same. This is because the samples are random and so they are different in each processing. Cross-validation method gives more stable result compared to holdout method. It is also observed that based on the datasets and the attributes, a classifier has to be chosen for reliably diagnose heart diseases.

6. REFERENCES

- [1] Go, A. S., Mozaffarian, D., Roger, V. L., Benjamin, E. J., Berry, J. D., Blaha, M.J., “Executive summary: heart disease and stroke statistics-2014 update: a report from the American heart association”, *Circulation*, vol. 129, no. 3, pp. 399-410, Jan. 2014. doi: 10.1161/01.cir.0000442015.53336.12.
- [2] Hearst, M. A., Dumais, S. T., Osuna, E., Platt, J. and Scholkopf, B., “Support vector machines”, in *IEEE Intelligent systems and their applications*, vol. 13, no. 4, pp. 18-28, July-Aug. 1998.
- [3] Wang, G., “A survey on training algorithms for support vector machine classifiers”, 2008 international conference on networked computing and advanced information management, pp. 123-128, Gyeongju, 2008.
- [4] Laaksonen, J. and Oja, E., “Classification with learning k-nearest neighbors”, *Proceedings of International Conference on Neural Networks (ICNN'96)*, Washington, DC, USA, 1996, pp. 1480-1483 vol.3.
- [5] Sanz, J.A., Galar, M., Jurio, A., Brugos, A., Pagola, M. and Bustince, H., “Medical diagnosis of cardiovascular diseases using an interval-valued fuzzy rule-based classification system”, *Applied Soft Computing*, vol. 20, pp. 103-111, July 2014. doi: 10.1016/j.asoc.2013.11.009.
- [6] Setiawan, N.A., “Fuzzy decision support system for coronary artery disease diagnosis based on rough set theory”, *International Journal of Rough Sets and Data Analysis*, vol. 1, no. 1, pp. 65-80, Jan. 2014. doi: 10.4018/ijrsda.2014010105.
- [7] Shouman, M., Turner, T. and Stocker, R., “Using decision tree for diagnosing heart disease patients”, *Proceedings of the Ninth Australian Data Mining Conference*, Australia, 2011, pp. 23-30.
- [8] Marateb, H.R. and Goudarzi, S., “A noninvasive method for coronary artery disease diagnosis using a clinically interpretable fuzzy-rule based system”, *Journal of Research in Medical Sciences*, vol. 20, no. 3, pp. 214-223, March 2015.
- [9] Goni, M. Osman, “Development of a web based expert system for diagnosis of heart disease using fuzzy logic”, M. Engg. Project, Institute of Information and Communication Technology, BUET, 2019.
- [10] Latha, C.B.C and Jeeva, S.C., “Improving the accuracy of prediction of heart disease risk based on ensemble classification techniques”, *Informatics in Medicine Unlocked*, vol. 16, 2019, 100203. doi: 10.1016/j.imu.2019.100203.
- [11] Raihan-Al-Masud M, Mondal MRH. Data-driven diagnosis of spinal abnormalities using feature selection and machine learning algorithms. *PLOS ONE*. 2020; 15(2): e0228422.
- [12] Heart disease dataset, UCI machine learning repository, <https://archive.ics.uci.edu/ml/datasets/Heart+Disease> [Last accessed on 31 Mar. 2020].
- [13] Cardiovascular disease, <https://www.kaggle.com/yassinehamdaoui1/cardiovascular-disease> [Last accessed on 31 Mar. 2020].
- [14] Cardiovascular disease dataset, <https://www.kaggle.com/sulianova/cardiovascular-disease-dataset> [Last accessed on 31 Mar. 2020].
- [15] Anaconda distribution website, <https://www.anaconda.com/distribution/> [Last accessed on 12 Feb. 2020].
- [16] Saeys Y, Inza I, and Larranaga p. 2007. A review of feature selection techniques in bioinformatics. *Bioinformatics* 23, 19 (2007), 2507–2517.
- [17] He J, Hu HJ, Harrison R, Tai PC, & Pan Y (2006). Transmembrane segments prediction and understanding using support vector machine and decision tree. *Expert Systems with Applications*, 30(1), 64–72.
- [18] Witten IH, & Frank E (2005). *Data mining: Practical machine learning tools and techniques*.
- [19] Keerthi SS, & Lin CJ (2003). Asymptotic behaviors of support vector machines with Gaussian kernel. *Neural Computation*, 15(7), 1667–1689.
- [20] Lin HT, & Lin CJ (2003). A study on sigmoid kernels for SVM and the training of non-PSD kernels by SMO-type methods. Taipei: Department of Computer Science and Information Engineering, National Taiwan University.
- [21] Chen M, Hao Y, Hwang K, Wang L, and Wang L, *Disease Prediction by Machine Learning Over Big Data From Healthcare Communities*, IEEE Access.
- [22] Bharati S., Podder P., Mondal M. R. H., and Robel M. R. A., “Threats and Countermeasures of Cyber Security in Direct and Remote Vehicle Communication Systems”, *Journal of Information Assurance and Security*, MIR Labs, USA, vol. 15 (2020), pp. 153-164, May 2020.

- [23] Bharati S., Podder P., and Mondal M. R. H., Diagnosis of Polycystic Ovary Syndrome Using Machine Learning Algorithms. Presented at 2020 IEEE Region 10 Symposium (TENSYP), 5-7 June 2020, Bangladesh.
- [24] Mondal M. R. H., Bharati S., Podder P., Podder P., "Data Analytics for Novel Coronavirus Disease", Informatics in Medicine Unlocked, Elsevier, Early version available in June 2020.
- [25] Khanam F., Nowrin I., and Mondal M. R. H., "Data Visualization and Analyzation of COVID-19", Journal of Scientific Research and Reports, vol. 26, no. 3, pp. 42-52, Apr. 2020.
- [26] Bharati, S., Podder, P., "Adaptive PAPR Reduction Scheme for OFDM Using SLM with the Fusion of Proposed Clipping and Filtering Technique in Order to Diminish PAPR and Signal Distortion". Wireless Personal Communication (2020). <https://doi.org/10.1007/s11277-020-07323-0>.
- [27] Mondal, M. R. H., and Armstrong, J., "Analysis of the effect of vignetting on MIMO optical wireless systems using spatial OFDM", Journal of Lightwave Technology, IEEE & OSA, vol. 32, no. 5, pp. 922-929, March 2014.
- [28] Sarker, N., Islam, M. A., and Mondal, M. R. H., "Two Novel Multiband Centimetre-Wave Patch Antennas for a Novel OFDM Based RFID System", Journal of Communications (JCM), ISSN: 1796-2021, vol. 13, no. 6, Jun. 2018.