Use of Efficient Machine Learning Techniques in the Identification of Patients with Heart Diseases

Pronab ghosh
Daffodil International University,
Dhaka, Bangladesh
pronab1712@gmail.com

Sami Azam*,
sami.azam@cdu.edu.au
Charles Darwin University, NT,
Australia
sami.azam@cdu.edu.au

Asif Karim
Charles Darwin University, NT,
Australia
asif.karim@cdu.edu.au

Mirjam Jonkman
Charles Darwin University, NT,
Australia
mirjam.jonkman@cdu.edu.au

MD. Zahid hasan
Daffodil International University,
Dhaka, Bangladesh
zahid15-4729@diu.edu.bd

ABSTRACT

Cardiovascular disease has become one of the world's major causes of death. Accurate and timely diagnosis is of crucial importance. We constructed an intelligent diagnostic framework for prediction of heart disease, using the Cleveland Heart disease dataset. We have used three machine learning approaches, Decision Tree (DT), K-Nearest Neighbor (KNN), and Random Forest (RF) in combination with different sets of features. We have applied the three techniques to the full set of features, to a set of ten features selected by "Pearson's Correlation" technique and to a set of six features selected by the Relief algorithm. Results were evaluated based on accuracy, precision, sensitivity, and several other indices. The best results were obtained with the combination of the RF classifier and the features selected by Relief achieving an accuracy of 98.36%. This could even further be improved by employing a 5-fold Cross Validation (CV) approach, resulting in an accuracy of 99.337%.

CCS CONCEPTS •Applied computing •Life and medical sciences •Health informatics

KEYWORDS

Relief, Pearson Correlations, K-Nearest Neighbor, Decision Tree and Random Forest

ACM Reference Format:
Pronab ghosh, Sami Azam*, sami.azam@cdu.edu.au, Asif Karim, Mirjam Jonkman, and MD. Zahid hasan. 2021. Use of Efficient Machine Learning Techniques in the Identification of Patients with Heart Diseases. In *2021 the 5th International Conference on Information System and Data Mining (ICISDM 2021), May 27–29, 2021, Silicon Valley, CA, USA*. ACM, New York, NY, USA, 7 pages. https://doi.org/10.1145/3471287.3471297

1 INTRODUCTION

Cardiovascular disease is a leading cause of death, accounting for more than 17.3 million [1] deaths per year globally. Machine learning tools could be useful in detecting heart disease, in particular coronary artery disease, which is the main cause of death but does not always have clear and distinctive symptoms. Accurate detection, distinguishing people with heart disease from healthy people, is important. The aim of this research is to use machine learning techniques in order to identify patients with heart disease. There are several challenges associated with this approach. Datasets often have missing values for certain variables or features which makes the application of some techniques difficult. Another difficulty is the selection of features to be used. The number and type of features used can affect both the accuracy of the algorithm and its computational complexity.

In this research we compare three classifiers: DT, KNN and RF, using the Cleveland Dataset. We combine this with feature selection techniques and compare three different sets of features: the full set of relevant features from the dataset, and two sets of features selected by different feature selection algorithms. The novelty of our work is as follows:

- The missing values issue is resolved using the KNN imputation method.
- We apply classifiers to three different sets of features: all relevant features, a set of six features selected by the Relief algorithm and a set of ten featured selected by the Pearson's Correlation technique.
- The outcomes of the three different classifiers are evaluated based on their performance measure indices and statistical results.
- A five-fold CV approach is applied to the selected features of Relief.
- The computational complexity of each algorithm was also investigated.

The remaining parts of the paper are structured as follows: in Section II, related work is briefly discussed. In Section III, various machine learning models are explained. In Section IV, the model is discussed, including missing data handling techniques and Feature Selection algorithms. A short explanation of different performance

ICISDM 2021, May 27–29, 2021, Silicon Valley, CA, USA
© 2021 Association for Computing Machinery.
ACM ISBN 978-1-4503-8954-9/21/05…$15.00
https://doi.org/10.1145/3471287.3471297

indices is also given. In section V, the results are discussed. Conclusions are described in section VI.

2 BACKGROUND

Heart disease analysis using machine-learning-based systems has been described in different research studies. Detrano et al. [2] proposed a logistic regression classifier for heart disease classification based on decision support system and acquired a classification accuracy of 77%. Kahramanli and Allahverdi [3] designed a classification system using a hybrid technique integrating an artificial neural network with a fuzzy neural network. They achieved a classification accuracy of 87%. Palaniappan and Awang [4] proposed an expert heart disease diagnostic system and applied machine-learning techniques such as Naïve Bayes, DT, and ANN. The Naive Bayes predictive model obtained an accuracy of 86%. The second-best prescient model was ANN which obtained an accuracy of 88%. The DT classifier accomplished 80% accuracy. Olaniyi and Oyedotun [5] proposed a three-phase model based on ANN to diagnose heart disease in angina patients and achieved a classification accuracy of 88.9%. Moreover, the proposed system could be easily deployed in healthcare information systems. Das et al. [6] proposed an ANN ensemble-based predictive model that diagnoses heart disease and used statistical analysis system enterprise miner 5.2, achieving an accuracy of 89%, a sensitivity of 80.9%, and a specificity of 95.9%. Jabbar et al. [7] designed a diagnostic system for heart disease using a multilayer perceptron ANN-driven back propagation learning algorithm together with a FS algorithm. The proposed system had a good performance in terms of accuracy. An integrated decision support medical system based on ANN and Fuzzy AHP has also been described [8], achieving a classification accuracy of 91%. An accuracy of 78.9% was reported in another study where the authors Gennari made use of CLASSIT conceptual clustering system [9]. This study was performed on a heart disease dataset that was created and maintained by the University of California Irvine UCI [10]. Machine Learning have also been used for several other studies [11, 12], and could be used for other automated health processes [13].

Comparing different machine learning algorithms in combination is a promising approach. For this research, three machine learning algorithms are proposed that are implemented on data of cardiovascular disease patients. The KNN algorithm is used to handle the missing data issue. Feature Selection is done following the Relief and Pearson Correlations approaches. The machine learning algorithms, DT, RF, and KNN are implemented after Feature Selection. A 5 – fold CV approach was also evaluated.

3 VARIOUS MACHINE LEARNING CLASSIFIERS

3.1 Decision Tree

Decision Tree, which utilized only 5 *numClasses*, is amongst the most useful algorithms in the field of machine learning. It employs an upside-down tree-based recursive progression process to resolve both classification and regression issues [11]. It chooses a root node tor starting the process followed by the splitting technique. The evaluation process of DT is calculated based on two measurements. 'Entropy' and 'Information Gain' which are explained in equations 1) and (2). [14]

$$\begin{aligned} \mathrm{E\,(D)} &= -\,\mathrm{P\,(positive)\,log2P\,(positive)} \\ &\quad -\,-\,\mathrm{P\,(negative)\,log2P\,(negative)} \end{aligned} \tag{1}$$

Entropy E takes both positive and negative attributes of D dataset.

$$\begin{aligned} \mathrm{Gain\,(AttributeX)} &= \mathrm{Entropy\,(DecisionAttributeY)} \\ &\quad -\mathrm{Entropy\,(X,Y)} \end{aligned} \tag{2}$$

For each of the features or attributes, the gain is measured and the one with the highest value is chosen because it provides the most knowledge. The whole cycle is repeated for tree sub-branches to eventually complete the DT. ID3 (Iterative Dichotomiser) is one of most popular variants of Decision Tree these days.

All types of Decision Tree are constructed based on the following steps:

1) Identify the most suitable attribute, $\boldsymbol{X}$, through the application of Attribute Selection Measures (ASM).

2) Mark $\boldsymbol{X}$ a decision node and divide the dataset into smaller segments or subsets.

3) Initiate the tree building process by repeating the previous two steps recursively for each child until one of the conditions below matches:

3.1) All the tuples belong to the same attribute value.

3.2) No other remaining attributes can be found.

3.3) No other instances can be found.

3.2 K-Nearest Neighbor

K-Nearest Neighbor (n_neighbors = 5) is a commonly used classification technique [13] which tends to balance several important parameters such as predictive performance, intuitive interoperability, and calculation time. While algorithms such as RF can have higher predictive capability, they are lagging behind in other parameters. Unsurprisingly, the adoption of KNN by industry is very widespread. KNN uses 'Euclidian Distance', see equation 3) [15], to evaluate the distance between two data points (X^n and X^m).

$$Dist\,\left(X^n - X^m\right) = \sqrt{\sum\nolimits_{i=1}^{D}\left(X_i^n - X_i^m\right)^2} \tag{3}$$

The basic workflow of the algorithms is stated below:

1. Initialize K to a corresponding number of neighbours

2 For each instance within the data:

2.1 Calculate the distance between the current instance in the dataset and the query instance.

2.2 Add the instance index and the calculated distance to an ordered set of collection, let us say $\boldsymbol{S}$.

3. Sort $\boldsymbol{S}$ in an ascending order (from smallest to largest)

4. Select the first K entries from this sorted collection

5. Retrieve the labels of the chosen K entries

6. Return the Mode of the K labels (Classification problems)

3.3 Random Forest

Random Forest (RF), is one of the most common supervised classification and regression techniques [15]. It works by building a forest from a multitude of random and unconnected DTs during the training phase. Ensemble strategies use multiple learning algorithms to create efficient descriptive analytics which can perform

Table 1: Value range of datasets

No	Attributes	Description	Value Range	No	Description	Value Range
1	Age	Age in years	29 to 79	8	Maximum heart rate achieved	71 to 202
2	Sex	Gender instance	0 and 1	9	Exercise induced angina	0, 1
3	Cp	Chest pain type	1 to 4	10	ST depression induced by exercise relative to rest	1 to 3
4	Trestbps	Resting blood pressure in mm Hg	94 to 200	11	Slope of the peak exercise ST segment	1,2, 3
5	Chol	Serum cholesterol in mg/dl	126 to 564	12	Number of major vessels colored by fluoroscopy	0 to 3
6	Fbs	Fasting blood sugar > 120 mg/dl	0 and 1	13	Defect types	3,6,7
7	Restecg	Resting ECG results	0 to 2	14	Diagnosis of heart disease	0 to 4

better than any of the individual models in that system [16]. RF can incur additional complexity in its calculation because it uses more features than a standalone DT, but it usually achieves higher precision when dealing with unknown datasets. Generally, RF follows the following steps:

1. From the training set, select k number of random subsets.
2. Carry out the training of k Decision Trees. 1 random subset is utilized for training exactly 1 decision tree.
3. Each of the individual trees independently provides a prediction on records of the test set.
4. Provide the final prediction for each of the candidates with the test set. Note that Random Forest utilizes the class with majority vote for deciding the candidate's final prediction.

3.4 Methodology

Some elements of the research have been briefly explained to provide an overview. The selection of the dataset is one of the most crucial issues in machine learning. The data set used for this research is described below.

3.5 Dataset

The dataset used for this research is part of a popular machine learning data repository named UCI [17]. The dataset has a total of 303 samples where data characteristics are multivariate, and we have used 14 of the attributes. Table 1 gives a description of the variables used and the range of their values. As can be seen, the 'num' attribute can have values from 0 to 4 where 0 means no heart disease and 1 to 4 means heart disease of different levels of severity. For this research, we have combined all cases (1 to 4) of heart disease and made the 'num' value equal to 1, using Label encoding [18].

3.6 The outcomes from different feature selection algorithms

Feature Selection (FS) is the method in which we choose the features to be used. Feature Selection is used for three reasons: it can increase the accuracy of the classification algorithm, it reduces time and computational complexity and it helps to prevent overfitting and underfitting issues associated with machine learning.

3.7 Relief feature selection algorithm

Relief [19] is an algorithm that follows a filter-based approach to select the features. Relief is also capable of determining feature dependencies. The highest scoring features are chosen. The algorithm works out a score for each feature. This scoring depends on the difference of feature values between closest neighbor pairs. Relief is mainly used for binary classification problems. In Table 2 score values for the different features selected by Relief have been compared. Of these features, *Cp* has the highest value of 0.28, and Thalach the lowest, 0.12.

3.8 Pearson's Correlation feature selection algorithm

Pearson Correlation [19] is called "product moment correlation coefficient" also known as just "correlation". Normally it is used to measure the association between the class feature and the continuous features. It has a value between - 1 and +1 and shows how closely two variables are related to each other linearly. Pearson Correlation is only useful for quantitative variables. We compared the Pearson Correlation values of the different features (see Table 3). Trestbps has the lowest value of 0.048, and Thal has the highest value of 0.157.

3.9 Descriptive analysis of the proposed model

The dataset used for this research is the Cleveland dataset containing 303 data with 13 input attributes. However, there are some missing values. This was resolved using the KNN imputation [20] technique. As the different attributes have very different ranges of values, a standard scaler [21] system has been used. Feature selection techniques are then applied, resulting in three sets of features: all features, features selected by Pearson's correlation, and features selected by the Relief FS. After that, datasets were separated into training and testing data. Three machine learning algorithms are then applied, DT, RF, and KNN. A 5 − fold CV approach was also evaluated. Figure 1 gives an overview of the model.

3.10 Performance measure indices

To test the efficiency of different classifiers, various performance assessment metrics are utilized [19]. The outcomes of the classifiers can be divided into:

Table 2: Features selected by Relief algorithms and their rankings.

Feature name	Feature code	Score
Type chest pain	Cp	0.287
Maximum heart rate	Thalach	0.120
Exercise induced angina	Exang	0.250
Slope of the peak exercise ST segment	Slope	0.168
Number of major vessels (0–3) colored by fluoroscopy	Ca	0.138
Thallium scan	Thal	0.267

Table 3: Features selected by Pearson's correlation algorithms and their rankings

Feature name	Feature code	Score
Age	Age	0.052
Chest Pain Type	Cp	0.137
Trestbps	Trestbps	0.048
Chol	Chol	0.049
Resting Electrocardiographic Results	Restecg	0.138
Thalach	Thalach	0.054
Oldpeak	Oldpeak	0.071
Slope	Slope	0.068
CA	Ca	0.115
Thal	Thal	0.157

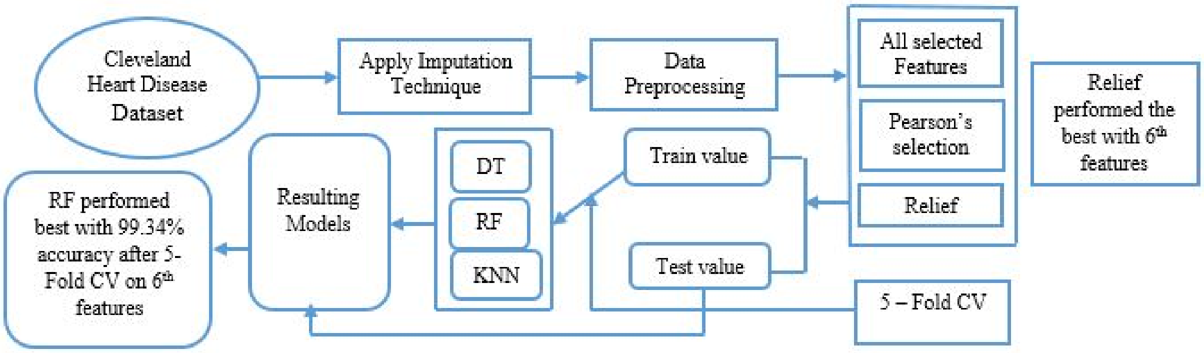
Cleveland Heart Disease Dataset
Apply Imputation Technique
Data Preprocessing
All selected Features
Pearson's selection
Relief
Relief performed the best with 6th features
Train value
Test value
5 – Fold CV
DT
RF
KNN
Resulting Models
RF performed best with 99.34% accuracy after 5-Fold CV on 6th features

Figure 1: The graphical representation of our proposed model.

TP = True Positive (Correctly Identified); FP = False Positive (Incorrectly Identified).

TN = True Negative (Correctly Rejected); FN = False Negative (Incorrectly Rejected).

Using the above, Accuracy, Sensitivity, Precision, Error rate, Specificity (SPE), Negative Predictive Value (NPV), False Positive Rate (FPR), False Discovery Rate (FDR), False Negative Rate (FNR) and Mean Squared Error (MSE) were calculated.

4 RESULTS AND DISCUSSION

4.1 Results of all features

Table 4 shows the results of the three machine learning algorithms when all 13 features are used. For DT we have observed a 0.93 SPE score and a 0.81 NPV score. RF algorithm is better than the other machine learning algorithms in terms of accuracy and precision. RF provides an SPE and NPV of 0.90, which is higher than the other classifiers we have used.

4.2 Results of features selected by Relief

Table 5 displays the outcomes of the machine learning algorithms using the six features that are ranked as important features by the Relief feature selection method.

DT demonstrated an SPE score of 0.87 and an NPV of 0.81. KNN achieves a 1.0 SPE score and a 0.81 NPV score. The RF technique resulted in a SPE value of 1.0 with an NPV of 0.9.

Table 4: Generated outcomes on 13 input features

Dimensions	Decision Tree	K-Nearest Neighbor	Random Forest
SPE	0.9285	0.8666	0.9062
NPV	0.8125	0.8125	0.9062
FPR	0.0714	0.1333	0.0937
FDR	0.0689	0.1379	0.1034
FNR	0.1818	0.1935	0.1034
MSE	0.1311	0.1639	0.0983

Table 5: Outcomes of six features selected by Relief

Dimensions	Decision Tree	K-Nearest Neighbor	Random Forest
SPE	0.8787	1.0	1.0
NPV	0.8125	0.8125	0.9062
FPR	0.0714	0.1333	0.0937
FDR	0.0689	0.1379	0.1034
FNR	0.1818	0.1935	0.1034
MSE	0.1311	0.1639	0.0983

Table 6: Displayed outcomes on ten different features

Dimensions	Decision Tree	K-Nearest Neighbor	Random Forest
SPE	0.9062	0.8787	0.967
NPV	0.9062	0.9062	0.9375
FPR	0.0937	0.1212	0.032
FDR	0.1034	0.1379	0.034
FNR	0.1034	0.1071	0.06
MSE	0.0983	0.1147	0.0480

4.3 Results of features selected by Pearson's Correlation

Table 6 shows the outcomes of the machine learning algorithms when ten features are selected using the Pearson correlation algorithm. By applying DT, we achieved an SPE and NPV of approximately 0.91. KNN resulted in an 0.88 score for SPE and a 0.90 score for NPV; whereas RF obtained a SPE of 0.97 and an NPV of 0.94, which is higher than the other classifiers.

4.4 Comparison results of different feature selection techniques

Figure 2 shows the accuracy, sensitivity, and precision of the three different machine learning algorithms when applied to the different sets of features. It can be seen that the best results (approximately 98.36% accuracy) are obtained with the six features selected by Relief and the RF classifier, while the KNN achieves an accuracy of 96% on the same features. For the Pearson's correlation technique, the RF algorithm is also the best classifier with 95% accuracy. The DT and KNN techniques obtained 90% and 88% accuracy. On the other hand, all of the classifiers show moderate performance when applied to the full set of features. The best results obtained with the six features selected by the Relief and the KNN and RF algorithms, achieving a sensitivity of 94%, however, these DT only achieved 89% accuracy for these features.

When using the 10 features of the Pearson Correlation, the DT classifier achieved better results than the other two classifiers (91% vs 89% and 89%). For all features, RF has the best sensitivity (89%), while the sensitivity

of DT and KNN are 83% and 81% respectively. RF is the classifier with the best precision for all sets of features with the best results achieved for the six features selected by Relief (97%) followed by the Pearson's correlation features (96%). For all features, the precision of RF is still 90%. The DT classifier outperforms KNN in terms of precision for all features (87% vs 84%) and for the 10 Pearson features (90% vs 89%) but not for the six features selected by the Relief FS algorithm (KNN 91%, DT 89%). Figure 3 displays the Error Rates of the three machine learning algorithms. It can be seen that the lowest error rates are achieved by the RF algorithm and the six Relief Features (1.6%).

For these features KNN has an error rate of 3.9% and DT of 11%. For the Pearson features RF also has a low error rate (4.9%) but KNN has an error rate of 11.59%. The error rate is DT is only slightly better (9.8%). The highest error rates occur when applying KNN

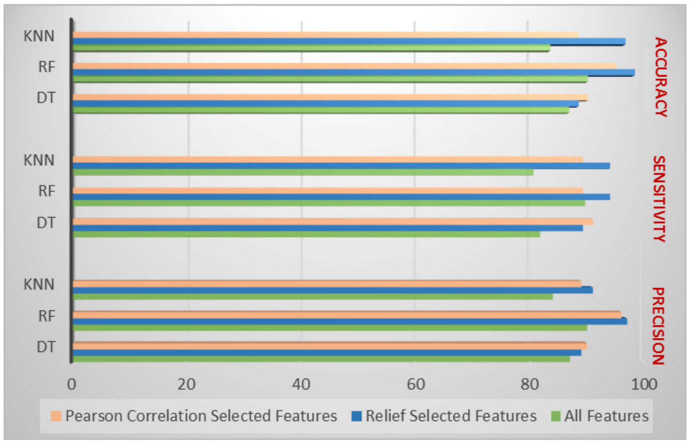

Figure 2: Best features selection processes through accuracy, sensitivity and precision.

Table 7: Displayed outcomes of 5- fold CV on the 6th Relief features

Algorithms	Validation Accuracy	Validation Standard Deviation	Validation Run Time
DT	97. 17%	1.67%	0.007447
RF	99.34%	1.85%	0.013764
KNN	90.92%	1.61%	0.017895

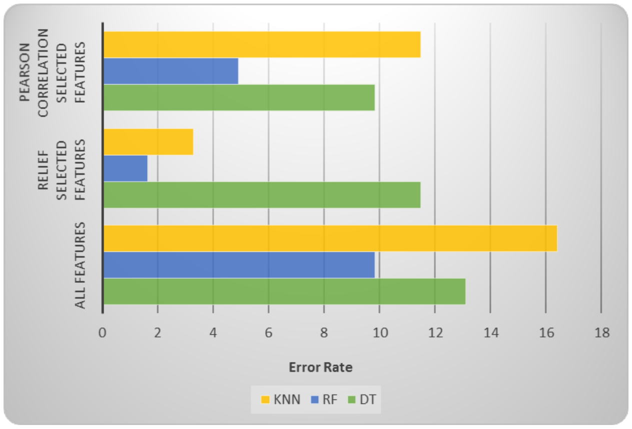

Figure 3: Error Rate.

to all features (16%) while for all features the DT classifier has an error rate of 13%. RF is again best but for all features, but the error rate is still 9%.

4.5 The recorded results of 5-fold cross validation technique on the 6th chosen features of Relief including slope

To show the flexibility of our proposed method, a 5-fold CV approach is examined using the features chosen by Relief. The outputs of the selected algorithms are shown in Table 7. The best result was recorded for the RF algorithm which had a validation accuracy of approximately 99. 34%. The worst result was obtained for the KNN classifier. The lowest execution time was noticed for the DT approach while the KNN method took more time (about 0.017895) than the others to generate the displayed outcome. Interestingly, the standard deviation of RF (1.85%) was larger than for the other methods.

4.6 Computational complexity of the introduced algorithms on 6th features

Table 8 shows the Computational Complexity for the three selected classifiers. Two types of computational complexities have been included: Training complexity and Prediction complexity, denoting n as the number of training samples, p as the number of features, n_{trees} as the number of trees (for methods based on various trees), and k as the number of neighbors.

5 CONCLUSIONS

Three machine learning algorithms have been applied to different feature-sets to establish the best combination for the identification of heart disease. The best result, is achieved with the RF classifier in combination with the six features selected by the Relief. It is noted that the results of the classifiers were dependent on the selection of the features and that the preprocessing strategy is also important. In the future we intend to apply these techniques to larger datasets so that options for fine tuning the model can be further investigated.

REFERENCES

[1] "WHO: global cause of death due to heart disease," Available: who.int/cardiovascular_diseases/resources/atlas/en/, Accessed: 10/7/20.
[2] R. Detrano *et al.*, "International application of a new probability algorithm for the diagnosis of coronary artery disease," Am. J. Cardiol., vol. 64, no. 5, pp. 304–310, 1989.

Table 8: Displayed outcomes of computational complexity

Models	Training	Training Complexity Calculation	Prediction	Predicted Complexity Calculation
DT	$O(n^2p)$	O(459045)	$O(p)$	O(5)
RF	$O(n^2pn_{trees})$	O(4590450)	$O(pn_{trees})$	O(50)
KNN	$O(knp)$	O(7575)	$O(np)$	O(1515)

[3] H. Kahramanli and N. Allahverdi, “Design of a hybrid system for the diabetes and heart diseases,” Expert Systems with Applications, vol. 35, no. 1-2, pp. 82–89, 2008.

[4] S. Palaniappan and R. Awang, “Intelligent heart disease prediction system using data mining techniques,” in Proceedings of IEEE/ACS International Conference on Computer Systems and Applications (AICCSA 2008), pp. 108–115, Doha, Qatar, March-April 2008.

[5] E. O. Olaniyi and O. K. Oyedotun, “Heart diseases diagnosis using neural networks arbitration,” International Journal of Intelligent Systems and Applications, vol. 7, no. 12, pp. 75–82, 2015.

[6] R. Das, I. Turkoglu, and A. Sengur, “Effective diagnosis of heart disease through neural networks ensembles,” Expert Systems with Applications, vol. 36, no. 4, pp. 7675–7680, 2009.

[7] M. A. Jabbar, B. L. Deekshatulu, and P. Chandra, “Classification of heart disease using artificial neural network and feature subset selection,” Global Journal of Computer Science and Technology Neural & Artificial Intelligence, vol. 13, no. 11, 2013.

[8] O. W. Samuel, G. M. Asogbon, A. K. Sangaiah, P. Fang, and G. Li, “An integrated decision support system based on ANN and Fuzzy_AHP for heart failure risk prediction,” Expert Systems with Applications, vol. 68, pp. 163–172, 2017.

[9] J. H. Gennari, P. Langley, and D. Fisher, “Models of incremental concept formation,” Artif. Intell., vol. 40, no. 1–3, pp. 11–61, 1989.

[10] K. Bache and M. Lichman, “UCI Machine Learning Repository,” University of California Irvine, vol. 2008, no. 14/8. 2013.

[11] A. R. Beeravolu, S. Azam, M. Jonkman, B. Shanmugam, K. Kannoorpatti and A. Anwar, "Preprocessing of Breast Cancer Images to Create Datasets for Deep-CNN," IEEE Access, vol. 9. 2021.

[12] P. Ghosh, S. Azam, M. Jonkman, A. Karim *et al.*, "Efficient Prediction of Cardiovascular Disease Using Machine Learning Algorithms with Relief and LASSO Feature Selection Techniques," in IEEE Access, vol. 9, 2021. DOI: 10.1109/ACCESS.2021.3053759.

[13] M. M. Alam, S. Saha, P. Saha, F. N. Nur, N. N. Moon, A. Karim, and S. Azam, “D-CARE: A Non-invasive Glucose Measuring Technique for Monitoring Diabetes Patients”, Proceedings of International Joint Conference on Computational Intelligence. Algorithms for Intelligent Systems. Springer, vol. 1, pp. 443-453, 2019. DOI: 10.1007/978-981-13-7564-4_38.

[14] Hegelich, “Decision trees and random forests: Machine learning techniques to classify rare events, ”Eur. Policy Anal. 2(1), 2016.

[15] L. Y. Hu, M.-W. Huang, S.-W. Ke, and C.-F. Tsai, “The distance function effect on K-Nearest Neighbor classification for medical datasets, ”Springer-Plus, vol. 5, no. 1, p. 1304, Aug. 2016

[16] P. Ghosh, A. Karim, S. T. Atik, S. A., M. Saifuzzaman," Expert Model of Cancer Disease Using Supervised Algorithms with a LASSO Feature Selection Approach," International Journal of Electrical and Computer Engineering, Vol. 11 (3), 2020.

[17] “UCI machine learning repository: Chronic kidney disease data set,” archive.ics.uci.edu/ml/datasets/heart+disease, Accessed on 10/7/20.

[18] “Smarter ways to encode categorical data for machine learning,” https://towardsdatascience.com/smarter-ways-to-encode-categoricaldata-for-machine-learning-part-1-of-3-6dca2f71b159, (Accessed on 10/07/2020).

[19] “A Quick Introduction on Pearson Correlations,” Available at. www.spss-tutorials.com/ , Accessed on 10/07/2020).

[20] S. Oehmcke, O. Zielinski and O. Kramer, “KNN ensembles with penalized DTW for multivariate time series imputation”, International Joint Conference on Neural Networks (IJCNN), 2016.

[21] “Standard Scaler Approach of Machine Learning,” Available: https://scikitlearn.org/stable/modules/generated/sklearn.preprocessing.StandardScaler.html, (Accessed on 25/07/2020).