A Taxonomy on Machine Learning Based Techniques to Identify the Heart Disease

Anand Kumar Srivastava[1,2(✉)], Pradeep Kumar Singh[2(✉)], and Yugal Kumar[2(✉)]

[1] Department of Computer Science and Engineering, ABES Engineering College, Ghaziabad, UP, India
anand8355@gmail.com

[2] Department of Computer Science and Engineering, Jaypee University of Information Technology, Waknaghat, Solan, HP, India
pradeep_84cs@yahoo.com, yugalkumar.14@gmail.com

Abstract. Every year average death of human being is 17.7 million caused by Heart Disease or Cardiovascular diseases (CVDs), which is 31% of all global deaths reflected in Survey of World Heart Day 2017 [33]. In September 2016, many countries have taken the various Global Hearts Initiative in prediction and diagnosis of heart Diseases at earlier stages so that it can be cure perfectly [5]. Many authors have studied in this filed to optimize the performance of various ML techniques using various approaches. In latest studies, many groups have uncovered that many optimization algorithm like Differential Evolution, Genetic Variants, and Particle Swarm optimization are associated with prediction algorithm like K-Nearest neighbor, Decision Tree, Neural Network, Support Vector machine, Logistic Regression etc. to make efficient medical system for CVDs. The Objective of our current study is to analyze the comparatively study of the ML techniques in terms of performance measure of different ML techniques that have been used by various authors in their research work in earlier studies of heart disease prediction and diagnosis process.

Keywords: Heart disease · Data mining · Machine learning · Heart failure · Logistic regression (LR) · K- means (KM) · Decision Tree (DT) · Support vector machine (SVM) · Multivariate adaptive regression splines (MARS) · Rough sets (RS) · K-Nearest neighbor (KNN) · Genetic algorithm (GA) · Neural network (NN) · Differential Evolution (DE) · Dictionary-based keyword spotting (DBS) · Natural language Process (NLP) · Random forest (RF) · Classification and regression tree (CART) · Naïve Bayes (NB) · Classification by clustering (CC) · Multilayer perceptron (MLP) · Radial basis function (RBF) · Single conjunctive rule learner (SCRL) · Bagged Tree (BT) · AdaBoost (AB) · Particle swarm optimization (PSO)

1 Introduction

In human body heart is very crucial part, through blood vessels it pumps the blood to complete body parts, so other body parts get the oxygen and essential nutrients from bloods as well as heart assists in the removal of metabolic wastes.

M. Prateek et al. (Eds.): NGCT 2018, CCIS 922, pp. 13–25, 2019.
https://doi.org/10.1007/978-981-15-1718-1_2

So it's very important that heart works properly. According to WHO on the basis of one survey as data available in 2015 [32], maximum human deaths (appox 17.7 million) are caused by the heart diseases and there are some factors that put people at increase risk for heart disease like Smoking, high cholesterol, high blood pressure, genetic effect, diabetes, obesity, etc. and few symptom that indicate the heart disease Chest pain (angina), Shortness of breath, Sweating, Irregular heartbeat, Nausea, heart burn, heart attacks.

Machine learning techniques can be used in heart disease prediction at earlier stages so that it can be cure within a desirable time period. In last 15 years many studied carried out in this field. Objective of each studied is to optimize the implemented algorithm with some suitable proposal.

1.1 Data Mining Techniques

Data Mining techniques are the process to convert raw data into useful information or we can say that it is the process to find some useful hidden patterns of data to predict some outcomes. These are the important DM techniques:

- Decision Tree
- Clustering
- Association
- Classification
- Prediction

In data mining techniques Association method works on the rule if-else procedure. It helps to discover the patterns between two items or data. Association rules are useful for analyzing and predicting the different diseases.

Classification data mining technique is the example of supervised learning technique. Classification algorithm predicts categorical class level. Prediction algorithm is also example of supervised learning technique but it predicts continuous valued function. For example we can suggest a classification model that predicts that heart is working correctly or not, that means answer will be yes or not on the basis of symptoms of patients. And in same case prediction algorithm predict the probability of chances of heart failure.

Clustering is the example of unsupervised learning techniques. It is the process of creating the group or classes of similar behavioral objects or data.

In clustering process on the basis of similar characteristics of objects or data and using any automatic techniques we makes meaningful cluster or classes.

J. Ross Quinlan in 1980 developed decision tree technique also known as ID3 later he presented C4.5. Decision tree algorithm concept is based on greedy approach and follows the tree structure.

1.2 Machine Leaning Algorithm

ML uses the DM techniques and other learning techniques to build model, so that we can predict the future outcomes. If we compare the DM and ML, Dm is the process to extract the hidden information or patterns from a particular dataset whereas ML deals

the algorithm improves automatically through experience based on data. Basically ML uses the learning techniques, on the basis of leaning algorithm ML can be categorize into 3 category Supervised Learning, unsupervised learning and reinforcement learning.

1.2.1 Supervised Learning

It can be explained that it uses training or labeled data to learn the mapping function F() from the (Z) input variable with (Y) output variable.

$$Y = F(Z)$$

Again supervised leaning problem can be divided into 2 category classification and regression.

Examples of Supervised learning are Logistic Regression, KNN, CART, blog–Linear Regression, Naïve Bayes. Ensembling is one type of supervised learning, In which we use the multiple combinations of weal ML models to predict future outcomes on new sample. Some examples are Bagging with random forest, Boosting with XGBoost. In [4], Author used the different ensembling methods in prediction of of heart disease and computed the accuracy of each techniques.

1.2.2 Unsupervised Learning

It only processes the input variable (Z) with no corresponding output variables and uses the unlabeled data to build mode for future prediction. Unsupervised learning problems can be divided into 3 types Association, clustering and Dimensionality Reduction. Dimensionality Reduction can be done using Feature selection methods and Extraction methods. In the process of Feature Extraction actually data transform from multi dimensional space to small or low dimensional space. K- means, Apriori, and PCA are examples of unsupervised learning.

2 Related Work

Since the last decade, many studies have been implemented and evaluated the performance measures in process of heart disease prediction and diagnosis. Different Authors used different algorithm with different feature set in prediction and diagnosis of heart diseases especially in heart failure.

Vivekandan et *al.* implemented a scheme to used the modified Differential Evolution algorithm for optimize feature selection [1]. Author has identified 13 features for the heart disease prediction, diagnosis and dataset taken from UCI repositories [2], further reduced these attributes to nine attributes for improving the accuracy. In this study author adapt the Fuzzy AHP algorithm to obtain the relative weight of each feature. Further author implemented feed forward NN algorithm for prediction and diagnosis of heart disease.

In [3], the author adapt the SVM classifier along with forward feature selection, back elimination feature and forward feature inclusion and implemented SPECTF dataset from UCI. Author concluded that, current approach obtained smaller subsets

and increases the performance accuracy of diagnosis compared to back elimination techniques and forward inclusion.

Shao et *al.* [4] proposed as feature selection algorithm rough set, multivariate adaptive regression splines and logistic regression to reduce the set of critical features for heart disease prediction and diagnosis with higher accuracy. In this study author analyzed hybrid multivariate adaptive regression splines artificial neural network algorithm is the best alternative because this method contained the less number of explanatory variables (features) and provided the efficient classification accuracy on prediction. On the other hand Oluwarotimi et *al.* [6] suggested hybrid algorithm based on Fuzzy-AHP and ANN mythology for Heart Failure and risk prediction using medical dataset [23]. This prediction process consists of two sequential stages. At first stage author implemented Fuzzy-AHP to choose rank and obtained the local weights of Heart Failure features and for set of given attributes author computed global weight, Author's implanted algorithm and declared that an average accuracy is 91.10%, which is more better than conventional Artificial neural network method. (4.40% higher in comparison)

Jabbar et *al.* implemented a classification algorithm which is the combination of genetic and KNN algorithm to predict and diagnose the heart disease of a patient for Andhra Pradesh population [7]. Author used 7 data sets computed the accuracy with and without GA along with KNN and concluded.

In [19], Performance Analysis of Ensemble classifier Random Forest, AdaBoost, Bagged Tree with PSO on parameter like accuracy, sensitivity, specificity, PPV, NPV are performed. Author used the Heart Disease Dataset which contains 270 instances and 14 attributes and concluded that Particle Swarm Optimization (PSO) is computationally efficient (inexpensive) in context to speed and memory.

Pouriyeh et *al.* [16] compared MLP, Decision Tree, NB, K-NN, SVM, SCRL and RBF classifiers in two experiments. In first experiment author evaluated the accuracy with the help of 10 fold cross validation method and further to estimate the efficiency performance of each algorithm. Author used K = 1, 3, 9 and 15 In case of KNN classifier and obtained the performance of KNN is best at K = 9 using whole data set [17]. In second experiments author use the bagging, boosting and stacking concept for evaluation of algorithm efficiency.

In this study Author found that, no improvement in support vector machine performance but, maintained its performance (accuracy) level in case of Bagging. SCRL improved 10.56%, it increased from 69.96% to 80.52% and Decision tree improved 0.99% (from 77.55%–78.54%). In case of Boosting SCRL improved 11.22% and Decision Tree improved 4.62% (from 77.55% to 82.17%). The others algorithm performance remained same as earlier. Where the combination of SVM and MLP has the best accuracy 84.15% in case of stacking.

In [15], Soni et *al.* analyzed the performance efficiency of Naïve Bayes, Decision Tree, ANN, classification via clustering Algorithm in heart disease prediction and diagnosis. In this study, research is divided into two phases, the first part author used the 25 set of Attributes and implemented the algorithm, in second phase author used the Genetic Algorithm for optimizing the feature set 25 to 15 while applying the classifying algorithm and found that efficiency is optimized as compare with first case.

In optimization of feature set from data set, Yan et *al.* used the genetic algorithm in selection of 24 critical features among 40 features available from data set [14]. With these features implemented system provides better diagnostic performance.

Anbarasi et *al.* compared the performance measure in the context of accuracy in percentage of Decision Tree, classification via clustering, Naïve bayes, in prediction of heart disease [12]. Author used the Genetic Algorithm to select the critical attributes which participate more closely in the heart risk, HF prediction and diagnosis. This result reduces the required no of tests taken by patient indirectly. 13 attributes (features) are compressed to six attributes using genetic search with dataset Cleveland Heart Disease database [31]. Further apply the classifiers and investigated the performance.

Guidi et *al.* implemented Heart Failure clinical decision support system along with a portable kit to acquire the set of clinical attributes [11]. This system provided various functioning with the using of different ML techniques. Author selected 12 attributes from database, data taken in period of 2001 to 2008 from Hospital of Maria Nuova, Florence, Italy.

Author compared different types of ML techniques and declared that the CART method provide the best result among all ML. Also give the humanly understandable decision making process, Classification and regression tree provides the accuracy of 87.6% in type prediction and 81.8% in severity assessment

Olaniyi et *al.* compared the efficiency performance efficiency of SVM and Back propagation Neural Network in heart disease prediction and diagnosis [9]. In this research implementation works are divided into two phases, the first part author used back propagation training and testing of the network while in second part prediction applied using support vector machine.

In [26], Author implemented a automatic system that automate the feature design for numerical sequence classification using Genetic programming, system is called Autofeed which is fully data driven.

Fen et al. designed comprehensive risk model using a modified method of random survival forest technique for predicting and diagnose the heart failure mortality with a high level of performance accuracy [20]. Author worked on MIMIC II database With 32 variables and achieved 82.1% performance.

In 2017 [21], Jin et *al.* designed robus and effective paradigm for heart failure prediction at earlier stage using short-term memory network (LSTM) model and author compared with methods such as Adaboost, Rnadom Forest and LR. Author showed that these methods performs best in the prediction and diagnose of heart failure.

Pu et *al.* proposed the method, with this we predict and prevent all kinds of diseases including cardiovascular diseases on the basis of Genes detection [22]. Because in human body Genes generally stable for long time after birth. Author finds that gene detection plays an important role in prediction of heart diseases.

Hui et *al.* [8] worked with data set [25] and proposed a system which extract the information and identify the risk factor for heart disease from clinical text using NLP techniques. Author analyzed the characteristics of clinical evidence and categorized into 3 main types, sentence level clinical measurements, token level clinical entities and sentence level clinical fact.

Nahar et *al*. Analyzed the promising techniques for feature selection for prediction for heart disease author proved that motivated feature selection process (MFS) and

MFS along with computerized feature selection process (CFS) are the best suited feature selection techniques [28]. Author used the SVM and Naïve Base classifier prediction and found that SVM predict more efficient then Naïve Base.

3 Research Methodology

This section discusses the different approaches or techniques used by various author in their studies. Broadly on the basis of literature work we conclude that in heart disease prediction and diagnosis process every author used the feature extraction phase using any feature selection algorithm, prediction and diagnosis phase using any machine learning algorithm. Each author worked on any particular medical dataset [2, 10, 18, 24, 27], which contain various instances and attributes.

3.1 Feature Extraction Phase

Prediction of any diseases is completely based on the features selections. Feature selection is the process to define few important and critical features from available feature dataset. We also know the 80–20 rule means 80% effects comes from the 20% effect [29]. So it's very important to selection of critical variables or features in the process of prediction and diagnosis. In the progress of diseases prediction expert system many authors used different optimization algorithm in feature selection process in last decades.

Optimization algorithm provides the optimal feature set because minimal feature set provide efficient result and this set consist of no irrelevant attributes. In today's scenario there are many optimization techniques available, all techniques work using heuristic based and gradient based search techniques in stochastic and deterministic contexts correspondingly. To make large acceptability of the optimization approach, medical expert systems are desired to develop using robust and efficient optimization algorithms. There are few example of optimization algorithm used in heart diseases prediction and diagnosis mainly are particle swarm optimization (PSO), Evolutionary algorithms simulated annealing and ant colony optimization.

3.1.1 Differential Evolution Optimization

Differential Evolution (DE) algorithm is one type of heuristic approach; it has mainly 3 advantages [30]. First, whatever initial parameter values is it can find true global minimum value. Second, is its speed, robustness and third, is less no of control parameters. DE developed in 1997 by Price and Storm. It is the branch of evolutionary programming.

The Coverage speed is the main issues in Evolutionary Algorithm, there are many studies have carried out to increase the speed with DE. It is mutation scheme that make self adaptive and self selection process. It is the stochastic optimization method that minimizes an objective function.

DE Algorithm is population based same as genetic algorithm using same operators like crossover, mutation and selection. But DE mainly relies on mutation in search operation.

3.1.2 Genetic Optimization

Genetic Algorithm (GA) is also population based same as the DE algorithm. It deals with constrained and unconstrained optimization problem that is completely based on natural selection [30]. GA iteratively modify the population of individual solutions. While creating the next generation from current population GA use the following three Rules

(a) Selection Rules
(b) Crossover Rules
(c) Mutation Rules

The best point in the population approaches an optimal solution and GA follows the population approaches.

The fundamental difference between GA and De is GA based on crossover whereas DE or Evolutionary Algorithm based on mutation for search mechanism.

Generally Genetic Algorithm needs the fitness function for evaluation of solution domain and genetic representation for solution domain.

3.1.3 Particle Swarm Optimization

PSO is also population based optimization technique. PSO developed in 1995 by Dr. Kennedy and Dr. Eberhart. Both Scientist were inspired with fish schooling or bird flocking. PSO has multiple advantages over GA like PSO is easy to implement and needs few parameter to adjust.

Algorithm start with set of random particle i.e. solution and then searches for best particle by updating generations.

PSO and GA both algorithm have many similarities like PSO and GA start with group of random generated population and search for best by using random technique and both have memory. Only difference is that PSO does not have operators like crossover and mutation. Particle updates them with internal velocity.

PSO have been applied in many research and application area in past several years. The best part with PSO is it does require few parameters to adjust, which makes PSO is more attractive for researchers. If we compare with other optimization algorithm following 2 points make stronger to PSO

(a) Fast (speed)
(b) Less costly (cheaper)

3.2 Prediction and Diagnosis Phase

Various ML Techniques involved in prediction and diagnosis of CVDs in different studies with different dataset. Each ML techniques have own strength and limitations.

Concept of **Neural network** is inspired from human brain concept. NN is based on a collection of connected nodes called artificial neurons. Nodes are called processing element. NN learn from itself from examples. There can be 3 layer input layer, hidden layer (not in all cases), output layer. In some simple cases only input and output layer can provide the result.

K- Means clustering is one type of unsupervised learning technique [34]. It used for unlabeled data. It finds the different group on basis of similarities of objects or data. K shows the no of groups. Algorithm iteratively works and rearranges the attributes in matched groups at each time.

Logistic Regression is used when the dependent variable (target) is categorical.

Example, Suppose we want to predict email is spam or not means yes (1) or No (0), Whether the HF is malignant (1) or not (0).

Multivariate Adaptive Regression Splines Jerome H. Friedman introduced MARS in 1991 in a form of regression analysis. It is also called non parametric regression technique. It helps to solve regression type problems because with the help of MARS we can predict the values of continuous dependent from a set of independent variable.

Rough set theory Zdzislaw Pawlak proposed RS theory, with this theory we can approximate the lower and upper boundary of the original set. **K-Nearest neighbor** is also non parametric technique used in both regression and classification. KNN classify the input data into defined groups on the basis of prior data that is called training data.

Support Vector Machine is works with labeled data means it is the example of supervised learning technique. SVM can perform efficiently with linear classification as well as non linear classification. Vapnik and Chervonenkis invented SVM in 1963. SVM performance is not good with large data set, when data set contain more noise.

Classification and regression tree is one type of Decision tree algorithm, in which we use classification tree and regression tree technique.

Decision Tree is the powerful tool used in ML. it is the graph based model classification tree and regression tree technique. It is the graph based model that represents possible consequences including all event outcomes. Tree consist of nodes always start with root node; mainly three types of node can be possible end node, chance node and decision nodes. Sometime in decision tree predictor values are continuous or infinite possible outcomes is called regression tree. Many authors used several of ensemble methods with DT to increase the performance of DT mainly are Boosted tree, Bagging and Random forest classifier.

Bagging used in regression and statistical classification, It avoid the over fitting means reduces variance.

Boosting helps in, to builds a series of trees is a sequential form in regression problem.

Random Forest is just like Bagging only difference is that bagging takes 1 parameter that is no of tree with all possible features whereas RF takes 2 parameter one is no of tree and second is how many features participate to search to find optimal or best feature.

Naïve Bayes is a one type of probabilistic classifier based on Bayes theorem. It is conditional probability model. In many situations in supervised learning with some type of probability model, NB classifier can train very optimally and efficiently.

Classification by Clustering is the technique by which we can find homogeneous subgroup and it is the type unsupervised learning technique. When we identify clusters or group inside the classified group is called classification by clustering. CC model gives best result at compare with individual classification or clustering technique.

4 Analysis

In order to further explore the relation between the heart disease prediction and techniques applied for prediction, efforts have been and same is mapped with the help of Table 1. It is clearly shown from Table 1 that Neural Network is the technique which is explored in maximum number of studies.

Table 1. Machine learning techniques used in different studies for heart disease identification

Features / References	DE	LR	MARS	GA	RS	SVM	DT	KNN	NN	KM	Fuzzy	DBS + NLP	CART	RF	NB	CC	MLP	SCRL	RBF	BT	AB	PSO
[1]	√								√		√											
[3]						√																
[4]		√	√		√				√													
[6]									√		√											
[7]				√				√														
[8]												√										
[9]						√			√													
[11]				√		√			√		√		√	√								
[13]							√								√	√						
[14]				√																		
[15]							√	√	√						√	√						
[16]							√	√							√	√	√	√	√			
[19]														√						√	√	√
[20]														√								

Table 2. Performance of algorithm along with set of attributes used for heart disease prediction and diagnosis in different studies

Reference No	Data set name	No of features used in experiment	Algorithm name	Algorithm accuracy (%)
[1]	UCI Repository	13	NN	83.0
		09		85.0
[3]	SPECTF	44	SVM	75.0
		19		78.0
	heart disease dataset	10		81.0
		4		85.0
[4]	UCI datasets	13	MARS + NN	82.14
[6]	Cleveland Heart Disease Dataset	13	Fuzzy + NN	91.10

(*continued*)

Table 2. (*continued*)

Reference No	Data set name	No of features used in experiment	Algorithm name	Algorithm accuracy (%)
[7]	Heart disease A.P	12	KNN + GA	100.00
[8]	i2b2 corpus	06	DBS + NLP	91.5
[9]	UCI ML datasets	13	NN + SVM	87.5
[11]	Maria Nuova Hospital, Florence, Italy, (2001–2008)	12	NN	77.8
			SVM	80.3
			Fuzzy + GA	69.9
			CART	81.8
			RF	83.3
[13]	Cleveland Heart Disease database	6	NB	96.5
			DT	99.2
			CC	88.3
[14]	352 diagnosed heart disease dataset	24	GA	92.3
[15]	CAD data set	15	NB + GA	96.5
			DT + GA	99.2
			CC + GA	88.3
[16]	Cleveland Data Set	76	SVM + MLP	84.15
[19]	Heart Disease Dataset	14	BT + PSO	100%
			RF + PSO	90.37
			AB + PSO	88.89
[20]	MIMIC II database	32	RF	82.1

In the current study we compared the accuracy measurement of different classifier KM, DT, LR, MARS, RS, KNN, GA, SVM, NN, DE, DBS, NLP, CART, RF, NB, CC, MLP, SCRL, RBF, BT, AB, PSO and combination of these. Table 1 show the various ML techniques used in various studies. Table 2 shows the accuracy of ML techniques in percentage along with associated no of features.

On the basis of Table 2 we found, accuracy of algorithm is depends upon the number of features selected for heart disease analysis.

When studies used the GA or PSO optimization techniques for feature selection, classifier worked efficiently.

Figure 1 Represents the various algorithms along with the number of attributes used in several studies.

Table 3. Performance comparison of ML algorithm used in different studies

Sr. no	Algorithm name	No of attributes	Accuracy
01	NN	9	85.0
		12	77.8
		13	83.0
02	MARS + NN	13	82.14
03	Fuzzy + NN	13	91.10
04	SVM + NN	13	87.5
05	SVM	4	85.0
		10	81.0
		12	80.3
		19	78.0
		44	75.0
06	SVM + MLP	76	84.15
07	GA	24	92.3
08	NB	06	96.5
09	DT	06	99.2
10	CC	06	88.3
11	NB + GA	15	96.5
12	DT + GA	15	99.2
13	CC + GA	15	88.3
14	KNN + GA	12	100.0
15	RF	12	83.3
		32	82.1
16	BT + PSO	14	100%
17	RF + PSO	14	90.37
18	AB + PSO	14	88.89
19	CART	12	81.8
20	DBS + NLP	06	91.5

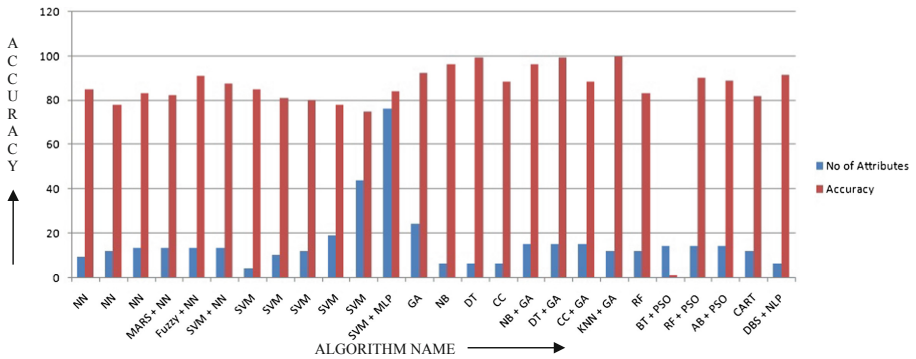

Fig. 1. Algorithm performance graph

5 Conclusions

Everyone knows if Heart is not functioning properly then how it can affects our whole body and. so Detection of heart diseases at early state is essential to prevent from other diseases affected from heart and save lives. In this study we are showing the comparatively analysis of various ML algorithm performances in the prediction and diagnosis of heart disease. Table 3 shows the comparatively study of all ML Techniques and on the basis of this table we analyze following conclusion.

First Conclusion, NB, DT, CC along with GA has shown better performance as NB, DT and CC performed without GA.

Second conclusion, BT, RF and AB along with PSO performed better as compared with performance of BT, RF, and AB without PSO.

Third Conclusion, Accuracy of NN along with Fuzzy or SVM is optimal as compared to accuracy of NN algorithm alone.

Summary of this study (Fig. 1) that classifier algorithm gives optimal result along with any feature optimization algorithm like GA or PSO.

References

1. Vivekanandan, T., Iyengar, N.C.S.N.: Optimal feature selection using a modified differential evolution algorithm and its effectiveness for prediction of heart disease. Comput. Biol. Med. **90**, 125–136 (2017). ScienceDirect
2. Heart Disease Data set from UCI Repositories. http://archive.ics.uci.edu/ml/datasets/HeartþDisease
3. Shilaskar, S., Ghatol, A.: Feature selection for medical diagnosis: evaluation for cardiovascular diseases. Exp. Syst. Appl. Sci. Direct **40**(10), 4146–4153 (2014)
4. Shao, Y.E., Hou, C.D., Chiu, C.C.: Hybrid intelligent modeling schemes for heart disease classification. Appl. Soft Comput. **14**(Part A), 47–52 (2014)
5. World Health Organization. http://www.who.int/mediacentre/factsheets/fs310.pdf. Accessed 03 Feb 2017
6. Samuel, O.W., Asogbon, G.M., Sangaiah, A.K., Fang, P., Li, G.: An integrated decision support system based on ANN and Fuzzy AHP for heart failure risk prediction. Int. J. Exp. Syst. Appl. **68**, 163–172 (2017)
7. Jabbar, M.A., Deekshatulu, B.L, Chandra, P.: Classification of heart disease using K- nearest neighbor and genetic algorithm. In: International Conference on Computational Intelligence: Modeling Techniques and Applications, vol. 10, pp. 85 – 94 (2013)
8. Hui, Y., Jonathan, M.G.: A hybrid model for automatic identification of risk factors for heart disease. Int. J. Biomed. Inform. **58**, S171–S182 (2015)
9. Olaniyi, E.O., Oyedotun, O.K., Khashman, A.: Heart diseases diagnosis using neural networks arbitration. Int. J. Intell. Syst. Appl. **12**, 75–82 (2015)
10. Heart Disease Dataset from UCI Repositories. http://archive.ics.uci.edu/ml/datasets/HeartþDisease
11. Guidi, G., Pettenati, M.C., Melillo, P., Iadanza, E.: A machine learning system to improve heart failure patient assistance. IEEE J. Biomed. Health Inform. **18**, 1750–1756 (2014)
12. Anbarasi, M., Anupriya, E., Iyengar, NChSN: Enhanced Prediction of Heart Disease with Feature Subset Selection using Genetic Algorithm. Int. J. Eng. Sci. Technol. **2**(10), 5370–5376 (2010)

13. Hongmei, Y., Jun, Z., Yingtao, J., Chenglin, P., Shouzhong, Xo: Selecting critical clinical features for heart diseases diagnosis with a real-coded genetic algorithm. Appl. Soft Comput. **8**, 1105–1111 (2008). ScienceDirect
14. Soni, J., Ansari, U., Sharma, D., Soni, S.: Predictive data mining for medical diagnosis: an overview of heart disease prediction. Int. J. Comput. Appl. **17**(8), 0975–8887 (2011)
15. Pouriyeh, S., Vahid, S., Sannino, G., De Pietro, G., Arabnia, H., Gutierrez, J.: A Comprehensive investigation and comparison of machine learning techniques in the domain of heart disease. In: 22nd IEEE Symposium on Computers and Communication (ISCC 2017), Workshops - ICTS4eHealth (2017)
16. Nahar, J., Imam, T., Tickle, K.S., Chen, Y.-P.P.: Computational intelligence for heart disease diagnosis: a medical knowledge driven approach. Exp. Syst. Appl. **40**, 96–104 (2013)
17. http://archieve.ics.uci.edu/ml/datasets
18. Yekkala, I, Dixit, S., jabbar, M.A.: Prediction of heart disease using ensemble learning and particle swarm optimization. In: IEEE International Conference on Smart Technologies For Smart Nation (2017). Accession Number 17766551
19. Miao, F., Cai, Y.P., Zhang, Y.X., Fan, X.M., Li, Y.: Predictive modeling of hospital mortality for patients with heart failure by using an improved random survival forest. IEEE Access **6**, 7244–7253 (2018). ISSN 2169-3536
20. Jin, B., Che, C., Liu, Z., Zhang, S., Yin, X., Wei, X.: Predicting the risk of heart failure with EHR sequential data modeling. IEEE Access **6**, 9256–9261 (2018). ISSN 2169-3536
21. Pu, L.N., Zhao, Z., Zhang, Y.T.: Investigation on cardiovascular risk prediction using genetic information. IEEE Trans. Inf. Technol. Biomed. **16**(5), 795–808 (2012)
22. Lafta, R., et al.: Coupling a fast fourier transformation with a machine learning ensemble model to support recommendations for heart disease patients in a tele health environment. IEEE Access. **5**, 10674–10685 (2017). Accession Number 16950418
23. http://archive.ics.uci.edu/ml/datasets/Heart+Disease
24. Pecchia, L., Melillo, P., Bracale, M.: Remote health monitoring of heart failure with data mining via CART method on HRV features. IEEE Trans. Biomed. Eng. **58**(3), 800–804 (2011)
25. Harvey, D.Y., Todd, M.D.: Automated feature design for numeric sequence classification by genetic programming. IEEE Trans. Evol. Comput. **19**(4), 474–489 (2014)
26. http://dx.doi.org/10.1016/j.jbi.2015.09.006
27. Nahar, J., Imam, T., Tickle, K.S., Chen, Y.P.P.: Computational intelligence for heart disease diagnosis: a medical knowledge driven approach. Exp. Syst. Appl. **40**(1), 96–104 (2013). ScienceDirect
28. Bunkley, N.: Joseph Juran, 103, pioneer in Quality Control, Dies. N. Y. Times (2008)
29. Water Resources Research Report, Optimization Using Differential Evolution, Report No: 060 Date: July 2008. [https://ir.lib.uwo.ca/cgi/viewcontent.cgi?article=1022&context=wrrr]
30. Blake, C.L., Mertz, C.J.: UCI Machine Learning Databases (2004). http://mlearn.ics.uci.edu/databases/heartdisease/
31. Beaglehole, R., Bonita, R.: Global public health: a scorecard. Lancet **372**(9654), 1988–1996 (2008)
32. http://www.who.int/cardiovascular_diseases/world-heart-day-2017/en/
33. https://www.datascience.com/blog/k-means-clustering
34. https://www.researchgate.net/publication/315800218_Rough_Sets_in_Machine_Learning_A _Review [accessed Jul 27 2018]