Journal of Engineering and Applied Sciences 14 (7): 2249-2257, 2019
ISSN: 1816-949X


# Heart Diseases Prediction Using Accumulated Rank Features Selection Technique

Mohammad-Ashraf Ottom and Walaa Alshorman
Department of Computer Information System,
Faculty of Information Technology and Computer Science,
Yarmouk Univeristy, Shafiq-Irshidat Street, Irbid, Jordan

**Abstract:** Diagnosis of heart disease is a very critical and challenging task in medical science field. Healthcare providers are collecting a massive amount data including heart disease data which unfortunately contains relevant, irrelevant and redundant features. Selecting optimal features set is a critical and important process for such high dimensional data analysis. Data mining and features selection techniques can assist in decision making and for better diagnostics of many diseases, however, traditional features selection techniques provide a limited contribution to classification. This research proposes a new approach for features selection. The new approach suggests collaboration between well-known features selection techniques by accumulating features rank for all selected features selection techniques, then, elects features which produce the highest rank. The optimal features are used to form a subset of dataset and features with lower rank will be eliminated. The full dataset and new dataset will be evaluated on five well known machine learning algorithms to evaluate the result of the proposed technique according to accuracy, recall, precision and f-measure. The proposed approach shown better prediction rate based where kNN recorded the best enhancement rate in accuracy (+3.7%), precision (+8.1%), recall and (+0.3.5%) f-measure (+0.5.7%).

**Key words:** Features selection, classification, data mining, accumulated rank and cleveland heart disease dataset, accumulating, eatures

## INTRODUCTION

Heart disease has appreciably increased over the last decade and become one of the most leading death causation around the world (Long *et al*., 2015). Heart disease rises the demand on public health providers and increase healthcare cost worldwide (Fowkes *et al*., 2013). Due to the presence of many heart disease features, doctors may face some challenges to diagnose heart disease quickly and accurately. Therefore, information technologies advancement could assist doctors in diagnosing such diseases (Kavitha and Kannan, 2016). The literature showed many attempts to utilize soft computing in diagnosing heart disease such as hybrid models. Typically, hybrid models consist of two phases, the first phase is features selection techniques to select the significant features, then feed selected features to data mining techniques and algorithms to produce knowledge and patterns (Shao *et al*., 2014; Ashraf *et al*., 2012).

Heart disease datasets usually contain relevant features, irrelevant data or redundant features. Irrelevant and redundant features may influence the prediction process, thus, features selection techniques stimulate data mining techniques to focus on useful features and ignore impractical features based on the idea of reducing the initial features set from n features to m features where $m<n$ to reduce the risk of over fitting, provide more accurate prediction, improve model generalization ability, and to reduce the computation cost (Long *et al*., 2015; Inbarani *et al*., 2015).

**Motivation:** Electronic medical records usually contain relevant, irrelevant and redundant features which create challenges for data mining tools for predicting and diagnosing diseases such as heart disease. Finding a set of features that best describe the dataset could enhance the mining process and produce a better outcome.

**Objectives:** Given the importance of features selection techniques in diagnosis and predicting diseases, the aim is to propose a new approach for diagnosing heart disease based on data mining techniques and features selections. The approach suggests cooperation between

**Corresponding Author:** Mohammad-Ashraf Ottom, Department of Computer Information System,
Faculty of Information Technology and Computer Science, Yarmouk Univeristy,
Shafiq-Irshidat, Street Irbid, Jordan

well-known features selection techniques to elect subset of features. The proposed approach will use features that best predict the class label based on features rank. A new dataset will be formed based on the elected features. The full-dataset and the new-dataset will be evaluated using five well known machine learning algorithms (Naive-Bayes, support vector machine, k-nearest neighbor, artificial neural network and decision tree) based on classification measures such as accuracy, recall, precision and f-measure.

**Scope:** This research investigates features selection domain which is frequently used data mining technique to find the optimal features set to overcome the problem of high dimensionality, redundant and irreverent features.

The rest of the study is organized as the following, description about some features selection techniques and data mining classifiers, research methodology, experiment and results and finally, conclusion and future works.

**Features selection techniques:** In general, features selection or dimensionality reduction techniques are based on the idea of reducing the initial features set from n features to m features where m<n, this technique leads to a better data visualization, data compression, better classification accuracy, fast and efficient data retrieval. Features extraction finds a new m dimensions that belong to the original dimensions n (Alhaj *et al*., 2016). Features selection retails a subset of m best features from an n original set of features and eliminates the remaining features according to some norms such as information gain, correlation based features selection, relief, gain ratio, symmetrical uncertainty and one-R.

**Information Gain (IG):** Measures the ability of a certain feature f to split the dataset instances into disjoint subset based on entropy. Decision tree classifier finds the best split feature (s) (Chuang *et al*., 2008; Ashraf *et al*., 2011).

**Correlation based feature selection technique:** Rank features that are greatly correlated with the class label but not correlated with each other, irrelevant features with low correlation with the class label should be ignored and redundant features also ignored because they are highly correlated with each other (Zhang *et al*., 2016; Wald *et al*., 2014).

**Relief:** A features selection technique takes a sample instance from the data randomly and then locating its nearest neighbor from class labels. Feature value of the nearest neighbors are compared to the sample instance in order to be used to update the relevance score of every feature, since, the useful features must be differentiate between instances of different classes and in the same time have the same value for instances from the same class (Canedo *et al*., 2014).

**Gain ratio:** A feature selection technique considered as an extension to information gain algorithm. Gain ratio features selection technique solve information gain bias problem by normalizing the contribution of all features (Karegowda *et al*., 2010; Ashraf *et al*., 2012).

**Symmetrical uncertainty:** A normalized form of information gain in which compensates for information gain's bias toward features with more values and normalizes its values to the range [0, 1] (Kumar *et al*., 2014; Ashraf *et al*., 2010).

**One-R:** A simple algorithm builds on one-rule of each attribute in the training data to select the smallest error rule. One-R treats all features with numerical values as continuous and use a straightforward method to split the values range into multiple disjoint intervals. This primitive scheme produce a simple rule based on features only and considered as a minimal form of classifier, One-R can be useful for pinpointing a baseline performance as a benchmark for other learning schemes (Yildirim, 2015) (Fig. 1).

**Machine learning classification techniques:** Machine learning is a branch of artificial intelligence in which systems and models learn from data, predict future unseen cases, make decisions and identify useful patterns.

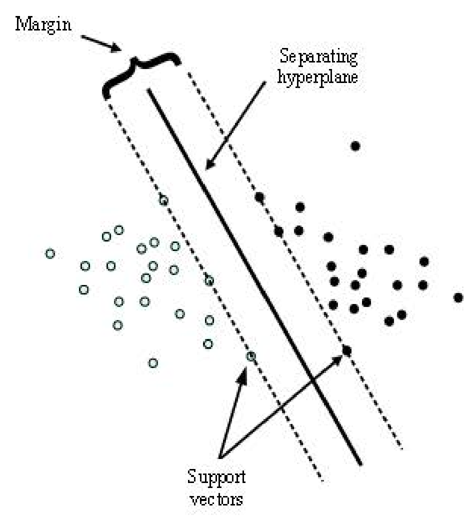


Fig. 1: SVM linear classes sepration (classfication) (Demidova *et al*., 2015)

Machine learning algorithms usually divided in two main types; Supervised and un-supervised learning methods (Dua and Du, 2016). Supervised machine learning is based on the idea of training and tuning algorithms that reason from known class labels instances in order to produce a general hypothesis to enhance the prediction of future cases (Kotsiantis, 2007). Supervised learning algorithms are categorized based on the learning algorithms structure and objective. The objective of supervised learning technique is to categorized data from prior information and obtaining common patterns (Singh *et al*., 2016). Example of machine learning algorithms are Support Vector Machine classifier (SVM), Artificial Neural Network (ANN) and decision tree. On the other hand, unsupervised machine learning algorithm is to capture and exploit the dataset structure to classify data into groups and pattern based on distance measure without the need to know class labels or the target. Unsupervised algorithms designed to summarize the data key features to decide the normal cluster of input patterns given a specific cost function. k-mean clustering, self-organization map and hierarchical clustering are examples of unsupervised learning algorithms (Breve and Pedronette, 2016).

This research uses the following machine learning classification techniques for training and testing the model and features selection methods to find the most performing features that predict the class label.

**Naive-Bayes:** Probabilistic model based on Baye's theorem that count the frequency and values combinations in each dataset to calculate a set of probabilities assuming that all features are independent which mean the presence or absence of a given feature is unrelated to the presence or absence of other features.

Naive-Baye considered an efficient supervised learning algorithm in addition to its robustness for noisy data and its applicability for huge dataset also Gaussian-Naive-Bayes classifiers shown the ability to perform well in many real-world applications such as document classification and spam filtering where the classifier requires a small amount of training data to estimate the necessary parameters (Krishnaiah *et al*., 2014; Ahmad, 2013).

**Support Vector Machine (SVM):** A boundary supervised learning algorithm can determine which object belong to which class with the boundaries of areas and implements a supervised learning. SVM is a mathematical function even being linear or non-linear classifier which means it can distinguish two different types of object fall into classes by transforming original features space into a high-dimensional space then search for separating hyper plane, the maximum margin in that area separating the class on both hyper plane sides. Two parallel hyper planes defining the class boundaries and constructing laying the highest possible distance from each other in linear model (Demidova *et al*., 2015).

**Artificial Neural Network (ANN):** Computing algorithm inspired by the biological neural networks in brains. ANN performs tasks by learning from examples without being previously programmed.

Generally, neural networks consist of artificial neurons connected in a network by some way to store information and to model the input-output of processes in addition to layers and activation functions as shown in Fig. 2. ANN has been used on a variety of tasks including, medical diagnosis, machine translation, speech recognition, computer vision and social network filtering (Pal and Mather, 2005).

**k-Nearest Neighbor (kNN):** A classifier model based on the idea of finding unknown data point category based on its nearest neighbors point which class is known. The value of k is important to determine the number of neighbors to be considered in sample data point class identification process. kNN calculate training points weights base on their distance from sample data point but this algorithm still has some drawback such as the computational complexity and required memory in addition, its sensitivity to noisy and irrelevant features which can clearly decrease its accuracy. Memory requirement drawback can overcomes by reducing the size of data specifically by eliminating the repeated patterns and data points (Krishnaiah *et al*., 2014).

**Decision tree J48:** Decision trees generated by C4.5 can be used for classification. J48 classifier is a C4.5 implementation which is the successor of the ID3 algorithms. ID3 algorithm is based on inductive logic programming language models that building a decision tree based on the training data using an entropy measure to specify the important features of training cases to populate the tree leaves.

J48 sets the training data set dominant attributes as tree root and create a leaf training dataset for each possible value the root can take then repeat this process for each leaf using training data classified by this leaf. J48

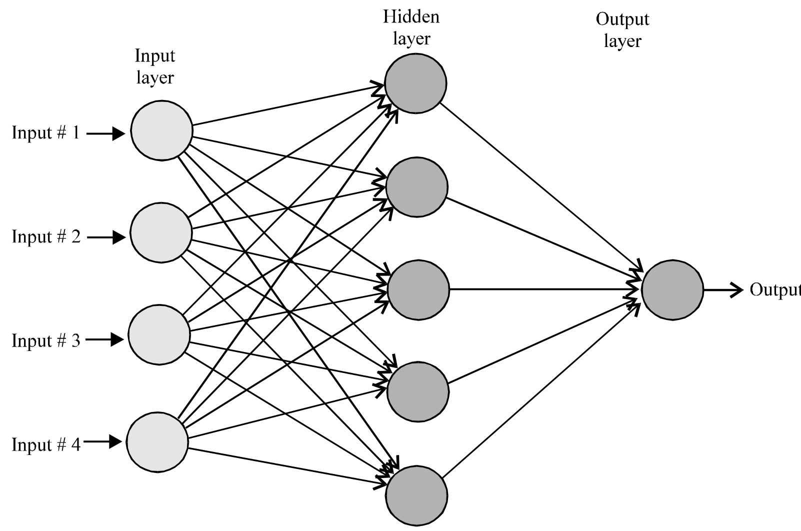


Fig. 2: Artificial networks layers

Table 1: Sample previous work results

| Classifiers | Accuracy | References |
|---|---|---|
| J48 | 0.531 | Fernandez-Delgado *et al.* (2014) |
| SVM | 0.5860 | |
| Naive Bayes | 0.8150 | Cheung (2001) |
| ANN backpropagation | 0.7840 | Kangwanariyakul *et al.* (2010) |
| Bayesian | 0.7840 | |
| ANN probabilistic | 0.7060 | |
| SVM linear | 0.7450 | |
| SVM polynomial | 0.7060 | |
| SVM kernel | 0.7080 | |
| Naive Bayes | 0.7860 | Andreeva (2006) |
| Decision tree | 0.7570 | |
| Neural network | 0.8280 | |
| Decision tree | 0.7760 | Pouriyeh *et al.* (2017) |
| Naive Bayes | 0.8350 | |
| kNN | 0.7620 | |
| ANN | 0.8230 | |
| SVM | 0.8420 | |
| J48 | 0.7820 | Kinge and Gaikwad (2018) |
| Naive | 0.8260 | |
| ANN | 0.7940 | |

ignores the missing values such as values of items that can be predicted using what is known about the feature values for other records (Saravanan *et al.*, 2008). The major function of ID3 algorithm is to identify the most appropriate feature to best disjoint data into multiple classes. ID3 uses entropy and information gain to measure the data item impurity. The smaller entropy means full data membership to one class while the highest value refers to the existence of more than one class. information gain measure the decreasing of weighted average entropy of feature compared with the completed dataset entropy (Dimitoglou *et al.*, 2012). Table 1 shows sample classification accuracy from previous works.

## MATERIALS AND METHODS

The phases of current research are summarized in Fig. 2 and 3:

**Data collection and classification:** The dataset is a publicly available data from UCI machine learning repository website. The dataset consist of 76 attributes and 303 records (instances), only 14 of them are actually used because all published previous work used the 14 features. Few missing values (six values), all missing values were replaced by "?". Table 2 shows dataset description.

We used Weka tool to evaluate the classification accuracy, recall, precision and f-measure for the data set before starting any step in evaluating the efficiency of the proposed approach using Naive Bayes, k-nearest neighbor, support vector machine, artificial neural network and tree J48 classification techniques. The process of splitting the dataset into training and testing data was accomplished by Weka where 2/3 of dataset used for trainings and the rest for testing.

We used Weka tool (Waikato Environment for Knowledge Analysis) to perform the classification accuracy, recall, precision and F-measure for the dataset before applying the proposed approach. Weka is an open source machine learning and data mining tool, it contains an extensive collection of algorithms and data pre-processing methods for data extraction and the experimental comparison of different machine learning techniques (Frank *et al.*, 2004).

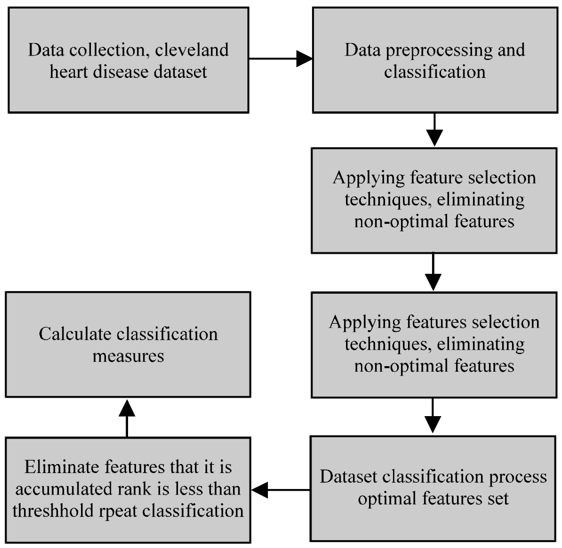


Fig. 3: Resaerch method

Table 2: Cleveland clinic foundation dataset description

| Feature name | Feature description |
|---|---|
| Age | Age in years |
| Sex | Sex (1 = male; 0 = female) |
| cp | Chest pain type |
| trestbps | Resting blood pressure |
| chol | Serum cholesterol in mg/dL |
| fbs | Fasting blood sugar>120 mg/dL |
| restecg | Resting electrocardiographic |
| thalach | Maximum heart rate achieved |
| exang | Exercise induced angina |
| oldpeak | ST depression induced by Exercise relative to rest |
| slope | The slope of the peak exercise ST segment |
| ca | Number of major vessels |
| thal | Normal, fixed or reversable defect |
| num | Diagnosis of heart disease (0 or 1) |

**Applying features selection techniques:** We applied information gain, gain ratio, relief F, symmetrical uncertainty and one-R features selection techniques on heart disease dataset. We recorded the resulted rank for all features using the selected features selection technique to determine which features should be eliminated and to enhance classification accuracy. Using this traditional way, features ranks were near and hardly to determine the less important features.

**Features selection techniques collaboration ranks:** Since, individual features selection techniques ranks produced near results, we suggest combining features ranks for all selected features selection techniques using multiplication, averaging and cumulation. Multiplying features ranks produced by all techniques shown similar outcome to individual features selection techniques. the median and average of features ranks were not appropriate, since, it did not provide a clear indicator about the non-optimal features set. Finally, we calculated the accumulated rank for each feature. Each feature accumulated rank equals the summation of its assigned ranks using features selection techniques. The accumulated rank was success to give as a clear cursor about which features should be removed.

**Eliminating features with accumulating rank less than threshold:** Based on the result of previous phase we adopted a certain indicator to decide which features are optimal and which are not. All features with accumulating rank <1 are considered as non-optimal features and accordingly, they have been removed. Using this procedure four features were eliminated to remain ten features out of fourteen original features. Finally, we calculate the changing rate of classification accuracy, recall, precision and f-measure between the accumulating rank features selection technique and traditional features selection techniques to produce conclusion and results. Equation 1-4 show the calculation procedure of classification accuracy, recall, precision and f-measure, respectively:

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

$$\text{Recall} = \frac{\text{TP}}{\text{TP+FN}} \quad (2)$$

$$\text{Precision} = \frac{\text{TP}}{\text{TP+FP}} \quad (3)$$

$$\text{F-measure} = 2*\frac{\text{Precision}*\text{recall}}{\text{Precision}+\text{recall}} \quad (4)$$

## RESULTS AND DISCUSSION

The experiment applied to publicly available heart disease data set from UCI machine learning repository called Cleveland Heart Disease-CHD. Figure 4 shows density for CHD features. CHD contains fourteen features and 303 instances. We applied Naive Bayes, k-nearest neighbor, support vector machine, artificial neural network and tree J48 classification techniques on the original CHD before applying any features selection technique, classification measures shown in Table 2.

Table 3 shows the total rank for all features using FSTs (correlation based, gain ratio, information gain, One-R, reflief and symmetrical analysis). The proposed approach suggests eliminating features with accumulated rank less than threshold (<1) wherefore, four features have been eliminated (trestbps, chol, fbs and restecg).

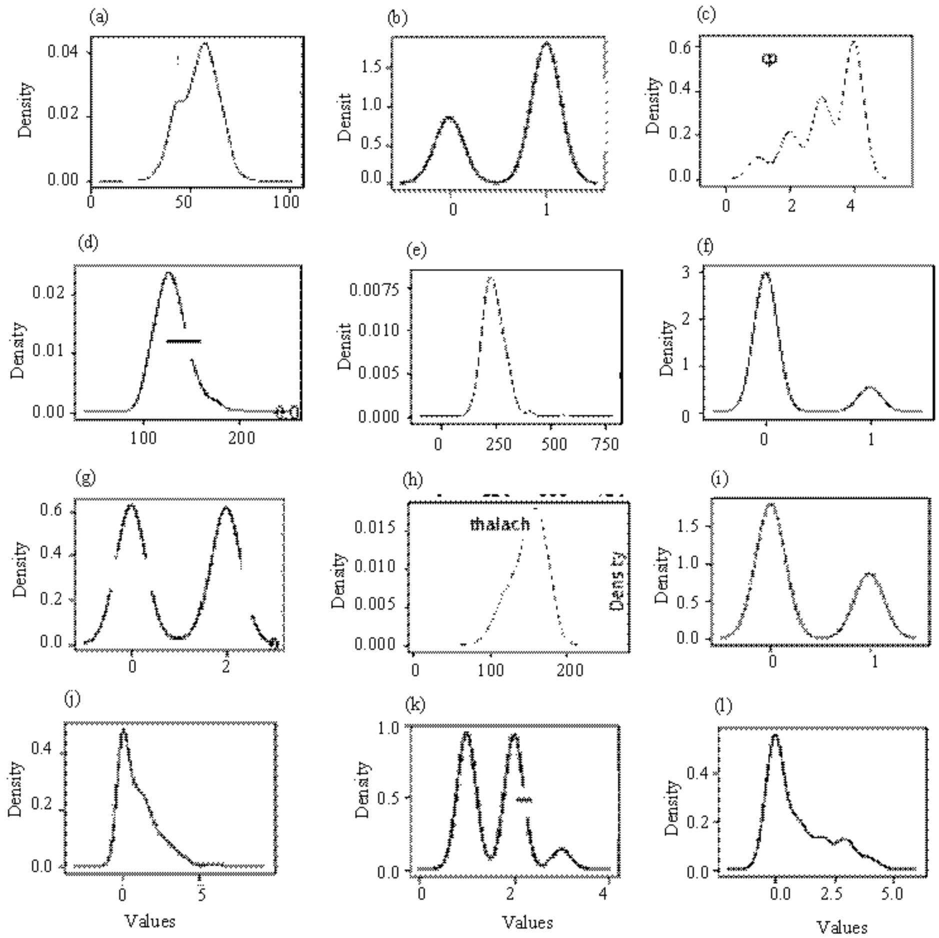


Fig. 4: CHD features density: a) Age; b) Sex; c) cp; d) trestops; e) chol; f) fbs; g) restecg; h) thalach; i) exang; j) oldpeak; k) slope and l) num

Table 3: Accumulative features ranking according to age and sex

| FST\features | Age | Sex | cp | trestbps | chol | fbs | restecg | thalach | exang | oldpeak | slope | ca | thal |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Correlation Based (CB) | 0.22 | 0.28 | 0.41 | 0.15 | 0.09 | 0.03 | 0.17 | 0.42 | 0.43 | 0.42 | 0.34 | 0.46 | 0.52 |
| Gain Ratio (GR) | 0.06 | 0.06 | 0.20 | 0.00 | 0.000 | 0.00 | 0.020 | 0.130 | 0.150 | 0.10 | 0.110 | 0.17 | 0.21 |
| Info Gain (IG) | 0.06 | 0.06 | 0.20 | 0.00 | 0.000 | 0.00 | 0.020 | 0.130 | 0.140 | 0.15 | 0.110 | 0.17 | 0.20 |
| OneR (OR) | 0.63 | 0.61 | 0.76 | 0.56 | 0.530 | 0.51 | 0.590 | 0.640 | 0.720 | 0.69 | 0.680 | 0.74 | 0.76 |
| Relief (Re) | 0.02 | 0.10 | 0.08 | 0.02 | 0.000 | 0.01 | 0.090 | 0.020 | 0.080 | 0.03 | 0.05 | 0.09 | 0.08 |
| Symmetrical (Sy) | 0.06 | 0.06 | 0.20 | 0.00 | 0.0 0 | 0.00 | 0.020 | 0.130 | 0.150 | 0.12 | 0.110 | 0.17 | 0.21 |
| Total | 1.06 | 1.17 | 1.86 | 0.73 | 0.610 | 0.55 | 0.920 | 1.460 | 1.670 | 1.52 | 1.400 | 1.80 | 1.98 |

Table 4 shows a list of features with accumulating rank greater than the threshold those features will form a new dataset containing same number of instances as the original dataset but less number of features (9 features), in addition to the class feature. Features with accumulating rank less than threshold have been considered as non-optimal features and accordingly, they have been removed from the dataset. Table 5 and 6 show results of applying classification techniques on newly constructed dataset. The results of applying classification techniques based on results of FSTs were as follow: all classification techniques produced better prediction rate based on accuracy, precision, recall and f-measure. kNN recorded the best enhancement rate in accuracy (+3.7%), precision (+8.1%), recall and (+0.3.5%) f-measure (+.5.7%), then ANN, J48, SVM and Naive Bayse as shown in Fig. 5. Figure 5 shows classification measures using the original dataset (black column) and classification measures using the new constructed dataset on gray.

Table 4: Classification results on original CHD

| Classifiers | Accuracy | Precision | Recall | F-measure |
|---|---|---|---|---|
| N-Bayes | 0.834 | 0.836 | 0.835 | 0.834 |
| SVM | 0.838 | 0.834 | 0.838 | 0.838 |
| kNN | 0.752 | 0.754 | 0.752 | 0.753 |
| ANN | 0.788 | 0.784 | 0.789 | 0.788 |
| TREE J48 | 0.784 | 0.785 | 0.785 | 0.785 |

Table 5: Accumulative features ranking

| Features/FTS | CB | GR | IG | OR | Re | Sy | Total |
|---|---|---|---|---|---|---|---|
| age | 0.22 | 0.06 | 0.06 | 0.63 | 0.02 | 0.06 | 1.06 |
| sex | 0.28 | 0.06 | 0.06 | 0.61 | 0.10 | 0.06 | 1.17 |
| cp | 0.41 | 0.20 | 0.20 | 0.76 | 0.08 | 0.20 | 1.86 |
| thalach | 0.42 | 0.13 | 0.13 | 0.64 | 0.02 | 0.13 | 1.46 |
| exang | 0.43 | 0.15 | 0.14 | 0.72 | 0.08 | 0.15 | 1.67 |
| oldpeak | 0.42 | 0.10 | 0.15 | 0.69 | 0.03 | 0.12 | 1.52 |
| slope | 0.34 | 0.11 | 0.11 | 0.68 | 0.05 | 0.11 | 1.40 |
| ca | 0.46 | 0.17 | 0.17 | 0.74 | 0.09 | 0.17 | 1.80 |
| thal | 0.52 | 0.21 | 0.20 | 0.76 | 0.08 | 0.21 | 1.98 |

Table 6: Classification results on CHD selected features

| Classifier | Accuracy | Precision | Recall | F-measure |
|---|---|---|---|---|
| N-Bayes | 0.835 | 0.878 | 0.828 | 0.852 |
| SVM | 0.842 | 0.896 | 0.826 | 0.8600 |
| kNN | 0.789 | 0.835 | 0.787 | 0.8100 |
| ANN | 0.819 | 0.854 | 0.819 | 0.8360 |
| TREE J48 | 0.792 | 0.842 | 0.789 | 0.8140 |

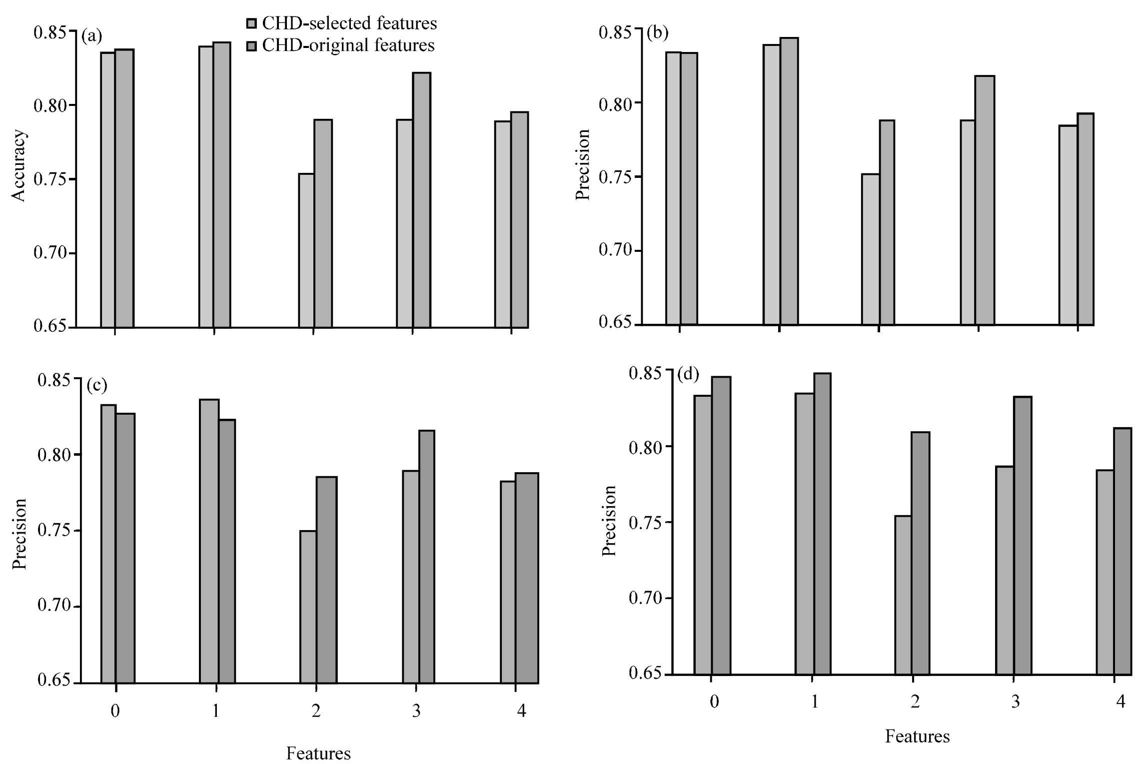


Fig. 5: CHD features 0: N-Bays: SVM 2: kNN 3: ANN 4: J48

## CONCLUSION

Features selection algorithms aim to eliminate irrelevant features from the original dataset on certain evaluation criteria to improve machine learning techniques predictions. In this research, we used several machine learning classification techniques (Naive Bayes, k-nearest neighbor, support vector machine, artificial neural network and tree J48 classification techniques) and several supervised features selection techniques (information gain, gain ratio, relieff, symmetrical uncertainty and oneR features selection techniques). We attempted to find and apply the most appropriate features selection techniques collaboration method which was the accumulated rank. We applied our approach to a public dataset, cleveland heart disease dataset, the results of applying classification techniques based on the results of features selection techniques were as follow, kNN recorded the best enhancement rate in accuracy (+3.7%) precision (+8.1%) recall and (+0.3.5%) F-measure (+0.5.7%). Finally, we are aspire to expand the use of accumulated rank as features selection technique by evaluating our approach over more datasets from different domains.

## REFERENCES


Ahmad, M.A.B., 2013. Mining health data for breast cancer diagnosis using machine learning. Ph.D Thesis, University of Canberra, Canberra, Australia.

Alhaj, T.A., M.M. Siraj, A. Zainal, H.T. Elshoush and F. Elhaj, 2016. Feature selection using information gain for improved structural-based alert correlation. PloS One, 11: 1-18.

Andreeva, P., 2006. Data modelling and specific rule generation via. data mining techniques. Proceedings of the International Conference on Computer Systems and Technologies-CompSysTech, June 15-16, 2006, University of Ruse, Ruse, Bulgaria, pp: IIIA.17-1-IIIA.17-6.

Ashraf, M., G. Chetty, D. Tran and D. Sharma, 2012. Hybrid approach for diagnosing thyroid, hepatitis and breast cancer based on correlation based feature selection and Naive Bayes. Proceedings of the International Conference on Neural Information Processing, November 12-15, 2012, Springer, Berlin, Germany, ISBN:978-3-642-34477-0, pp: 272-280.

Ashraf, M., K. Le and X. Huang, 2010. Information gain and adaptive neuro-fuzzy inference system for breast cancer diagnoses. Proceedings of the 2010 5th International Conference on Computer Sciences and Convergence Information Technology (ICCIT), November 30-December 2, 2010, IEEE, Seoul, South Korea, ISBN:978-1-4244-8567-3, pp: 911-915.

Ashraf, M., K. Le and X. Huang, 2011. Iterative weighted k-NN for constructing missing feature values in Wisconsin breast cancer dataset. Proceedings of the 2011 3rd International Conference on Data Mining and Intelligent Information Technology Applications (ICMiA), October 24-26, 2011, IEEE, Macao, China, ISBN:978-1-4673-0231-9, pp: 23-27.

Breve, F.A. and D.C.G. Pedronette, 2016. Combined unsupervised and semi-supervised learning for data classification. Proceedings of the 2016 IEEE 26th International Workshop on Machine Learning for Signal Processing (MLSP), September 13-16, 2016, IEEE, Vietri sul Mare, Italy, ISBN:978-1-5090-0747-9, pp: 1-6.

Canedo, V.B., N.S. Marono, A.A. Betanzos, J.M. Benitez and F. Herrera, 2014. A review of microarray datasets and applied feature selection methods. Inf. Sci., 282: 111-135.

Cheung, N., 2001. Machine learning techniques for medical analysis. BSc Thesis, School of Information Technology and Electrical Engineering, University of Queenland, Brisbane, Australia.

Chuang, L.Y., C.H. Ke and C.H. Yang, 2008. A hybrid both filter and wrapper feature selection method for microarray classification. Proceedings of the International MultiConference on Engineers and Computer Scientists (IMECS'2008) Vol. 1, March 19-21, 2008, International Association of Engineers, Hong Kong, ISBN:978-988-98671-8-8, pp: 1-5.

Demidova, L., Y. Sokolova and E. Nikulchev, 2015. Use of fuzzy clustering algorithms ensemble for SVM classifier development. Intl. Rev. Mod. Simul., 8: 446-457.

Dimitoglou, G., J.A. Adams and C.M. Jim, 2012. Comparison of the C4.5 and a Naive Bayes classifier for the prediction of lung cancer survivability. J. Comput., 4: 1-9.

Dua, S. and X. Du, 2016. Data Mining and Machine Learning in Cybersecurity. CRC Press, Boca Raton, Florida, USA., ISBN-13:978-1-4398-3943-0, Pages: 224.

Fernandez-Delgado, M., E. Cernadas, S. Barro and D. Amorim, 2014. Do we need hundreds of classifiers to solve real world classification problems?. J. Mach. Learn. Res., 15: 3133-3181.

Fowkes, F.G.R., D. Rudan, I. Rudan, V. Aboyans and J.O. Denenberg *et al*., 2013. Comparison of global estimates of prevalence and risk factors for peripheral artery disease in 2000 and 2010: A systematic review and analysis. Lancet, 382: 1329-1340.

Frank, E., M. Hall, L. Trigg, G. Holmes and I.H. Witten, 2004. Data mining in bioinformatics using Weka. Bioinformatics, 20: 2479-2481.

Inbarani, H.H., M. Bagyamathi and A.T. Azar, 2015. A novel hybrid feature selection method based on rough set and improved harmony search. Neural Comput. Appl., 26: 1859-1880.

Kangwanariyakul, Y., T. Naenna, C. Nantasenamat and T. Tantimongcolwat, 2010. Data mining of magnetocardiograms for prediction of ischemic heart disease. Excli J., 9: 82-95.

Karegowda, A.G., A.S. Manjunath and M.A. Jayaram, 2010. Comparative study of attribute selection using gain ratio and correlation based feature selection. Int. J. Inform. Technol. Knowledge Manage., 2: 271-277.

Kavitha, R. and E. Kannan, 2016. An efficient framework for heart disease classification using feature extraction and feature selection technique in data mining. Proceedings of the 2016 International Conference on Emerging Trends in Engineering, Technology and Science (ICETETS), February 24-26, 2016, IEEE, Pudukkottai, India, ISBN:978-1-4673-6726-4, pp: 1-5.

Kinge, D. and S.K. Gaikwad, 2018. Survey on data mining techniques for disease prediction. Intl. Res. J. Eng. Technol., 5: 630-636.

Kotsiantis, S.B., 2007. Supervised machine learning: A review of classification techniques. Informatica, 31: 249-268.

Krishnaiah, V., G. Narsimha and N.S. Chandra, 2014. Survey of classification techniques in data mining. Intl. J. Comput. Sci. Eng., 2: 65-74.

Kumar, A., S. Roy and P. Ranjan, 2014. Dimensionality reduction for high dimensional data: A CFS and Chi Square approach. Proceedings of the 2014 International Conference on Information and Communication Technology for Competitive Strategies, November 14-16, 2014, ACM, New York, USA., ISBN:978-1-4503-3216-3, pp: 1-6.

Long, N.C., P. Meesad and H. Unger, 2015. A highly accurate firefly based algorithm for heart disease prediction. Expert Syst. Appl., 42: 8221-8231.

Pal, M. and P.M. Mather, 2005. Support vector machines for classification in remote sensing. Intl. J. Remote Sens., 26: 1007-1011.

Pouriyeh, S., S Vahid, G. Sannino, G. De Pietro and H. Arabnia *et al*., 2017. A comprehensive investigation and comparison of machine learning techniques in the domain of heart disease. Proceedings of the 2017 IEEE International Symposium on Computers and Communications (ISCC), July 3-6, 2017, IEEE, Heraklion, Greece, ISBN:978-1-5386-1630-7, pp: 204-207.

Saravanan, N., V.N.S.K. Siddabattuni and K.I. Ramachandran, 2008. A comparative study on classification of features by SVM and PSVM extracted using Morlet wavelet for fault diagnosis of spur bevel gear box. Exp. Syst. Appl., 35: 1351-1366.

Shao, Y.E., C.D. Hou and C.C. Chiu, 2014. Hybrid intelligent modeling schemes for heart disease classification. Appl. Soft Comput., 14: 47-52.

Singh, A., N. Thakur and A. Sharma, 2016. A review of supervised machine learning algorithms. Proceedings of the 2016 3rd International Conference on Computing for Sustainable Global Development (INDIACom), March 16-18, 2016, IEEE, New Delhi, India, ISBN:978-1-4673-9417-8, pp: 1310-1315.

Wald, R., T.M. Khoshgoftaar and A. Napolitano, 2014. Using correlation-based feature selection for a diverse collection of bioinformatics datasets. Proceedings of the 2014 IEEE International Conference on Bioinformatics and Bioengineering (BIBE), November 10-12, 2014, IEEE, Boca Raton, Florida, USA., ISBN:978-1-4799-7502-0, pp: 156-162.

Yildirim, P., 2015. Filter based feature selection methods for prediction of risks in hepatitis disease. Intl. J. Mach. Learn. Comput., 5: 258-263.

Zhang, W., J. Tang and N. Wang, 2016. Using the machine learning approach to predict patient survival from high-dimensional survival data. Proceedings of the 2016 IEEE International Conference on Bioinformatics and Biomedicine (BIBM), December 15-18, 2016, IEEE, Shenzhen, China, ISBN:978-1-5090-1612-9, pp: 1234-1238.