

Disease Prediction Improvement Based on Modified Rough Set and Most Common Decision Tree

Eman Al-Shamery and Ali Rahoomi Al-Obaidi

Department of Software, Faculty of Information Technology, University of Babylon, Hillah, Iraq

Abstract: In the health care sector predict the type or severity of the diseases is important for helping people to know their health stability and find solution for any negative indicator. This study aims to improve prediction of the diseases by exploiting Modified Rough Set (MRS) for features selection which it is proposed as a new method and employing a Most Common Decision Tree (MCDT) which is suggested as a modified of decision tree method for making decision. In addition to pre-processing stage which contains the Mode-Relation-Average (MRA) method for filling missing value and grouping. The system consists of 2 main stages, first is features selection based on MRS and the second disease prediction using MCDT. The output of MRS Model is three subsets of features graduate according to the importance degree of features: Most Important (MI), Important (I) and Unimportant (UM). The MCDT is used to predict disease type. Subsequently, 3-cross validation is used in the testing process. The proposed system has been applied on binary and multi classes heart disease (5 classes). Finally, the proposed methods have been compared with the other methods such as Naive Bayes and Bayes Net. The Max correctly classified instances obtained using the proposed methods are better than 2 previous methods.

Key words: Heart disease, mode-relation-average, modified rough set, most common decision tree, output, people

INTRODUCTION

The global ambition increases to achieve the best quality in the field of health care. There became a global boom in the field of information systems and electronic processing systems. The reason for this is the progress in the field of artificial intelligence which generated an increase in analytical methods and the promotion of knowledge representation. The role of artificial intelligence in the field of health care includes diagnosis and decision-making, robots and robotic surgery (Ashrafian *et al.*, 2015). Intelligent systems uses Artificial Intelligence (AI) for solving complex problems, generally AI assume such a system on machine or computer.

There are 2 strategies of learning model: supervised and unsupervised learning. In supervised learning, model parameters need to learn by using training set in presence a class of attributes whereas unsupervised no presence of attributes class for training set, e.g., neural network, naïve bayes and decision tree are supervised and k-means clustering is unsupervised. Each technique serves a different purpose depending on the modelling objective (Sanap and Ishtake, 2013).

Health care systems began to provide patients and doctors with personal information about the unique health

experiences of kind for everyone in both sickness and health. Thus, the people can take concrete action to improve their health also note the impact of lifestyle decisions. Working together in the family networks and the convergence of digitally-powered, the person will be able to reduce the prevalence of complex chronic diseases which currently represent 75% of care costs of the disease in the United States of America (Flores *et al.*, 2013). Diseases of the heart and blood vessels produce enormous health and economic burden in globally. A range of disease states and clinical home including stroke and congenital heart disease and hardening of the arteries subclinical and coronary heart disease, heart failure and the associated results including economic costs, procedures and quality of care (Mozaffarian *et al.*, 2015).

Literature review: For features selection methods: feature selection via. dependence maximization is proposed as a method for features selection by Song *et al.* (2012). This method select features based on dependence maximization between the selected features and the labels of an estimation problem. This method suggests weighting scheme for backward and forward selection. Forward trying to increase the potential candidates for feature selection as possible for each insert features and eliminate backward trying to achieve this for each deletion of

features. The limitation of these method is that when the feature eliminated (selected) cannot be selected (eliminated) again, although, it may be needed later (Song *et al.*, 2012). Genetic Algorithm (GA) for features selection has been explained in survey on feature selection methods by Chandrashekar and Sahin (2014). This method can be used to find a subset of features. The chromosome bits indicate if the feature included or not. The objective function of global maximum can be found which gives the best sub-optimal subset where the objective function is the performance prediction. The drawbacks of this method, first it required a number of computations to obtain the feature subset and the second it cannot store the accuracies of evaluated subsets for future retrieval (Chandrashekar and Sahin, 2014). Correlation Feature Selection (CFS) is explained by Canedo *et al.* (2014). This method ranks feature subsets according to a correlation. The most important subsets of features that are uncorrelated with each other and highly correlated with the class. Irrelevant features should be ignored because that have a low correlation with the class. The drawbacks of this method is that the selected subset might contain redundant features and thus, the subset is not optimal as well as a smaller subset can be highly correlated variables might be obtained (Canedo *et al.*, 2014). Hierarchical Feature Selection (HFS) is described by Wan (2016). This method exploits hierarchical relationships between features and then remove hierarchical redundancy between features in order to improve the predictive accuracy of classification algorithms. The limitation of these method makes single comparison only between child and its parent as well as it uses greedy top down approach for visiting hierarchy that means features selected in vertical manner and not utilize whole hierarchy (Wan, 2016).

In the prediction process, 3 supervised machine learning algorithms (K-NN, Naive Bayes and decision list algorithm) applied on Cleveland heart disease database are used by Soni *et al.* (2011). Weighted fuzzy rule is proposed as a method for clinical decision support system by Anooj (2012) to diagnosis heart disease 0 (no presence) to 4. Computational intelligence using medical knowledge driven approach is proposed as a method for heart disease diagnosis by Nahar *et al.* (2013). This method used One-Over-All (OVA) approach that applied on heart diseases data set have five classes. OVA approach dealing with multi class label as binary using binary classifier which suppose one of the class labels is considered as positive and the rest as negative. This system uses different algorithms such Naive Bayes and J48 (Nahar *et al.*, 2013). A survey of data mining techniques on risk prediction of heart disease has been accomplished by Purusothaman and Krishnakumari (2015).

MATERIALS AND METHODS

In this study, we will review the features selection method and the prediction methods which represent the theoretical background for proposed system.

Rough set theory: Rough set is proposed by Pawlak to deal with uncertainty and incompleteness. It offers mathematical tools to discover patterns hidden in data and identifies partial or total dependencies in data based on indiscernibility relation. The features selection is calculated to determine the relevant feature. The terms reducts and core are important instruments for feature selection. From all features rough set theory can achieve a subset of features that maintain the remarkable ability of the original features, using only the data without any additional information. The lower approximation of set is the set of objects of the information system table which certainly belongs to the class. The upper approximation of a set includes all objects of information system table which possibly belongs to the class. In this step assign people may be normal or abnormal. In other words union lower approximation and Boundary Region. Boundary region is the difference between upper approximation set and lower approximation set that is referred to as Bnd. Boundary contains the patients have no decision making, people may be normal or abnormal. Positive region is the set of all objects that belong to lower approximation. In this step union lower approximation consist of union of lowers approximation sets. Reducts is the minimum range representation of the original data without loss of information. It is minimal subsets of attributes that can be interested in finding all possible. When indiscernibility relation of set of attributes and its superset are same, then any attribute that found in superset and not found in the set is redundant and called reduct on subset. Core the intersection of the elements of reducts is called core or indispensable attributes, i.e., core attribute is a feature that appears in all reducts (Anitha, 2016).

Decision tree: It is one of the most important and popular classifier which is simple and easy to implement. It based on dividend-conquer scheme. It does not have parameter setting or domain knowledge. It deals with the huge amount of data dimensions. It is appropriate for exploring knowledge discovery. It is easy to read and interpret the results that obtained from the decision tree (Soni *et al.*, 2011).

Naive Bayes classifier: It is based on Bayes theorem. It uses conditional independence that means a value of attribute is an independent of others attributes values of a given class. Naive Bayes is an algorithm that is

supervised learned which assumes that they have equal importance features and they are independent. For many machine learning problems, Naive Bayes algorithm still works well although, the features are not independent (Dangare and Apte, 2012).

Bayesian network: Bayesian network is a method that has been used to represent dependencies in a probability distribution graphically by using structure of a directed acyclic graph. Each feature represent as node and dependencies between these features which represent as arcs. Bayesian network is a statistical model computes that for any subset of unobserved stochastic variables find the joint probability distribution given that the variables in the complementary subset are observed. Bayesian network is a statistical classifier by using winner-takes-all rule to the posterior probability distribution for the unobserved class node (Sucar *et al.*, 2014) (Fig. 1).

The proposed system: Where (*) refer to developing or modifying of original methods.

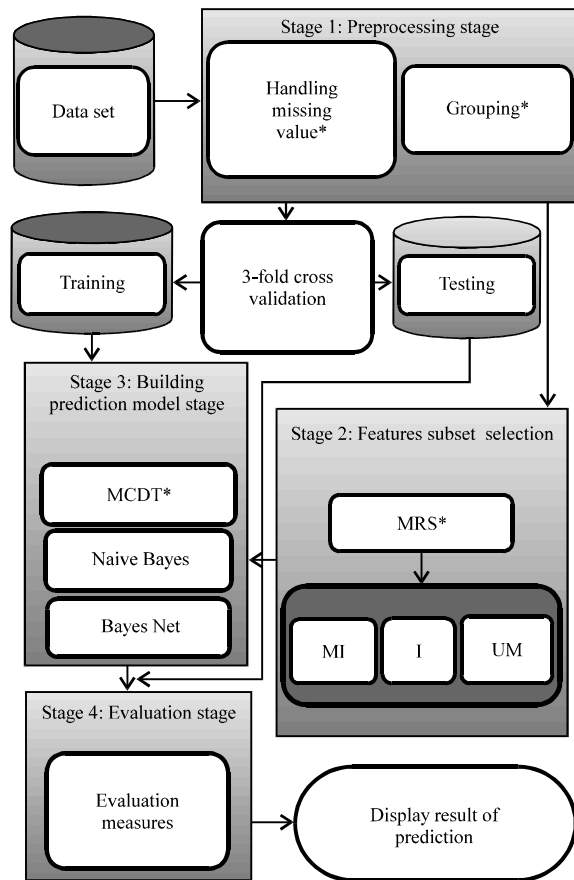


Fig.1: Block diagram of proposed system

Prepressing: Prepressing stage is divided into three phases: the Grouping phase, handling the missing values in a data set and normalization phase.

Grouping: For any certain feature, The Grouping is a process that give same code for each interval under specialist consultant i.e., coding range of near values in same code. The Grouping is an essential process in the features selection stage especially when applying indiscernibility relation of rough set method, suppose that the age of the first patient’s and the age of the second patient equal to 30 and 31, respectively: the result of elementary set for two different sub set group {p1}, {p2} but in fact age values are very close. The Grouping has been applied to solve this problem. For example coding ages from 30-39-1 where the age of Patient one and the age of patient 2 coding with same code and then the result of elementary set for age feature is same group for patient one and Patient 2 {p1, p2}. In heart data set using opinion of doctor’s to covert data value for some attributes for quantitative to qualitative and then coding to quantitative to use it in rough set feature selection.

Handling missing values: Handling missing value using particular procedure using MAR method. This method consists of 3 steps:

- Determining class type
- Finding mode of group (most frequent group belongs to this class)
- Replacing missing values by interval average of group

MRA is calculated for each feature separately. The mode is the most common value (repeat) in the feature at that class.

Normalization: Normalization is required in the prediction stage essentially when using statistical learning methods such as Bayes Net and Naive Bayes. In this work a normalization has been applied before using statistical prediction methods (Bayes Net and Naive Bayes). Normalization is used to scale of each numerical values of the numerical features to range symbolized [0, 1] to obtain new value, the formula of normalization is show in Eq. 1:

$$n_{val} = \frac{v - \min_A}{\max_A - \min_A} \quad (1)$$

where, the maxA and the minA is the maximum and minimum original values of features, respectively (Panse and Kathuria, 2016).

Modified rough set for features selection: Modified Rough Set for Feature Selection (MRS) is organized in nine steps. In begging the lower approximation is determined for classes that make the following process for each certain data value of features sub set must give same class, i.e., for each patient (303 in heart disease) must determine which of them is certain normal and which of them is certain abnormal. The record is certain class (normal/abnormal) when all features value of it have unique values or the record is similar to other records values that have same class (records have same values features and same class). The main steps of MRS is shows in Algorithm 1.

Algorithm 1; Modified rough set for feature selection:

Name: Modified rough set

Input: Data set (U, Att) where U is universal set of discussion and Att is an attribute set that must be not empty and has limited range (finite), values of θ_1 and θ_2

Output: Three subset of features MI, I and UM

Begin

1: Find lower approximation:

$$-AX = \{x_i | i \in U - |[x_i]_{L_i} \text{ind}(A) \subset X\}, X \in \text{Att}$$

2: Find upper approximation:

$$-AX = \{x_i | i \in U - |[x_i]_{L_i} \text{ind}(A) \cap X \neq \emptyset\}$$

3: Find boundary:

$$\beta = \overline{AX} - AX$$

4: Find positive reign $\rho = \cup AX$

Union of all lowers set

5: Find indiscernibility of positive reign for any $G \subseteq \text{Att}$ there are is an associated equivalence relation:

$$\text{IND}(G) = \{(x, y) \in p \times p : \forall \alpha(x) = \alpha(y)\}$$

6: Find reducts $\delta = \text{minIND}$

If set of attributes and its superset = indiscernibility relation then small set, Super set = redundant

End if

7: Compute reliability degree for each attributes using the following formula:

$$\forall \text{Attribute } R_i = \left(\frac{\sum_{j=1}^{\text{Total number of reducts}} |R_i|}{\text{Total number of Redu}} \right) \%$$

8: Rate reducts:

$$\forall \text{Red Att } R_{\text{Rate}} = \left(\frac{1}{\text{RedLength}} \sum_{i=1}^{\text{RwREEDLength}} R_i \right) *$$

$$\text{AV certain} = \left(\sum_{i=1}^{\text{Rv}} \frac{R_i}{\text{RedLength}} \right) * (\text{AV one} *)$$

$$\text{RedLength} = \left(\frac{1}{\text{RedLength}} \sum_{i=1}^{\text{RedLength}} R_i \right) *$$

$$\left(\left(\frac{1}{\text{Total numbers of features}} \right) \right) * \text{RedLength}$$

9: Assign impotence degree for features Most Important (MI), Important (I) and Unimportant (UM)

If $R_i \geq \theta_1$ and ϵ_{Red} then

Feature is MI

Elseif $R_i \geq \theta_2$ and $R_i < \theta_1$ then

Feature is I

Else

Feature is UM

Endif

End

Next, upper approximation is computed for classes that make the following process for each data value of sub set of features values possible give same class, i.e., for all patients determining any of them is possible normal and which of them is possible abnormal. The record is possible class (normal/abnormal) when all features values of it has unique values or the record is similar to other records values have same class (records have same values features and same class) or the record is similar to other records values that have different class (records have same values features and different class). The upper bound of for class is the union of lower bound approximation and boundary region. Then, finding boundary region of classes. For each data value of sub set of features have no decision. When features values of a record are similar to other record values but they different in class value, put both records in boundary region, i.e., when records have equivalent features values but they different in decision class (one give yes and other give no) this features values lead to no decision and the set of this records called boundary region. Later, positive region of classes is computed. Positive region consists of union of elements of lower bound approximation. Positive region consist of set of records that give certain decision, i.e., ignoring record that gives no decision for prediction class. Then, indiscernibility relation is applied for positive region to find elementary set. Where indiscernibility relation is equivalence relation. Applying indiscernibility relation for all possible subsets of attributes. There are sub set of attributes where n is number of attributes. In step 6, reducts are calculated for sub set of attributes. reduct is the minimum range representation of the original data without loss of information. It is minimal subsets of attributes can be interested in finding all possible. If indiscernibility relation of set of attributes and its superset are same then any attribute that found in superset and not found in the set is redundant and called reduct on subset. Additionally, Reliability degree is computed for each attributes depending on the appearance number in reducts where reliability degree for each attribute is appearance average of feature in reducts on total number of reducts. Assigning weight for each reduct is implemented in

step 8 based on reliability degree of attributes, sub set of reducts attributes and No. of attributes in reduct. The reducts rate is calculated by multiplying Average Weight of Certain Length (AWCL) with average reliability degree of reduct. AWCL is calculated by multiplying Average Weight of One Length (AWOL) with number of features in reducts where AWOL is calculated by division the value one on total number of attributes (13 in heart disease). Finally assigning impotence degree of features for 3 groups Most Important (MI), Important (I) and Unimportant (UM) based on reliability degree. If Reliability degree between [75-100] is called MI subset else if reliability degree between [35-75] is called I subset else if reliability degree <35 is called UM.

Most Common Decision Tree (MCDT): Many stopping criteria have been used in MCDT such as number of iterations and number of used attributes. The main steps of MCDT is shows in Algorithm 2. After assigning the type of class as leaf node. The direction of leaf node is determined either left or right. Then linking the leaf node with its parent.

Algorithm 2; MCDT Algorithm:

Name: Most Common Decision Tree
 Input: Diseases Dataset, u and i where u is Number of used attributes and i is iteration count
 Output: Class type

```

Begin
  Step 1: If stop criteria is true then
    1.1 Create New Leaf node
    1.2 Increase iteration count
    1.3 Set type of node as Leaf node
    1.4 Set direction of node Left or Right
    1.5 Link leaf node with parent
    1.6 Predicated class using maximum probability of instance
  Else
    1.7 Create New Internal node
    1.8 Increase iteration count
    1.9 Set direction of node Left or Right
    1.10 Link Internal node with parent
    1.11 Compute information before and after splitting for unused attributes using:

      Entropy (t) = - ∑j p(j|t) log2 p(j|t)

      GINI (t) = 1 - ∑j [p(j|t)]2

      Classification error (t) = 1 - maxi p(i|t)

    Gain-en = info-before-splitting-en
    info-after-splitting-en
    Name-en = get name of attributes
    Gain-gi = info-before-splitting-gi-info-after-splitting-gi
    Name-gi = get name of attributes
    Gain-er = info-before-splitting-er-info-after-splitting-er
    Name-er = get name of attributes
    1.12 Select best gain for three measure as possible choice attribute for spiting (Name-en, Name-gi or Name-er)
    1.13 Use most common to select best attributes
    If Name-en = Name-gi = Name-er then
      Set selected attributes as Name-en
      Mark attributes as used attributes
  
```

```

Else if Name-en = Name_gi then
  Set selected attributes as Name-en
  Mark attributes as used attributes
Else if Name-en = Name-er then
  Set selected attributes as Name-en
  Mark attributes as used attributes
Else if Name-gi = Name-er then
  Set selected attributes as Name-gi
  Mark attributes as used attributes
Else
  Set selected attributes as Name-en
  Mark attributes as used attributes
End if
1.14 Use most common to select best category for splitting of selected attributes
For each category compute information for three measure
Select best value for three measures as possible choice category for splitting
(CatName-en, CatName-gi or CatName-er)
If CatName-en = CatName-gi = CatName-er then
  Set selected category as CatName-en
Else if CatName-en = CatName-gi then
  Set selected category as CatName-en
Else if CatName-en = CatName-er then
  Set selected category as CatName-en
Else if CatName-gi = CatName-er then
  Set selected category as CatName-gi
Else
  Set selected category as CatName-en
End if
End for
1.15 According Split point split data into two parts: Tree-Left, Tree-Right
If Val <= Split point category
  Tree-Left
Else
  Tree -Right
End if
Step 2: Return tree binary spiting
End
  
```

Max probability of instances is used in order to predicate the class type. Create new internal node until stopping criteria satisfying. For all unused attributes, the gain is computed before and after splitting to select next internal node has best value of gain by three measures: entropy, Gini index and classification error:

$$\text{Entropy (t)} = \sum_j p(j|t) \log_2 p(j|t)$$

Where:

$$\text{GINI (t)} = 1 - \sum_j [p(j|t)]^2$$

And:

$$\text{Classification error (t)} = 1 - \max_i p(i|t)$$

One of 3 selections is nominated as next internal node that has the best gain for these measures with maximum vote. If each measure indicated different node the winner node of Entropy measure will be chosen. The same above approach is used to compute best category splitting for selected node to trace the tree and go down left or right. This approach represents the most common for making decision.

Table 1: Dataset description

Disease types	No. of features	No. of classes	No. of patients
Binary class heart disease	13	2	303
Multi class heart disease	13	5	303

Table 2: Summary for core and reducts of heart disease

Rough set	No. of reducts	No. of core
Original data	446	0
Grouping	102	2

Table 3: The subset level measuring in original and grouping

Level sub set	Original		Grouping	
	Attributes name	No.	Attributes name	No.
MI	Non	0	Age, sex, cp, trestbps, restecg, oldpeak, ca	7
I	Sex, CP, threstbps, chol, restecg, thalach, exang, oldpeak, slope, ca, thal	11	Chol, fbs, exang, slop, thal	5
UM	Age, fbs	2	Thalach	1

RESULTS AND DISCUSSION

About 3 cross validation has been used for evaluating. In this study data set of binary and multi classes have been applied, the dataset description are shown in Table 1. Finally, the system is compared with the other methods such as Bayes Net and Naive Bayes.

Reducts: Heart disease dataset has 8190 (213-2) subset of MRS reduced this complexity by selecting reducts that represent most relevant subsets. The total number of reducts of heart disease is 446 using original data and 102 using grouping data.

Core attribute: Core attributes are a features that appears in all reducts (Table 2).

Level subset of attributes

Most Common Decision (MCDT): The evaluation measures have been used for binary and multi classification are accuracy and Precision. Accuracy (Acc) is a measure used to measure the proximity of the true value. Precision (Pre) or Confidence or True Positive Accuracy (TPA) indicates the percentage of positive cases predicted that are a real positive properly (Sim *et al.*, 2012; Powers, 2011) (Table 3-5):

$$Acc = \frac{(TP+TN)}{(TP+TN+FP+FN)} \quad (2)$$

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

Table 4: Binary classes prediction heart disease using different techniques

Prediction method	DT (J48)	MCDT	Naive Bayes	Bayes Net
Full attributes				
Acc	76.56	80.10	81.51	81.18
Pre	78.52	78.39	80.68	80.22
MI and I using original data				
Acc	74.25	89.01	78.87	76.23
Pre	76.54	86.06	79.41	77.38
MI using grouping				
Acc	74.25	94.06	77.55	75.90
Pre	75.90	92.22	77.58	76.30
MI and I using grouping				
Acc	76.56	94.06	82.5	82.17
Pre	78.52	92.54	82.45	81.97

Table 5: Multi classes prediction heart disease using different techniques

Prediction method	DT (J48)	MCDT	Naive Bayes	Bayes net
Full attributes				
Acc	54.12	88.68	55.11	52.47
Pre	54.12	87.88	52.50	51.20
MI and I using original data				
Acc	53.13	88.82	55.44	56.10
Pre	44.90	90.57	51.00	51.50
MI using grouping				
Acc	55.44	94.06	53.13	49.50
Pre	40.10	92.54	46.30	45.80
MI and I using grouping				
Acc	54.12	93.67	55.77	49.50
Pre	29.30	90.40	52.60	45.80

Binary class of heart disease: Table 4 discussed in binary class.

Multi class of heart disease: Table 5 discussed in multiclass of heart diseases.

CONCLUSION

In this study, the type or severity of the heart diseases are discussed. The system is implemented on binary and multi classes data set. In the preprocessing stage, MAR has been proposed as a new method for replacing missing values in data set. About 2 main methods have been suggested, MRS for features selection and MCDT for disease prediction. MRS indicated a significant impact for reducing the number of features and assigning the features that are directly related to the disease. MRS determined reliability degree for each feature. The results show the MRS is efficiency. In general MRS gives precise description for each feature. MCDT satisfied a better result because it depend on 3 different impurity measures instead of single impurity measure for building tree model. MCDT can help the specialists in early diagnosing of heart disease. The accuracy is improved when discarding UM features. Finally the proposed system is compared with others prediction methods Naive Bayes and Bayes Net, the accuracy obtained of proposed system shows that it is better than 2 previous methods.

REFERENCES

- Anitha, K., 2016. Rough neural network. GSTF. J. Math. Stat. Oper. Res., 3: 59-62.
- Anooj, P.K., 2012. Clinical decision support system: Risk level prediction of heart disease using weighted fuzzy rules. J. King Saud Univ. Comput. Inf. Sci., 24: 27-40.
- Ashrafian, H., A. Darzi and T. Athanasiou, 2015. A novel modification of the turing test for artificial intelligence and robotics in healthcare. Intl. J. Med. Rob. Comput. Assisted Surg., 11: 38-43.
- 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.
- Chandrashekar, G. and F. Sahin, 2014. A survey on feature selection methods. Comput. Electr. Eng., 40: 16-28.
- Dangare, C.S. and S.S. Apte, 2012. Improved study of heart disease prediction system using data mining classification techniques. Intl. J. Comput. Appl., 47: 44-48.
- Flores, M., G. Glusman, K. Brogaard, N.D. Price and L. Hood, 2013. P4 medicine: How systems medicine will transform the healthcare sector and society. Pers. Med., 10: 565-576.
- Mozaffarian, D., E.J. Benjamin, A.S. Go, D.K. Arnett and M.J. Blaha *et al.*, 2015. Executive summary: Heart disease and stroke statistics-2015 update. Circulation, 131: 434-441.
- Nahar, J., T. Imam, K.S. Tickle and Y.P.P. Chen, 2013. Computational intelligence for heart disease diagnosis: A medical knowledge driven approach. Expert Syst. Appl., 40: 96-104.
- Panase, R. and V. Kathuria, 2016. Role of policy in deployment of wind energy: Evidence across states of India. Renewable Sustainable Energy Rev., 53: 422-432.
- Powers, D.M.W., 2011. Evaluation: From precision, recall and F-measure to ROC, informedness, markedness and correlation. J. Mach. Learn. Technol., 2: 37-63.
- Purusothaman, G. and P. Krishnakumari, 2015. A survey of data mining techniques on risk prediction: Heart disease. Indian J. Sci. Technol., Vol. 8, 10.17485/ijst/2015/v8i12/58385
- Sanap, S.A. and S.H. Ishtake, 2013. Intelligent heart disease prediction system using data mining techniques. Intl. J. Healthcare Biomed. Res., 1: 94-101.
- Sim, N.L., P. Kumar, J. Hu, S. Henikoff and G. Schneider *et al.*, 2012. SIFT web server: Predicting effects of amino acid substitutions on proteins. Nucleic Acids Res., 40: W452-W457.
- Song, L., A. Smola, A. Gretton, J. Bedo and K. Borgwardt, 2012. Feature selection via dependence maximization. J. Mach. Learn. Res., 13: 1393-1434.
- Soni, J., U. Ansari, D. Sharma and S. Soni, 2011. Predictive data mining for medical diagnosis: An overview of heart disease prediction. Int. J. Comput. Appl., 17: 43-48.
- Sucar, L.E., C. Bielza, E.F. Morales, P. Hernandez-Leal and J.H. Zaragoza, 2014. Multi-label classification with Bayesian network-based chain classifiers. Pattern Recognit. Lett., 41: 14-22.
- Wan, C., 2016. Novel hierarchical feature selection algorithms for predicting gene's aging-related function. AI. Matters, 2: 23-24.