

Significant Feature Selection Method for Health Domain using Computational Intelligence: A Case Study for Heart Disease

M. D. Ananda Raj

Research scholar, Bharathiyar University,
Coimbatore, TamilNadu, India
anandmd@rediffmail.com

Dr. P. Suresh

HOD, Department of Computer Science,
Sowdeswari College,
Salem, TamilNadu, India
sur_bhoo71@rediffmail.com

Abstract—In the medical field, the diagnosing of cardiovascular disease is that the most troublesome task. The diagnosis of heart disease is difficult as a decision relied on grouping of large clinical and pathological data. Due to this complication, the interest increased in a very vital quantity between the researchers and clinical professionals regarding the economical and correct heart disease prediction. In case of heart disease, the correct diagnosis in early stage is important as time is the very important factor. Heart disease is the principal supply of deaths widespread, and the prediction of Heart Disease is significant at an untimely phase. Machine learning in recent years has been the evolving, reliable and supporting tools in medical domain and has provided the best support for predicting disease with correct case of training and testing. The main idea behind this work is to find relevant heart disease feature among the large number of feature using rough computational Intelligence approach. The proposed feature selection approach performance is better than traditional feature selection approaches. The performances of the rough computation approach is tested with different heart disease data sets and validated with real-time data sets.

Keywords—heart disease, prediction, machine learning, feature selection

I. INTRODUCTION

Heart disease (cardiovascular disease) is a disease involving heart or blood vessels. According to WHO, heart disease is one of the leading causes for the universal death. According to a recent study, An estimated 17.7 million people died from CVDs in 2015, representing 31% of all worldwide deaths and If current trends are allowed to continue, 23.6 million people will die from heart disease in coming 2030 [1]. The diagnosis of Heart disease in most cases depends on a complex combination of clinical and pathological data. Because of this complication, there exists a significant amount of interest among clinical professionals and researchers regarding the efficient and accurate prediction of Heart disease. According to the statistical data from WHO, one third population worldwide died from Heart disease; Heart disease is found to be the leading cause of death globally. Computational biology is often applied through the process of translating biological knowledge into clinical practice, as well as in the understanding of biological phenomena from the clinical data. This process involves the development of a predictive model and the integration of distinctive types of data and knowledge for diagnostic purposes. Furthermore, this process requires the design and combination of distinct methodologies from statistical analysis and data mining.

Heart disease prediction system can assist medical professionals in predicting the state of Heart, based on the clinical data of patients provided into the system. Doctors may sometimes fail to take accurate decisions while diagnosing the Heart disease of a patient, therefore, Heart disease prediction systems, which use machine learning algorithms assist in such cases to get accurate results. Early detection and correct diagnosis of heart disease are needed using appropriate counseling and medicines. Machine learning approaches have

the potential to generate a knowledge-rich environment which can help to significantly help in Medical diagnosis [14]. Various supervised machine learning approaches like Decision tree, SVM, Naive Bayes, Random forest and neural network can be used for prediction of heart disease, and all need to be analyzed in terms of performance of heart disease prediction. The main risk features of Heart disease include diabetes, family history of Heart disease, smoking, obesity, high LDL cholesterol and low HDL cholesterol. Finding the key features which causes the heart disease is the key challenge in health domain. The idea of this research work is to identify the significant features which are used to find the heart disease to resolve the heart attack problem for healthy society.

Effective and efficient automated heart disease prediction systems can be beneficial in the healthcare sector for heart disease prediction. This automation will also reduce the number of tests to be taken by a patient. Hence, it will not only save cost but also the time for both, Doctors and patients. Good heart disease prediction models always depend on the optimal feature selection method used to find the key features which cause the heart disease. In the existing studies, principle component analysis, wavelet transformation, single value decomposition method, Linear discriminate analysis, Minimum noise ratio methods were used for key feature selections from the large dimensional textile data. To overcome the drawback of existing methods, Rough Computational Intelligence based Significant Attribute Selection (RCISAS) algorithm is proposed to find the important attributes which play the key role in predicting heart disease. The importance of the attributes is identified using the rough set soft computing approach.

II. BACKGROUND

The “medicine of the future” for cardiovascular disease prevention and prediction was proposed using ubiquitous computing and machine learning techniques. This process done through 5 different stages. In sensor stage, the ECG signal is passed through an ADC and record into memory. In the second-stage, preprocessing is done by noise removal and then passed to annotation stage. In Annotation stage it found the fiducial points from the ECG signal. Then features are selected from ECG and then applied machine learning algorithms. ANN gives 90% of sensitivity and 90% specificity [1]. Principal component analysis (PCA) framework [2] was proposed for feature extraction and information gain ratio for feature selection. Comparing with the other scoring function like Euclidean distance and Pearson correlation coefficients, the performance of the system is improved. This framework helps in the prediction of the heart disease by showing the reduced test attributed needed for the diagnosis of the heart diseases. Information gain [3] was used to filter features which do not contribute much in the heart disease classification task. Classification is done using ANN. This research showed that the feature selection helps in increasing computational efficiency while improving classification accuracy.

Identification of the heart cycles [4] are the important step in heart beats identification. The heart cycle features are extracted in various analysis domains, such as time, frequency, time-frequency, nonlinear high order statistics, state space, etc and these parameters were utilized to construct a SVM classifier for heart beats identification. The proposed feature set has the ability to capture the discriminative morphological and spectral characteristics of heart murmurs. These results are reached with a smaller complexity, since only 10 features are applied. Associative classification model [5] was proposed for heart disease prediction using subset of features. This method removes irrelevant, redundant attributes and generates compact rule set. These generated rules were used to build classifier. Feature selection measures were used to determine the attributes which contribute more towards the prediction of heart disease. This approach indirectly reduces the no. of diagnosis tests which are needed to be taken by a patient. A GUI based system was developed for faster prediction of heart disease. This system shows the diagnosis for given patient which requires accurate information on symptoms and lab report of present and past body conditions. BPNN [6] was used to improve the performance of the prediction system. The proposed system with ANN and five-fold cross validation provides 83% classification accuracy.

Statistics, social network and RAD [7] techniques were used to perform feature selection for heart disease. The symptoms to the syndromes were used for classification. The experimental result shows that the prediction accuracy was improved, and also the interpretability is enhanced. Comparing with minimum redundancy maximum relevance (MRMR) feature selection method, RAD gets better balance between the positive and negative classes. Measure was proposed to correctly classify PVC beats. The accuracy, sensitivity, and specificity were calculated in order to compare the training algorithms. k-NN, NN and SVM classification methods was used with different schemes. According to the experiments k-NN classification algorithm using GA [8] features

accomplished the best performance. This study showed that high classification accuracy can be achieved without implementing any feature extraction method and by using time series of the signal for input. GA can be used to reduce the size of the input vectors representing the data. An abnormal heart sounds detection method was proposed based on scaled time-frequency representation (TFR) and feature selection. The sizes of TFRs [9] are scaled to the same size through the interpolation algorithm which enables the direct comparison. Then they compared the performance of different TFRs. STFT achieved the best overall score 85.40% after incorporation of feature selection. Relief ranking method [10] was used to get a relevant set of features and SBS is used to achieve a set of non-redundant features. After applying the Relief, all the features are ranked. Then classification accuracy is measured with the help of Support Vector Machine (SVM) classifier. The proposed method uses a combination of a relevancy based feature selection method and a correlation based feature selection method. The accuracy of the ensemble proposed feature selection method is found to be greater than accuracy of both the methods (Relief and SBS) individually.

To identify the best feature and for the removal of redundant and irrelevant attributes, ranking algorithm [11] was proposed. SVM turned out to be the best classifier after performing feature selection, but before attribute selection Naïve Bayes gave best result. Area under ROC curve analysis showed better improvements after feature selection in all the three classifiers. A hybrid approach of feature selection [12] was used to optimize the classification problem. Consolidated results of SVM-RFE and gain-ratio were used to get subset of features and remove irrelevant or redundant feature. On subset of features naïve bayes and random forest are applied to classify them into presence or absence of disease. It has been shown in the results that accuracy improved for both classifiers when applied to selected features. Proposed feature selection approach reduced the size of dataset and enhanced the performance of both the classifiers models. Machine learning [13] techniques were applied to predict how much a heart patient follows medication. They have used eleven classification algorithms for prediction. Those are Random Forest (RF), Logistics Model trees (LMT), Random Tree (RT), J48, SVM, Rotation Forest (RF), Radial Basis Function (RBF), Naive bayes, Bayesian Network (Bayesnet), multilayer perceptron (MLP), simple classification and regression tree (CART). In LMT they have combined tree induction classification and linear logistics regression. Using logistics regression they have generated class estimate along with the classification. J48 algorithm results in decision tree where each node is splitted using attribute with highest information gain and then repeat this for smaller sub list.

Dynamic Bayesian network [14] with temporal abstraction was presented for prognosis of coronary heart disease. This model used incomplete data values for prognosis of CHD. Four techniques named random oversampling, SMOTE-N, a combination of SMOTE-N with clustering under sampling, and a combination of random oversampling with clustering under sampling used for resampling of imbalance in training data. Shilaskar et al. [15] identified a heart disease diagnosis system consisting of feature selection and classification techniques. SVM classifier of forwarding feature inclusion combined with back-elimination and forward feature selection

for three datasets namely Arrhythmia, and heart disease datasets and ECG datasets. Experiential results trot out that the feature selections enhanced classification techniques accurately and scaled down the number of input variables. Wang et al[16] proposed a heart disease prediction system which identifies the risk of cardiovascular diseases in adults using naive Bayes classifier. First, it analyses the medical records and identifies the probability of heart disease. This work mainly focuses on improvement of sensitivity, accuracy and specificity measures of the heart disease classification and prediction process. It focuses on primary factors of heart disease such as diabetes, blood lipid levels, the functioning of kidney and arteries for earlier detection of the risk factors. Research community proposed various machine learning techniques for the development of heart disease prediction model and discussed the importance of feature selections methods to improvise the accuracy of prediction models. In this paper, rough computation model is proposed for signification feature selection to improvise the prediction model.

III. COMPUTATIONAL INTELLIGENCE APPROACH FOR FEATURE SELECTION

Computational Intelligence (CI) is the theory, design, application and development of biologically and linguistically motivated computational paradigms. Traditionally the three four pillars of CI have been Rough Set, Neural Networks, Fuzzy Systems and Evolutionary Computation. CI plays a major role in developing successful intelligent systems, including medical and cognitive developmental systems. The idea of this research work is develop good feature selection model using rough set approach.

A. Overview of Rough Set

Soft Computing plays major role in scientific applications. Rough set and fuzzy are the two hands of soft computing. Today world is driven by data. Data are generated and collected using sensors, cameras and smart devices. The presence of vagueness, incompleteness, and granularity in the generated data gives unreliable solution in the data analysis. Managing these data is a challenging task for the users as well as researchers. Various methods have been developed by the experts to address this issue. Fuzzy Set (Zadeh, 1965) and Rough Set (Pawlak, 1982) Theory are two important models to handle imperfect knowledge. Several innovative techniques have been proposed to address the problem in the field of uncertainty and knowledge discovery, which are based on Rough set theory. These techniques include frameworks for the representation of uncertainty. Rough set theory is one of the important non-statistical approaches in data analysis.

The presence of imperfect knowledge in any information system is a vital issue for a researcher.

Over the years, Rough set theory (Pawlak, 1982 and Pawlak, 1991) has become a precious tool for the motion of various problems, such as representation of uncertain or imprecise knowledge; evaluation of quality and availability of information with respect to consistency; reasoning based on uncertainty and knowledge analysis. The primary notions of the theory of rough sets are the approximation space and lower and upper approximations of a set. The approximation space is a classification of the domain of interest into disjoint

categories as shown in figure1. The classification formally represents our knowledge about the domain.

Let U_s be the universe and let $RS \subseteq U_s \times U_s$ be an equivalence relation on U_s , called an indiscernibility relation. The pair $A = (U_s, RS)$ is called an approximation space. The lower and upper approximation of set X with respect to RS can be written as:

$$\underline{RS}(X) = \{x \in U_s : [x]_{RS} \subseteq X\} \text{ -----(1)}$$

$$\overline{RS}(X) = \{x \in U_s : [x]_{RS} \cap X \neq \emptyset\} \text{ , -----(2)}$$

where $[x]_{RS} = \{y \in U_s \mid x R y\}$ is the equivalence class of x . if $[x]_{RS} \subseteq X$, Then it is certain that $x \in X$. if $[x]_{RS} \subseteq X - \overline{RS}(X)$ then it is clear that $x \notin X$. $[x]_{RS} \subseteq X$ is said to be rough with respect to RS if $\underline{RS}(X) \neq \overline{RS}(X)$. Otherwise X is RL-discernible.

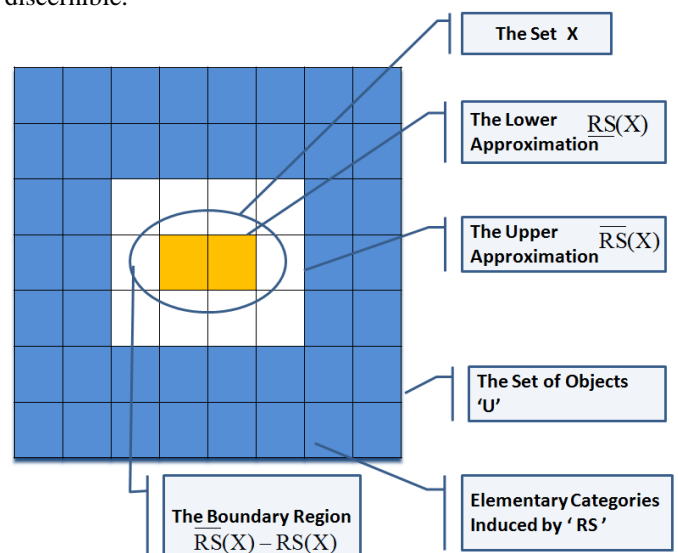


Figure 1 Representation of Rough Set

Rough set theory is a precious tool for data mining. In the past few years, the concept of basic rough set has been extended in many different directions. Generalization of rough set theory can be done mainly in three ways, namely, set-theoretic framework using non-equivalence binary relations; granule based definition using coverings and subsystems based definition using other subsystems.

B. Rough Set Attribute Dependency

Attribute Dependency in rough set can be interpreted as the combination of objects in knowing the values of attributes. This checks the degree of dependency between the two attributes. Every attribute produces the indiscernibility among the given two rough sets. Indiscernibility relation concept in the rough sets defines the association between the set of attributes. The same indiscernible objects can be represented redundantly. A

set of attributes A depends on set of attributes B is given as follows:

$$A \Rightarrow B \quad \text{-----}(3)$$

Where, the attributes of A are uniquely determined by the values of attributes of B. A completely depends on the values of B if there exists in the functional dependency relationship between them. The functional dependency depicts that both the A and B are subsets with each other. Suppose if B depends completely on C then with a degree of k then it can be as $k(0 \leq k \leq 1)$ denoted by the positive region,

$$A \Rightarrow_k B \quad \text{-----}(4)$$

Where the positive region c is given as,

$$POS_R(A) = \bigcup_{B \in I / Ind A} RB \quad \text{-----}(5)$$

So for all the elements in the universe, I in terms of A and B can be given as,

$$\gamma(A, B) = |P OSA(B)|/|I| \quad \text{-----}(6)$$

Here the dependency of B on A depends on the value of k. if $k=1$ then we can say that B depends on A completely and if $k < 1$ then it states that B depends on A but partially. The k expresses the blocks of I/B with respect to A and it is termed as degree of dependency.

C. Significant Features Selection using Rough Set

The significant attributes are determined using the concept of reducts and core in rough set theory. Due to this indiscernibility relation, it becomes ease in finding the redundant values or the redundant attributes of a set. The several set approximation subsets of attributes, which exists in minimal are termed as reducts. The set of all the conditional attributes of set approximations that exists as a set is defined as a core, and in simple, it is the intersection of all the reducts to a set or a system considered. For example, let T be a set of attributes, the attribute R be the subset of attribute C then it is depicted as below:

$T = (U, R, D)$ // independent of C Then the positive relation is given as,

$$POS_R(D) = POS_C(D) \quad \text{-----}(7)$$

$$\text{i.e. CORE}(C) = \cap \text{RED}(C) \quad \text{-----}(8)$$

where, CORE (C) defines all the conditional attributes, and RED (C) defines all the set of reducts of attribute C. One of the approaches for these reducts or the conditional attributes is done dynamically and depicted by the decision tables. In these decision tables, the attributes are classified in two ways as relevant and often. When the attributes are more often repeated it is given as majority or relevant and the set of attributes appearing as common to the original sets in decision tables are given through as often. Based on the rough set theory concepts

core and reduct, the Rough Computational Intelligence based Significant Attribute Selection (RCISAS) algorithm is proposed.

Algorithm: Rough Computational Intelligence based Significant Attribute Selection (RCISAS)

Input Att: of attributes, n: number of attributes

Output : Optimal Attributes

1. Calculate approximation values for all attributes using eq 16
 $\mu(P, Z)(a) = (\text{Card}(\text{POSP}(Z)) - \text{Card}(\text{POSP}(a)(Z)) / \text{Card}(U)$ (16)
2. Compare the approximation values of each attribute a_j ,
3. If both attributes are highly associated, remove the less approximation value Attribute(s).
4. Use the selected significant attribute(s) value as input to heart disease prediction model

The concept of significant attributes can be given as elimination of unnecessary information from the decision table or information table without lose in any other data from the table. So this reduction of redundant attributes is generalized by significance of attributes. To discover the significance of attributes, the attributes need to be evaluated initially. So for this evaluation dispensability and indispensability is used. This can also be given by a very closed interval [0,1]. The process of obtaining significant attributes in a decision table is processed by eliminating the attributes from the attribute set.

For a set considered as $\gamma(R, S)$ let the attribute be a in a set. And when the attribute a is removed from the set $\gamma(R, S)$ then it can be given as,

$$\gamma((R \Leftarrow \{a\}, S)) \quad \text{-----}(9)$$

Then by the above conditions and processes the significance of attributes can be given as by normalizing the basic difference in between the coefficient and the set obtained after removing the attribute i.e. $\gamma(R, S)$ and $\gamma((R \Leftarrow \{a\}, S))$ is given as below:

$$\sigma(R, S)(a) = (\gamma(R, S)) \Leftarrow (\gamma(R \Leftarrow \{a\}, S)) / (\gamma(R, S)) = 1 \quad \text{----}(10)$$

$$\sigma(R, S)(a) = (\gamma(R \Leftarrow \{a\}, S)) / (\gamma(R, S)) \quad \text{-----}(11)$$

Thus here, The coefficient $\sigma(a)$ is termed as error of classification. This error of classification in general occurs when the attribute is removed from set considered. And so the significance of the attributes can be protracted by the other remaining attributes of a set and can be given as,

$$\sigma(R, S)(B) = (\gamma(R, S)) \Leftarrow (\gamma(R \Leftarrow B, S)) / (\gamma(R, S)) = 1 \quad \text{--}(12)$$

$$\sigma(R, S)(B) = (\gamma(R \Leftarrow B, S)) / (\gamma(R, S)) \quad \text{-----}(13)$$

Here, $\sigma(B)$ is given as the coefficient obtained from the extension of an attribute significance. Also B is considered as subset of R i.e. B is reduct of the set of attributes in R. The

attribute of any subsets B and R is deliberated as the reduct of R and so after removing the attribute this can be given as,

$$\sigma(R,S)(B) = (\gamma(R, S)) \Leftrightarrow \gamma(B, S)/(\gamma(R, S)) = 1 \text{ ---(14)}$$

$$\sigma(R,S)(B) = (\gamma(B, S))/(\gamma(R, S)) \text{ -----(15)}$$

Thus, $\sigma(R, S)$ is defined as the reduct approximation or error of reduct approximation which depicts the significance of attributes of B relatively in the set R. The minimum error of approximation results in the increase of accuracy in a set through classification approach.

The proposed the Rough Computational Intelligence based Significant Attribute Selection (RCISAS) algorithm is applied on heart disease data sets to find the most significant features which causes heart diseases in health domain.

IV. EXPERIMENT AND RESULT ANALYSIS

The main idea behind this work is to find relevant heart disease feature among the large number of feature using rough computational Intelligence approach. The Rough Computational Intelligence based Significant Attribute Selection (RCISAS) algorithm is proposed. The performance of the algorithm is tested with important heart disease data sets Cleveland, and Hungarian. The dataset used in this work is obtained from UCI machine learning repository [20]. Heart Disease database contains 76 features. The data set consist many missing values and noisy. The data preprocessing is applied and removed the missing values present in the data set. The idea of the proposed algorithm is to find the most significant features which are mainly considered for the prediction of heart disease among the 76 features. Sample data set with the list of attributes and their values are shown in table 1 and the number of instances available for each of the datasets is shown in TABLE 2

Sr No.			Description
1	Age	Continuous	Age in years
2	Sex	Discrete	value 1: Male value 0: Female
3	Chest Pain(C P)	Discrete	1 = typical angina 2 = atypical angina 3 = non-anginal pain 4 = asymptomatic
4	Trestbps	Continuous	Resting blood pressure (in mmHg)
5	Htn	Discrete	1=presence 0=absence
6	Chol	Continuous	Serum cholesterol in mg/dl
7	Fbs	Discrete	Fasting blood sugar \geq 120 mg/dl : 1 = true 0 = false

8	Fmchis	Discrete	1=presence 0=absence
9	Restecg	Discrete	Resting result: 0 = normal 1 = having ST-T abnormality 2 = showing probable or define left ventricular hypertrophy by Estes criteria
10	Thalach	Continuous	Maximum heart rate achieved
11	Exang	Discrete	Exercise induced angina: 1 = yes, 0 = no
12	Oldpeak	Continuous	Depression induced by exercise relative to rest
13	Slope	Discrete	The slope of the peak exercise segment: 1 = up sloping 2 = flat 3 = down sloping
14	Ca	Discrete	Number of major vessels colored by fluoroscopy that ranged between 0 to 3
15	Thal	Discrete	3 = normal 6 = fixed defect 7 = reversible defect
16	num	Discrete	0 = healthy 1 = patient who is subject to possible heart disease(1,2,3,4)

Table.1 Sample Heart Disease Data Set

	Cleveland	Hungarian
Total number Of instance	287	283
Class 0	155	180
Class 1	49	35
Class 2	30	26
Class 3	32	27
Class 4	12	15

Table.2 Number of Instance in Heart Disease Data Set

The Rough Computational Intelligence based Significant Attribute Selection (RCISAS) algorithm is developed using mat lab for finding the significant attributes among the various conditional attributes using rough set

techniques. The attributes are to be classified based on the attribute significant value which we obtained through the computational program. Table 3, shows the attribute significant values of Hungarian data set.

Feature	RCISAS Significance Value
Age	0.367347
Sex	0.321639
Cp	0.353741
trestbps	0.369048
chol	0.355442
fbs	0.479592
restecg	0.377551
thalach	0.348639
Exang	0.688776
Oldpeak	0.369048
Slope	0.367347
Ca	0.380952
Thal	0.37585

Table 3. Significance Value of Hungarian Data Set

The figure 2 shows the graphical representation of each heart disease features significance value with respect to Hungarian data sets. The graph shows the contribution of each attribute in the determination of heart disease. The exang (exercise induced angina) and fbs (fasting blood sugar) plays major contribution than other features

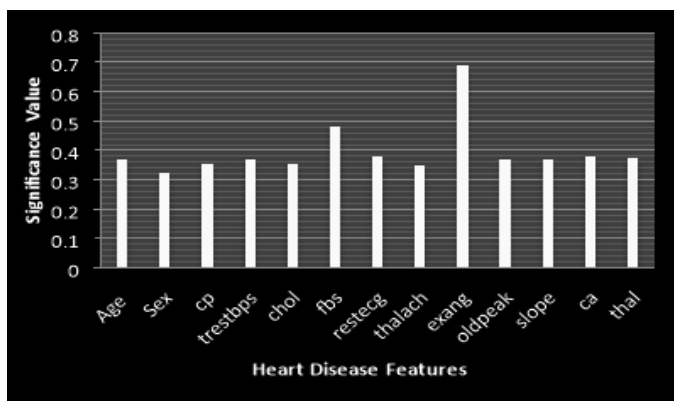


Figure 2. Hungarian Heart disease data set features with significance value.

Table 4, shows the attribute significant values of Cleveland data set.

Feature	RCISAS Significance Value
Age	0.30363
Sex	0.28563
Cp	0.29868
Trestbps	0.306931

chol	0.29538
fbs	0.356436
restecg	0.326733
thalach	0.306931
exang	0.330033
oldpeak	0.321782
slope	0.326733
ca	0.331683
thal	0.325083

Table 4. Significance Value of Cleveland Data Set

The figure 3 shows the graphical representation of each heart disease features significance value with respect to Cleveland data sets. The graph shows the contribution of each attribute in the determination of heart disease. The fbs (fasting blood sugar), exang (exercise induced angina), ca (number of major vessels) and restecg (resting electrocardiographic results) plays major contribution in heart disease prediction than other features.

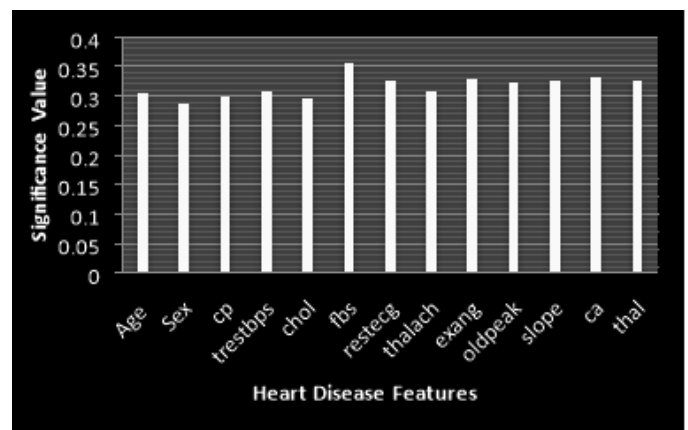


Figure 3. Cleveland Heart disease data set features with significance value

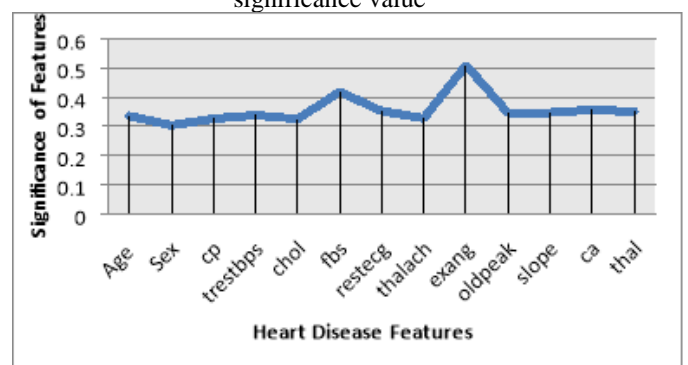


Figure 4. Significance of Features based on both the data sets

Based on the RCISAS algorithm result analysis on Hungarian and Cleveland data set, fig.5 shows that, the significant attributes, are exang(exercise induced angina),fbs(fasting blood sugar),ca (number of major vessels),restecg (resting electrocardiographic results),thal,slope (the slope of the peak exercise ST segment),oldpeak (ST depression induced by exercise relative to rest) and trestbps (resting blood pressure).The significant attribute will be supplied to the data mining prediction algorithm to predict heart disease.

The main idea of this work is to handle the problem of high dimensional data. This problem is solved by using proposed RCISAS algorithm in the health domain.

V. CONCLUSION

The diagnosis of heart disease is difficult as a decision relied on grouping of large clinical and pathological data. The main focus of this research work is to find significant heart disease feature among the large number of features using rough computational Intelligence approach. The Rough Computational Intelligence based Significant Attribute Selection (RCISAS) algorithm is proposed and the performance of the algorithm is tested with important heart disease data sets Cleaveland, and Hungarian. The proposed algorithm also compared with diverse feature selection approaches used for the heart disease prediction. Observation shows that in most of the cases RCISAS algorithm performance is high compared with other feature selection methods. The RCISAS feature selection technique suggests that the most important features for heart diseases are exang, fbs, ca, restecg, thal, slope, oldpeak and trestbps. Good feature selection algorithm will be more useful for the development of novel heart disease prediction models.

REFERENCES

- [1] S. Nikolaiev, Y. Timoshenko, "Reinvention of the Cardiovascular Diseases Prevention and Prediction Due to Ubiquitous Convergence of Mobile Apps and Machine Learning", *Information Technologies in Innovation Business (ITIB)*, 7-9 October, 2015, Kharkiv, Ukraine
- [2] R. Kavitha and E. Kannan, "An efficient framework for heart disease classification using feature extraction and feature selection technique in data mining," 2016 International Conference on Emerging Trends in Engineering, Technology and Science (ICETETS), Pudukkottai, 2016, pp. 1-5.
- [3] A. Khemphila and V. Boonjing, "Heart Disease Classification Using Neural Network and Feature Selection," 2011 21st International Conference on Systems Engineering, Las Vegas, NV, 2011, pp. 406-409.
- [4] D. Kumar, P. Carvalho, M. Antunes, R. P. Paiva and J. Henriques, "Heart murmur classification with feature selection," 2010 Annual International Conference of the IEEE Engineering in Medicine and Biology, Buenos Aires, 2010, pp. 4566-4569.
- [5] M. A. jabbar, P. Chandra and B. L. Deekshatulu, "Prediction of risk score for heart disease using associative classification and hybrid feature subset selection," 2012 12th International Conference on Intelligent Systems Design and Applications (ISDA), Kochi, 2012, pp. 628-634.
- [6] M. A. jabbar, P. Chandra and B. L. Deekshatulu, "Prediction of risk score for heart disease using associative classification and hybrid feature subset selection," 2012 12th International Conference on Intelligent Systems Design and Applications (ISDA), Kochi, 2012, pp. 628-634.
- [7] Y. Wang, G. Li, W. Xu, G. Liu and Y. Wang, "Symptom selection of inquiry diagnosis data for coronary heart disease in Traditional Chinese Medicine by using social network techniques," 2010 IEEE International Conference on Bioinformatics and Biomedicine Workshops (BIBMW), Hong Kong, 2010, pp. 785-789.
- [8] Y. Kaya and H. Pehlivan, "Feature selection using genetic algorithms for premature ventricular contraction classification," 2015 9th International Conference on Electrical and Electronics Engineering (ELECO), Bursa, 2015, pp. 1229-1232.
- [9] W. Zhang, J. Han and S. Deng, "Abnormal heart sounds detection based on the scaled time-frequency representation and feature selection," 2016 Computing in Cardiology Conference (CinC), Vancouver, BC, 2016, pp. 1177-1180.
- [10] S. Tiwari and B. Singh, "A hybrid approach for feature selection," 2015 Third International Conference on Image Information Processing (ICIIP), Wagnaghat, 2015, pp. 277-280.
- [11] J. S. A. Sabab, M. A. R. Munshi, A. I. Pritom and Shihabuzzaman, "Cardiovascular disease prognosis using effective classification and feature selection technique," 2016 International Conference on Medical Engineering, Health Informatics and Technology (MediTec), Dhaka, 2016, pp. 1-6.
- [12] K. Pahwa and R. Kumar, "Prediction of heart disease using hybrid technique for selecting features," 2017 4th IEEE Uttar Pradesh Section International Conference on Electrical, Computer and Electronics (UPCON), Mathura, 2017, pp. 500-504.
- [13] Karanasiou GS, Tripoliti EE, Papadopoulos TG, et al. "Predicting adherence of patients with HF through machine learning techniques," 2016 Healthcare Technology Letters, 2016, pp.165-170.
- [14] KaliaOrphanou, Athena Stassopoulou, and ElpidaKeravnou, "DBN-Extended: A Dynamic Bayesian Network Model Extended With Temporal Abstractions for Coronary Heart Disease Prognosis", IEEE JOURNAL OF BIOMEDICAL AND HEALTH INFORMATICS, VOL. 20, NO. 3, MAY 2016
- [15] Shilaskar, S. and Ghatol, A. (2013), 'Expert systems with applications feature selection for medical diagnosis : evaluation for cardiovascular Diseases', Journal of expert sy stem with application 4(10), 4146–4153.
- [16] Wang (2014), 'Study on qi-deficiency syndrome identification modes of Coronary heart disease based on metabolomic biomarkers', Journal of evidence-based complementary and alternative medicine 24(16), 192–198.
- [17] Pawlak, Z. "Rough Sets", Int. Jour. Inf. Comp. Sc., vol.11, pp.341-356, 1982
- [18] Pawlak, Z. "Rough Sets: Theoretical Aspects of Reasoning about Data", Kluwer Academic Publishers, 1991.
- [19] L.A. Zadeh, Fuzzy sets, Information and Control, Vol 8, Issue 3, pp.338-353, 1965
- [20] Heart-disease data," 2017, \url{https://archive.ics.uci.edu/ml/datasets/Heart+Disease}