FEATURE SELECTION METHODS FOR CLASSIFICATION – A COMPARISON

Kavitha C.R[1], Mahalekshmi T[2]

[1]*Research Scholar, R&D, Bharathiar University, Coimbatore, Tamil Nadu, India*
[2]*Principal, Sree Narayana Institute of Technology, Kollam, Kerala, India*

Abstract

Feature selection refers to the process of choosing the most significant features for a given task, while discarding the noisy, irrelevant and redundant features of the data set. These noisy feature set might mislead the classifier. Feature selection technique reduces the dimensionality of the feature set of the data set. The main aim of this work is to perform binary and multiclass classification more accurately using reduced number of attributes. This paper proposes two different feature selection methods. The first feature selection method is done using Information Gain and Forward Selection ($IG_{fw}S$). The second feature selection is performed using Recursive Feature Elimination with SVM **(***SVMRFE). Then rough set theory was applied to both the feature selection methods to obtain hybrid feature selection methods RST+SVMRFE and RST+ $IG_{fw}S$. Further, a comparative study of all the four feature selection methods was performed. From the results of the study, it is found that the feature selection is a very important data mining technique which helps to achieve the good classification accuracy with the reduced number of attributes. Based on the comparative analysis conducted the feature selection methods SVMRFE and RST+SVMRFE shows better performance than other feature selection methods considered under the study. And the random forest classifier achieves the maximum accuracy with all the datasets on which SVMRFE and RST+SVMRFE feature selection methods were applied.*

Keywords: *Feature Selection, Classification, Rough Set, Information Gain, Ranking,*

---***---

1. INTRODUCTION

Feature Selection is a pre-processing technique which helps in removing the irrelevant features from the dataset which contains too many features. Feature selection helps the data mining algorithm zero in on relevant features so that the hypothesis space can be reduced. [1] Basically the feature selection is done in two ways; one is to rank the features based on certain criteria and the top K features are selected and the other is to select a minimal feature subset without any decrease of the learning performance.

Filter, Wrapper and Embedded approaches are used for feature selection. In filter approach the feature selection is performed without considering the classification algorithm that will be applied to the selected attributes.[2] Here a subset of attributes that preserves the possible relevant information found in the entire set of attributes is selected. [3] In wrapper approach feature selection is performed by taking into account the classification algorithm that will be applied to the selected attributes. Here an optimized subset of attributes for a given classification algorithm is selected. [3] The embedded approach incorporates variable selection as a part of model fitting and the selection technique is specific to the model. The external search algorithms that are used in the filter and wrapper approaches cannot cover all possible variable combinations, excluding problems with only a few variables. Thereby, their solutions are likely to be suboptimal. [4]

This paper is organized as follows. The data sets used for the study is described in section 2. The literature review that are relevant to the work are presented in section 3 followed by the description of the experiment in section 4. The Results and discussions are stated in section 5. And the conclusion is given in section 6 followed by the references.

2. DATASET

Four datasets were used for conducting the experiment. They are hypothyroid (HT) dataset, Wisconsin Breast Cancer (BC) dataset, Dermatology (DT) Dataset and Liver Disorder (LD) dataset. These datasets were downloaded from UCI repository ("http://mlearn.ics.uci.edu/ ML Repository.html"). [5] The characteristics of the four data sets are summarized in the Table 1. The HT dataset contains 500 observations and 29 variables with missing values reported. The BC dataset contains 569 instances (357 benign – 212 malignant), where each one represents FNA test measurements for one diagnosis case. [6] DT dataset contains 34 attributes, 366 instances and with missing values. LD dataset contains 345 instances without any missing values.

Table-1: Data Set's Characteristics

Data Set	HT	BC	DT	LD
No of Example	500	569	366	345
Input Attributes	28	31	33	6
Output Classes	3	2	6	2
Total No. of Attributes	29	32	34	7

3. LITERATURE REVIEW

3.1 Ranking by Information Gain

It is a measure of how good an attribute is for predicting the class of each of the training data. [7] It is the reduced amount of desired information or information entropy caused by partitioning the instances according to a feature. [8] Entropy determines the randomness or impurity of a collection of instances. The entropy of the collection of instances S is given by

$$\text{Entropy}(S) = -\sum_{i=1}^{m} p_i \log_2(p_i) \qquad [9]$$

Where p_i is the probability that an arbitrary instance in S belongs to class w_i.

Information gain, gain(S, F) of a feature F, relative to a collection of examples S is defined by

$$\text{Gain}(S, F) = \text{Entropy}(S) - \sum_{j=1}^{\vartheta} \frac{|S_j|}{|S|} \times \text{Entropy}(S_j) \qquad [9]$$

Where ϑ is the set of all possible values for feature F and Sj contains instances in S that feature A has value ϑ.

3.2 Forward Selection

In forward selection approach, each attribute is added to the model one at a time. At each step, each feature that is not already in the model is tested for inclusion in the model. Those variables are added to the model if the P-value is equal or above the predetermined level.

The forward selection approach starts with an empty set of attributes and it adds each unused attribute of the given input data set. The performance is estimated for each added attribute using 10 fold cross validations. The attribute which shows the highest performance is only added to the selection. Then next iteration is initiated with the modified selection. [10]

3.3 Support Vector Machine—Recursive Feature Elimination (SVMRFE)

SVM-RFE [11] is a feature selection approach in which feature selection begins with all the features and eliminates one feature at a time in a sequential backward elimination manner. SVM-RFE can be used with binary and multi class problems. The squared coefficients (j = 1, … , p) of the weight vector **w** obtained from binary problem are employed as feature ranking criteria. Those features having the largest weights are considered the most informative. In SVM-RFE an iterative procedure, the SVM classifier is trained, ranks for all features are calculated, and the feature with the smallest rank is discarded. The procedure is repeated until a small subset of attributes is obtained. **[12]**

3.4 RST

Rough set theory (RST) developed by Pawlak in the1980s, is used for data analysis, discovering inter data relationships, finding interesting patterns and decision making. [13] In RST, an information system (IS) is defined as a system IS = (U, A, V_a) where U ={x_1, x_2,….x_n}, A is the set of attributes a: U $\rightarrow V_a$ and V_a is a set of values for the attribute a. [14] The attributes set can be divided into two subsets C and D which are subsets of A. C and D represents the condition and decision attributes. [15] [16] [17]

Let B⊂A be a subset of attributes, the indiscernibility relation Ind(B) is defined as IND(B)={(x,y) /(x,y)∈UxU, a(x)=a(y), ∀a∈B} where a(x) is the value of an attribute a of object x. x and y are said to be indiscernible with respect to a if a(x)=a(y). [16]

For any concept X$\subseteq U$, the attribute subset P $\subseteq A$, X could be approximated by the P-Upper and P-Lower approximation using the knowledge of P. The lower approximation of X is the set of objects of U that are surely in X ,where as the upper approximation of X is the set of objects of U that are possibly in X. The upper and Lower approximations are defined as follows P*(X) = $x\epsilon U\ \{P\ x : P(x) \subseteq X\}$ & P*(X) = $x\epsilon U\ \{P\ x : P(x) \cap X \neq \emptyset\}$. The boundary region is defined as: BNP (X) = P*(X) – P*(X). [18]

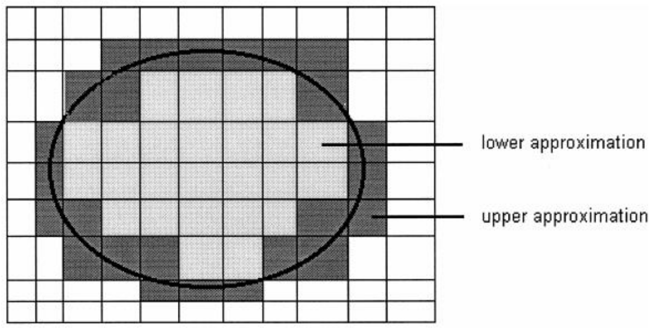

Fig-1: Schematic representation of the upper and lower approximation of set X. [19]

In an information system, the reduct is an essential part of an IS which is a subset of attributes by itself that can fully characterize the knowledge in the database. There can be many subsets of attributes that preserve the equivalence-class structure. The core is the common attributes set of all reducts. The discernibility matrix is used for computing reducts and core.

4. EXPERIMENT

4.1 Data Pre-Processing

The data pre-processing is the first important step that is carried out in order to remove the missing values from the dataset. The input datasets such as hypothyroid and liver-disorder has missing values, which was removed by using the option WEKA. filters. unsupervised. attribute. Replace Missing Values in WEKA.[20] This replaces all missing values for nominal and numeric attributes in a dataset with the modes and means from the training data.

4.2 Feature Selection with Information Gain and Forward Selection

In this study, first feature selection was done using information gain ranking of the attributes along with the forward selection approach. The block diagram for the feature selection process of $IG_{fw}S$ is shown in the figure 2.

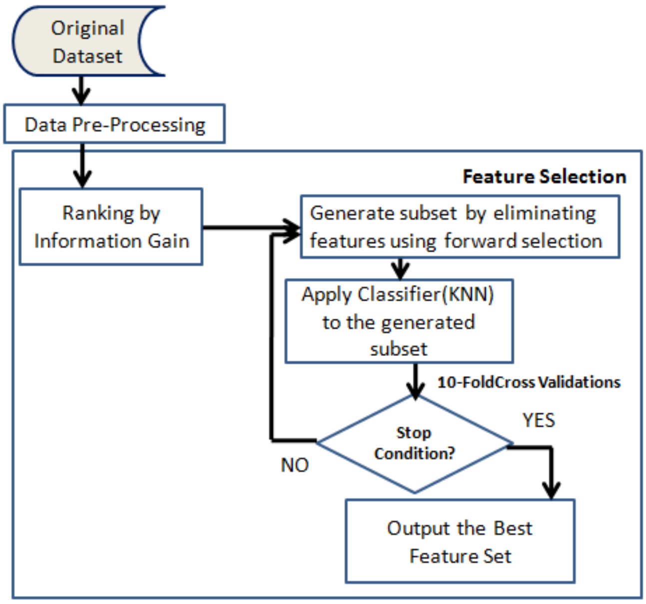

Fig-2: Block Diagram for the feature selection process of $IG_{fw}S$

All the attributes in a dataset were ranked in a descending order based on information gain. The forward selection begins with an empty selection of attributes. Then in each round it includes each unused attribute of the given input dataset and for each added attribute, the performance is estimated using KNN model [21] with 10-fold cross-validations. The attribute which gives the highest increase of performance is only added to the selection list. Then next iteration begins with the modified selection. The iteration continues until an addition of a new attribute does not increase the performance of the model.

4.3 Feature Selection with SVMRFE

The second feature selection was done using Recursive Feature Elimination with SVM [22]. The block diagram for the feature selection process of SVMRFE is given in the figure 3. SVMRFE is an iterative procedure of the backward elimination of features. The datasets is used to train the SVM classifier. The weights of all the features are calculated and ranked. The attributes are then sorted based on the weight vectors as the classification basis. The attribute having the smallest weight will be deleted in each iteration.

The iterative procedure stops when there is only one feature remaining in the dataset. The output of this process provides a list of features in the order of their weight. The algorithm will eliminate the attribute with smallest ranking weight, while retaining the attributes of significant impact. Finally, the attributes will be listed in the descending order of explanatory difference degree. The sorted feature list is used for training SVM and is evaluated based on SVM prediction accuracy to obtain the optimum feature subsets.

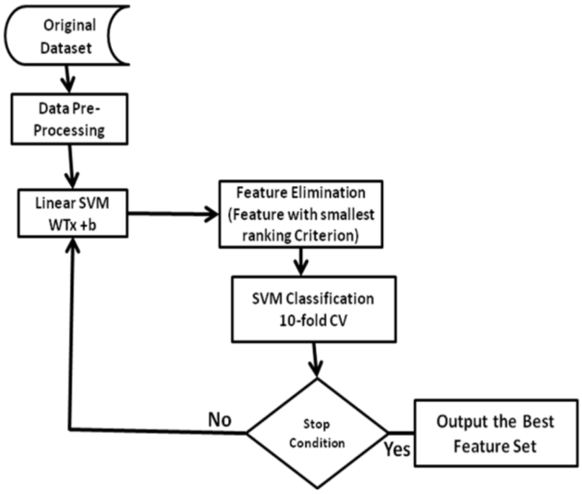

Fig-3: Block Diagram for the feature selection process of SVMRFE

4.3.1 Application of RST with ($IG_{fw}S$) and (SVMRFE)

4.3.1.1 Generation of Reducts

After preprocessing is completed, the reducts of the different data sets such as hypothyroid, Breast Cancer, Dermatology and Liver Disorder were generated using RSES 2.0 tool. Ten reducts were generated using genetic algorithm.

4.3.1.2 Selection of Reduct Occurred Attributes

All the attributes that occurred in the reducts were selected for the further feature selection process. From Table 6, it is found that from the dataset hypothyroid, 13 attributes; Breast Cancer, 16 attributes; Dermatology, 16 attributes and Liver Disorder, 7 attributes were selected from the ten reducts which were generated using genetic algorithm in RSES 2.0 tool. [23]

4.3.1.3 Feature selection Process

The feature selection process is carried out by giving the reduct occurred attributes as the input dataset. This input dataset is applied to the two Feature selection approaches **$IG_{fw}S$** and **SVMRFE.** The feature selection process using RST and **$IG_{fw}S$** is shown in the figure 4. The feature selection process using RST and SVMRFE is shown in the figure 5.

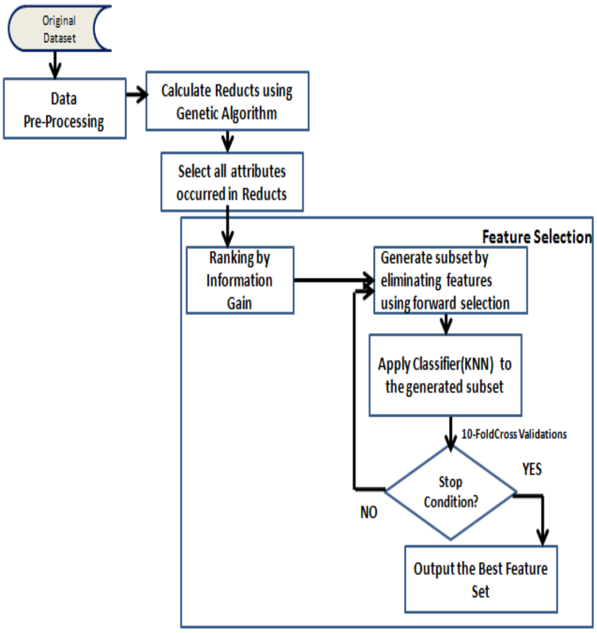

Fig-4: Block Diagram for the feature selection process using RST and $IG_{fw}S$

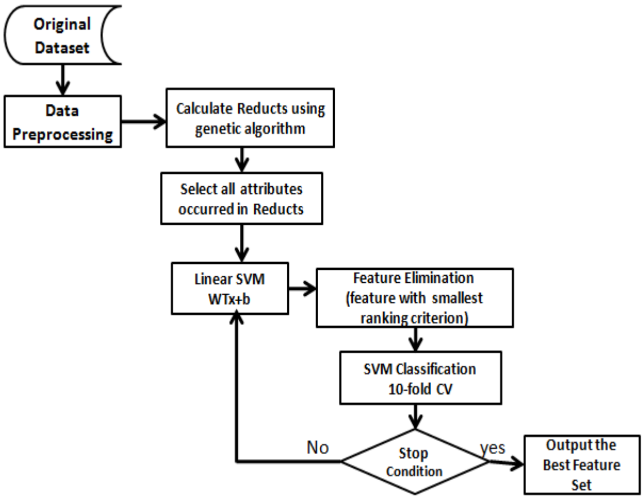

Fig-5: Block Diagram for the feature selection process using RST and SVMRFE.

5. RESULTS AND DISCUSSION

Feature selection is used for reducing the number of attributes by removing the irrelevant and noisy attributes and thereby increases the learning accuracy. In this section, the experimental result is used to make a comparison and find out the effectiveness of the feature selection approaches used in this paper. The classification accuracy was found out for the attributes selected by using the feature selection methods like $IG_{fw}S$, RST+ $IG_{fw}S$, SVMRFE and RST+SVMRFE using four classifiers like Random Forest[24], IBK[25], J48[26]and JRip [27]. Table 2 shows the details of the number of attributes that were selected after performing different feature selection methods.

From this comparative study, we can understand the importance of feature selection. It is evident that all the attributes in a dataset is not required for classification. Noisy and unimportant attributes can be removed from the data set.

5.1 Comparison of Feature Selection Methods on Different Datasets

From the experiment conducted, it is quite clear that Feature selection methods can eliminate more than 90% of the attributes without affecting the classification accuracy as shown in Table 2. By applying feature selection techniques, the number of attributes is reduced considerably.

Table -2: Details of the number of attributes of different datasets

Data Set	Attribute Count Details	Feature Selection	No of Attributes Selected
Hypothyroid	Total No of Attributes =30	$IG_{fw}S$	3
	Reduct Occurred Attributes =13	RST+$IG_{fw}S$	3
	Total No of Attributes=30	SVMRFE	10
	Reduct Occurred Attributes=13	RST+SVM RFE	8
Breast Cancer	Total No of Attributes =31	$IG_{fw}S$	4
	Reduct Occurred Attributes =16	RST+$IG_{fw}S$	3
	Total No of Attributes=31	SVMRFE	11
	Reduct Occurred Attributes=16	RST+SVM RFE	11
Dermatology	Total No of Attributes =35	$IG_{fw}S$	6
	Reduct Occurred Attributes =16	RST+$IG_{fw}S$	5
	Total No of Attributes=35	SVMRFE	11
	Reduct Occurred Attributes=16	RST+SVM RFE	11
Liver Disorder	Total No of Attributes =7	$IG_{fw}S$	3
	Reduct Occurred Attributes =7	RST+$IG_{fw}S$	3
	Total No of Attributes=7	SVMRFE	7
	Reduct Occurred Attributes=7	RST+SVM RFE	7

Table 3: Performance Analysis of Feature Selection (FS) Methods on Data Sets using Random Forest Classifier

Feature Selection Methods	BC	HT	DT	LD
RST+SVMRFE	95	94	87	63
SVMRFE	96	95	94	63
RST+ $IG_{fw}S$	90	98	83	61
$IG_{fw}S$	91	98	89	61
Before FS	96	92	95	63

Table 3 shows the performance analysis of the different feature selection methods based on accuracy on datasets such as breast cancer, hypothyroid, dermatology and Liver disorder using random forest classifier. From this table it is clear that the feature selection methods RST+SVMRFE and SVMRFE show better performance than $IG_{fw}S$ and RST+ $IG_{fw}S$. Since RST+SVMRFE achieve higher accuracy with Breast Cancer, Hypothyroid and Liver-Disorder datasets and SVMRFE achieves higher accuracy with the Hypothyroid, Dermatology and Liver-Disorder datasets.

Table 4 shows the performance analysis of the different feature selection methods based on accuracy on datasets such as breast cancer, hypothyroid, dermatology and Liver disorder using IBK classifier. With the IBK classifier, the feature selection method SVMRFE shows higher accuracy with the Breast Cancer, Dermatology and Liver Disorder datasets.

Table 4: Performance Analysis of Feature Selection Methods on Data Sets using IBK Classifier

Feature Selection Methods	BC	HT	DT	LD
RST+SVMRFE	96	99	89	70
SVMRFE	94	99	93	70
RST+ $IG_{fw}S$	92	98	84	64
$IG_{fw}S$	93	98	86	64
Before FS	95	99	98	70

Table 5: Performance Analysis of Feature Selection Methods on Data Sets using J48 Classifier

Feature Selection Methods	BC	HT	DT	LD
RST+SVMRFE	92	98	87	69
SVMRFE	92	98	89	69
RST+ $IG_{fw}S$	69	87	84	62
$IG_{fw}S$	69	89	88	62
Before FS	93	100	96	69

Table 5 shows the performance analysis of the different feature selection methods based on accuracy on datasets such as breast cancer, hypothyroid, dermatology and Liver disorder using J48 classifier. With the J48 classifier, the feature selection method SVMRFE achieves the highest accuracy with all the datasets followed by the feature selection method RST+SVMRFE.

Table 6: Performance Analysis of Feature Selection Methods on Data Sets using Jrip Classifier

Feature Selection Methods	BC	HT	DT	LD
RST+SVMRFE	93	99	84	65
SVMRFE	94	99	87	65
RST+$IG_{fw}S$	92	98	82	62
$IG_{fw}S$	93	98	89	62
Before FS	93	99	92	65

Table 6 shows the performance analysis of the different feature selection methods based on accuracy on datasets such as breast cancer, hypothyroid, dermatology and Liver disorder using Jrip classifier. In this case too, the feature selection method SVMRFE achieves the highest accuracy with all the datasets followed by the feature selection method RST+SVMRFE.

From the comparison study of the different feature selection methods as shown in the figure 11, it is found that the feature selection i.e. selection of relevant attributes and removal of unwanted attributes can achieve higher or equivalent performance accuracy with the performance accuracy of all the attributes before feature selection. From this experiment, it is found that the feature selection method SVMRFE outperforms all the other feature selection methods used in this study. The next better feature selection method is RST+SVMRFE which are almost equivalent to SVMRFE.

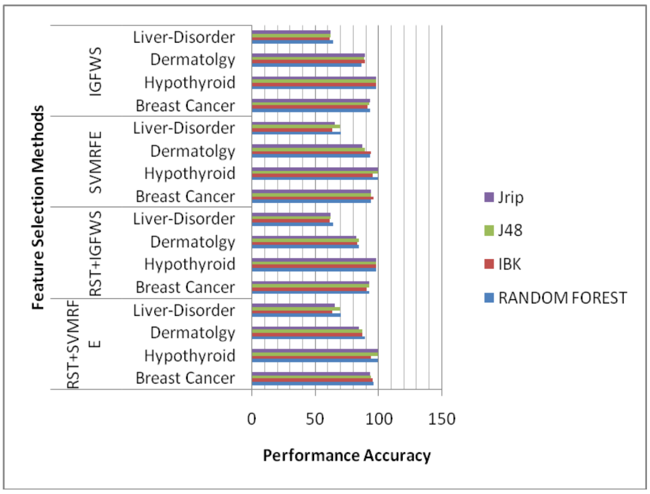

Fig 11: Graph showing Comparison of Performance Accuracy of Feature selection methods on different datasets with the 4 classifiers

5.2 Comparison of Feature Selection Methods on Different Classifiers

In this section, the four datasets after applying different feature selection methods (SVMRFE, RST+SVMRFE, $IG_{fw}S$ and RST+ $IG_{fw}S$) are used for performing classification using different classification models like random forest, IbK, J48 and JRip. The results of the classification are shown in the tables 7, 8, 9 and 10 given below.

From Table 7 we can note that even with 3 attributes out of 30 attributes from the hypothyroid dataset, we could achieve the same or improved accuracy than the accuracy obtained without applying any feature selection. With the hypothyroid dataset, random forest classifier was able to achieve 99% of accuracy with SVMRFE and RST+SVMRFE whereas the IBK Classifier was able to achieve maximum accuracy with $IG_{fw}S$ and RST+$IG_{fw}S$ and JRip classifier achieved maximum accuracy of 99% with SVMRFE and RST+SVMRFE.

Table 7: Results on Hypothyroid Dataset

Feature Selection	**#Attributes**	**RFAccuracy**	**IBKAccuracy**	**J48Accuracy**	**JRipAccuracy**
Before Feature Selection	30	**99**	92	**100**	**99**
SVMRFE	3	**99**	95	99	**99**
RST+SVMRFE	3	**99**	94	99	**99**
$IG_{fw}S$	4	98	**98**	98	98
RST+$IG_{fw}S$	4	98	**98**	98	98

Table 8: Results on Breast Cancer Dataset

Feature Selection	**#Attributes**	**RFAccuracy**	**IBKAccuracy**	**J48Accuracy**	**JRipAccuracy**
Before Feature Selection	31	95	**96**	93	93
SVMRFE	11	94	**96**	**94**	**94**
RST+SVMRFE	11	**96**	95	**94**	93
$IG_{fw}S$	4	93	91	92	93
RST+$IG_{fw}S$	3	92	90	92	92

From Table 8, random forest classifier achieves the maximum accuracy of 96% with RST+SVMRFE, IBK classifier achieves the highest accuracy of 96% with SVMRFE, J48 classifier achieves the highest accuracy of 94% with SVMRFE and RST+SVMRFE and JRip classifier achieves 94% accuracy with SVMRFE.

Table 9: Results on Dermatology Dataset

Feature Selection	**#Attributes**	**RFAccuracy**	**IBKAccuracy**	**J48Accuracy**	**JRipAccuracy**
Before Feature Selection	35	**98**	**95**	**96**	**92**
SVMRFE	11	**93**	**94**	**89**	87
RST+SVMRFE	11	89	87	87	84
$IG_{fw}S$	7	86	89	88	**89**
RST+$IG_{fw}S$	6	84	83	84	82

From Table 9, In the case of Dermatology dataset, we were able to find an exceptional case that the classifiers were not able to achieve even the accuracy obtained without applying any feature selection methods.

Table 10: Results on Liver-Disorder Dataset

Feature Selection	**#Attributes**	**RFAccuracy**	**IBKAccuracy**	**J48Accuracy**	**JRipAccuracy**
Before Feature Selection	7	65	**63**	**69**	**65**
SVMRFE	7	**70**	**63**	**69**	**65**
RST+SVMRFE	7	**70**	**63**	**69**	**65**
$IG_{fw}S$	3	64	61	62	62
RST+$IG_{fw}S$	3	64	61	62	62

In the case of Liver-Disorder dataset, the table 10 shows that random forest classifier achieved 70% as the highest accuracy with SVMRFE and RST+SVMRFE. The classifiers IBK, J48 and JRip was able to maintain the same accuracy as that of the accuracy obtained without applying any feature selection methods for SVMRFE and RST+SVMRFE.

6. CONCLUSION

Feature selection is very important since the datasets considered for a research study may contain thousands of attributes, out of which some attributes might be unnecessary. In this research work we have presented four feature selection methods that were applied to different datasets. These four feature selection algorithms were evaluated on the data sets used in this experiment and the selected features from each of the algorithm were used to develop a classification model using random forest, IBK, J48 and Jrip. The results reveal that feature selection preserves classification accuracy i.e. redundant attributes can be removed efficiently from the different dataset without giving up the classification performance. Usually it is preferred that a small number of variables are used for performing classification because it is less computationally expensive to run. [28]. Also a smaller set of features helps in interpreting and discovering knowledge easily. From this study, the feature selection methods SVMRFE and RST+SVMRFE shows better performance based on classification accuracy with most of the datasets considered under study. And also the random forest classifier shows the maximum performance accuracy with all the datasets.

REFERENCES

[1] Pinar Yildirim,"Filter Based Feature Selection Methods for Prediction of Risks in Hepatitis Disease", International Journal of Machine Learning and Computing, Vol. 5, No. 4, August 2015

[2] Dehuri, S. and Ghosh, A. (2013), Revisiting evolutionary algorithms in feature selection and nonfuzzy/fuzzy rule based classification. WIREs Data Mining Knowl Discov, 3: 83–108. doi:10.1002/widm.1087

[3] Alex A. Freitas,"Data Mining and Knowledge Discovery with Evolutionary Algorithms",Springer-Verlag Berlin Heidelberg 2002

[4] Timo Simila,"Advances in Variable Selection and Visualization Methods or Analysis of Multivariate Data", Helsinki University of Technology, Dissertations in Computer and Information Science Espoo 2007

[5] Lichman, M. (2013). UCI Machine Learning Repository[http://archive.ics.uci.edu/ml]. Irvine, CA: University of California, School of Information and Computer Science.

[6] Ioannis Anagnostopoulos,Christos Anagnostopoulos, Angelos Rouskas1,George Kormentzas,Dimitrios Vergados, 'The Wisconsin Breast Cancer Problem: Diagnosis and DFS time prognosis using probabilistic and generalised regression neural classifiers, special issue Computational Analysis and Decision Support Systems in Oncology', last quarter 2005

[7] G.K Gupta, Introduction to Data Mining with Case Studies, Second Edition, PHI Learning Private Limited, 2014

[8] Yogendra Kumar Jain, Upendra, Intrusion Detection using Supervised Learning with Feature Set Reduction, International Journal of Computer Applications (0975 – 8887) Volume 33– No.6, November 2011

[9] Seyed Jalaleddin Mousavirad, Hossein Ebrahimpour-Komleh, A Comparative Study On Medical Diagnosis Using Predictive Data Mining: A Case Study

[10] P. Bagavathi Sivakumar. "Performance Comparison of Attribute Set Reduction Algorithms in Stock Price Prediction - A Case Study on Indian Stock Data", Lecture Notes in Computer Science, 2010

[10] [10] I. Guyon, J. Weston, S. Barnhill, and V. Vapnik. Gene selection for cancer classification using support vector machines.Machine Learning , 46(1-3):389–422, 2002.

[11] Xin Zhou, David P. Tuck; MSVM-RFE: extensions of SVM-RFE for multiclass gene selection on DNA microarray data. Bioinformatics 2007; 23 (9): 1106-1114. doi: 10.1093/bioinfor matics /btm036

[12] Guyon, I., Weston, J., Barnhill, S. et al. Machine Learning (2002) 46: 389. doi:10.1023/A:1012487302797

[13] K. Anitha, Dr. P.Venkatesan, 'Feature Selection by Rough –Quick Reduct Algorithm', International Journal of Innovative Research in Science, Engineering and Technology, ISSN: 23198753 Vol. 2, Issue 8, August 2013

[14] J. Z. Pawlak, Grzymala-Busse, R. Slowinski, and W. Ziarko, "Rough sets," in Communications of the ACM, vol. 38, New York, NY, USA, November 1995.

[15] B. Walczak , D.L. Massart, Rough sets theory, Chemometrics and Intelligent Laboratory Systems 47 1999 1–1

[16] Sengupta N., Sil J. (2011) Comparison of Performance for Intrusion Detection System Using Different Rules of Classification. In: Venugopal K.R., Patnaik L.M. (eds) Computer Networks and Intelligent Computing. Communications in Computer and Information Science, vol 157. Springer, Berlin, Heidelberg

[17] Singh Kavita, Zaveri Mukesh, Raghuwanhi Mukesh, A Rough Neurocomputing Approach for Illumination Invariant Face Recognition System, Rough Sets And Knowledge Technology, 7th International Conference, RSKT 2012, Chengdu, China, Augut 2012 Proceedings

[18] K.Anitha , Dr.P.Venkatesan, Feature Selection By Rough –Quick Reduct Algorithm, International Journal of Innovative Research in Science, Engineering and Technology, (ISO 3297: 2007 Certified Organization), ISSN: 2319-8753, Vol. 2, Issue 8, August 2013

[19] WEKA: Machine learning software, Available at www.weka.net.nz/
[20] Jun Wang, Zucker, Jean-Daniel: Solving Multiple-Instance Problem: A Lazy Learning Approach. In: 17th International Conference on Machine Learning, 1119-1125, 2000.
[17] M.O. Stitson, J.A.E. Weston, A Gammerman, V.Vork, V. Vapnik, "Theory of Support Vector Machines", Technical Report CSD-TR-96-17, Department of Computer Science, Royal Holloway College, University of London, Dec. 1996
[18] Breiman, L. Machine Learning (2001) 45: 5. doi:10.1023/A:1010933404324
[19] D. Aha, D. Kibler (1991). Instance-based learning algorithms. Machine Learning. 6:37-66.
[20] Ross Quinlan (1993). C4.5: Programs for Machine Learning. Morgan Kaufmann Publishers, San Mateo, CA.
[21] William W. Cohen: Fast Effective Rule Induction. In: Twelfth International Conference on Machine Learning, 115-123, 1995.
[22] M.B. Seasholtz, B. Kowalski, Analytica Chimica Acta 277 (1993) 165.
[23] Jan G. Bazan and Marcin Szczuka, RSES and RSESlib – A Collection of Tools for Rough Set Computations, Proc. of RSCTC'2000, LNAI 2005, 2001, Springer Verlag
[24] Breiman, L. Machine Learning (2001) 45: 5. doi:10.1023/A:1010933404324
[25] D. Aha, D. Kibler (1991). Instance-based learning algorithms. Machine Learning. 6:37-66.
[26] Ross Quinlan (1993). C4.5: Programs for Machine Learning. Morgan Kaufmann Publishers, San Mateo, CA.
[27] William W. Cohen: Fast Effective Rule Induction. In: Twelfth International Conference on Machine Learning, 115-123, 1995.
[28] M.B. Seasholtz, B. Kowalski, Analytica Chimica Acta 277 (1993) 165.