A novel filter based feature selection for image and signal classification

D. ASIR ANTONY GNANA SINGH[a], S. APPAVU ALIAS BALAMURUGAN[b], E. JEBAMALAR LEAVLINE[c]
[a]*Department of Computer Science and Engineering, Anna University, BIT Campus, Tiruchirappalli – 620 024, Tamilnadu, India*
[b]*Department of Information Technology, K.L.N College of Information Technology, Sivagangai – 630 611, Tamilnadu, India*
[c]*Department of Electronics and Communication Engineering, Anna University, BIT Campus, Tiruchirappalli – 620 024, Tamilnadu, India*

In the digital era, the application of image and signal classification influences several areas including medical, engineering, science, and technology. Due to the advancements in digital imaging and signal acquisition, images and signals are generated massively through various image and signal acquisition devices. Processing these massive images and signals for classification is a very challenging task to the researchers due to the high-dimensional space that contains irrelevant and redundant features. The irrelevant and redundant features reduce the performance of the classification algorithms in terms of classification accuracy. Therefore, the feature selection plays a significant role in the image and signal classification in order to reduce the irrelevant and redundant features from the high-dimensional space to improve the accuracy of the classifiers. This paper proposes a novel filtering approach with clustering based feature selection (FACFS) for image and signal classification. The performance of the proposed method is tested on various real-world image and signal datasets and compared with various state-of-the-art feature selection methods in terms of classification accuracy and redundancy rate. The experimental results show that the proposed method is very promising than the other methods compared.

(Received December 4, 2015; accepted August 3, 2016)

Keywords: Image and signal classification, Image acquisition devices, Digital imaging, Signal processing

1. Introduction

In the recent past, thanks to the advancements in digital imaging and signal acquisition technologies and sophisticated signal and image processing techniques, massive volume of images and signals are acquired through various image capturing devices. This includes imaging devices such as digital cameras, X-ray, radiography, fluoroscopy, computed tomography (CT) [1], digital gamma ray imaging devices such as digital scintigraphy, single-photon emission computed tomography (SPECT), positron emission tomography (PET) and signal acquisition techniques such as electroencephalography (EEG), magneto encephalography (MEG) [2], electrocardiography (ECG), electromyography (EMG), medical ultrasonography, radar, sonar and satellite signals, etc [3]. Processing these massive images and signals in order to build the classifier with high accuracy is the challenging task among the researchers since they contain high-dimensional space with irrelevant and redundant features. The image and signal classification tasks are carried out in our day-to-day life in many areas as follows.

Geological and meteorological signals are used to predict weather condition, cyclone and rain using satellite or weather images [4]. In medical field, the classifier is used as the medical diagnosis system to predict the diseases using the various types of medical images. In banking sector, the classifier is used as the currency detector to identify the genuineness and the denomination of the currency using the images captured through scanners. In authentication and security surveillance systems, classifier is used as face and object recognition system to predict or identify the face or object from the surveillance images. In handwriting recognition system, classifier is used to recognize the digits and alphabets from the handwritten images. In voice recognition system, it is used to predict the voice (speech) signal of the person or stranger using the voice (speech) signal [5]. In satellite signal prediction, the classifier is used to predict the interference of the signal using satellite signals. In medical signal prediction system, the classifier is used to diagnose the diseases using the biomedical signals such as EEG, ECG, and EMG [6].

In general, classifier is built using classification algorithms with the features extracted from images and signals in order to carry out the classification task. The extracted features contain relevant, irrelevant, and redundant features. The irrelevant and the redundant features reduce the accuracy of the classifier. Therefore, feature selection is employed to remove the irrelevant and redundant features to develop a highly accurate classifier.

This paper proposes a novel filtering approach with clustering based feature selection (FACFS) for image and signal classification using ranking with clustering approach in order to improve the accuracy of the classification algorithms. The performance of the proposed method is tested on various real-world image and signal datasets and compared with various state-of-the-art feature selection methods. The experimental results show that the proposed method outperforms the other methods compared. The remainder of this paper is organized as follows: Section 2 discusses the literature review and Section 3 explains the proposed feature selection method for image and signal classification. Section 4 details the implementation and experimental setup. Section 5 discusses the results obtained and Section 6 concludes the paper.

2. Literature review

This section reviews the research works that are related to the proposed method. Feature selection is a process of removing irrelevant and redundant features from a given dataset to improve the accuracy of the classifier. The feature selection methods are categorized into four namely wrapper, embedded, filter, and hybrid methods based on how the classifier is used in feature selection process.

Wrapper method [7] uses the classifier to validate the feature subset in feature subset selection process. It is computationally costlier and does not possess generality since it produces higher accuracy for the classifier which is used to validate the feature subsets. The embedded method [8] uses a part of training phase of the classifier. It is computationally efficient than wrapper method and does not possess generality since it produces high accuracy only for the classifier which is used for the feature selection process. The filter method [9] uses any one of the statistical or mathematical measures to identify the significant features from a given dataset. This requires very less computation time and space compared to the wrapper and embedded methods. The hybrid method combines the filter and wrapper methods [10].

On the other hand, feature selection methods are categorized into two based on how the features are combined in feature selection process namely feature subset selection and feature ranking methods.

The feature subset selection combines all features into a maximum number of possible combinations using any one of the searching algorithms to select the best feature subst. It is computationally costly and space complexity is high. This includes correlation-based feature selection, consistency-based feature subset selection, etc [11-14].

The feature ranking method uses any one of the statistical, probabilistic, or mathematical measures to weight each feature of a given dataset. Then, the weighted features are ranked based on their weight. The top ranked features are chosen as the selected significant features. The feature ranking-based feature selection includes information gain-based feature selection (IGFS), chi-squared-based feature selection (CQFS), Gain ratio-based feature selection (GRFS), and symmetric uncertainty (SUFS) feature selection method [15]. These methods do not yield higher classification accuracy, since the ranking method identifies only the relevant features but fails to identify the redundant features.

Therefore, the proposed FACFS method takes the advantage of the ranking approach to identify relevant features and removes irrelevant features and uses clustering approach in order to discover and remove redundant features to improve the accuracy of the classifiers.

3. Filtering approach with clustering based feature selection (FACFS)

The FACFS algorithm comprises of two phases. In phase 1, the algorithm receives a dataset ‘D’ as input and calculates the information gain weight for all the features with the corresponding class-target-attribute. The features that have ‘0’ information gain weight are removed as they are treated as irrelevant features. In phase 2, the relevant features are clustered using k-means clustering. The clusters of features contain redundant features. Then the representative features from each cluster are chosen using the information gain weight with the threshold value. The resulting representative features from each cluster are considered as the selected significant features from the dataset.

3.1 Algorithm with theoretical analysis of FACFS

Consider a real-world dataset ‘D’ namely ‘Mfeat-pix’ taken from UCI repository [16]. It contains 2000 patterns (instances) with 240 features $F = \{f_1, f_2, \ldots, f_{240}\}$ where F is a set of whole features of dataset D and $f_i \in F$, to predict the hand written digits from 0 to 9 by one target-class contains 10 labels (0 to 9) which indicates the possible hand written digits 0 to 9.

Step1: Removal of irrelevant features

In order to identify the relevant features to the target-class attribute, information-gain-weight (IG_{Wi}) is calculated for all features as follows: Initially, the expected information ‘I’ needed for classification is computed using the target-class attribute of ‘D’ as shown in Equation 1, where ‘z’ is the total number of distinct labels.

$$I(D) = -\sum_{l=1}^{z} P_l \log_2(P_l) \quad (1)$$

Then the information required for each feature $I_f(D)$ of the dataset 'D' is calculated as shown in Equation 2, where 'd' denotes the number of distinct values of the feature f.

$$I_f(D) = -\sum_{m=1}^{d} \frac{|D_m|}{|D|} \times I(D_m) \tag{2}$$

$$IG(f) = I(D) - I_f(D) \tag{3}$$

Information gain is computed for each feature as shown in Table 1 by subtracting the required information for a variable $I_f(D)$ from the expected information for classifying the dataset I(D) as shown in Equation 3. The features having '0' information gain weights are treated as irrelevant features and they are removed. In the 'Mfeat-pix' dataset, all the features have information weight greater than 0 therefore the entire feature (240) set is kept for redundancy removal process.

Table 1. Sample table of features and their information gain of D

Feature index	IG weight
1	0.2482
2	0.3488
3	0.3705
. . .	. . .
240	0.2023

Step2: Removal of redundant features

In order to identify and remove redundant features, 'F' is grouped into 'k' number of clusters, where k=3, using Equation 4, where $\{f_1, f_2, \ldots, f_n\}$ are the features and 'n' is the total number of features 240, (k <= n), z = $\{z_1, z_2, \ldots, z_k\}$ reduces the sum of squares within the cluster and 'μ_p' is the mean of 'z_p' [17, 18]. Each cluster contains redundant features and these redundant features are ranked based on their IG weight as shown in the Table 2.

$$\underset{z}{argmin} \sum_{p=1}^{k} \sum_{v_q \in z_p} || z_q - \mu_p ||^2 \tag{4}$$

Table 2. Sample table to represent 3 clusters with redundant features and their IG weight.

Cluster1		**Cluster2**		**Cluster3**	
Total number of features =73		**Total number of features =72**		**Total number of features =96**	
Feature index	**IG weight**	**Feature index**	**IG weight**	**Feature index**	**IG weight**
153	0.6970	68	0.4468	57	0.6230
138	0.6910	67	0.4327	72	0.6130
168	0.6400	83	0.4305	58	0.6050
154	0.6370	230	0.4239	73	0.5830
139	0.6290	143	0.4113	42	0.5090
152	0.6250	69	0.4001	59	0.5060
123	0.6070	128	0.3998	215	0.5000
169	0.5970	166	0.3743	87	0.4940
124	0.5830	17	0.3730	56	0.4920
167	0.5620	3	0.3705	43	0.4760
. . .	. . .	. . .	. . .	. . .	. . .
235	0.1710	226	0.0405	36	0.1100

Then, the threshold value (T_{vd}) that the total number of features to be selected from the dataset is calculated as T_{vd} = (ceil(N/a)+b), where N is the total number of features in the dataset. The arbitrary constants a, b are chosen as 10 and 3, respectively.

Hence, T_{vd}=(ceil(240/10)+3)=27. Then the number of features selected from each cluster is determined as $T_{vc}=T_{vd}/k=9$ since T_{vd} is divisible by k. Therefore, the top ranked (T_{vc}=9) features are chosen from each cluster and combined together as the selected significant features Fs=$\{f_{153}, f_{138}, f_{168}, f_{154}, f_{139}, f_{152}, f_{123}, f_{169}, f_{124}, f_{68}, f_{67}, f_{83}, f_{230}, f_{143}, f_{69}, f_{128}, f_{166}, f_{17}, f_{57}, f_{72}, f_{58}, f_{73}, f_{42}, f_{59}, f_{215}, f_{87}, f_{56}\}$ thereby redundancy is eliminated. Then, these features are given to the classification algorithms namely NB, IB1, and J48 and the classification accuracies are obtained.

4. Implementation and experimental setup

The FACFS is implemented and experiments were conducted using WEKA and MATLAB12b with the system configuration of Intel® Core™ 2 CPU T5300 @ 1.73GHz Processor, 4 GB Memory (RAM) and 32-bit Windows vista Home Premium Operating system. The performance of the proposed method is tested on various image and signal datasets collected from UCI database repository [16] as tabulated in Table 3. Further, the performance of the FACFS is compared with various state-of-the-art feature selection methods namely information gain-based feature selection (IGFS), chi-squared-based feature selection (CQFS), Gain ratio-based feature selection (GRFS), and symmetric uncertainty-based feature selection (SUFS). Three different classification algorithms namely probabilistic-based Naïve Bayes classifier (NB), instance-based classifier (IB1), and tree-based classifier J48 are used to validate the performance of the FACFS with other state-of-the-art feature selection methods in terms of classification accuracy.

Table 3. Details of image and signal datasets.

Dataset	Nature	Number of instances	Number of Features
Dermatology	Image	366	34
Kdd Japanese vowels	Signal	4275	14
Kdd synthetic control	Signal	600	60
Letter	Image	2000	16
Mfeat-pix	Image	2000	240
Segmentation	Image	2310	19
Sonar	Signal	208	60
Spectrometer	Signal	531	93
Vehicle	Image	846	18
Waveform	Signal	5000	40

The proposed method also is validated in terms of redundancy rate [15] as shown in Equation 5 by calculating the redundancy rate of the selected features using the proposed and other feature selection methods.

$$\mathrm{RED}(f) = \frac{1}{N(N-1)} \sum_{x_i, x_j \in X, i>j} c_{i,j} \tag{5}$$

where N is the total number of features in the dataset ‘D’. $C_{i,j}$ denotes the correlation between two features x_i and x_j.

Further, in order to statistically analyze the performance of the feature selection methods on the datasets in terms of classification accuracy and redundancy, the following statistical tests are conducted. The Friedman test [19] is performed with the null hypothesis that “*all the methods perform equivalently*”. If the result of Friedman test is p=0, at α=0.10 where p is the probability that the null hypothesis is accepted and α is the level of significance, the null hypothesis rejected, meaning that the algorithms compared are statistically different. In that case, the analysis of means (ANOM) test is conducted to rank each state-of-the-art feature selection method based on their performance and to identify the significant difference among the methods at α = 0.05 and α=0.10 with the critical line (CL), upper and lower decision or critical lines (UDL and LDL) as discussed in [20].

For the experimental setup, the number of clusters k is set as 3 and the threshold value (T_{vd}) is calculated to determine the total number of features to be selected from the dataset using the formula T_{vd} = (ceil(N/a)+b), where N is the total number of features in the dataset. The arbitrary constants a, b are chosen as 10 and 3, respectively. Then the number of features to be selected from each cluster is determined by $T_{vc}=T_{vd}/k$ if T_{vd} is divisible by k, else ceil(T_{vd}/k) number of features are selected from the k–1 number of larger clusters and T_{vd} – (ceil(T_{vd}/k) × (k–1)) number of features are selected from the smallest cluster.

5. Experimental results and discussion

Initially, T_{vd} numbers of features are selected from each dataset using the feature selection methods as shown in Table 4. Then the selected features are given to the classification algorithms namely NB, IB1, and J48. Then, the classification accuracies for each classifier are obtained with10-fold cross validation test mode as tabulated in Table 5, Table 6, and Table 7, respectively.

Table 4. Number of selected features from the datasets using all the feature selection methods.

Dataset	Number of features selected
Dermatology	6
Kdd Japanese vowels	5
Kdd synthetic control	9
Letter	5
Mfeat-pix	27
Segmentation	5
Sonar	9
Spectrometer	13
Vehicle	5
Waveform	7

Table 5. Classification accuracy of NB classifier with respective feature selection methods.

Dataset	FACFS	IGFS	CQFS	GRFS	SUFS
Dermatology	83.33	76.50	69.12	69.12	76.50
Kdd Japanese vowels	71.87	71.15	71.15	74.80	71.15
Kdd synthetic control	84.83	78.16	78.00	78.00	77.16
Letter	45.09	45.95	45.95	45.95	45.95
Mfeat-pix	79.00	70.40	69.45	68.90	69.90
Segmentation	74.67	69.04	69.04	69.69	69.04
Sonar	70.67	68.26	68.26	65.38	64.42
Spectrometer	42.18	38.79	40.67	40.11	39.54
Vehicle	43.97	40.30	40.30	40.30	40.30
Waveform	71.80	69.48	69.48	71.76	71.76
Average Accuracy	66.74	62.80	62.14	62.40	62.57

Table 6. Classification accuracy of IB1 classifier with respective feature selection methods.

Dataset	FACFS	IGFS	CQFS	GRFS	SUFS
Dermatology	78.96	73.47	63.11	63.38	73.49
Kdd Japanese Vowels	89.05	87.38	87.38	88.93	87.38
Kdd synthetic Control	90.66	77.16	80.00	77.00	85.00
Letter	73.52	64.91	64.91	64.91	64.91
Mfeat-pix	79.70	69.80	68.50	67.65	68.65
Segmentation	90.64	89.65	89.65	89.35	89.69
Sonar	78.36	73.55	73.55	83.65	74.03
Spectrometer	51.78	49.34	52.16	40.86	49.90
Vehicle	65.36	58.51	58.51	57.09	58.51
Waveform	68.82	67.00	67.00	68.98	68.98
Average accuracy	76.68	71.08	70.48	70.18	72.05

Table 7. Classification accuracy of J48 classifier with respective feature selection methods.

Dataset	FACFS	IGFS	CQFS	GRFS	SUFS
Dermatology	78.41	75.13	67.75	67.48	75.13
Kdd Japanese vowels	81.84	78.56	78.52	79.64	78.56
Kdd synthetic control	83.00	75.66	77.16	79.50	81.16
Letter	74.76	67.77	67.77	67.76	67.77
Mfeat-pix	74.50	68.85	67.95	66.10	66.35
Segmentation	91.64	88.09	88.00	88.00	87.96
Sonar	73.55	67.30	67.30	71.15	69.71
Spectrometer	46.51	44.44	49.34	37.28	41.80
Vehicle	65.13	57.32	57.32	55.08	57.32
Waveform	71.94	70.88	70.90	72.56	72.60
Average accuracy	74.13	69.40	69.20	68.45	69.84

The redundancy rate is calculated on the selected features from the dataset and tabulated in Table 8. The classification accuracy of NB, IB1, and J48 classifiers with respective datasets and the feature selection methods are illustrated in Fig. 1. The redundancy rate with respective datasets and feature selection methods are illustrated in Fig. 2. The average classification accuracy of NB, IB1, and J48 classifiers with respect to the feature selection methods are shown in Fig. 3. The average redundancy rate with respect to the feature selection methods is depicted in Fig. 4.

Table 8. Redundancy rate of all dataset with respective feature selection methods.

Dataset	FACFS	IGFS	CQFS	GRFS	SUFS
Dermatology	0.373	0.484	0.433	0.435	0.484
Kdd Japanese vowels	0.234	0.191	0.191	0.191	0.191
Kdd synthetic control	0.631	0.787	0.786	0.482	0.783
Letter	0.177	0.184	0.184	0.184	0.184
Mfeat-pix	0.277	0.370	0.387	0.369	0.372
Segmentation	0.403	0.578	0.578	0.588	0.581
Sonar	0.228	0.326	0.326	0.287	0.341
Spectrometer	0.563	0.714	0.644	0.548	0.788
Vehicle	0.516	0.771	0.771	0.703	0.771
Waveform	0.513	0.544	0.544	0.513	0.513

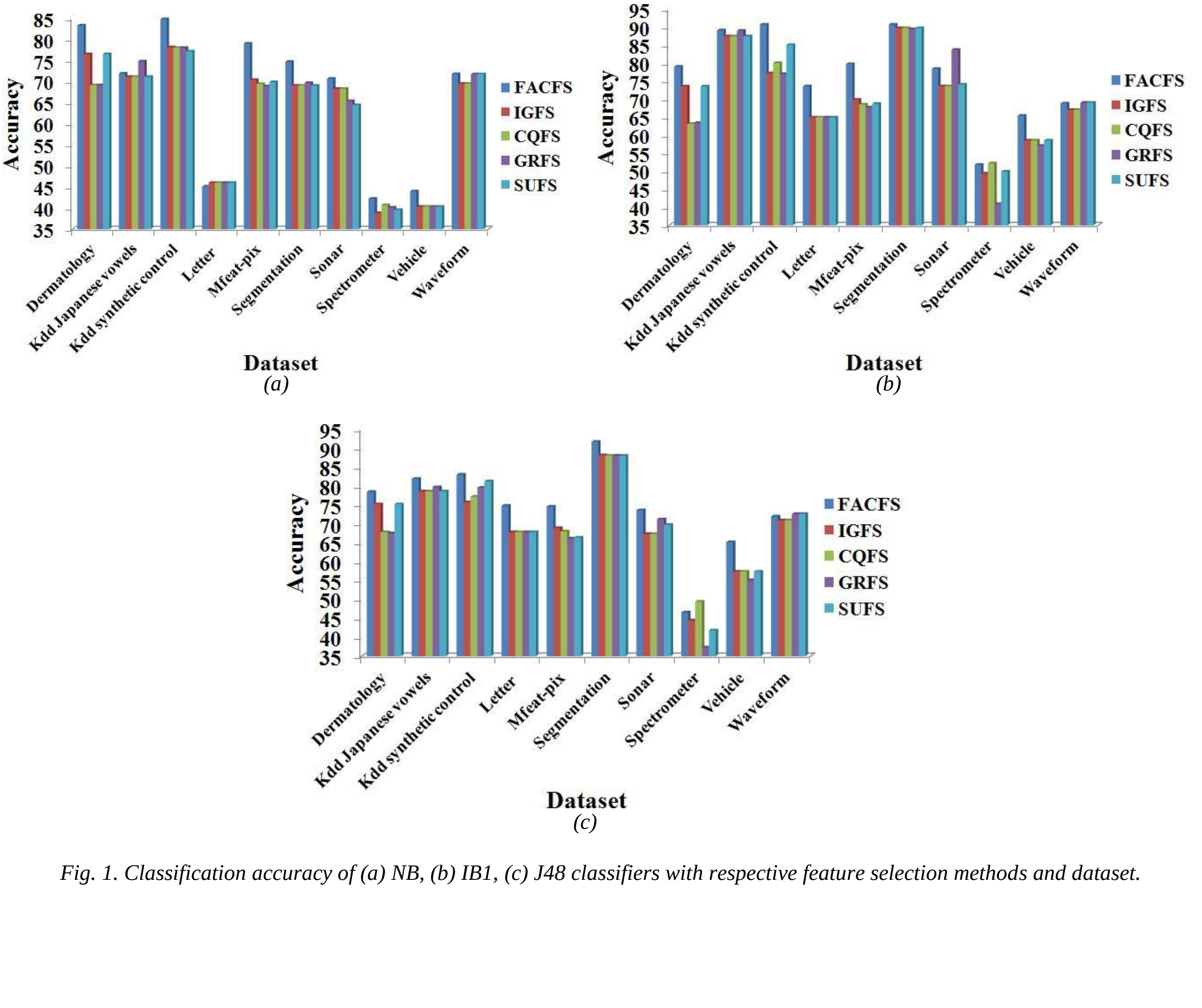


Fig. 1. Classification accuracy of (a) NB, (b) IB1, (c) J48 classifiers with respective feature selection methods and dataset.

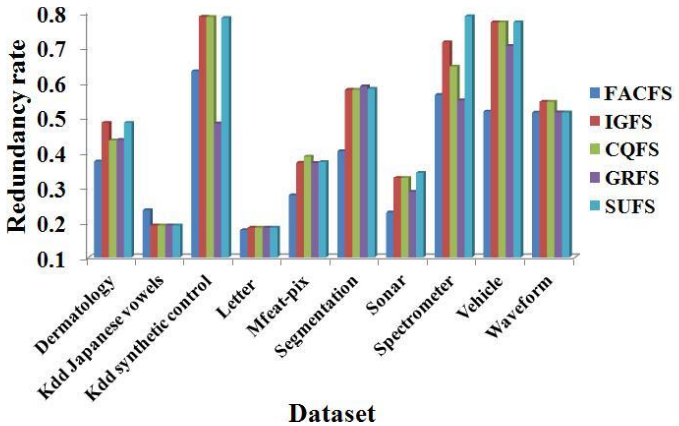
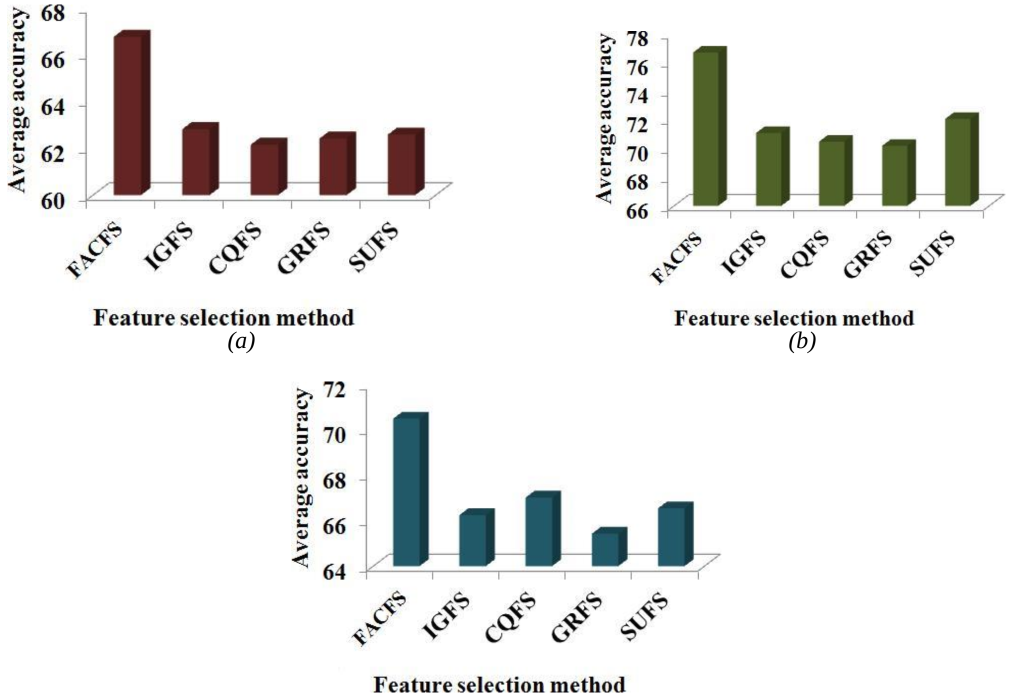


Fig. 2. Redundancy rate with respective feature selection methods and datasets.



(a) *(b)*

(c)

Fig. 3. Average accuracy of (a) NB, (b) IB1, (c) J48 classifiers with respective feature selection methods.

In order to observe the statistical significance of the FACFS in terms of classification accuracy and redundancy rate, the Friedman test is conducted on the obtained results of classification accuracy and redundancy rate and the results of Friedman test is found to be p = 0 at α= 0.10. Hence, the null hypothesis is rejected meaning that all the state-of-the-art feature selection methods significantly differ from each other in terms of classification accuracy and redundancy rate. Hence, ANOM mean rank test is conducted on the obtained classification accuracy of NB, IB, and J48 with respect to feature selection methods and the results are shown in Fig. 5. Further, the ANOM mean rank test is conducted on the obtained redundancy rate with respect to the feature selection methods and the same depicted in Fig. 6.

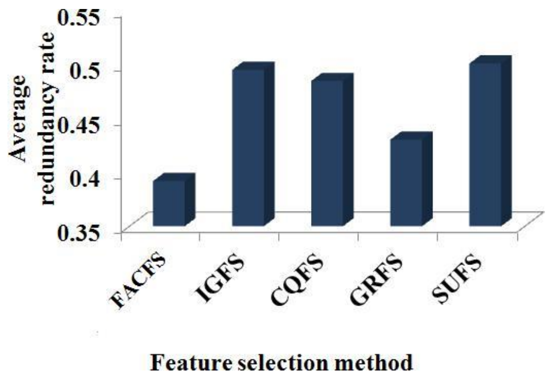


Fig. 4. Average redundancy rate with respective feature selection methods.

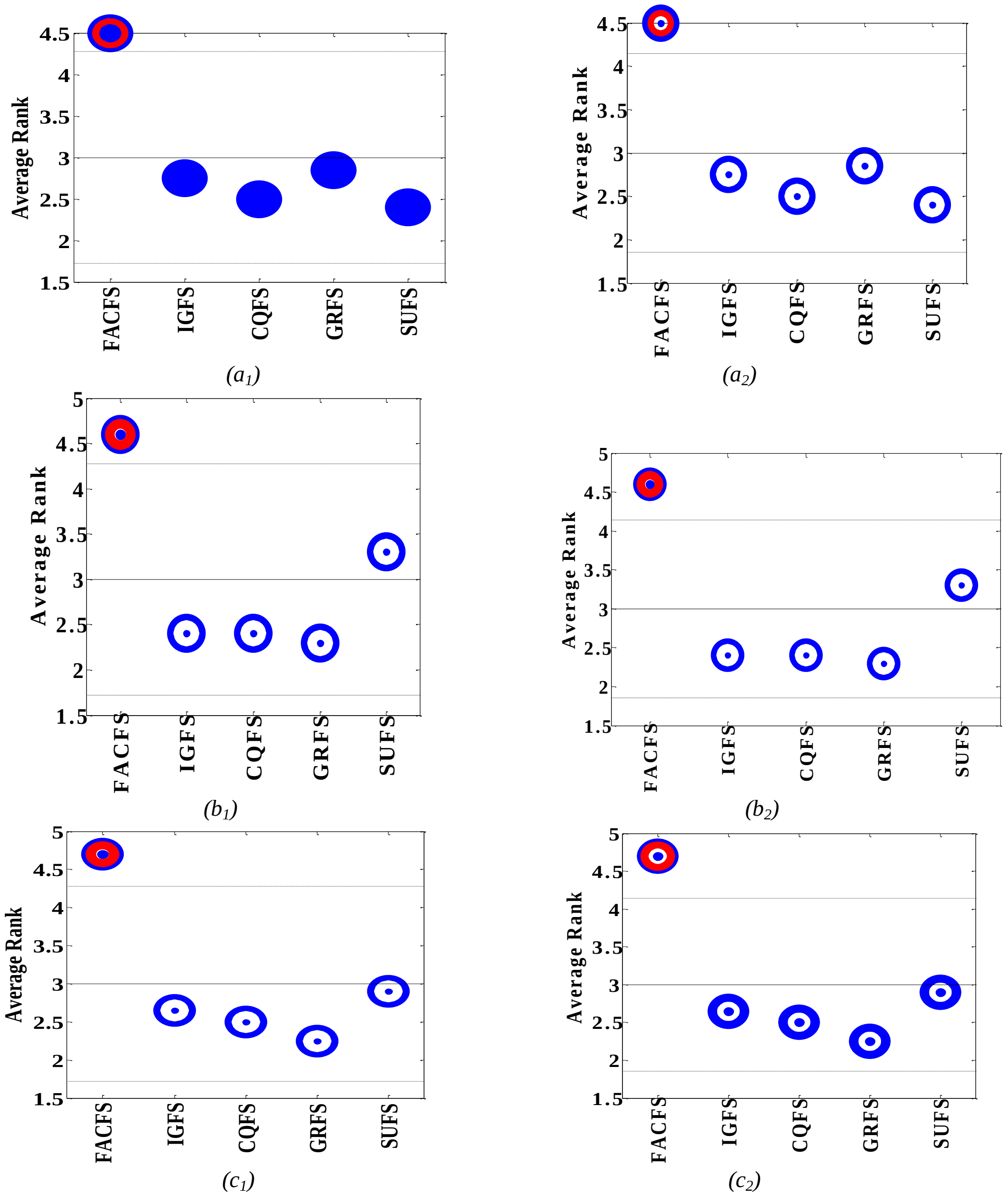


Fig. 5. ANOM charts indicating average ranks of each feature selection method in terms of classification accuracy with the selected features from all the datasets (a_1) NB at α=0.05, (a_2) NB at α=0.10, (b_1) IB1 at α=0.05, (b_2) IB1 at α=0.10, (c_1) J48 at α=0.05 (c_2) J48 at α=0.10

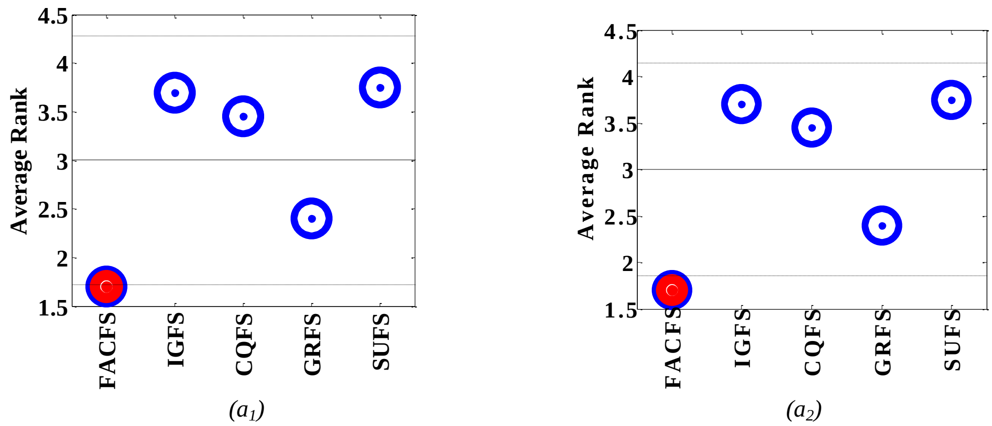


Fig. 6. ANOM charts indicating average ranks of each feature selection method in terms of redundancy rate of the selected features from all the datasets (a_1) at α=0.05, (a_2) at α=0.10.

From Table 5 and Fig. 1 (a), it is observed that FACFS produces better classification accuracy with NB classifier than other methods compared for all datasets except Kdd Japanese vowels and Letter dataset. From Table 6 and Fig. 1 (b), it is observed that FACFS produces better accuracy with IB1 than other methods compared for all the datasets except Sonar, Spectrometer, and Waveform. From Table 7 and Fig. 1 (c), it is observed that FACFS produces better accuracy with J48 than other methods for all datasets except Spectrometer and Waveform.

From Table 8 and Fig. 2, it is obvious that FACFS reduces the redundancy rate for all datasets except Kdd Japanese vowels, Kdd synthetic control and spectrometer. From Fig. 3 it is observed that FACFS performs better for the classifiers NB, IB1, and J48, respectively in terms of average accuracy. From Fig. 4, it is observed that FACFS performs better in reducing the redundancy than all other methods compared.

The ANOM charts shown in Fig. 5 indicate that FACFS achieves higher average rank in terms of classification accuracy of NB, IB1, and J48. In Fig. 5(a_1) and (a_2), FACFS lies above UDL indicating that FACFS achieves significantly greater classification accuracy than the overall average accuracy with NB. The feature selection methods IGFS, CQFS, GRFS and SUFS fall below CL indicating that they achieve less classification accuracy than the overall average accuracy with NB.

In Fig. 5(b_1) and (b_2), FACFS lies above UDL, this indicates that FACFS achieves significantly greater classification accuracy than the overall average classification accuracy with IB1. The feature selection methods namely, IGFS, CQFS and GRFS fall below CL which indicates that they achieve less classification accuracy than the overall average accuracy with IB1. SUFS falls between UDL and CL therefore it achieves fair classification accuracy with respect to the overall average accuracy with IB1.

In Fig. 5(c_1) and (c_2) FACFS lies above UDL indicating that it achieves significantly greater classification accuracy than the overall average accuracy with J48. The feature selection methods IGFS, CQFS, GRFS, and SUFS fall below CL indicating that they achieve less classification accuracy than the overall average accuracy with J48. From Fig. 6, it is found that the FACFS lies below LDL indicating that FACFS significantly reduces the redundancy than the overall average redundancy rate. The feature selection methods IGFS, CQFS, and SUFS fall above CL indicating that they do not perform well in reducing the redundancy rate. GRFS falls between CL and LDL therefore it fairly reduces the redundancy with respect to the overall average redundancy rate.

6. Conclusion and future work

This paper proposed a novel filter based feature selection approach for image and signal classification using the ranking with clustering approach in order to improve the accuracy in image and signal classification. The proposed method works with two phases. In the first phase, it removes irrelevant features using the raking approach. In the second phase, it removes the redundant features using clustering approach. The performance of the proposed method is compared with four state-of-the-art feature selection methods in terms of classification accuracy and redundancy rate. The performance of these methods is also statistically analyzed and it is observed that the proposed method outperforms the other methods compared. In future, different statistical measures can be used to obtain the relevant features and other clustering techniques can be employed for redundancy analysis.

References

[1] K. Araki, K. Maki, K. Seki, K. Sakamaki, Y. Harata, R. Sakaino, T. Okano, K. Seo, Dentomaxillofacial Radiol, **33**(1), 51 (2014).
[2] Y. Wang, B. Liu, L. Fu, Z. Cui, J. Neurol. Sci., **355**(1-2), 120 (2015).
[3] K. Najarian, R. Splinter, Biomedical signal and image processing, CRC press (2012).
[4] J. D. Mollard, Encyclopedia of Natural Hazards (2013).
[5] A. Tsanas, M. A. Little, P. E. McSharry, J. Spielman, Ramig LO, IEEE Trans. Biomed. Eng., (2012)
[6] C. H. Chen, L. F. Pau, P. S. P. Wang (Eds.) Handbook of pattern recognition and computer vision, Imperial College Press (2010).
[7] G Chen, J. Chen, Neurocomputing, **159**(2), 219 (2015)
[8] Imani, Maryam Bahojb, Mohammad Reza Keyvanpour, Reza Azmi, Appl. Artif. Intell., **27**(5), 408 (2013).
[9] Freeman, Cecille, Dana Kulić, Otman Basir, Pattern Recogn., (2015).
[10] Guyon, Isabelle, A. Elisseeff, J Mach Learn Res., (2003).
[11] A. Mark Hall, Correlation-based feature selection for machine learning. Ph.D Dissertation, The University of Waikato, 1999.
[12] O. Akande, Kabiru, O. Owolabi Taoreed, O. Sunday Olatunji, J Nat Gas Sci. Eng. **22**, 515 (2015).
[13] R. Liu Huan, Setiono, Proc. International Conference on Machine Learning, Bari, Italy, 1996.
[14] Aytuğ Onan, Expert Syst. Appl. **42(**20), 6807 (2015).
[15] Z. Zhao, F. Morstatter, S. Sharma, S. Alelyani, A. Anand, H. Liu, ASU feature selection repository (2010).
[16] M. Lichman, UCI Machine Learning Repository [http://archive.ics.uci.edu/ml]. Irvine, CA: University of California, School of Information and Computer Science., (2013).
[17] L. Ahmad Amir, Dey, Data & Knowledge Engineering, (2007).
[18] M. W. Ayech, D. Ziou, Expert Syst. Appl., **42**(6), 2959 (2015).
[19] Xu Jianbo, Guogen Shan, Amei Amei, Jiwei Zhao, Daniel Young, Sheila Clark. Commun. Stat Simul. Comput., (2015).
[20] M. Hibon, S. F. Crone, N. Kourentzes, Proc. 32nd Annual international Symposium on Forecasting, Boston, 2012.

*Corresponding author: asirantony@gmail.com