# Effective hybrid feature subset selection for multilevel datasets using decision tree classifiers

S. Dinakaran*

T. John College,
Bangalore, India
Email: dinakaran77@gmail.com
*Corresponding author

P. Ranjit Jeba Thangaiah

Department of Information Technology,
Karunya Institute of Technology and Sciences,
Coimbatore, India
Email: prjt@live.com

**Abstract:** Feature selection is one of the most significant procedures in machine learning algorithms. It is particularly to improve the performance and prediction accuracy for complex data classification. This paper discusses a hybrid feature selection technique with the decision tree-based classification algorithm. The feature selected using information gain (IG) is combined with the features selected from ReliefF which generates the resultant feature subset. Then the resultant feature subset is in turn combined with a correlation-based feature selection (CFS) method to generate the aggregated feature subset. To perform classification accuracy on the aggregated feature subset, different decision trees-based classification algorithm such as C4.5, decision stumps, naive Bayes tree, and random forest with ten-fold cross-validation have been deployed. To check the prediction accuracy of the proposed work eight different multilevel University of California, Irvine (UCI) machine learning datasets have been used with minimum to maximum numbers of features. The main objective of the hybrid feature selection is to improve the classification accuracy, prediction and to reduce the execution time using standard datasets.

**Keywords:** feature selection; decision tree; information gain; ReliefF; correlation-based feature selection; CFS; naïve Bayes tree; random forest; C4.5; decision stump; exclusive OR; intersection; ranker.

**Reference** to this paper should be made as follows: Dinakaran, S. and Thangaiah, P.R.J. (2023) 'Effective hybrid feature subset selection for multilevel datasets using decision tree classifiers', *Int. J. Advanced Intelligence Paradigms*, Vol. 24, Nos. 1/2, pp.206–228.

**Biographical notes:** S. Dinakaran did his BSc (Computer Science) at Kongunadu Arts and Science, Coimbatore, M.C.A. at Karunya University, Coimbatore, He completed his full time PhD at Karunya Institute of Technology and Sciences, Coimbatore. He has published and presented many national and international journals/conferences. His areas of research include data mining and machine learning.

P. Ranjit Jeba Thangaiah did his BSc (Physics) at P.S.G. College of Arts and Science, Coimbatore, M.C.A. at Karunya Institute of Technology, Coimbatore, MPhil at Manonmaniam Sundaranar University, Tirunelveli, He completed his full time PhD at Bharathiar University, Coimbatore. He has been working with Karunya University for the past 15 years. He has published more than 30 research articles in international journals and conferences. His areas of research include networking, data mining and machine learning. He is a life member of CSI.

## 1 Introduction

Feature selection plays a major role in applicable fields such as data mining, image processing, artificial intelligence, pattern recognition, medical data, Bioinformatics (Wang and Wu, 2013). The high-dimensional data categorised by a large number of features seriously reduces the performance of learning algorithms. In the high-dimensional classification tasks, the features are usually grouped into three types such as redundant, relevant, irrelevant features (Stanczyk and Jain, 2015). The feature selection techniques are divided into three main types: filters (Liu and Dougherty, 2005) wrappers (Kohavi and John, 1997) and embedded method (Guyon et al., 2006). The filter approach, without dependency on any machine learning algorithm, provides faster result, but provides less reliability when compared to the wrapper approach. Alternatively, the wrapper approach is dependent on any one of the classifying algorithms, but performs very slowly for large feature sets that contain more than thousands of features but it generates better results. Menghour and Souici-Meslati (2014) have proposed algorithms for feature selection based on classical feature ranking approaches, furthermore different variants of ACO and binary PSO are used for e-mail classification.

The hybrid approach is a decent combination of the two methods to overcome these issues. In the hybrid method (Hsu et al., 2011; Yang et al., 2011), attributes are filtered initially, and then determined by the wrapper approach. It is frequently initiated that, the hybrid method is capable of finding a better solution, while a single technique frequently traps into an unformed solution (Liu et al., 2013). The hybrid feature selection based on multi-filter weights and multi-feature weights is presented (Wang and Feng, 2019). The hybrid methods are highly successful than a single filter approach for dimensionality reduction because they were capable of generating a higher reduction rate without the damage of classification precision in several circumstances (Ghareb et al., 2016).

Decision trees are dominant learning methods which are able to organise the information extracted for a training dataset in a hierarchical structure. The medical and scientific fields have extensively used decision tree method, of different computational technique that uses flowchart identical tree structures (Tung et al., 2005). The decision tree is a popular and comprehensive classification algorithm in machine learning since it is the precision and flexibility in representing the classifying procedure (Jiang and Cai, 2009). Decision trees are a linear technique which is easy to understand and recognise (Chen et al., 2014). There are several specific decision-tree algorithms, Iterative Dichotomiser 3 (ID3) proposed by Quinlan in 1986 and its enhanced version C4.5 (Quinlan, 1993) are two of the well-known classifiers (Tan and Liang, 2012).

The performance of the proposed hybrid feature subset selection methods is evaluated against those of existing Decision tree-based classification algorithm (Jiang and Cai, 2009) such as C4.5 (Quinlan, 1993, 1996; Cheng et al., 2010), decision stumps (Iba and Langley, 1992; Jian et al., 2007), naïve Bayes tree (Kohavi, 1996; Wang et al., 2015) and random forest (Breiman, 2001; Calderoni et al., 2015) are classifiers using the classification accuracy, precision, and 10-fold cross validation on 8 datasets from UCI machine learning repository.

The UCI (University of California, Irvine) machine learning repository is a collection of databases, domain models, and information generators that are used by the machine learning community for the experiential analysis of machine learning algorithms. The archive was formed as an ftp archive in 1987 by David Aha and fellow scholar at UC Irvine (Asuncion and Newman, 2007). Multilevel datasets are the datasets with mixtures of various datasets like medical datasets, forecasting datasets, chemical datasets and written numerals (‘0’–‘9’) extracted from a collection of Dutch utility maps datasets. The dataset consists of a minimum numbers to maximum numbers of attributes and instances. The attribute characteristics are also categorised into nominal and numeric datasets.

The objective of this research is to develop a hybrid feature subset selection method, combining both feature subsets from information gain and ReliefF. The resultant feature subset is combined with correlation-based feature subset selection for getting aggregated feature subset. The main purpose of this hybrid feature selection is to increase the classification accuracy, predictability and reduction in the execution time of the repository datasets. This article has been outlined as follows: Section 2 entails the details about related work and literature survey conducted. Section 3 describes the methodology of the hybrid feature selection, framework and algorithm design of the proposed hybrid feature selection approach. Section 4 provides the details on the datasets used and the result is discussed. Section 5 concludes the proposed work.

## 2 Literature review

In this section, we review the recent research on hybrid feature selection techniques in various datasets and also review some of the existing feature selection methods and weakness of a method.

The rapid increase in the numbers of large datasets within many domains allows extraordinary challenges to data mining (Han and Kamber, 2011). From the above, a few studies have examined the feature selection issue in high dimensional data and applied hybrid methods with two feature selection algorithms. Most of the feature selection approaches discussed above does not converse about the computational time. This motivates why not to go with three feature selection algorithms to select features subsets, and address the computational time with each datasets which is proposed in this paper.

## 3 Methodology

In this section, the hybrid feature subset selection is displayed in Figure 1. The parts of the feature selection methods are explained in the next subsections.

## *3.1 Feature selection methods*

The proposed work has considered three various attribute evaluators such as, correlation-based feature selection (CFS) (Koprinska et al., 2015), information gain (IG) and ReliefF (Zeng et al., 2015).

### *3.1.1 Information gain*

Information gain (IG) is a univariate technique that selects attributes on the basis of the information input related to the class variable without reflecting feature interaction (Kullback, 1952). Information gain has been calculated using Shannon's entropy measurement. High entropy states that the distribution is uniform, and low entropy states that the distribution is grouped around a point (Dag et al., 2012). Information gain pays mutual consideration of data distribution (Lin, 2013). Information, measures the predictable reduction in entropy of the class before and after perceiving the attributes. Information is frequently used to estimate the applicable degree of feature when building a decision tree (Wu et al., 2008).

**Figure 1** Hybrid feature selection framework

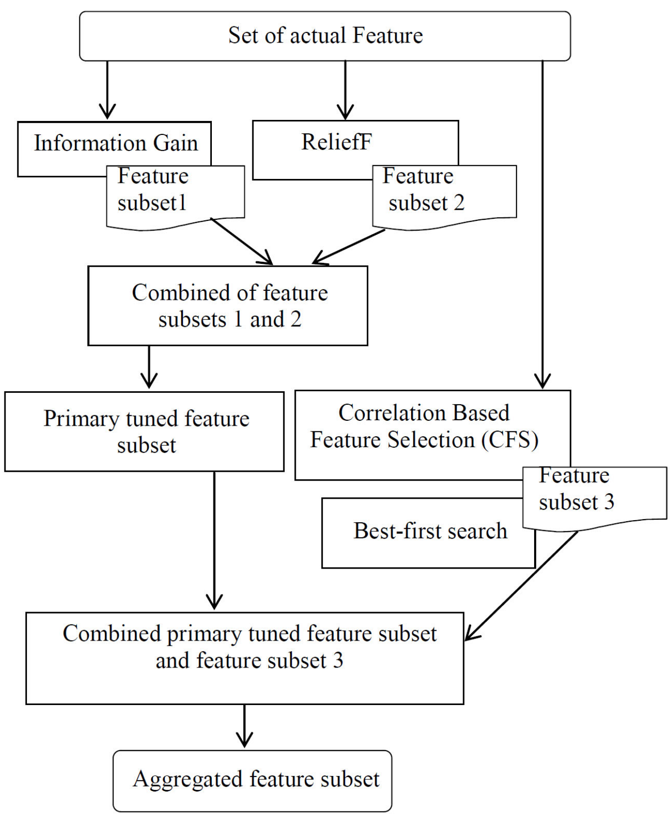


**Table 1** Literature review

| *Papers* | *Methods* | *Weakness* | *Datasets* |
|---|---|---|---|
| Liu et al. (2014) | Hybrid method (HBM) combines the document frequency and term frequency in the feature selection process. | Parameter optimisation combines the high accuracy of wrapper metric in feature selection. | Four benchmark corpora datasets were used: PU1, LingSpam, SpamAssassin, and Trec2007 |
| Apollonia et al. (2016) | BDE-XRank and BDE-XRankf are two algorithms which combine a wrapper feature selection technique based on a binary differential evolution (BDE) algorithm with a rank-based filter feature selection technique. | The fitness function is described as a weighted linear aggregation function in order to handle the two objectives as a single objective problem. | Six datasets were used: colon tumour, leukemia, lung cancer, lymphoma-DLBCL, ovarian cancer, prostate cancer |
| Hsu et al. (2011) | Hybrid with F-score and information gain (IG) are the most vital of preliminary screening. Inversed sequential floating search approach is used as fine tuning. | Time complexities have not been addressed in this method. Accuracy with 5-fold cross validation in percentage is used here. | Two datasets were used: ALL/AML, lung cancer |
| Sangaiah and Kumar (2018) | ReliefF feature reduction with entropy-based genetic algorithm for breast cancer recognition | Ability to make reduced-size subset of relevant features while yielding substantial classification accuracy for huge datasets | Wisconsin breast cancer dataset |
| Ben Brahima and Limam (2016) | Filter stage: hybrid instance-based candidate feature subset selection, wrapper stage: cooperative subset search (HIB-CSS). | The challenge in this method is that it changes the problem of the small sample size to a tool that allows selecting only limited subsets of variables to be evaluated. | Eight datasets were used: DLBCL, bladder, lymphoma, prostate, breast, CNS, lung, Gisette |
| Chuang et al. (2011) | Hybrid using correlation-based feature selection (CFS) and Taguchi-genetic algorithm (TGA). | Wrapper approaches normally outperform filter approaches in terms of prediction accuracy. | Eleven datasets were taken from http://www.gems-system.org: 9_tumors, 11_tumors, 14_tumors, brain_tumor1, brain_tumor2, DLBCL, Leukemia1, Leukemia2, Lung_cancer, SRBCT, Prostate_tumor |
| Chiew et al. (2019) | HEFS, a novel CDF-g algorithm is used to generate primary feature subsets. The next phase produces a set of baseline features, from the secondary feature subsets by using a function perturbation ensemble. | HEFS performs not much accuracy while using the classifiers like decision stump and naive Bayes. | Phishing dataset from the University of California Irvine (UCI) repository. |
| Xie et al. (2013) | The generalised F-score applies to rank features, and SFS and SFFS and SBFS are used to select the essential features. | Computation costs are high | Eryrhenato-squamous diseases dataset from UCI |
| Kannan and Ramaraj (2010) | Genetic algorithm (GA) and local search (LS) with correlation-based filter ranking are combined. | The ranking of features based on the filter approaches has linear time complexity in terms of feature. Dimensionality. | Eight gene expression datasets from UCI repository: Breast, CNS, Leukemia, Leukemia_3c, Leukemia_4c, Ovarian, SRBCT, MLL |

### *3.1.2 ReliefF*

Relief algorithm is dealing with the incomplete, multi-class and noisy data. In ReliefF algorithm, k nearest HIT and the k nearest MISS instance are identified instead of finding one nearest HIT and MISS instance. To understand with the multi-class issue, the algorithm discovers one near miss instance for each various class and medians the updated weight (Kononenko, 1994). ReliefF calculates the weights of attributes by averaging the margin of every attribute (Wei et al., 2014). The ReliefF algorithm pursues the filter method; meanwhile it does not use the response from the classifier to assign the weights to the attributes (Kononenko et al., 1997).

### *3.1.3 Correlation-based feature selection*

CFS is a common filter approach that ranks attribute subsets according to a correlation-based experiential evaluation utility. The bias of the experiential evaluation utility is toward subsets that include attributes that are extremely correlated with the class and uncorrelated with instances. Irrelevant attribute must be ignored since they will have a low correlation with the class. Redundant attributes must monitor out because they will be highly correlated with the remaining attributes (Hall, 1999). The CFS method scores and ranks subsets of attributes, rather than individual attributes (Hall, 2000). CFS is a multivariate subset filter approach. It uses a search algorithm combined with an evaluation utility to estimate the value of attribute subsets. The implementation of CFS is used by Bolon-Canedo et al. (2014) and applied as the forward best first search as its search algorithm.

## *3.2 Hybrid feature selection methodology*

The hybrid feature selection is attained by combining both filter and wrapper approach aiming at the better classification accuracy and reduced computational time. The hybrid feature subset selection techniques frequently compiled with two parameters viz., attribute evaluator and search methods. The attribute evaluator is the strategy by which subset of features are allocated. For instance, they might be allocated by creating a model and assessing the precision of the model.

Three different attribute evaluators are used in hybrid model such as, correlation-based feature selection (CFS), information gain (IG) and ReliefF. Meanwhile, two various search methods such as, best-first (Russell and Norvig, 2003) and ranker (Witten and Frank, 2009) are applied. The search method is designed in which the search area of conceivable attribute subsets is traversed based on the subset assessment. The information gain generates a set of feature subset and on the other hand ReliefF also generates another set of feature subsets. Similarly, the correlation-based feature selection generates a set of feature subsets. All generated feature subsets from different methods are merged with two various techniques such as intersection (INS) and exclusive OR. Figure 1 shows the framework of the hybrid feature selection.

### 3.3 First stage: information gain with ReliefF

Information gain is applied to find the finest split feature in the decision tree classifier. Let $p_i$ be the probability that an arbitrary tuple in $A$ belongs to class $C_j$, estimated by $|C_{j,A}|/|A|$ predictable information (entropy) needed to classify a tuple in $A$:

$$Info(A) = -\sum_{j=1}^{m} p_i\left(c_j\right)\log_2 p_i\left(c_j\right) \tag{1}$$

The entropy of $A$ after perceiving another variable $b$ is defined as:

$$Info(A|b) = \sum_{j=1}^{m} p_i\left(c_j|b\right)\log_2 p_i\left(c_j|b\right) \tag{2}$$

The gain showing the amount of extra information about $A$ delivered by $b$, is measured by

$$Gain(A|b) = Info(A) - Info(A|b) \tag{3}$$

Information gain is an attribute evaluator that evaluates the value of a feature by weighing the attributes with respect to the class. InformationGain(Class, Feature) = B(Class) – B(Class | Feature) where B is information entropy.

Ranking is a search method that ranks features by their individual assessments. The ranker involves the parameter with the number of attributes to retain and the default value (–1) indicates that all the attributes are to be retained; hence to reduce the attribute set, either use attribute retainer or a threshold. Threshold is a parameter by which attributes can be discarded as the basics of the threshold value. The default value is –1.798 that states that no attributes are rejected. In the first stage of the attribute selection method, information gain is used as attribute evaluator and ranker is used as a search method to generate feature subsets. The evaluation classes will provide a score for each attribute and ranks the attribute's base on the scores. Therefore, the generated feature subset is considered as the feature subset 1.

ReliefF evaluates the value of a feature through frequent sampling, instance and reflecting the value of the particular feature of the nearest instance of the similar and dissimilar class. It operates both on discrete and continuous class data.

To deal with the multi-class issue, the algorithm finds one near miss instance for each different class and averages the updated weight using the formula:

$$W(A) = W(A) - diff(A, Yi, H)/s + \sum[P(C)Y\, diff(A, Yi, M(C))]/s$$

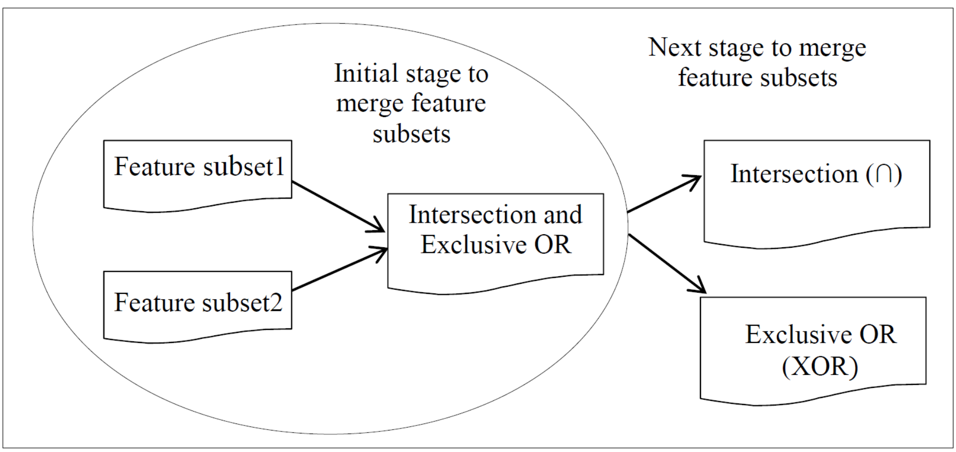
Let $A$ and $C$ be the attributes and classes respectively. Let the sample be $S$ with size '$s$'. Let the weights for all the attributes is denoted as $W(A)$, nearest hit $H$ and nearest miss $M$, and continuous attribute $A$.

Next, the attribute selection method ReliefF is used as attribute evaluator and ranker is used as a search method to generate feature subsets; this feature subset is considered as the feature subset 2.

ReliefF will perform with various parameters such as number of nearest neighbours is counted as 10 for attribute estimation without disturbing the default value. Next the sample size, i.e., number of instances to sample at –1 indicates that all instances will be used for attribute estimation. Seed is another parameter; the size of the seed is taken as 1 which is used for random sampling the instance. Sigma is a distance scaling factor, for observation *i*, influence on the attribute weights from its nearest neighbour *j* is multiplied by $\exp((-rank(i, j)/sigma)^2)$, where $rank(i, j)$ is the position of *j* in the list of nearest neighbours of *i* sorted by distance in the ascending order. The finishing parameter WeightByDistance is used to weigh the nearest neighbours by their distance.

Now these feature set 1 and feature set 2 are combined to yield a new feature subset, considered as a primary tuned feature set. Intersection and XOR are the two unification techniques applied to combine the feature subset. Based on the ranker, all the features are ranked by the weights and several features are selected. To select the most significant feature subsets the threshold value is assigned. The threshold value assigned for information gain is 0.02 and ReliefF respectively 0.06. At the bases of the threshold values, the attributes are weighed and the attributes are ranked. The value is assigned by using an iteration method, and is checked for various data sets. With respect to a threshold value, two sets of feature subsets are obtained in the initial stage.

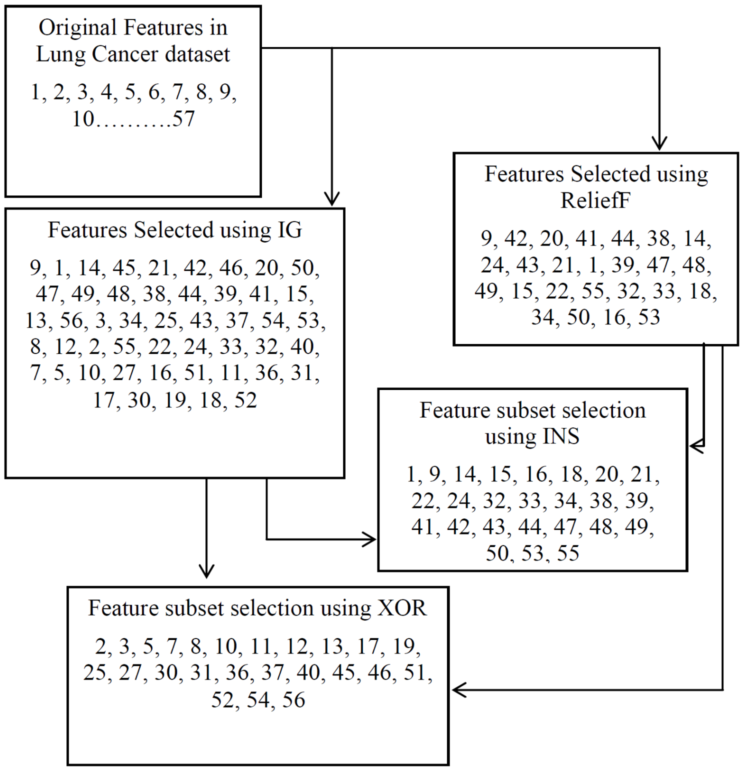
**Figure 2** Combined model with intersection and exclusive OR




Thereby the selected feature subsets are merged with intersection and XOR to generate primary tuned feature subset for the further process. Feature subset is assigned by the particular values viz, first feature subsets as FS1 = {}, Second feature subset as FS2 = {}, and the feature original subset as FS = {}. Revising the description of the entire set FS succeeds as a subset of FS1 and FS2, therefore, some set is a subset of itself. The combining model of Exclusive OR and intersection are shown in Figure 2. Thus ‘FS1 is a subset of FS’ denoted as $FS1 \subset FS$.

Figure 3 shows how the intersection and XOR are generated from primary tuned feature subset selection for lung cancer. The lung cancer datasets contain 57 features which are defined as original feature. Information gain, a feature selection method is introduced to select a set of feature subsets and 49 features are selected, followed by the ReliefF method and 25 features are selected. Then feature subsets are merged with intersection and XOR. The subsets are stored in FSIG = { } as an information gain subset and FSReF = {} as ReliefF subset. Intersection method is performed with both feature subsets (FS1 and FS2) then generate 25 feature subsets as resultant subset and the results of the subsets are stored in Res1ISN = intersect(FSIG, FSReF). Similarly XOR method is executed, then generate 24 feature subsets as resultant subset and the results of the subsets are stored in Res1XoR = setxor(FSIG, FSReF) for further process.

**Figure 3** Generation of primary tuned feature subset for lung cancer




### 3.4 *Second stage: primary tuned FS with CFS*

CFS is a multivariate subset filter algorithm. Correlation-based feature selection weighs the subset of features with the separate predictive competence of every feature using the degree of redundancy amongst them. However, having the low inter correlation is preferred for feature subsets that are extremely associated with the class (Hall, 1999). The estimation function considers the convenience of individual features for predicting the class label also the level of correlation between them. It is expected that the best feature subset will have features much correlated with the class and interrelated with one another. The evaluation function can be seen in the following equation:

$$r_S = \frac{a\overline{r_{cf}}}{\sqrt{a + a(a-1)\overline{r_{ff}}}} \tag{4}$$

where $r_S$ is the $r$ of an attributes subset $S$ having '$a$' attributes, $r_{cf}$ is the average attribute and class correlation, and $r_{ff}$ is the average attribute of feature inter-correlation. (Asuncion and Newman, 2012). To boost the classification accuracy again, one more feature selection technique CFS is used. The Correlation-based feature selection is used as attribute evaluator and best-first search is used as search method; to generate feature subsets; this feature subset is considered as the feature subset 3.

### *3.4.1 Algorithm for hybrid feature subset selection*

*Begin*
 *Set original feature subset = FS*
 *Set training data and test data = t*
  *Rank the features according to the weights*
  $t \in FS$
  *then add to the intersection list*
 *Select* $FS_1$ *and* $FS_2$ *set threshold* $V = 0.02$ *and* $0.06$
 *Training subset t is trained with features in the selected feature subset*
*If* $V > FS$
 *For each* $t \in FS$
 *Feature sets are* $FS_1 = \{\}$, $FS_2 = \{\}$, *and* $FS_3 = \{\}$
 *To perform intersection*
  $FS1 \cap FS2 = \{x \mid x \wedge FS1 \wedge x \in FS2\}$
*To perform exclusive OR*
*The XOR operation FS1.FS2 is identical to non-equivalence* $FS1 \neq FS2$.
$FS1 \oplus FS2$ *can be implemented using AND and OR as*
$FS1 \oplus FS2 = (FS1 \wedge !FS2) \vee (!FS1 \wedge FS2)$
$= (FS1 \vee FS2) \wedge (!FS1 \vee !FS2)$
 *PTFS = Primary tuned feature subset is equipped for the next execution*
*To perform with CFS feature subset*
 *Consider CFS FS as FS3*
 *To perform intersection*
  $PTFS \cap FS3 = \{x \mid x \in PTFS \wedge x \in FS3\}$
*To perform exclusive OR*
  $PTFS.FS3 = (PTFS\ .!FS3)\ .(!PTFS\ .FS3)$
  $= (PTFS\ .FS2)\ .\ (!PTFS\ .!FS3)$
*AFS =Aggregate feature subset is generated*
*End*

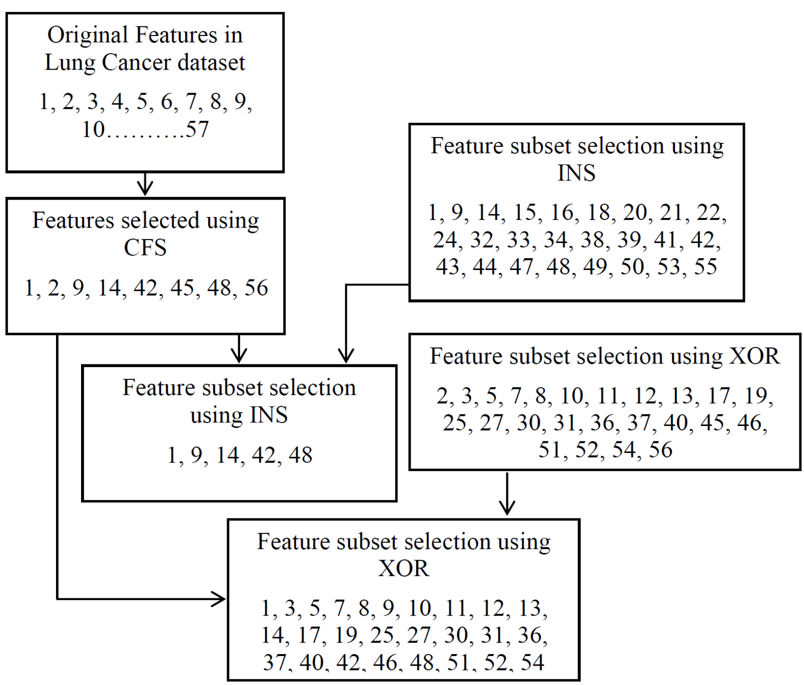
The correlation subset evaluator works with two parameters, i.e., locally predictive and missing separate. Locally predictive is used to identify locally predictive attributes. The extreme limitation of correlation-based feature selection is its failure to select attributes that have locally predictive values when they are overshadowed by robust, globally predictive attributes. Some locally predictive attributes that are genuinely useful will help in predicting instances that it learns from the previous iteration.

Missing separate data values are treated as a separate value when calculating correlations. But, if the missing represents truly missing data, then a sophisticated structure such as distributing the counts related to missing entries across the values of an attribute might be more suitable. Best-first may start with the empty set of attributes and search forward, or start with the full set of attributes and search backward, or start at any point and search in both directions (by considering all possible single attribute additions and deletions at a given point).

The search direction for a CFS with best-first is search by forward search direction. Lookup cache size is set at 1 of the lookup cache to evaluate subsets. This is expressed as a multiplier of the number of attributes in the dataset. Such search term is used to set the amount of backtracking here value is set at 5. Starter set is used to set the starting point for the search. This is specified as a comma separated list of attribute indexes starting at 1.

Figure 4 shows how the intersection and XOR are generated for the aggregated feature subset selection for lung cancer. The lung cancer datasets contain 57 features which are defined as original feature. Correlation-based feature selection, a feature selection approach is introduced to select a set of feature subset and 8 features are selected, followed by intersection method in which 25 features are selected. Then feature subsets are merged with intersection and XOR. The subsets are stored in FSCFS = { } as a correlated feature selection subset and Res1ISN = { } as intersecting subset. Intersection method is performed with both subsets then a resultant subset with 5 features is generated and the results of the subsets are stored in Res2ISN=intersect(FSCFS, Res1ISN). Correlation-based feature selection, a feature selection approach is introduced to select a set of feature subsets and 8 features are selected, followed by exclusive OR method and 24 features are selected.

**Figure 4** Generation of aggregated feature subset for lung cancer



Then feature subsets are merged with intersection and XOR. The subsets are stored in FSCFS = { } as a correlated feature selection subset and Res1XoR = { } as XOR subset. Exclusive OR method is performed with both subsets then a resultant subset with 26 features is generated and the results of the subsets are stored in Res2XoR = setxor(FSCFS, Res1XoR). Now the yielded feature subset 3 is combined with the primary tuned feature sets, therefore another new feature subset is generated which is considered as an aggregated feature subset. The unification method intersection and exclusive OR are used with the full training set to combine the feature subsets. The aggregated feature subset is now considered as a final feature subset to find the classification accuracy. Decision tree-based classification algorithms such as C4.5, Decision Stumps, Naïve Bayes Tree, and Random Forest with 10 fold cross-validations are used to perform the experimental results and to obtain the significant and statistical outcomes.

## 4 Results

In this section, we calculate the performance of the proposed algorithm, and show the experimental results compared with the other three different types of feature subset selection algorithms using eight various UC Irvine datasets correspondingly.

### *4.1 Datasets*

To perform a comparative evaluation of the various feature selection methods viz., information gain, ReliefF, correlation-based feature subsets and proposed hybrid feature selection approach, eight different multilevel UCI Machine Learning Repository datasets were used with minimum to maximum numbers of features, instances and classes. The description of the dataset is given in Table 2.

**Table 2** Dataset descriptions

| *S. no.* | *Name of dataset* | *No. of features* | *No. of instances* | *No. of classes* |
|---|---|---|---|---|
| 1 | Lung cancer | 57 | 32 | 3 |
| 2 | Prostate_tumorVSNormal | 12,601 | 34 | 2 |
| 3 | Ovarian | 15,155 | 253 | 2 |
| 4 | Ozone | 73 | 2,536 | 2 |
| 5 | SRBCT | 2,309 | 83 | 4 |
| 6 | Gas | 130 | 136 | 6 |
| 7 | lsvt-voice-rehabilitation | 311 | 87 | 2 |
| 8 | mfeat Fourier | 77 | 688 | 10 |

### *4.2 Experimental results*

The classification accuracy is enumerated with a various decision tree-based classifiers such as C4.5, Decision Stump, Random forest and Naive Bayes tree classifiers. The complete experiment is tested with 10 fold cross validation for all the feature selection methods.

**Table 3** Classification accuracy of various feature selection methods with C4.5 classifier

| *Dataset* | */CA-IG/ (%)* | */CA-ReliefF/ (%)* | */CA-CFS/ (%)* | */CA-XOR/ (%)* | */CA-ISN/ (%)* |
|---|---|---|---|---|---|
| Lung cancer | 75 | 75 | 78.13 | 71.88 | 93.75 |
| Prostate_tumorVSNormal | 85.29 | 85.29 | 94.12 | 94.12 | 94.12 |
| Ovarian | 96.44 | 97.63 | 96.04 | 96.05 | 96.05 |
| Ozone | 96.85 | 96.84 | 96.88 | 96.65 | 97.12 |
| SRBCT | 85.54 | 86.75 | 85.54 | 86.75 | 86.75 |
| Gas | 92.65 | 91.91 | 94.87 | 94.18 | 94.87 |
| lsvt-voice-rehabilitation | 73.56 | 82.76 | 80.46 | 79.31 | 85.06 |
| mfeat Fourier | 88.01 | 87.94 | 88.37 | 88.23 | 88.66 |

Table 3 shows the comparative results of classification accuracy of various feature selection methods with C4.5 classifiers. Different abbreviations are used, a number of original features /OF/, number of features selected using information gain /FS-IG/, number of features selected in relief /FS-ReliefF/, number of features selected in correlation-based feature selection /FS-CFS/, number of features selected in exclusive OR /FS-XOR/, number of features selected in intersection /FS-ISN/ and classification accuracy /CA/. From the given eight datasets the hybrid intersection method yields better classification accuracy for five datasets viz., lung cancer, ozone, gas, lsvt-voice-rehabilitation and mfeat Fourier datasets. Accuracy remains same for the other datasets such as, prostate_tumor VSNormal and SRBCT datasets. Classification accuracy remains same for XOR and CFS methods in the prostate_tumor VSNormal dataset, similarly for XOR and ReliefF methods in the SRBCT datasets. SRBCT dataset yields better classification accuracy in ReliefF, XOR and Intersection methods. Prostate_tumor VSNormal dataset yields better classification accuracy for CFS, XOR and Intersection methods.

**Table 4** Classification accuracy of various feature selection methods with decision stump

| *Dataset* | */CA-IG/ (%)* | */CA-ReliefF/ (%)* | */CA-CFS/ (%)* | */CA-XOR/ (%)* | */CA-ISN/ (%)* |
|---|---|---|---|---|---|
| Lung cancer | 75 | 75 | 75 | 71.88 | 90.63 |
| Prostate_tumorVSNormal | 100 | 97.06 | 91.18 | 91.18 | 91.18 |
| Ovarian | 97.23 | 96.84 | 97.23 | 97.23 | 97.23 |
| Ozone | 97.12 | 97.12 | 97.12 | 97.12 | 97.12 |
| SRBCT | 54.22 | 54.22 | 54.22 | 54.22 | 54.22 |
| Gas | 60.29 | 61.03 | 61.03 | 61.03 | 61.03 |
| lsvt-voice-rehabilitation | 75.86 | 75.86 | 75.86 | 75.86 | 75.86 |
| mfeat Fourier | 56.98 | 56.98 | 56.98 | 56.98 | 56.98 |

Table 4 shows the comparative results of classification accuracy of various feature selection methods with decision stump classifiers. From the given eight datasets, the hybrid intersection method yields a better classification accuracy of the lung cancer dataset and meanwhile classification accuracy remain same for six datasets: ovarian, ozone, Gas, SRBCT, lsvt-voice-rehabilitation and mfeat Fourier datasets. Significantly XOR yields same classification accuracy for six datasets: Ovarian, Ozone, Gas, SRBCT, lsvt-voice-rehabilitation and mfeat Fourier datasets. Prostate_tumor VSNormal dataset gives the better classification accuracy in Information Gain. Ozone, SRBCT, lsvt-voice-rehabilitation and mfeat Fourier datasets yield same classification accuracy for all selected features selection methods.

**Table 5** Classification accuracy of various feature selection methods with random forest

| *Dataset* | */CA-IG/ (%)* | */CA-ReliefF/ (%)* | */CA-CFS/ (%)* | */CA-XOR/ (%)* | */CA-ISN/ (%)* |
|---|---|---|---|---|---|
| Lung Cancer | 81.25 | 81.25 | 84.36 | 71.88 | 90.63 |
| Prostate_tumorVSNormal | 97.06 | 100 | 100 | 97.06 | 100 |
| Ovarian | 94.86 | 98.02 | 98.81 | 90.12 | 98.81 |
| Ozone | 96.92 | 97.16 | 97.12 | 97.12 | 96.81 |
| SRBCT | 98.8 | 96.39 | 96.39 | 89.16 | 98.8 |
| Gas | 96.32 | 97.79 | 97.79 | 97.06 | 97.79 |
| lsvt-voice-rehabilitation | 77.01 | 86.21 | 82.76 | 89.66 | 83.91 |
| mfeat Fourier | 91.57 | 92.73 | 92.44 | 93.75 | 93.17 |

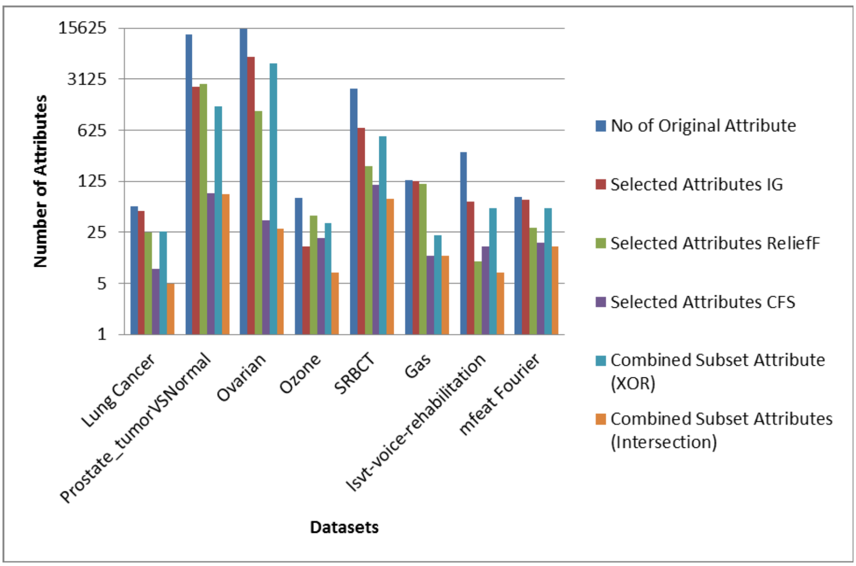
Table 5 illustrates the comparative results of classification accuracy of various feature selection methods with random forest classifiers. From the given eight datasets, the hybrid intersection method yields better classification accuracy for two datasets: lung cancer and SRBCT dataset. Subsequently, the classification accuracy remains the same for the other three datasets such as ovarian, prostate_tumor VSNormal and gas datasets. The hybrid XOR method yields better classification accuracy for two datasets, lsvt-voice-rehabilitation and mfeat Fourier datasets.

**Table 6** Classification accuracy of various feature selection methods with NB tree

| *Dataset* | */CA-IG/ (%)* | */CA-ReliefF/ (%)* | */CA-CFS/ (%)* | */CA-XOR/ (%)* | */CA-ISN/ (%)* |
|---|---|---|---|---|---|
| Lung cancer | 78.13 | 78.13 | 90.63 | 68.75 | 90.63 |
| Prostate_tumorVSNormal | 100 | 100 | 100 | 100 | 100 |
| Ovarian | 97.81 | 98.42 | 99.6 | 98.14 | 99.6 |
| Ozone | 96.96 | 96.53 | 96.85 | 96.81 | 97 |
| SRBCT | 100 | 98.8 | 100 | 98.8 | 100 |
| Gas | 97.06 | 91.18 | 94.85 | 92.63 | 94.85 |
| lsvt-voice-rehabilitation | 77.01 | 81.61 | 79.31 | 78.16 | 85.06 |
| mfeat Fourier | 87.36 | 89.54 | 88.66 | 87.35 | 88.81 |

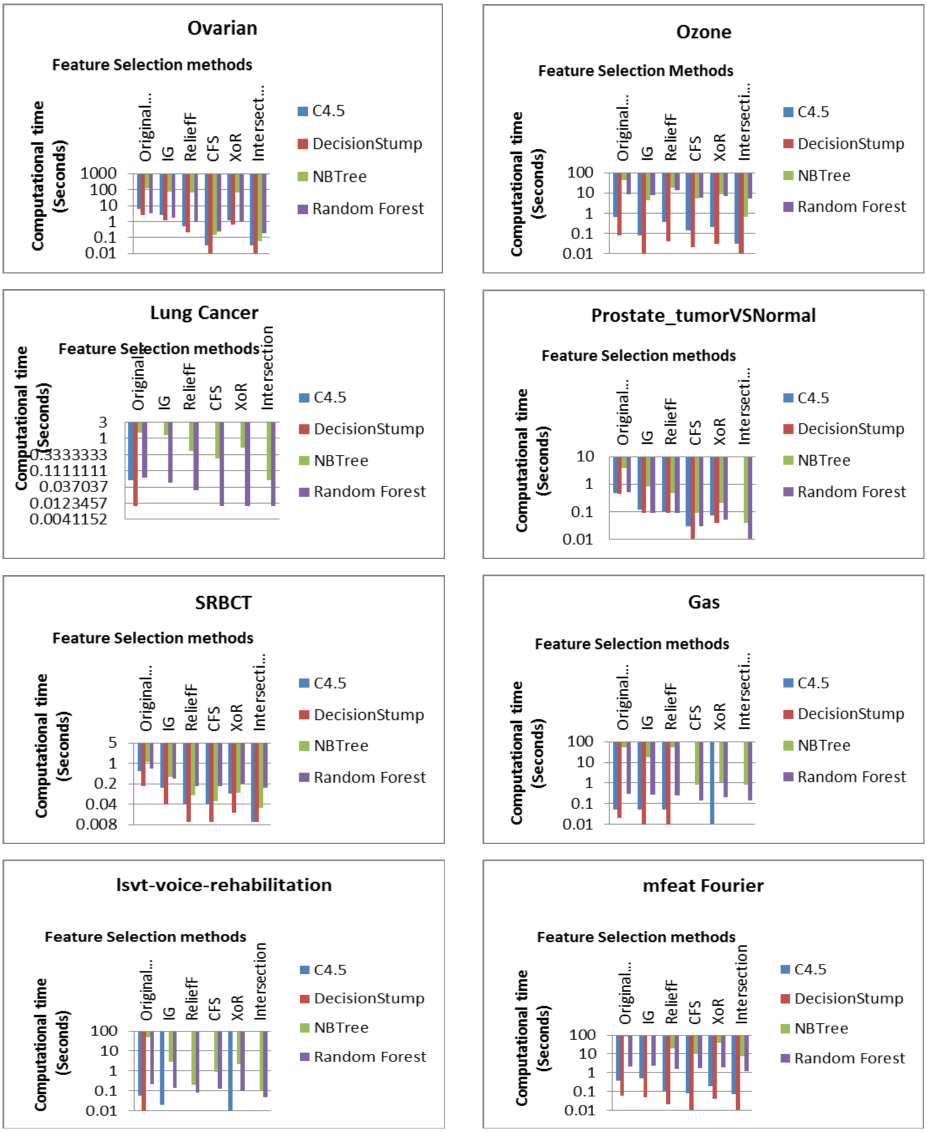
Table 6 show the comparative results of classification accuracy of various feature selection methods with naive Bayes tree classifiers. From the given eight datasets, the hybrid intersection method yields better classification accuracy for two datasets: ozone and lsvt-voice-rehabilitation dataset. Consequently the classification accuracy remains same for four datasets such as lung cancer, Prostate_tumorVSNormal, ovarian and SRBCT datasets. Then hybrid XOR yields the same classification accuracy for Prostate_tumorVSNormal. Gas dataset yields better classification accuracy in information gain and respectively mfeat Fourier dataset yields better classification accuracy in ReliefF. The Prostate_tumorVSNormal is the only dataset that generates 100% classification accuracy for all the feature selection methods.

**Figure 5** Number of feature selected with various datasets (see online version for colours)




The hybrid feature selection approach shows better classification accuracy in overall decision tree classifier methods. Figure 5 shows the number of features selected in various feature selection methods such as IG, ReliefF, CFS, XOR and intersection. The graph illustrates that the minimum five features are selected in intersection methods and maximum 15,155 features are in the ovarian dataset. The proposed hybrid method Intersection selects very limited features in every datasets. Likewise hybrid method XOR also selects less features, however, both these methods yields the better classification accuracy for all the selected classifiers. Figure 6 shows the computational time in seconds, for various datasets with different feature selection methods also with various classifiers such as C4.5, decision stump, random forest and naive Bayes tree classifiers. The figure clearly shows that the hybrid method intersection always generates very low computational time while comparing to other feature selection methods.

**Figure 6** Computational time (seconds) for various datasets with different feature selection methods (see online version for colours)




## 4.3 *Statistical comparison of multiple feature selection methods: performing all pairwise comparison*

Computation of the rankings is done with various feature selection methods and statistical test of all pairwise comparison is done with Hypotheses ordered by *p*-value and adjusting of a by Holm and Shaffer static method, considering an initial $\alpha = 0.05$.

A set of pairwise comparisons can be connected with a set or family of hypotheses. The test statistics for comparing the $i^{\text{th}}$ and $j^{\text{th}}$ classifier are $Z = (R_i - R_j) / \sqrt{k(K+1)/6N}$

where $R_i$ is the average rank calculated from the Friedman test for the $i^{th}$ classifier, $k$ is the number of classifiers to be associated and $N$ is the number of datasets used in the comparison. The Z value is used to find the consistent p-value, i.e., probability from the table of the normal distribution, which is then compared with a suitable level of significance $\alpha$. The tests differ in the way they adjust the value of $\alpha$ to compensate for multiple comparisons.

**Table 7** Rankings for the feature selection methods using C4.5 classifier

| | *IG* | *ReliefF* | *CFS* | *XOR* | *ISN* |
|---|---|---|---|---|---|
| Lung cancer | 0.75 (3.5) | 0.75 (3.5) | 0.781(2) | 0.719 (5) | 0.938 (1) |
| Prosts | 0.853 (4.5) | 0.853 (4.5) | 0.941 (2) | 0.941 (2) | 0.941 (2) |
| Ovarian | 0.964 (2) | 0.976 (1) | 0.96 (5) | 0.961 (3.5) | 0.961(3.5) |
| Ozone | 0.969 (2.5) | 0.968 (4) | 0.969 (2.5) | 0.967 (5) | 0.971 (1) |
| SRBCT | 0.855 (4.5) | 0.868 (2) | 0.855 (4.5) | 0.868 (2) | 0.868 (2) |
| Gas | 0.927 (4) | 0.919 (5) | 0.949 (1.5) | 0.941 (3) | 0.949 (1.5) |
| lsvt-voice-rehabilitation | 0.736 (5) | 0.828 (2) | 0.805 (3) | 0.793 (4) | 0.851 (1) |
| mfeat Fourier | 0.88 (4) | 0.879 (5) | 0.884 (2) | 0.882 (3) | 0.887 (1) |
| Average rank | 3.750 | 3.375 | 2.813 | 3.438 | 1.625 |

**Table 8** Hypotheses ordered for C4.5classifier by p-value and adjusting by Nemenyi, Holm and Shaffer static method, considering an initial $\alpha = 0.05$

| *i* | *Hypothesis* | *Z = (R₀ –Rᵢ)/SE* | *P* | $\alpha_{NM}$ | $\alpha_{HM}$ | $\alpha_{SM}$ |
|---|---|---|---|---|---|---|
| 1 | IG vs ISN | 2.686 | 0.007 | 0.005 | 0.005 | 0.005 |
| 2 | ISN vs XOR | 2.675 | 0.008 | 0.005 | 0.0055 | 0.0083 |
| 3 | ISN vs ReliefF | 2.212 | 0.027 | 0.005 | 0.0063 | 0.0083 |
| 4 | ISN vs CFS | 1.502 | 0.133 | 0.005 | 0.0071 | 0.0083 |
| 5 | IG vs CFS | 1.185 | 0.236 | 0.005 | 0.0083 | 0.0083 |
| 6 | XOR vs CFS | 0.790 | 0.430 | 0.005 | 0.01 | 0.0125 |
| 7 | CFS vs ReliefF | 0.710 | 0.478 | 0.005 | 0.0125 | 0.0125 |
| 8 | IG vs ReliefF | 0.474 | 0.636 | 0.005 | 0.0167 | 0.0167 |
| 9 | IG vs XOR | 0.394 | 0.694 | 0.005 | 0.025 | 0.025 |
| 10 | XOR vs ReliefF | 0.080 | 0.936 | 0.005 | 0.05 | 0.05 |

The standard error in the pairwise comparison between two classifiers is considered as $SE = \sqrt{k(K+1)/6N}$. Table 8 presents the family of hypotheses ordered by their p-value.

Holm and Shaffer are the two procedures used as hypothesis ordered by p-value and adjusting $\alpha$. The same procedures were also used by Demsar (2006) and Garcia and Herrera (2008) for comparisons of multiple classifiers including a control technique. Using step down approach it adjusts the value of $\alpha$. Let $p_1, \ldots, p_m$ be the ordered p-values (ascending order) and $H_1, \ldots, H_m$ be the consistent hypotheses. Holm's procedure rejects $H_1$ to $H_{(i-1)}$ if $i$ is the least integer such that $pi > \alpha = (m - i + 1)$ where $m = k(k - 1)/2$. From Table 8, both Nemenyi and Holm's test rejects null hypothesis, and Shaffer test rejects the Hypothesis 2 subsequently the corresponding p-values are smaller than the adjusted $\alpha$'s

**Table 9** Rankings for the feature selection methods using decision stump classifier

| | *IG* | *ReliefF* | *CFS* | *XOR* | *INS* |
|---|---|---|---|---|---|
| Lung cancer | 0.75 (3) | 0.75 (3) | 0.75 (3) | 0.719 (5) | 0.906 (1) |
| Prosts | 1.0 (1) | 0.971 (2) | 0.912 (4) | 0.912 (4) | 0.912 (4) |
| Ovarian | 0.972 (2.5) | 0.968 (5) | 0.972 (2.5) | 0.972 (2.5) | 0.972 (2.5) |
| Ozone | 0.971 (3) | 0.971 (3) | 0.971 (3) | 0.971 (3) | 0.971 (3) |
| SRBCT | 0.542 (3) | 0.542 (3) | 0.542 (3) | 0.542 (3) | 0.542 (3) |
| Gas | 0.603 (5) | 0.61 (2.5) | 0.61 (2.5) | 0.61 (2.5) | 0.61 (2.5) |
| lsvt-voice-rehabilitation | 0.759 (3) | 0.759 (3) | 0.759 (3) | 0.759 (3) | 0.759 (3) |
| mfeat Fourier | 0.57 (3) | 0.57 (3) | 0.57 (3) | 0.57 (3) | 0.57 (3) |
| Average rank | 2.938 | 3.063 | 3.000 | 3.250 | 2.750 |

**Table 10** Hypotheses ordered for decision stump classifier by p-value and adjusting by Nemenyi, Holm and Shaffer static method, considering an initial $\alpha = 0.05$

| *i* | *Hypothesis* | $Z = (R_0 - R_i)/SE$ | *P* | $\alpha_{NM}$ | $\alpha_{HM}$ | $\alpha_{SM}$ |
|---|---|---|---|---|---|---|
| 1 | ISN vs XOR | 0.632 | 0.527 | 0.005 | 0.005 | 0.005 |
| 2 | ISN vs ReliefF | 0.396 | 0.692 | 0.005 | 0.0055 | 0.0083 |
| 3 | IG vs XOR | 0.394 | 0.694 | 0.005 | 0.0063 | 0.0083 |
| 4 | ISN vs CFS | 0.316 | 0.752 | 0.005 | 0.0071 | 0.0083 |
| 5 | XOR vs CFS | 0.316 | 0.752 | 0.005 | 0.0083 | 0.0083 |
| 6 | IG vs ISN | 0.300 | 0.764 | 0.005 | 0.01 | 0.0125 |
| 7 | XOR vs ReliefF | 0.236 | 0.813 | 0.005 | 0.0125 | 0.0125 |
| 8 | IG vs ReliefF | 0.158 | 0.875 | 0.005 | 0.0167 | 0.0167 |
| 9 | CFS vs ReliefF | 0.080 | 0.936 | 0.005 | 0.025 | 0.025 |
| 10 | IG vs CFS | 0.078 | 0.938 | 0.005 | 0.05 | 0.05 |

Shaffer's (1986) procedure has succeeded Holm's step down approach, at $j^{th}$ stage, instead of eliminating $H_i$ if $p_i \leq \alpha = (m - i + 1)$, eliminate $H_i$ if $p_i \leq \alpha = t_i$, where $t_i$ is the maximum number of hypotheses which can be true specified that any $(i - 1)$ hypotheses are false. Shaffer is a static method, that is, $t_1, \ldots, t_m$ the term entirely determined for the given Hypotheses $H_1, \ldots, H_m$, independent of the observed p-values. The probable numbers of true hypotheses, and therefore the values of $t_i$ can be attained from the recursive procedure $S(k) = \bigcup_{j=1}^{k} \{\binom{2}{j} + x : x \in S(k-j)\}$ where $S(k)$ is the set of probable numbers of true hypotheses with $k$ classifiers being compared, $k = 2$, and $S(0) = S(1) = \{0\}$.

**Table 11** Rankings for the feature selection methods using random forest classifier

| | *IG* | *ReliefF* | *CFS* | *XOR* | *INS* |
|---|---|---|---|---|---|
| Lung cancer | 0.813 (3.5) | 0.813 (3.5) | 0.844 (2) | 0.719 (5) | 0.906 (1) |
| Prosts | 0.971 (4.5) | 1.000 (2) | 1.000 (2) | 0.971 (4.5) | 1.000 (2) |
| Ovarian | 0.949 (4) | 0.980 (3) | 0.988 (1.5) | 0.901 (5) | 0.988 (1.5) |
| Ozone | 0.969 (4) | 0.972 (1) | 0.971 (2.5) | 0.971 (2.5) | 0.968 (5) |
| SRBCT | 0.988 (1.5) | 0.964 (3.5) | 0.964 (3.5) | 0.892 (5) | 0.988 (1.5) |
| Gas | 0.963 (5) | 0.978 (2) | 0.978 (2) | 0.971 (4) | 0.978 (2) |
| lsvt-voice-rehabilitation | 0.770 (5) | 0.862 (2) | 0.828 (4) | 0.897 (1) | 0.839 (3) |
| mfeat Fourier | 0.916 (5) | 0.927 (3) | 0.924 (4) | 0.938 (1) | 0.932 (2) |
| Average rank | 4.063 | 2.500 | 2.688 | 3.500 | 2.250 |

**Table 12** Hypotheses ordered for random forest classifier by p-value and adjusting by Nemenyi, Holm and Shaffer static method, considering an initial $\alpha = 0.05$

| *i* | *Hypothesis* | $Z = (R_0 - R_i)/SE$ | *P* | $\alpha_{NM}$ | $\alpha_{HM}$ | $\alpha_{SM}$ |
|---|---|---|---|---|---|---|
| 1 | IG vs ISN | 2.292 | 0.022 | 0.005 | 0.005 | 0.005 |
| 2 | IG vs ReliefF | 1.976 | 0.048 | 0.005 | 0.0055 | 0.0083 |
| 3 | IG vs CFS | 1.738 | 0.082 | 0.005 | 0.0063 | 0.0083 |
| 4 | ISN vs XOR | 1.580 | 0.114 | 0.005 | 0.0071 | 0.0083 |
| 5 | XOR vs ReliefF | 1.264 | 0.206 | 0.005 | 0.0083 | 0.0083 |
| 6 | XOR vs CFS | 1.027 | 0.304 | 0.005 | 0.01 | 0.0125 |
| 7 | IG vs XOR | 0.712 | 0.477 | 0.005 | 0.0125 | 0.0125 |
| 8 | ISN vs CFS | 0.554 | 0.580 | 0.005 | 0.0167 | 0.0167 |
| 9 | ISN vs ReliefF | 0.316 | 0.752 | 0.005 | 0.025 | 0.025 |
| 10 | CFS vs ReliefF | 0.238 | 0.812 | 0.005 | 0.05 | 0.05 |

**Table 13** Rankings for the feature selection methods using NB tree classifier

| | *IG* | *ReliefF* | *CFS* | *XOR* | *INS* |
|---|---|---|---|---|---|
| Lung cancer | 0.781 (3.5) | 0.781 (3.5) | 0.906 (1.5) | 0.688 (5) | 0.906 (1.5) |
| Prosts | 1.000 (3) | 1.000 (3) | 1.000 (3) | 1.000 (3) | 1.000 (3) |
| Ovarian | 0.978 (5) | 0.984 (3) | 0.996 (1.5) | 0.981 (4) | 0.996 (1.5) |
| Ozone | 0.970 (1.5) | 0.965 (5) | 0.969 (3) | 0.968 (4) | 0.970 (1.5) |
| SRBCT | 1.000 (2) | 0.988 (4.5) | 1.000 (2) | 0.988 (4.5) | 1.000 (2) |
| Gas | 0.971 (1) | 0.912 (5) | 0.949 (2.5) | 0.926 (4) | 0.949 (2.5) |
| lsvt-voice-rehabilitation | 0.770 (5) | 0.816 (2) | 0.793 (3) | 0.782 (4) | 0.851 (1) |
| mfeat Fourier | 0.874 (4.5) | 0.895 (1) | 0.887 (3) | 0.874 (4.5) | 0.888 (2) |
| Average rank | 3.188 | 3.375 | 2.438 | 4.125 | 1.875 |

The results agree to average accuracy in test data and used with eight datasets. Tables 7, 9, 11 and 13 illustrate the complete process of calculation of average rankings. The average ranking for various feature selection methods such as Information gain, ReliefF, Correlation-based feature selection, XOR and intersection using with several classifiers namely C4.5, decision stump, random forest and NB tree classifiers. Friedman test is used to rank all the feature selection methods and average ranking are calculated.

**Table 14** Hypotheses ordered for NB Tree by p-value and adjusting by Nemenyi, Holm and Shaffer static method, considering an initial $\alpha = 0.05$

| *i* | *Hypothesis* | $Z = (R_0 - R_i)/SE$ | *P* | $\alpha_{NM}$ | $\alpha_{HM}$ | $\alpha_{SM}$ |
|---|---|---|---|---|---|---|
| 1 | ISN vs XOR | 2.845 | 0.004 | 0.005 | 0.005 | 0.005 |
| 2 | XOR vs CFS | 2.133 | 0.033 | 0.005 | 0.0055 | 0.0083 |
| 3 | ISN vs ReliefF | 1.896 | 0.058 | 0.005 | 0.0063 | 0.0083 |
| 4 | IG vs ISN | 1.660 | 0.097 | 0.005 | 0.0071 | 0.0083 |
| 5 | IG vs XOR | 1.185 | 0.236 | 0.005 | 0.0083 | 0.0083 |
| 6 | CFS vs ReliefF | 1.185 | 0.236 | 0.005 | 0.01 | 0.0125 |
| 7 | IG vs CFS | 0.948 | 0.343 | 0.005 | 0.0125 | 0.0125 |
| 8 | XOR vs ReliefF | 0.948 | 0.343 | 0.005 | 0.0167 | 0.0167 |
| 9 | ISN vs CFS | 0.712 | 0.477 | 0.005 | 0.025 | 0.025 |
| 10 | IG vs ReliefF | 0.236 | 0.813 | 0.005 | 0.05 | 0.05 |

Once p-value is within a multiple comparison, as per the illustration in Tables 8, 10, 12 and 14, it imitates the probability error of a definite comparison. Tables 8, 10, 12 and 14 display hypotheses ordered for various classifiers ordered by p-value and adjusting of a by Holm and Shaffer static method, considering an initial $\alpha = 0.05$. Table 14 shows the Hypothesis 1 is rejected for Nemenyi, Holm's and Shaffer test, subsequently the corresponding p-values are smaller than the adjusted $\alpha$'s. The hypotheses orders for table 10 and 12 null hypotheses will be rejected by Nemenyi, Holm and Shaffer static methods.

## 5 Conclusions

In this paper, the hybrid feature subset selection for multilevel datasets with the decision tree classifiers is proposed, where two different fusion methods, namely Intersection and Exclusive OR are used for combining the feature subsets to increase the classification accuracy. The results of the hybrid feature subset selection methods have exposed that the intersection method were highly active in reducing dimensionality and they could generate a better reduction rate in the second stage of tuning. The comparative results are accomplished using various feature selection methods such as IG, ReliefF, CFS, XOR and intersection also classified with four decision tree classifiers.

As a result, the hybrid feature subset selection ignores the less important features and selects the highly important features. The proposed hybrid feature selection method was used to improve the classification accuracy with the minimum number of feature in various datasets. The statistical test comparison to perform pairwise comparison is used and the intersection approach shows the higher ranking in all classifiers. The hybrid feature subset selection algorithm was proposed, based on the intersection and the XOR method to build the efficient and constant preprocessing techniques. The hybrid feature selection technique works effectively in high dimensional data and overcomes the problem in various decision tree classifiers. Experimental comparative effects show that the hybrid feature subset selection of the multilevel data, being highly stable, and with improved classification accuracy. The intersection method takes less computational time as compared to all other methods. Thus, the hybrid feature selection method can be adopted for the data preprocessing methods.

## References


Apollonia, J., Leguizamona, G. and Alba, E. (2016) 'Two hybrid wrapper-filter feature selection algorithms applied to high-dimensional microarray experiments', *Applied Soft Computing*, Vol. 38, pp.922–932.

Asuncion, A. and Newman, D. (2007) *UCI Machine Learning Repository* [online] https://archive.ics.uci.edu/ml/about.html (accessed 23 March 2020).

Asuncion, A. and Newman, D.J. (2012) *UCI Machine Learning Repository*, School of Information and Computer Sciences, University of California, Irvine [online] http://mlearn.ics.uci.edu/MLRepository.html (accessed April).

Ben Brahima, A. and Limam, M. (2016) 'A hybrid feature selection method based on instance learning and cooperative subset search', *Pattern Recognition Letters*, Vol. 69, pp.28–34.

Bolon-Canedo, V., Porto-Diaz, I., Sanchez-Marono, N. and Alonso-Betanzos, A. (2014) 'A frame work for cost-based feature selection', *Pattern Recognition*, Vol. 47, pp.2481–2489.

Breiman, L. (2001) 'Random forests', *Machine Learning*, Vol. 45, pp.5–32.

Calderoni, L., Ferrara, M., Franco, A. and Maio, D. (2015) 'Indoor localization in a hospital environment using random forest classifiers', *Expert Systems with Applications*, Vol. 42, No. 1, pp.125–134.

Chen, K-H., Wang, K-J., Tsai, M-L., Wang, K-M., Adrian, A.M., Cheng, W-C., Yang, T-S., Teng, N-C., Tan, K-P. and Chang, K-S. (2014) 'Gene selection for cancer identification: a decision tree model empowered by particle swarm optimization algorithm', *BMC Bioinformatics*, Vol. 15, p.49.

Cheng, J-H., Chen, H-P. and Lin, Y-M. (2010) 'A hybrid forecast marketing timing model based on probabilistic neural network rough set and C4.5', *Expert Systems with Applications*, Vol. 37, No. 3, ,pp.1814–1820.

Chiew, K.L., Tan, C.L., Wong, K.S., Yong, K.S.C. and Tiong, W.K. (2019) 'A new hybrid ensemble feature selection framework for machine learning-based phishing detection system', *Information Sciences*, May, Vol. 484, pp.153–166, Elsevier.

Chuang, L-Y., Yang, C-H., Wu, K-C. and Yang, C-H. (2011) 'A hybrid feature selection method for DNA microarray data', *Computers in Biology and Medicine*, Vol. 41, No. 4, pp.228–237, Elsevier.

Dag, H., Sayin, K.E., Yenidogan, I., Albayrak, S. and Acar, C. (2012) 'Comparison of feature selection algorithms for medical Data', *Innovations in Intelligent Systems and Applications (INISTA)*, pp.1–5, IEEE, doi: 10.1109/INISTA.2012.6247011.

Demsar, J. (2006) 'Statistical comparisons of classifiers over multiple data sets', *Journal of Machine Learning Research*, Vol. 7, pp.1–30.

Garcia, S. and Herrera, F. (2008) 'An extension on 'Statistical comparisons of classifiers over multiple data sets for all pairwise comparisons', *Journal of Machine Learning Research*, Vol. 9, pp.2677–2694.

Ghareb, A.S., Abu Bakar, A. and Hamdan, A.R. (2016) 'Hybrid feature selection based on enhanced genetic algorithm for text categorization', *Expert Systems with Applications*, Vol. 49, pp.31–47.

Guyon, I., Gunn, S., Nikravesh, M. and Zadeh, L. (2006) *Feature Extraction Foundations and Applications*, Springer-Verlag, Berlin Heidelberg, ISBN: 978-3-540-35488-8.

Hall, M.A. (1999) *Correlation-Based Feature Subset Selection for Machine Learning*, PhD thesis, Department of Computer Science, University of Waikato, Hamilton, New Zealand.

Hall, M.A. (2000) 'Correlation-based feature selection for discrete and numeric class machine learning', *Proceedings of the 17th International Conference on Machine Learning*, Stanford University/Morgan Kaufmann Publishers, California.

Han, J. and Kamber, M. (2011) *Data Mining: Concepts and Techniques*, 3rd ed., Morgan Kaufmann, MA, USA.

Hsu, H-H., Hsieh, C-W. and Lu, M-D. (2011) 'Hybrid feature selection by combining filters and wrappers', *Expert Systems with Applications*, Vol. 38, No. 7, pp.8144–8150, Elsevier.

Iba, W. and Langley, P. (1992) 'Induction of one-level decision trees', *ML92: Proceedings of the Ninth International Conference on Machine Learning*, July, pp.233–240, Morgan Kaufmann.

Jian, Z., Shirai, H., Takahashi, I., Kuroiwa, J., Odaka, T. and Ogura, H. (2007) 'Masquerade detection by boosting decision stumps using UNIX commands', *Computers & Security*, Vol. 26, No. 4, pp.311–318.

Jiang, L.X. and Cai, Z. (2009) 'Learning decision tree for ranking', *Knowl. Inf. Syst.*, Vol. 20, No. 1, pp.123–135.

Kannan, S.S. and Ramaraj, N. (2010) 'A Novel hybrid feature selection via Symmetrical Uncertainty ranking based local memetic search algorithm', *Knowledge-Based Systems*, Vol. 23, No. 6, pp.580–585, Elsevier.

Kohavi, R. (1996) 'Scaling up the accuracy of naive-bayes classifiers: a decision-tree hybrid', *Proceedings of the Second International Conference on Knowledge Discovery and Data Mining*, pp.202–207, AAAI Press.

Kohavi, R. and John, G.H., (1997) 'Wrappers for feature subset selection', *Artificial Intelligence*, Vol. 97, pp.273–324, Elsevier.

Kononenko, I. (1994) 'Estimating attributes: Analysis and extensions of RELIEF', *Proceeding of Euro Conference in Machine Learning*, pp.171–182.

Kononenko, I., Simec, E. and Sikonja, M.R. (1997) 'Overcoming the myopia of inductive learning algorithms with ReliefF', *Appl. Int.*, Vol. 7, No. 1, pp.39–55.

Koprinska, I., Rana, M. and Agelidis, V.G. (2015) 'Correlation and instance based feature selection for electricity load forecasting', *Knowledge-Based Systems*, July, Vol. 82, pp.29–40.

Kullback, S. (1952) 'An application of information theory to multivariate analysis', *Annals of Mathematical Statistics*, Vol. 23, No. 2, pp.88–102.

Lin, H-Y. (2013) 'Feature selection based on cluster and variability analyses for ordinal multi-class classification problems', *Knowledge-Based Systems*, January, Vol. 37, pp.94–104.

Liu, H. and Dougherty, E.R. (2005) 'Evolving feature selection', *IEEE Intelligent Systems*, Vol. 20, No. 6, pp.64–76.

Liu, H., Wei, R. and Jiang, G. (2013) 'A hybrid feature selection scheme for mixed attributes data', *Comp. Appl. Math*, April, Vol. 32, No. 1, pp.145–161.

Liu, Y., Wang, Y., Feng, L. and Zhu, X. (2014) 'Term frequency combined hybrid feature selection method for spam filtering', *Pattern Anal. Applic.*, May, Vol. 19, No. 2, pp.369–383, Springer-Verlag, London.

Menghour, K. and Souici-Meslati, L. (2014) 'Classical and swarm-based approaches for feature selection in spam filtering', *Int. J. of Advanced Intelligence Paradigms*, Vol. 6, No. 3, pp.214–234.

Quinlan, J.R. (1986) 'Induction of decision trees', *Machine Learning*, Vol. 1, No. 1, pp.81–106.

Quinlan, J.R. (1993) 'C4.5: programs for machine learning', *Machine Learning*, September, Vol. 16, No. 3, pp.235–240, Morgan Kaufmann Publishers, Inc.

Quinlan, J.R. (1996) 'Improved use of continuous attributes in C4.5', *Journal of Artificial Intelligence Research*, Vol. 4, pp.77–90.

Russell, S.J. and Norvig, P. (2003) *Artificial Intelligence: A Modern Approach*, 2nd ed., Upper Saddle River, New Jersey, pp.94, 95, note 3, Prentice Hall, ISBN 0-13-790395-2.

Sangaiah, I. and Kumar, A.V.A. (2018) 'Improving medical diagnosis performance using hybrid feature selection via ReliefF and entropy based genetic search (RF-EGA) approach: application to breast cancer prediction', *Cluster Computing*, Vol. 22, No. 3, pp.6899–6906, Springer.

Shaffer, J.P. (1986) 'Modified sequentially rejective multiple test procedures', *Journal of the American Statistical Association*, Vol. 81, No. 395, pp.826–831.

Stanczyk, U. and Jain, L.C. (2015) 'Feature selection for data and pattern recognition', *Studies in Computational Intelligence*, Springer, Chapter 1, pp.1–7, Springer-Verlag Berlin Heidelberg Publisher, ISBN 978-3-662-45619-4.

Tan, T-Z. and Liang, Y-Y. (2012) 'ASF/DT, adaptive step forward decision tree construction', *Proceedings of the 2012 International Conference on Wavelet Analysis and Pattern Recognition*, 15–17 July, Xian.

Tung, K.Y., Huang, I.C., Chen, S.L. and Shih, C.T. (2005) 'Mining the generation Xers' job attitudes by artificial neural network and decision tree-empirical evidence in Taiwan', *Expert Systems with Applications*, Vol. 29, No. 4, pp.783–794, Elsevier.

Wang, J. and Wu, L. (2013) 'Maximum weight and minimum redundancy: a novel frame-work for feature subset selection', *Pattern Recognition*, Vol. 46, pp.1616–1627.

Wang, S., Jiang, L. and Li, C. (2015) 'Adapting naive Bayes tree for text classification', *Knowl. Inf. Syst.*, July, Vol. 44, No. 1, pp.77–89.

Wang, Y. and Feng, L. (2019) 'A new hybrid feature selection based on multi-filter weights and multi-feature weights', *Applied Intelligence*, Vol. 49, No. 12, pp.1–25, Springer.

Wei, P., Ma, P., Hu, Q., Su, X. and Ma, C. (2014) 'Comparative analysis on margin based feature selection algorithms', *Int. J. Mach. Learn. & Cyber.*, June, Vol. 5, No. 3, pp.339–367.

Witten, I.H. and Frank, E. (2009) *Data Mining: Practical Machine Learning Tools and Techniques*, 2nd ed., Morgan Kaufmann.

Wu, X., Kumar, V., Quinlan, J., Ghosh, J., Yang, Q., Motoda, H., McLachlan, G.J., Ng, A., Liu, B., Yu, P..S, Zhou, Z., Steinbach, M., Hand, D.J. and Steinberg, D. (2008) 'Top 10 algorithms in data mining', *Knowl. Inf. Syst.*, Vol. 14, No. 1, pp.1–37.

Xie, J., Lei, J., Xie, W., Shi, Y. and Liu, X. (2013) 'Two-stage hybrid feature selection algorithms for diagnosing erythemato-squamous diseases', *Health Information Science & Systems*, December, Vol. 1, No. 1, pp.1–10, Springer.

Yang, Y., Liao, Y.X., Meng, G. and Lee, J. (2011) 'A hybrid feature selection scheme for unsupervised learning and its application in bearing fault diagnosis', *Expert Syst. Appl.*, Vol. 38, pp.11311–11320.

Zeng, Z., Zhang, H., Zhang, R. and Yin, C.X. (2015) 'A novel feature selection method considering feature interaction', *Pattern Recognition*, August, Vol. 48, No. 8, pp.2656–2666.