Estimation and Feature Selection by Application of Knowledge Mined from Decision Rules Models

Wieslaw Paja[1] and Krzysztof Pancerz[2,3]

[1] Department of Computer Science,
Faculty of Mathematics and Natural Sciences,
University of Rzeszow, Prof. St. Pigonia Str. 1, 35-310 Rzeszow, Poland
`wpaja@ur.edu.pl`
[2] University of Information Technology and Management
Sucharskiego Str. 2, 35-225 Rzeszow, Poland
[3] University of Management and Administration
Akademicka Str. 4, 22-400 Zamosc, Poland
`kpancerz@wszia.edu.pl`

Abstract. Feature selection methods, as a preprocessing step to machine learning, are effective in reducing dimensionality, removing irrelevant data, increasing learning accuracy, and improving result comprehensibility. However, the recent increase of dimensionality of data poses a severe challenge to many existing feature selection methods with respect to the efficiency and effectiveness. In this work, we introduce a novel concept, relevant feature selection based on information gathered from decision rule models. A new measure for a feature rank based on the feature frequency and rule quality is additionally defined. The efficiency and effectiveness of our method is demonstrated through exemplary use of five real-world datasets. Six different classification algorithms were used to measure the quality of learning models built on original features and on selected features.

Key words: Feature selection, feature ranking, decision rules, dimensionality reduction, relevance and irrelevance

1 Introduction

In the era of the acquisition of vast amounts of data, different domain information databases, efficient analysis and retrieval of regularity have become an extremely important task. The issue of classification and object recognition is applied in many fields of human activity. Data mining is fraught with many aspects which hinder it, like a very large number of observations, too many attributes, the insignificance of the part of variables for the classification process, mutual interdependence of conditional variables, the simultaneous presence of variables with different types, the presence of undefined values of variables, the presence of erroneous values of the variables, uneven distribution of categories for the target variable. Thus, the development of efficient methods for significant feature selection is valid.

Feature selection (*FS*) methods are frequently used as a preprocessing step to machine learning experiments. An *FS* method can be defined as a process of choosing a

subset of original features so that the feature space is optimally reduced according to a certain evaluation criterion. Feature selection has been a fruitful field of research and development since 1970's and it has been proven to be effective in removing irrelevant features, increasing efficiency in learning tasks, improving learning performance like predictive accuracy, and enhancing comprehensibility of the learned results [1].

The feature selection methods are typically divided into three classes based on how they combine the selection algorithm and the model building: filter, wrapper and embedded *FS* methods. Filter methods select features with respect to the model. They are based only on general features like the correlation with the variable to be predicted. These methods select only the most interesting variables. Then, a selected subset will be a part of the classification model. Such methods are effective in computation time and robust to overfitting [2]. However, some redundant, but relevant features can remain unrecognized. In turn, wrapper methods evaluate subsets of features which allow to detect the possible interactions between variables [1, 3]. However, the increase in overfitting risk, when the number of observations is insufficient, is possible. Additionally, the significant computation time, when the number of variables is large, highly increases. The third type, called embedded methods, is intended for reducing the classification of learning. Methods in this group try to combine the advantages of both methods mentioned previously. Thus, the learning algorithm takes advantage of its own variable selection algorithm. Therefore, it needs to know initially what a good selection is, which limits its exploitation [4].

Kohavi and John [1] observed that there are several definitions of relevance that may be contradictory and misleading. They proposed two degrees of relevance (*strong* and *weak*) that are required to encompass all notions usually associated with this term. In their approach the relevance is defined in the absolute terms, with the help of the ideal Bayes classifier. In this context, a feature X is *strongly relevant* when removal of X alone from the data always results in deterioration of the prediction accuracy of the ideal Bayes classifier. In turn, a feature X is weakly relevant if it is not strongly relevant and there exists a subset of features S, such that the performance of the ideal Bayes classifier on S is worse than the performance on $S \cup \{X\}$. A feature is irrelevant if it is neither strongly nor weakly relevant.

Nilsson et al. [5] introduced the formal definition of two different feature selection problems:

1. *Minimal Optimal Feature Selection (MOFS)* consisting in identification of minimal set of features to obtain the optimal quality of classification.
2. *All Relevant Feature Selection (ARFS)*), where the problem is to find all the variables that may, under certain conditions, improve the classification.

There are two important differences between these problems. The first one is detection of attributes with low importance (*ARFS*) [6], which may be completely obscured by other, more important attributes, from the point of view of the classifier (*MOFS*). The second difference is to find the boundary between the variables poorly, but realistically related to the decision and those for which such a relation is created as a result of random fluctuations. The formal definition of the problem of all relevant feature selection (*ARFS*) as a distinct problem from the classical minimal optimal feature selection (*MOFS*), was proposed recently in 2007 [5].

In our research, we used the *contrast variable* concept to distinguish between relevant and irrelevant features [6]. It is a variable that does not carry information on the decision variable by design that is added to the system in order to discern relevant and irrelevant variables. Here, it is obtained from the real variables by random permutation of values between objects. The use of contrast variables was, for the first time, proposed by Stoppiglia et al. [7] and then by Tuv et al. [8].

2 Methods and Algorithms

During experiments the following general procedure was applied:

1. *Step 1. Selection of dataset and features for investigation.*
   (a) *Application of a set of ranking measures to calculate importance for each feature:*
       i. *With set of contrast features.*
       ii. *Without contrast features.*
   (b) *Definition (selection) of the most important feature subset.*
2. *Step 2. Application of different machine learning algorithms for classification of unseen objects using the 10-fold cross validation method:*
   (a) *Using all original features.*
   (b) *Using only selected, important features.*
3. *Step 3. Comparison of gathered results using different evaluation measures.*

In the first step, a dataset as well as a feature for investigation were defined. Then, different ranking measures were applied to estimate importance of each feature. In order to check specificity of the feature selection, the dataset was extended by adding contrast variables. It means that each original variable was duplicated and its values were randomly permuted between all objects. Hence, a set of non-informative by design shadow variables was added to original variables. The variables that were selected as important more significantly than random, were examined further, using different tests. To define the level of feature importance, six well-known ranking measures were applied: *ReliefF*, *Information Gain*, *Gain Ratio*, *Gini Index*, *SVM weight*, and *RandomForest*. Additionally, our new measure, called *RQualityFS*, was introduced. It is based on the frequency of presence of different feature in a rule model generated from an original dataset and it also takes into consideration the quality of the rules in which this feature occurs. Rank quality of the i-th feature could be presented as follow:

$$Q_{A_i} = \sum_{j=1}^{n} Q_{R_j}\{A_i\} \tag{1}$$

where n is a number of rules inside the model, Q_{R_j} defines the classification quality of the rule R_j and $\{A_i\}$ describes the presence of the i-th attribute, usually it is equal to 1 (the feature occurred) or to 0 (the feature did not occur).

In turn, the quality of the rule is defined as follows:

$$Q_{R_j} = \frac{E_{corr}}{E_{corr} + E_{incorr}} \tag{2}$$

where E_{corr} depicts a number of correctly matched learning objects by the j-th rule and E_{incorr} depicts a number of incorrectly matched learning objects by this rule.

During the second step, a test probing the importance of variables was performed by analyzing the influence of variables used for model building on the prediction quality. Six different machine learning algorithms were applied to build different predictors for the original set of features and for selected features: *Classification Tree* (*CT*), *Random Forest* (*RF*), *CN2 decision rules algorithm* (*CN2*), *Naive Bayes* (*NB*), *k-Nearest Neighbors* (*kNN*), and *Support Vector Machine* (*SVM*). During this step, a 10-fold cross validation paradigm was used. Ten known evaluation measures were uti-lized in each predictor: *Classification Accuracy* (*CA*), *Sensitivity*, *Specificity*, *Area Under ROC curve* (*AUC*), *Information Score* (*IS*), *F1 score* (*F1*), *Precision*, *Brier measure*, *Matthew Coefficient Correlation* (*MCC*) *parameter*, and finally *Informadness* (*Inform.*) ratio [9].

3 Investigated Datasets

Our initial investigations focus on applying the developed algorithm on several re-al-world datasets. Five datasets have been used during experiments. Four of them are gathered from the UCI ML repository, while the fifth set has been developed earlier by the authors [10]. A summary of datasets is presented in Table 1. These datasets have diverse numbers of objects, features and their types as well as classes.

Table 1. A summary characteristic of benchmark datasets

Dataset	# instances	# features	# classes
Breast cancer	286	9	2
Heart disease	303	13	2
Lung cancer	32	56	3
Primary tumor	339	17	21
Skin cancer	548	13	13

4 Results and Conclusions

To illustrate the proposed methodology, only results for *Breast cancer* datasets will be presented in details. The first step of the experiment revealed six features, that were recommended as important by all or almost all ranking measures. In Table 2, we can observe that *deg-malig*, *node-caps*, *irradiat*, *inv-nodes*, *breast*, and *menopause* features create a stable and core set of features which have the highest values of seven measures of importance, particularly using *RQualityFS* measure, introduced in our investigation. In the same table, comparison with importance of contrast values (*italic* rows and *contrast* index) is also presented. The most important contrast feature is *tumor-size* (*contrast*) for which *RQualityFS* measure, defined earlier, is equal to 2.34. In this way, we

also treated a threshold that separates the core, relevant set of attributes from other less informative attributes. Most of the measures (except *SVM weight*) used in this approach show that the selected set of features has higher values of these parameters than the gathered threshold value (underlined values). These values are denoted in bold style in Table 2. Hereby, we can observe that different measures give different thresholds.

Table 2. Ranking of features using seven different measures

Feature	ReliefF	Inf. gain	Gain Ratio	Gini	SVM weight	RF	RQualityFS
deg-malig	0.08	**0.08**	**0.05**	**0.02**	0.07	**2.01**	**8.78**
node-caps	**0.15**	**0.06**	**0.08**	**0.02**	0.05	**1.21**	**7.23**
irradiat	**0.13**	**0.03**	**0.03**	0.01	0.00	**0.88**	**4.94**
inv-nodes	**0.15**	**0.07**	**0.05**	**0.02**	0.03	0.32	**3.78**
breast	-0.01	0.00	0.00	0.00	0.06	0.32	**3.66**
menopause	-0.03	0.00	0.00	0.00	0.02	0.00	**3.14**
tumor-size (contrast)	*-0.01*	*0.01*	*0.00*	*0.00*	*0.04*	*0.01*	***2.34***
Age	0.01	0.01	0.01	0.00	0.04	-0.12	2.24
breast (contrast)	*0.01*	*0.00*	*0.00*	*0.00*	*0.00*	*0.06*	*2.05*
tumor-size	0.07	**0.06**	0.02	0.01	0.10	0.04	1.74
age (contrast)	*0.05*	*0.01*	*0.00*	*0.00*	*0.01*	*-0.06*	*1.27*
deg-malig (contrast)	*0.06*	*0.00*	*0.00*	*0.00*	*0.01*	***0.46***	*1.23*
menopause (contrast)	***0.09***	*0.01*	*0.01*	*0.00*	*0.06*	*0.02*	*1.16*
irradiat (contrast)	*-0.01*	*0.00*	*0.00*	*0.00*	*0.00*	*-0.16*	*0.86*
breast-quad (contrast)	*0.07*	*0.00*	*0.00*	*0.00*	*0.03*	*-0.05*	*0.00*
inv-nodes (contrast)	*-0.02*	***0.02***	***0.02***	***0.01***	***0.14***	*0.07*	*0.00*
node-caps (contrast)	*-0.04*	*0.00*	*0.00*	*0.00*	*0.02*	*-0.03*	*0.00*
breast-quad	-0.05	0.01	0.01	0.00	**0.16**	0.13	0.00

The second step of the experiment was devoted to evaluation of prediction of the quality of utilized machine learning algorithms described in Section 2. During this step, six different algorithms were applied using the 10-fold cross validation method. The average results for the *Breast cancer* dataset are shown in Figure 1. This procedure was utilized for two specified sets:

1. The original set containing all descriptive features.
2. Only the set of selected features defined on the basis of its importance calculated in the first step.

Additionally, all average results for the *Breast cancer* dataset and for other investigated data are collected in Tables from 4 to 9. On the basis of these results, it could be stressed that the set of selected features, which contains only 6 from 9 attributes, has the similar (even better) prediction quality than it was observed with all 9 original attributes. For instance, the popular *AUC* and classification accuracy (*CA*) measures in data analysis increased (see Figure 1). Furthermore, all other measures also increased a little. With the exception of the *Brier* score, which decreased, but the lower the *Brier* score is for a set of predictions, the better the predictions are calibrated [11].
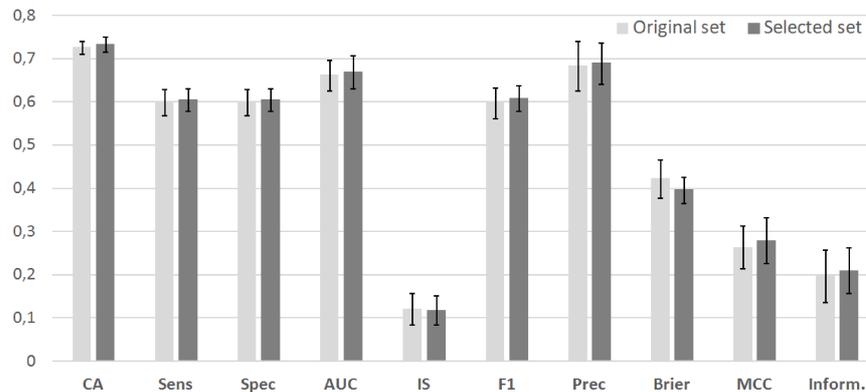
Fig. 1. Summary results and their standard deviations for Breast cancer dataset.

Similar results were obtained for other investigated datasets (see Figures from 2 to 5). Effectiveness of six learning algorithms used in this research can be observed on the basis of the results collected in Tables from 4 to 9. A summary of feature selection is presented in Table 3.

All numbers of features in the selected sets are significantly smaller than in the original one. The average is only about 37% of original features selected. Using these selected sets promising initial results of classification efficiency can be gained together with substantial reduction of the problem of dimmensionality.

References

1. Kohavi, R., John, G.H.: Wrappers for feature subset selection. Artif. Intell. 97, 273-324 (1997)

Table 3. Summary of selected features in benchmark datasets

Dataset	# original features	# selected features	% of original features
Breast cancer	9	6	67
Heart disease	13	5	38
Lung cancer	56	3	5
Primary tumor	17	5	29
Skin cancer	13	6	46
Average	**22**	**5**	**37**

Table 4. Average results of Classification tree on original (normal font) and selected sets (italic font)

Dataset	CA	Sens	Spec	AUC	IS	F1	Prec	Brier	MCC	Inform.
Breast cancer	0.71	0.58	0.58	0.63	0.10	0.58	0.63	0.44	0.21	0.17
	0.71	*0.58*	*0.58*	*0.64*	*0.11*	*0.58*	*0.63*	*0.40*	*0.20*	*0.16*
Heart disease	0.76	0.76	0.76	0.79	0.51	0.76	0.76	0.42	0.52	0.52
	0.74	*0.73*	*0.73*	*0.76*	*0.45*	*0.73*	*0.74*	*0.45*	*0.47*	*0.46*
Lung cancer	0.43	0.44	0.72	0.64	0.34	0.42	0.43	0.89	0.16	0.16
	0.60	*0.63*	*0.80*	*0.68*	*0.43*	*0.59*	*0.61*	*0.60*	*0.43*	*0.43*
Primary tumor	0.39	0.19	0.97	0.79	1.16	0.20	0.18	0.79	0.16	0.16
	0.35	*0.17*	*0.96*	*0.78*	*0.90*	*0.14*	*0.18*	*0.80*	*0.19*	*0.14*
Skin cancer	0.79	0.78	0.92	0.92	1.30	0.78	0.78	0.34	0.70	0.70
	0.77	*0.73*	*0.91*	*0.91*	*1.19*	*0.73*	*0.75*	*0.34*	*0.65*	*0.64*
Average	**0.62**	**0.55**	**0.79**	**0.76**	**0.68**	**0.55**	**0.56**	**0.58**	**0.35**	**0.34**
	0.63	***0.57***	***0.80***	***0.75***	***0.62***	***0.56***	***0.58***	***0.52***	***0.39***	***0.37***
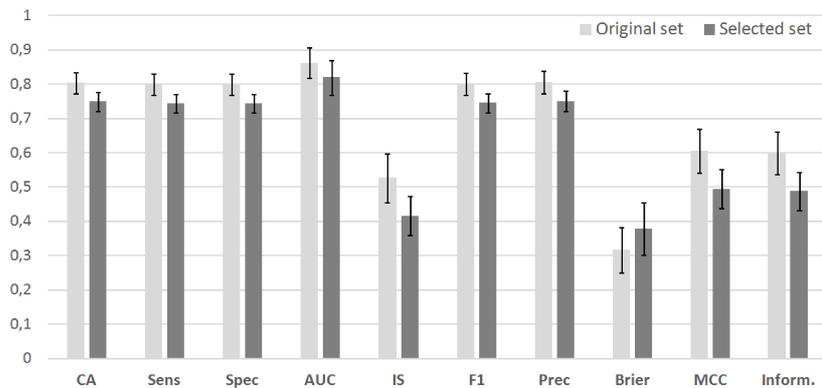
Fig. 2. Summary results and their standard deviations for Heart disease dataset.

Table 5. Average results of Random forest on original (normal font) and selected sets (italic font)

Dataset	CA	Sens	Spec	AUC	IS	F1	Prec	Brier	MCC	Inform.
Breast cancer	0.75	0.59	0.59	0.70	0.08	0.58	0.79	0.37	0.32	0.17
	0.75	*0.61*	*0.61*	*0.71*	*0.10*	*0.61*	*0.75*	*0.36*	*0.33*	*0.21*
Heart disease	0.81	0.81	0.81	0.89	0.41	0.81	0.82	0.27	0.63	0.62
	0.77	*0.76*	*0.76*	*0.86*	*0.35*	*0.76*	*0.77*	*0.31*	*0.53*	*0.53*
Lung cancer	0.35	0.33	0.65	0.73	0.17	0.34	0.43	0.61	0.01	-0.02
	0.50	*0.53*	*0.75*	*0.68*	*0.30*	*0.50*	*0.49*	*0.59*	*0.26*	*0.28*
Primary tumor	0.45	0.21	0.97	0.87	1.03	0.33	0.37	0.71	0.35	0.17
	0.33	*0.11*	*0.96*	*0.82*	*0.68*	*0.37*	*0.31*	*0.80*	*0.25*	*0.07*
Skin cancer	0.83	0.79	0.93	0.97	1.11	0.80	0.85	0.27	0.75	0.72
	0.75	*0.69*	*0.90*	*0.94*	*0.99*	*0.68*	*0.72*	*0.33*	*0.61*	*0.58*
Average	**0.64**	**0.54**	**0.79**	**0.83**	**0.56**	**0.57**	**0.65**	**0.44**	**0.41**	**0.33**
	0.62	*0.54*	*0.79*	*0.80*	*0.48*	*0.58*	*0.61*	*0.48*	*0.40*	*0.33*

Table 6. Average results of CN2 rules on original (normal font) and selected sets (italic font)

Dataset	CA	Sens	Spec	AUC	IS	F1	Prec	Brier	MCC	Inform.
Breast cancer	0.72	0.57	0.57	0.61	0.12	0.56	0.68	0.46	0.22	0.14
	0.75	*0.60*	*0.60*	*0.66*	*0.14*	*0.61*	*0.74*	*0.39*	*0.31*	*0.21*
Heart disease	0.82	0.81	0.81	0.84	0.58	0.81	0.83	0.33	0.64	0.62
	0.74	*0.73*	*0.73*	*0.76*	*0.43*	*0.74*	*0.75*	*0.43*	*0.48*	*0.47*
Lung cancer	0.44	0.44	0.71	0.65	0.39	0.44	0.46	0.72	0.15	0.14
	0.56	*0.55*	*0.77*	*0.69*	*0.25*	*0.68*	*0.58*	*0.60*	*0.46*	*0.32*
Primary tumor	0.45	0.21	0.97	0.87	1.03	0.33	0.37	0.71	0.35	0.17
	0.33	*0.11*	*0.96*	*0.82*	*0.68*	*0.37*	*0.31*	*0.80*	*0.25*	*0.07*
Skin cancer	0.82	0.79	0.93	0.94	1.32	0.81	0.84	0.27	0.75	0.72
	0.76	*0.70*	*0.90*	*0.92*	*1.09*	*0.72*	*0.78*	*0.34*	*0.64*	*0.60*
Average	**0.65**	**0.56**	**0.80**	**0.78**	**0.69**	**0.59**	**0.64**	**0.50**	**0.42**	**0.36**
	0.63	*0.54*	*0.79*	*0.77*	*0.52*	*0.62*	*0.63*	*0.51*	*0.43*	*0.33*

Table 7. Average results of Naive Bayes classifier on original (normal font) and selected sets (italic font)

Dataset	CA	Sens	Spec	AUC	IS	F1	Prec	Brier	MCC	Inform.
Breast cancer	0.73	0.66	0.66	0.69	0.16	0.66	0.67	0.43	0.33	0.31
	0.74	*0.65*	*0.65*	*0.70*	*0.17*	*0.66*	*0.68*	*0.40*	*0.33*	*0.31*
Heart disease	0.83	0.83	0.83	0.90	0.62	0.83	0.83	0.27	0.66	0.66
	0.78	*0.78*	*0.78*	*0.87*	*0.50*	*0.78*	*0.78*	*0.29*	*0.55*	*0.55*
Lung cancer	0.62	0.64	0.81	0.74	0.67	0.63	0.64	0.72	0.44	0.44
	0.60	*0.63*	*0.80*	*0.68*	*0.43*	*0.59*	*0.61*	*0.60*	*0.43*	*0.43*
Primary tumor	0.40	0.17	0.97	0.81	0.98	0.31	0.31	0.75	0.28	0.13
	0.38	*0.16*	*0.97*	*0.80*	*0.88*	*0.43*	*0.33*	*0.79*	*0.29*	*0.13*
Skin cancer	0.78	0.77	0.92	0.96	1.24	0.78	0.80	0.27	0.71	0.69
	0.73	*0.70*	*0.90*	*0.94*	*1.12*	*0.71*	*0.73*	*0.33*	*0.61*	*0.59*
Average	**0.67**	**0.61**	**0.84**	**0.82**	**0.74**	**0.64**	**0.65**	**0.49**	**0.48**	**0.45**
	0.65	***0.58***	***0.82***	***0.80***	***0.62***	***0.63***	***0.63***	***0.48***	***0.44***	***0.40***

Table 8. Average results of kNN classifier on original (normal font) and selected sets (italic font)

Dataset	CA	Sens	Spec	AUC	IS	F1	Prec	Brier	MCC	Inform.
Breast cancer	0.71	0.60	0.60	0.65	0.16	0.61	0.64	0.47	0.24	0.21
	0.72	*0.60*	*0.60*	*0.61*	*0.10*	*0.61*	*0.65*	*0.45*	*0.25*	*0.20*
Heart disease	0.77	0.76	0.76	0.85	0.51	0.76	0.76	0.36	0.53	0.53
	0.70	*0.70*	*0.70*	*0.80*	*0.40*	*0.70*	*0.70*	*0.46*	*0.40*	*0.40*
Lung cancer	0.43	0.46	0.72	0.66	0.35	0.44	0.44	0.68	0.17	0.18
	0.53	*0.53*	*0.76*	*0.62*	*0.35*	*0.49*	*0.47*	*0.66*	*0.27*	*0.29*
Primary tumor	0.49	0.26	0.98	0.84	1.48	0.41	0.27	0.75	0.26	0.24
	0.37	*0.19*	*0.97*	*0.82*	*1.13*	*0.36*	*0.24*	*0.78*	*0.20*	*0.16*
Skin cancer	0.81	0.82	0.93	0.94	1.40	0.82	0.81	0.29	0.75	0.76
	0.77	*0.75*	*0.91*	*0.92*	*1.26*	*0.75*	*0.76*	*0.34*	*0.67*	*0.66*
Average	**0.64**	**0.58**	**0.80**	**0.79**	**0.78**	**0.61**	**0.59**	**0.51**	**0.39**	**0.38**
	0.62	***0.55***	***0.79***	***0.76***	***0.65***	***0.58***	***0.57***	***0.54***	***0.36***	***0.34***

Table 9. Average results of SVM classifier on original (normal font) and selected sets (italic font)

| Dataset	CA	Sens	Spec	AUC	IS	F1	Prec	Brier	MCC	Inform.
Breast cancer	0.73	0.59	0.59	0.69	0.09	0.59	0.69	0.37	0.27	0.19
	0.73	*0.58*	*0.58*	*0.68*	*0.08*	*0.58*	*0.69*	*0.38*	*0.25*	*0.17*
Heart disease	0.83	0.83	0.83	0.90	0.53	0.83	0.83	0.24	0.65	0.65
	0.77	*0.76*	*0.76*	*0.86*	*0.36*	*0.76*	*0.77*	*0.32*	*0.53*	*0.52*
Lung cancer	0.62	0.56	0.79	0.71	0.17	0.46	0.40	0.62	0.33	0.35
	0.56	*0.55*	*0.77*	*0.61*	*0.38*	*0.68*	*0.58*	*0.57*	*0.46*	*0.32*
Primary tumor	0.36	0.22	0.97	0.79	1.19	0.32	0.22	0.94	0.20	0.19
	0.24	*0.12*	*0.96*	*0.71*	*0.85*	*0.26*	*0.17*	*1.05*	*0.11*	*0.09*
Skin cancer	0.84	0.83	0.94	0.97	1.37	0.84	0.85	0.21	0.78	0.78
	0.75	*0.71*	*0.90*	*0.93*	*1.09*	*0.72*	*0.74*	*0.33*	*0.63*	*0.61*
Average	**0.68**	**0.61**	**0.82**	**0.81**	**0.67**	**0.61**	**0.60**	**0.48**	**0.45**	**0.43**
	0.61	***0.55***	***0.79***	***0.76***	***0.55***	***0.60***	***0.59***	***0.53***	***0.40***	***0.34***
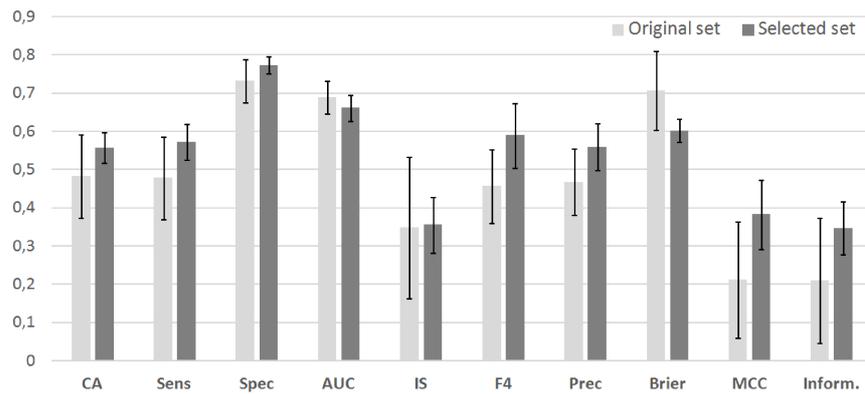
**Fig. 3.** Summary results and their standard deviations for Lung cancer dataset.
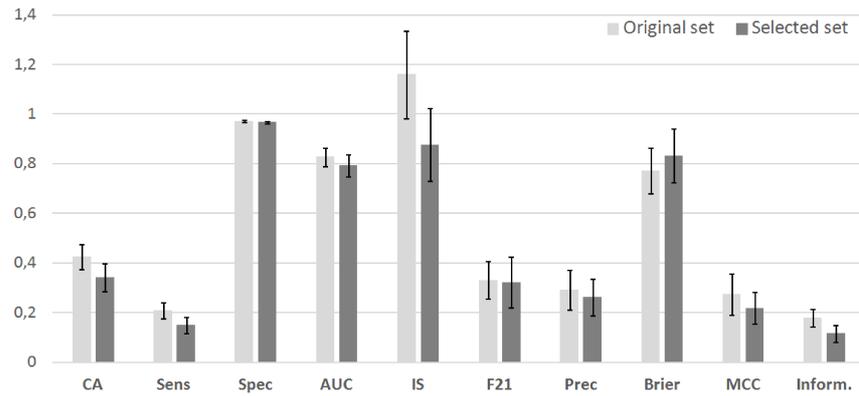
**Fig. 4.** Summary results and their standard deviations for Primary tumor dataset.
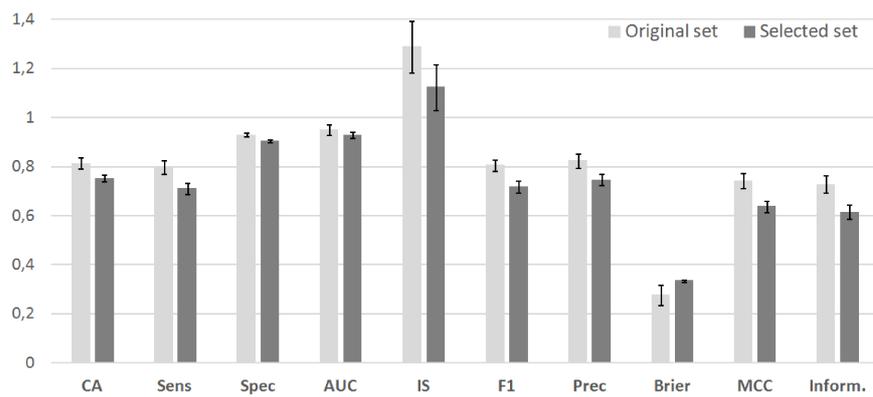
Fig. 5. Summary results and their standard deviations for Skin cancer dataset.

2. Bermingham, M.L., Pong-Wong, R., Spiliopoulou, A., Hayward, C., Rudan, I., Campbell, H., Wright, A.F., Wilson, J.F., Agakov, F., Navarro, P., Haley, C.S.: Application of high-dimensional feature selection: evaluation for genomic prediction in man. Sci. Rep. 5, (2015)

3. Phuong, T.M., Lin, Z., Altman, R.B.: Choosing SNPs using feature selection. Proceedings - 2005 IEEE Computational Systems Bioinformatics Conference, CSB 2005. pp. 301-309 (2005)

4. Zhu, Z., Ong, Y.S., Dash, M.: Wrapper-filter feature selection algorithm using a memetic framework. IEEE Trans. Syst. Man, Cybern. Part B Cybern. 37, 70-76 (2007)

5. Nilsson, R., Pena, J.M., Bj okegren, J., Tegner, J.: Detecting multivariate differentially expressed genes. BMC Bioinformatics. 8, 150 (2007)

6. Rudnicki, W.R., Wrzesień, M., Paja, W.: All Relevant Feature Selection Methods and Applications. In: Stańczyk, U. and Lakhmi, C.J. (eds.) Feature Selection for Data and Pattern Recognition. pp. 11-28. Springer-Verlag Berlin Heidelberg, Berlin (2015)

7. Stoppiglia, H., Dreyfus, G., Dubois, R., Oussar, Y.: Ranking a Random Feature for Variable and Feature Selection. J. Mach. Learn. Res. 3, 1399-1414 (2003)

8. Tuv, E., Borisov, A., Torkkola, K.: Feature Selection Using Ensemble Based Ranking Against Artificial Contrasts. International Symposium on Neural Networks. pp. 2181-2186 (2006)

9. Fawcett, T.: An introduction to ROC analysis. Pattern Recognit. Lett. 27, 861-874 (2006)

10. Hippe, Z.S., Bajcar, S., Blajdo, P., Grzymala-Busse, J.P., Grzymala-Busse, J.W., Knap, M., Paja, W., Wrzesien, M.: Diagnosing Skin Melanoma: Current versus Future Directions. TASK Q. 7, 289-293 (2003)

11. Hernandez-Orallo, J., Flach, P., Ferri, C.: A unified view of performance metrics: translating threshold choice into expected classification loss. J. Mach. Learn. Res. 13, 2813-2869 (2012)