

A New Gene Selection Method Based on Random Subspace Ensemble for Microarray Cancer Classification

Giuliano Armano¹, Camelia Chira², and Nima Hatami¹

¹ DIEE-Department of Electrical and Electronic Engineering,
University of Cagliari,

Piazza D'Armi, I-09123 Cagliari, Italy

² Department of Computer Science,
Babes-Bolyai University,

Kogalniceanu 1, Cluj-Napoca 400084, Romania

Abstract. Gene expression microarray data provides simultaneous activity measurement of thousands of features facilitating a potential effective and reliable cancer diagnosis. An important and challenging task in microarray analysis refers to selecting the most relevant and significant genes for data (cancer) classification. A random subspace ensemble based method is proposed to address feature selection in gene expression cancer diagnosis. The introduced Diverse Accurate Feature Selection method relies on multiple individual classifiers built based on random feature subspaces. Each feature is assigned a score computed based on the pairwise diversity among individual classifiers and the ratio between individual and ensemble accuracies. This triggers the creation of a ranked list of features for which a final classifier is applied with an increased performance using minimum possible number of genes. Experimental results focus on the problem of gene expression cancer diagnosis based on microarray datasets publicly available. Numerical results show that the proposed method is competitive with related models from literature.

Keywords: random subspace ensembles, multiple classifier systems, multivariate feature selection, gene expression data analysis, pairwise diversity.

1 Introduction

High-throughput technologies are able nowadays to produce huge amount of valuable information which can be used in the identification and classification of various diagnostic areas. The required analysis of this information creates a real challenge for machine learning and new robust models are still required to efficiently tackle this task.

Cancer diagnosis based on gene expression data is an important emerging medical application domain of microarray analysis tools [14]. Clinical-based cancer classification has been shown to have limited diagnostic ability [11]. On the other

hand, the classification of different tumor types using gene expression data is able to generate valuable knowledge for important tasks such as cancer diagnosis and drug discovery [11].

Classifying microarray samples (for example, cancer versus normal cells) according to their gene expression profiles represents an important and challenging task. The complexity of the problem rises from the huge number of features that contribute to a profile compared to the very low number of samples normally available in microarray analysis. Another challenge for classification is the presence of noise (biological or technical) in the dataset which further affects the classifier accuracies. Moreover, the inherent presence of a large number of irrelevant genes increases the difficulty of the classification task influencing the discrimination power of relevant features [11].

The problem of extracting significant knowledge from microarray data requires the development of robust methods able to address this task [9,12]. Lu and Han [11] emphasize that the most important aspects of classification and gene selection methods are their computation time, classification accuracy and ability to reveal biologically meaningful gene information. Many studies consider the selection of genes to be an important integral preprocessing step for the classification problem [11] able to reduce dimensionality, remove irrelevant or noisy genes and improve the learning accuracy [11,8,1]. Furthermore, gene selection can reduce the computational cost of the classifier and lead to more compact results easily interpretable in the diagnostics task [10].

Multiple classifier systems use base classifiers with complementary behaviour resulting in an efficient alternative to a complex and hard-to-manage single classifier. Among many well-known ensemble methods, Random Subspace Ensemble (RSE) [4] is an efficient model which obtained good results particularly for high-dimensional classification problems. RSE uses a number of base classifiers, each of them considering only a subset (randomly determined) of the original feature space.

In this paper, the *Diverse Accurate Feature Selection (DAFS)* method is introduced to deal with feature selection in gene expression cancer diagnosis. The main idea behind DAFS is to adapt RSE for feature selection by efficiently exploring accuracy and diversity information. To be more specific, the proposed DAFS method uses individual and ensemble accuracies of base classifiers specialized on random subspaces as well as pairwise diversity to rank features. This is achieved by assigning to each feature a score calculated by a metric which takes into account the mean ratio between individual and ensemble accuracies weighted by their diversity. This way, the introduced approach takes advantage of RSE to deal with the high-dimensionality of the addressed problem while in the same time building many different classifiers on the same samples to overcome the sample-size limitation.

The proposed method is evaluated for 11 cancer gene expression datasets [13] (from which nine are multiclass and two are binary classification problems). Computational experiments and comparisons to related methods indicate a good

performance of the proposed DAFS method and emphasize the potential of RSE to deliver a fast effective feature selection method.

The structure of the paper is as follows: section 2 presents the DAFS method detailing the feature-space coverage approach, the feature evaluation and ranking phase and the DAFS algorithm; section 3 presents the computational experiments discussing the obtained results and comparing them with related methods and section 4 contains the conclusions of the paper.

2 Diverse Accurate Feature Selection

In this section, we introduce the *Diverse Accurate Feature Selection (DAFS)* method to facilitate multivariate feature selection for gene expression classification. The proposed DAFS relies on the RSE method [4] to select features triggering a high classification accuracy and furthermore on certain diversity measures to minimize the redundancy among selected features. It should be emphasized that the proposed method is intended for feature selection and not for classification, which means that DAFS should be used rather as a preprocessing step of the classification task (approach suggested in many studies [11,8,1,10]).

Let us consider a sample set $X_{M \times N}$ of size N and dimensionality M . Instead of using all features for each classifier in the ensemble, RSE [4] samples the feature set. The ensemble assigns a class label by majority or average voting of L classifiers built on different feature subsets of size m (where $m \ll M$) sampled randomly and uniformly from the original feature set. Each base classifier is invariant to points that are different from the training points only in the unselected features, thus encouraging diversity. This approach results in each classifier generalizing in a different way. Hence, while most other classification methods suffer from the curse of dimensionality, the RSE method can take advantage of high dimensionality to build an ensemble of "diverse" classifiers, each specialized on an small subspace [4].

2.1 Feature-Space Coverage

Let us denote by γ the total number of relevant features returned by the selection algorithm. The *selection ratio* is defined as $\frac{\gamma}{M}$. As indicated in many research reports [3,11], the number of useful genes resulted from feature selection is very small (without loss of generality, we can consider this number to be as small as 10 out of thousands of genes). Considering the high number of features present in microarray gene expression datasets (normally more than 5000), we can state that *selection ratio* < 0.002 carrying a large amount of information by a small number of genes. Therefore, any feature selection strategy should evaluate each gene at least once for an effective selection.

The probability that a particular feature f_i is hit in m trials of RSE is m/M (a measure called *SR - Subspace Ratio*). Therefore, the probability of not selecting a feature in any of the L classifiers of the ensemble is $P(\bar{f}_i) = (1 - SR)^L$. The probability of f_i being *at least* in one of the L selections is $1 - P(\bar{f}_i)$. Furthermore,

the probability of all features to appear at least once in one of the L classifiers (called P_{cov} - feature coverage probability) is defined as follows:

$$P_{cov} = (1 - (1 - SR)^L)^\gamma \quad (1)$$

The assumption here is that the features within the selected subset of size m are sampled independently. The probability P_{cov} is considered in the proposed DAFS method to ensure the feature-space coverage by informing the selection of L and m parameters.

2.2 Feature Evaluation and Ranking

The core of any feature selection method is the process of evaluating each feature and selecting the most relevant ones for the problem at hand. Furthermore, this process should avoid redundant features because of their negative effect on the classification task (since redundant features do not bring any additional necessary information and further increase dimensionality).

The individual accuracy and the diversity among ensemble members are two key characteristics affecting ensemble accuracy [7]. The relation of the individual accuracy and diversity with the ensemble accuracy is still not clear but underlies the only way of designing an ensemble. The idea behind the DAFS method is to use the accuracy and diversity driven information not for designing an ensemble but for selecting the most relevant features from a very large set.

To minimize the redundancy between the selected features, the feature evaluation and ranking phase of DAFS makes use of an important ensemble characteristic i.e. diversity. As opposed to many standard methods which directly compare the similarity within a set of features (e.g. by Euclidean distance) to measure their redundancy, we propose to evaluate features by comparing the labels diversity of the base classifiers using a specific feature. The more diverse the classifiers to which a specific features contributes to, the less redundant the feature is.

The accuracy of a base classifier built on a specific feature subset is able to consider the contribution of those features to the classification task. Thanks to this important ensemble characteristic, the importance of a feature is *indirectly* measured by comparing the performances of the corresponding classifiers.

To be more specific, let us consider a feature f_i which appears in k random subspace sets $RS_{f_i} = \{rs_{i_1}, rs_{i_2}, \dots, rs_{i_k}\}$, each subspace being used in the training set of k base classifiers $L_{f_i} = \{l_{i_1}, l_{i_2}, \dots, l_{i_k}\}$. Intuitively, a feature is important if the corresponding classifiers using it get a high accuracy in the ensemble. Therefore, *Individual accuracy/RSE accuracy* ratio can be used for evaluating the feature *relevancy*. In the same way, $d(l_{ens}, l_{i_u}), u = 1 \dots k$ computes the *redundancy* of a feature using an individual classifier l_{i_u} compared to the ensemble l_{ens} (d refers to any pairwise diversity measure).

Considering all of the above mentioned issues, we propose a measure called *Diversity-Accuracy (DA)* associated with each feature. The DA measure uses the ensemble components to incorporate both accuracy (to assess feature relevancy)

and diversity (to assess feature redundancy) information. For a feature f_i present in the k subsets RS_{f_i} used in classifiers L_{f_i} , the DA measure is defined as follows:

$$DA(f_i) = \frac{1}{k} \sum_{u=1}^k d(l_{ens}, l_{i_u})^r \times \left[\frac{a(l_{i_u})}{a(l_{ens})} \right]^r, \quad (2)$$

where $a(l)$ represents the accuracy of classifier l and r is a parameter used for adjusting the contribution of an individual classifier versus the ensemble to the feature evaluation.

It should be noted that the ensemble performance represents the contribution of the entire set of features and majority of base classifiers. A feature will get a high DA value if the classifiers using that feature have more diversity and accuracy with respect to the ensemble (majority of other features). Therefore, the DA measure considers both the ensemble accuracy and diversity as reference points for evaluating the performance of the individual classifiers for each feature.

2.3 DAFS Algorithm

The main steps of proposed DAFS method refer to randomly sampling the feature set, building the RSE, calculating the DA value for each feature appearing in at least one base classifier and ranking the features according to the DA value.

The DAFS algorithm is detailed below (Algorithm 1). The input parameters are m - the subspace size, L - the ensemble size and G - the maximum number of features that will be selected. The sample set $X_{M \times N}$ is divided in training (X_t) and testing (X_e) subsets (by using for example cross-validation).

The *feature evaluation and selection* phase of the algorithm refers to the application of RSE method and the evaluation of features based on the DA measure. Ensemble accuracy is computed by majority voting strategy. Based on the individual and ensemble classifier results, the DA value is calculated for each feature using formula 2. Features are then sorted in descending order of the DA value and the first γ are selected. The final number γ of selected features is determined by considering all possible feature sets (of size up to the maximum G parameter) and choosing the one that triggers the best performance on the validation set by the final classifier h . The γ features are finally used in a *classification* phase to determine the performance of a classifier using the selected features projected on the training and testing sets.

3 Experimental Results for Gene Expression Data

This section presents computational experiments for various cancer gene expression datasets. Results are analysed from multiple perspectives (e.g. subset size vs. ensemble size, DA feature evaluation, different classification algorithms) and compared to those of related methods.

Algorithm 1. DAFS Algorithm

Parameters

- m : subspace size
- L : ensemble size
- G : maximum number of selected features

Input

- $\mathbb{X}_t, \mathbb{X}_e$ represent the training and testing sets respectively
- $l_i, i = 1 \dots L$ are the base classifiers in RSE
- h is the classifier used for the final classification

Feature evaluation and selection

- Select relevant m and L which satisfy the *feature-space coverage* P_{cov} condition (formula 1)
- Create L sets of feature subspace $rs_i, i = 1 \dots L$ randomly and independently sampled from the entire feature space
 - Build learners (subclassifiers) each using the training data projected on its selected subspace
 - Apply \mathbb{X}_v on the L classifiers and assign the labels by majority voting (\mathbb{X}_v is a subset of \mathbb{X}_t used for validation)
 - Calculate the DA value for each feature using formula 2
 - Sort the features in descending order of DA value
 - Select the first γ features from the sorted list, $1 \leq \gamma \leq G$, which trigger the best accuracy on \mathbb{X}_v by classifier h

Classification

- Train the classifier model h using \mathbb{X}_t projected on the selected feature space of size γ
 - Project the testing set on the selected feature space and classify them applying h
-

3.1 Dataset Description

The cancer gene expression datasets available in [13] are engaged for computational experiments. Nine out of the 11 datasets considered are multiclass while two of them are binary classification problems. Table 1 presents the different characteristics of these datasets. The column *Maximum prior* indicates the prior probability of the main diagnostic class for each problem.

The main properties of the considered datasets are as follows: *11_Tumors* contains 174 samples with 12533 genes with the task of finding 11 various human tumor types, *14_Tumors* refers to 14 various human tumor types and 12 normal tissue types, *9_Tumors* has 60 samples in 9 various human tumor types, *Brain_Tumor1* with 5 human brain tumor types, *Brain_Tumor2* contains 4 malignant glioma types, *Leukemia1* is a three-class task: Acute myelogenous leukemia (AML), acute lymphoblastic leukemia (ALL) B-cell, and ALL T-cell, *Leukemia2* with AML, ALL, and mixed-lineage leukemia (MLL) classes, *Lung_Cancer* contains 4 lung cancer types and normal tissues, *SRBCT* is for

Table 1. Cancer gene expression datasets used in computational experiments (from [13])

Dataset	Samples	Genes (features)	Classes	Features/Samples	Max. Prior
11_Tumors	174	12533	11	72	15.5 %
14_Tumors	308	15009	26	49	9.7 %
9_Tumors	60	5726	9	95	15.0 %
Brain_Tumor1	90	5920	5	66	66.7 %
Brain_Tumor2	50	10367	4	207	30.0 %
Leukemia1	72	5327	3	74	52.8 %
Leukemia2	72	11225	3	156	38.9 %
Lung_Cancer	203	12600	5	62	68.5 %
SRBCT	83	2308	4	28	34.9%
Prostate_Tumor	102	10509	2	103	51.0 %
DLBCL	77	5469	2	71	75.3 %

Small, round blue cell tumors (SRBCT) of childhood, *Prostate_Tumor* refers to a binary classification task Prostate tumor and normal tissues, and *DLBCL* has 77 samples with two possible classes: Diffuse large b-cell lymphomas (DLBCL) and follicular lymphomas.

3.2 DAFS Setup

The DAFS method needs the specification of the classifiers used in the RSE and of the diversity measure used in calculating the DA feature values.

We have used different types of classification algorithms - k-Nearest Neighbour (kNN), Support Vector Machine (SVM) and Multi-Layer Perceptron (MLP) - as base learners in order to show that the proposed method is independent of the particular base classifier. The error backpropagation algorithm is used for the training of the MLP base classifiers and the iterative estimation process is stopped when an average squared error of 0.9 over the training set is obtained, or when the maximum number of iterations is reached (adopted mainly for preventing networks from overtraining). We also varied the number of hidden neurons to experimentally find the optimal architecture of the MLPs for each problem. The other parameter values used for training are as follows: learning rate is 0.4 and momentum parameter is 0.6. In the case of SVM serving as base classifier, linear kernel is used. In the case of kNN classifier, the standard euclidean distance is used to calculate distances between samples. The value of k was varied from 3 to 9 in order to find the best neighbourhood size for each problem. All other parameters for these three algorithms have been chosen according to the standard setting of MATLAB Toolboxes.

The datasets are split into training and testing subsets by using Leave-One-Out Cross-Validation (LOOCV). Since the problems are low sample size, the evaluation of a method assessed by LOOCV provides realistic generalization for unseen data. Furthermore, 20% of the training set is used for the validation phase.

Several diversity measures are considered to compute the DA measure in the evaluation of DAFS. In [7], Kuncheva and Whitaker emphasize that there is no generally accepted formal definition for diversity and present several definitions from statistics to measure pairwise and non-pairwise diversities. Based on the analysis presented in [7], the diversity measures used in this paper are Q-statistics (*Q-sta*), Correlation coefficient (*Corr*), Disagreement measure (*Dis*) and Double Fault measure (*DF*). We briefly recall these measures below:

$$Q - sta_{la,lb} = \frac{N^{11}N^{00} - N^{01}N^{10}}{N^{11}N^{00} + N^{01}N^{10}} \quad (3)$$

$$Corr_{la,lb} = \frac{N^{11}N^{00} - N^{01}N^{10}}{\sqrt{(N^{11} + N^{10})(N^{01} + N^{00})(N^{11} + N^{01})(N^{10} + N^{00})}} \quad (4)$$

$$Dis_{la,lb} = \frac{N^{01} + N^{10}}{N^{11} + N^{10} + N^{01} + N^{00}} \quad (5)$$

$$DF_{la,lb} = \frac{N^{00}}{N^{11} + N^{10} + N^{01} + N^{00}} \quad (6)$$

where la, lb represent the two classifiers compared, N^{11} and N^{00} are the number of samples for which both classifiers la and lb concurrently make correct and incorrect decisions respectively while N^{10} and N^{01} are the number of samples for which la and lb do not agree on their labels.

3.3 Numerical Experiments

The proposed feature selection algorithm is compared to related methods based on the recognition rates of the final classifiers. The following (m, L) pairs have been tested for the values of the subspace size (m) and the ensemble size (L): (10, 10000), (50, 500), (500, 100), (1000, 20) and (1000, 10). All these (m, L) value pairs are chosen in accordance with the feature coverage probability P_{cov} (formula 1). For each dataset considered, all the (m, L) pairs have been tested and the best performing one is used in the reported results. In the same way, all four diversity measures given in the previous section (equations 3 to 6) are considered for each dataset and the one triggering the best performance is used to report the results of DAFS.

The last input parameter of DAFS refers to G , the maximum number of selected features, and is set to 30. The final number γ of selected features is determined by considering all possible feature sets (of size up to G) and choosing the one that leads to the best performance on the validation set (as explained in section 2.3). The value of γ is different for each dataset considered in the experiments. Finally, the parameter r used in the DA measure (equation 2) is set to 1 for all experiments.

Table 2 presents the comparative numerical results. The accuracies of single kNN , SVM and MLP are given in the first three columns of results. The next

two columns show the accuracy obtained by standard RSE and the average individual accuracy in RSE (column labelled *Avg. Ind.*). These five columns provide the baseline results of standard methods without any feature selection to comparatively assess the DAFS performance. Furthermore, two commonly used feature selection methods in machine learning i.e. Mutual information (*MI*) and *MRMR* [2] have been applied for all considered cancer gene expression datasets and the results are given in the columns labelled *MI* and *MRMR* respectively. Last column in Table 2 contains the accuracy reported by proposed DAFS method. Except for the single classifier results, SVM has been used as the final classifier to produce the recognition rates presented in Table 2.

The average rank of DAFS (computed based on the average results over all datasets) is the best among all considered methods compared in Table 2. The advantages of DAFS compared to the baseline methods are clear from the results obtained. Also, DAFS clearly obtains better recognition rates than mutual information. An overall better performance compared to MRMR can also be observed. For three out of the 11 considered datasets (i.e. 11_Tumors, 9_Tumors and Brain_Tumor1), DAFS and MRMR report the same accuracy while for the Lung_Cancer dataset (a five-class problem with a high maximum prior probability of the dominant class), MRMR obtains a better recognition rate of 96.5 (same accuracy with single SVM) compared to 95.4 reported by DAFS. However, the proposed DAFS outperforms MRMR for the rest of seven datasets considered.

The results presented in Table 2 show that SVM classifiers are robust against dimensionality and able to achieve a good performance even without feature selection. However, feature selection significantly improves the classification performance of non-SVM learners. The recognition rates of kNN and MLP increase dramatically (more than double) when DAFS is first used in feature selection.

Table 2. Recognition rates obtained by proposed DAFS (last column) compared to the results of some baseline methods as well as those of MI and MRMR feature selection methods

Dataset	single SVM	single kNN	single MLP	RS ensemble	Avg. Ind.	MI	MRMR	DAFS method
11_Tumors	94.6	73.2	55.8	68.1	52.8	92.2	95.3	95.3
14_Tumors	75.6	45.8	11.9	50.9	33.7	73.6	74.9	74.9
9_Tumors	62.7	44.1	20.5	44.9	31.6	60.0	65.6	65.6
Brain_Tumor1	91.2	85.3	83.9	84.4	79.1	89.9	91.0	90.6
Brain_Tumor2	77.8	65.1	62.6	63.0	54.9	77.5	78.8	78.8
Leukemia1	97.5	81.9	79.4	82.0	73.1	96.9	97.0	97.5
Leukemia2	95.2	85.9	88.6	87.7	80.3	94.8	95.0	97.2
Lung_Cancer	96.5	86.5	84.8	87.7	79.1	94.9	96.5	95.4
SRBCT	98.1	83.2	89.5	90.5	78.4	97.1	97.3	100.0
Prostate_Tumor	92.0	80.7	75.5	79.9	71.1	91.8	92.0	92.3
DLBCL	95.8	77.9	81.8	78.8	72.5	95.6	97.3	98.0
Avg. Rank	1.72	5.36	6	5	7.45	3.45	1.81	1.36

4 Conclusions

A simple and effective feature selection method has been proposed and evaluated for gene expression cancer diagnosis. The main strength of the introduced DAFS method is represented by an efficient combination of classification accuracy information extracted using RSE and diversity information generated by using certain measures. The first component facilitates the identification of relevant features in obtaining a good recognition rate in classification while the second component focuses on minimizing the redundancy among selected features. Experimental results for 11 microarray cancer diagnosis datasets support the conclusions that DAFS is a fast and efficient feature selection method. It has been shown that DAFS is able to reduce the computational cost of the classification by removing irrelevant features and improving the recognition rate. The best results are obtained using SVM as base classifier while DAFS performance induced by the diversity measure varies from one dataset to another.

Obviously, the DAFS functionality is independent from the particular classifier used while the recognition rate based on the selected features does depend on the final classifier. For gene expression cancer classification, experiments revealed that better accuracies are produced by SVM in comparison with kNN and MLP. Moreover, the non-SVM learners benefit more from the proposed feature selection method in the sense that classification accuracies are significantly improved when using DAFS as a preprocessing step.

Acknowledgments. This work has been partially supported by the Italian Ministry of Education - Investment funds for basic research, under the project ITALBIONET - Italian Network of Bioinformatics.

Camelia Chira acknowledges the support of CNCS Romania through grant PN II TE 320 - Emergence, auto-organization and evolution: New computational models in the study of complex systems.

References

1. Banerjee, M., Mitra, S., Banka, H.: Evolutionary Rough Feature Selection in Gene Expression Data. *IEEE Transactions on Systems, Man, and Cybernetics-Part C: Applications and Reviews* 37(4), 622–632 (2007)
2. Ding, C., Peng, H.: Minimum redundancy feature selection from microarray gene expression data. *J. Bioinform. Comput. Biol.* 3, 185–205 (2005)
3. Golub, T.R., Slonim, D.K., Tamayo, P., Huard, C., Gaasenbeek, M., Mesirov, J.P., Coller, H., Loh, M.L., Downing, J.R., Caligiuri, M.A., Bloomfield, C.D., Lander, E.S.: Molecular classification of cancer: class discovery and class prediction by gene expression monitoring. *Science* 286(5439), 531–537 (1999)
4. Ho, T.: The random subspace method for constructing decision forests. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 20(8), 832–844 (1998)
5. Huang, H.-L., Chang, F.-L.: ESVM: Evolutionary support vector machine for automatic feature selection and classification of microarray data. *BioSystems* 90, 516–528 (2007)

6. Kuncheva, L.I., Rodriguez, J.J., Plumpton, C.O., Linden, D.E.J., Johnston, S.J.: Random Subspace Ensembles for fMRI Classification. *IEEE Transactions on Medical Imaging* 29(2), 531–542 (2010)
7. Kuncheva, L.I., Whitaker, C.J.: Measures of diversity in classifier ensembles. *Machine Learning* 51, 181–207 (2003)
8. Larrañaga, P., Calvo, B., Santana, R., Bielza, C., Galdiano, J., Inza, I., Lozano, J.A., Armañanzas, R., Santafé, G., Pérez, A., Robles, V.: Machine learning in bioinformatics. *Briefings in Bioinformatics* 7(1), 86–112 (2006)
9. Lee, C.P., Leu, Y.: A novel hybrid feature selection method for microarray data analysis. *Applied Soft Computing* 11, 208–213 (2011)
10. Liu, H., Liu, L., Zhang, H.: Ensemble gene selection by grouping for microarray data classification. *Journal of Biomedical Informatics* 43, 81–87 (2010)
11. Lu, Y., Han, J.: Cancer classification using gene expression data. *Information Systems* 28(4), 243–268 (2003)
12. Maji, P., Paul, S.: Rough set based maximum relevance-maximum significance criterion and gene selection from microarray data. *International Journal of Approximate Reasoning* 52, 408–426 (2011)
13. Statnikov, A., Aliferis, C., Tsamardinos, I., Hardin, D., Levy, S.: (2004), <http://www.gems-system.org>
14. Statnikov, A., Aliferis, C.F., Tsamardinos, I., Hardin, D., Levy, S.: A Comprehensive Evaluation of Multicategory Classification Methods for Microarray Gene Expression Cancer Diagnosis. *Bioinformatics* 21(5), 631–643 (2005)