Proceedings of the 2013 Federated Conference on
Computer Science and Information Systems pp. 83–89

# Combining One-Class Support Vector Machines for Microarray Classification

Bartosz Krawczyk
Department of Systems and Computer Networks,
Wrocław University of Technology
Wrocław, Poland
bartosz.krawczyk@pwr.wroc.pl

***Abstract*—The advance of high-throughput techniques, such as gene microarrays and protein chips have a major impact on contemporary biology and medicine. Due to the high-dimensionality and complexity of the data, it is impossible to analyze it manually. Therefore machine learning techniques play an important role in dealing with such data. In this paper we propose to use a one-class approach to classifying microarrays. Unlike canonical classifiers, these models rely only on objects coming from single class distributions. They distinguish observations coming from the given class from any other possible states of the object, that were unseen during the classification step. While having less information to dichotomize between classes, one-class models can easily learn the specific properties of a given dataset and are robust to difficulties embedded in the nature of the data. We show, that using one-class support vector machines can give as good results as canonical multi-class classifiers, while allowing to deal with imbalanced distribution and unexpected noise in the data. To cope with high dimensionality of the feature space, we propose to form an ensemble, based on Random Subspace and prune it with the usage of diversity measure. Experimental investigations, carried on public datasets, prove the usefulness of the proposed approach.**

***Index Terms*—machine learning, one-class classification, multiple classifier systems, classifier ensembles, bioinformatics, microarray analysis, high dimensionality.**

## I. Introduction

CONTEMPORARY high-throughput technologies produce massive volumes of biomedical data. Transcriptional research and profiling, with the usage of microarray technologies are powerful tools to gain a deep insight into the pathogenesis of complex diseases that plague modern society, such as cancer. Recent works on cancer profiling showed without a doubt, that gene expression patters can be used for high-quality cancer subtype recognition [1] - leukemias [2], melanoma [3], breast cancer [4] or prostate cancer [5] to name a few.

Identifying cancer properties, based on their distinct expression profiles may provide necessary information for a breakthrough, that is required for patient-tailored therapy. Currently there are no distinct rules on how individuals respond to chemotherapy and existing chemotherapies have in most cases severe side-effects with varying medical efficiency.

Due to massive amounts of data generated by microarray experiments and their high complexity and dimensionality, one requires a decision support system to extract the meaningful information from them. Machine learning is widely used for this task [6], with two distinct areas - unsupervised [7] and supervised learning [8]. In this paper we will focus on the latter one.

Supervised machine learning is a promising approach for analyzing microarray results in context of predicting patients outcome. Support Vector Machines are among the most popular classifiers used for this task [9]. Multiple Classifier Systems [10], or classifier ensembles, have gained an significant attention of the bioinformatics community in recent years. Random Forest [11] and Rotation Forest [12] ensembles have displayed an excellent classification accuracy for small-sample, high dimensionality microarray datasets, outperforming single-model approaches.

Another important issue is the problem of curse of dimensionality. Microarray data suffer from a relatively small number of objects, in comparison to the feature space dimensionality, often reaching several thousands. This causes difficulties for machine learning algorithms, reducing their performance and increasing their computational complexity. Among this data flood a major number of parameters possess small discriminative power and is irrelevant to the classification process, which makes feature selection a crucial step in microarray analysis [13].

Although there are many applications of machine learning-based decision support systems in bioinformatics, there are still many unresolved problems, such as:

- How to integrate heterogeneous data sources to achieve better insight into the mechanism behind complex diseases?
- How to organize, store, analyze and visualize high-dimensionality data obtained from the biomedical data flood?
- How to deal with the problem of high-dimensionality, small sample size, which strongly affects the classification performance and may lead to overfitting, poor generalization and unstable predictors?
- How to cope with difficulties embedded in the nature of microarray data, such as noise or class imbalance, as canonical machine learning classifiers cannot cope with them easily?

In this paper the last two issues are addressed.

978-1-4673-4471-5/$25.00 

We propose to analyze microarray data with the usage of one-class classifiers, instead of commonly applied binary ones. Up to author's knowledge this is the first work on applying one-class ensembles and one-class classification in general, to the microarray classification.

To cope with the high dimensionality problem we apply an ensemble approach, based on Random Subspaces [14]. By decomposing the feature space we at the same time reduce the overall computational complexity of the classification model and assure initial diversity among the pool of individual classifiers in the committee. A diversity-based pruning method is applied to discard redundant classifiers and to chose mutually complementary one-class predictors. Experiments, based on a set of publicly available microarray datasets, show that the proposed approach maintains a good classification accuracy, while displaying an improved robustness to atypical data distribution and prevalent noise.

## II. One-class classification

The aim of one-class classification (OCC) is to recognize one specific class from the more broad set of classes (e.g., selecting horses from all animals). The given class is known as target class $\omega_t$, while the remaining are denoted as outliers $\omega_O$. During the learning only examples target class (known also as positive examples) are being presented to learner, while it is assumed that during the exploitation phase new, unseen objects from other classes may appear.

OCC problems are common in the real world where positive examples are widely available but negative ones are hard, expensive or even impossible to gather [15]. Let us consider an engine. It is a quite easy and cheap to collect data about its normal work. Collecting observations about failures it is expensive and sometimes impossible, because in this case we would have to spoil the engine.

Such approach is very useful as well for many practical cases especially when the target class is "stable" and outlier one is "unstable". To explain this motivation let us consider a computer security problem as spam filtering or intrusion detection (IDS/IPS) [16].

Among several types of classifiers dedicated to OCC, the most popular is one concentrating on estimation of a closed boundary for given data, assuming that such a boundary will describe sufficiently the target class [17]. The main aim of those methods is to find the optimal size of the volume enclosing given training points. Too small size could lead to overfitting the model, while too big size might lead to extensive acceptance of outliers into the target class. Those methods rely strongly on the distance between objects [18]. Boundary methods require smaller number of objects to properly estimate the decision criterion, which makes them a perfect tool for applications suffering from a small sample size,such as microarrays classification. The well-known boundary methods are one-class support vector machine (OCSVM) [19] and support vector data description (SVDD) [20]. In this work we will use the former one.

### A. One-class support vector machine

One-class SVM classifier (OCSVM) [19] can deal with datasets containing only patterns from one target class. OCSVM classification aims at discriminating one class of target samples from all other ones. It consists of learning the minimum volume contour that encloses most of the data in a given dataset. Its original application is the outlier detection finding data that differ from most of the data within a dataset.

Let $\chi = \{x_1, x_2, ..., x_m\}$ be a given dataset in $\mathbb{R}^d$. Each $x_j$ is a feature vector describing an object. OCSVM use the training data to learn a function $f_\chi : \mathbb{R}^d \mapsto \mathbb{R}$ such that most of the data in $\chi$ belong to the set $\mathcal{R}_\chi = \{x \in \mathbb{R}^d; f_\chi(x \geqslant 0)\}$ while the volume of $\mathcal{R}_\chi$ is minimal. This problem is known as $MinimalVolumeSet$ (MVS) estimation. Membership of $x$ to $\mathcal{R}_\chi$ indicates whether this estimated volume is overall similar to $\chi$ or not. Therefore when considering a $M$-class recognition problem we have to learn $M$ membership functions $f_{\chi_i}$ - one for each class.

OCSVM uses the following approach to estimate the MVS. A kernel function $k(\cdot,\cdot) : \mathbb{R}^d \times \mathbb{R}^d \mapsto \mathbb{R}$. In our research we use a Gaussian Radial Basis Function(RBF) kernel :

$$k(x, x') = exp[- \parallel x - x' \parallel^2 /2\sigma^2], \tag{1}$$

where $x'$ is the object after mapping to a hypersphere, $\| \cdot \|$ denotes the Euclidean norm in $\mathbb{R}^d$. The kernel induces a new, artificial feature space $\mathcal{H}$ by the usage of mapping $\phi : \mathbb{R}^d \mapsto \mathcal{H}$ dened by $\phi(x) \triangleq k(x, \cdot)$. It has been shown that $\mathcal{H}$ reproduces kernel Hilbert spaces of given functions, with dot product denoted as $\langle \cdot, \cdot \rangle_\mathcal{H}$. The reproducing kernel property implies that:

$$\langle \phi(x), \phi(x') \rangle_\mathcal{H} = \langle k(x, \cdot), k(x', \cdot) \rangle_\mathcal{H} = k(x, x'), \tag{2}$$

which makes the evaluation of $k(x, x')$ a linear operation in $\mathcal{H}$, while it is a nonlinear operation in $\mathbb{R}^d$.

Considering the RBF:

$$\parallel \phi(x) \parallel^2_\mathcal{H} \triangleq \langle \phi(x), \phi(x) \rangle_\mathcal{H} = k(x, x) = 1. \tag{3}$$

From this one may assume that all the data mapped into $\mathcal{H}$ are located on the hypersphere with radius equal to one, centered onto the origin of $\mathcal{H}$, which is denoted $S_{(o,R=1)}$. The OCSVM determines in $\mathcal{H}$ the hyperplane $\mathcal{W}$ that separates most of the data from the $S_{(o,R=1)}$, while at the same time maximizing the distance from it. This practically implements the solution to the MVS estimation problem.

Let:

$$\mathcal{W} = \{h(\cdot) \in \mathcal{H}; \langle h(\cdot), w(\cdot) \rangle_\chi - \rho = 0\}, \tag{4}$$

where parameters $w(\cdot)$ and $\rho$ are the results of the following optimization problem

$$\min_{w,\xi,\rho} \frac{1}{2} \parallel w(\cdot) \parallel^2_\mathcal{H} + \frac{1}{vm} \sum_{j=1}^{m} \xi_j - \rho, \tag{5}$$

subject to (for $j = 1, ..., m$)

$$\langle w(\cdot), k(x_j, \cdot) \rangle_\chi \geqslant \rho - \xi_j, \tag{6}$$

where $\xi_j \geqslant 0$, $v$ is a control parameter for the fraction of the data that are allowed to be located on the wrong side of the $\mathcal{W}$ (outliers which do not belong to the $\mathcal{R}_\chi$) and $\xi_j$ are slack variables.

It can be shown that a solution to Eq. (5,6) can be expressed by the following:

$$w(\cdot) = \sum_{j=1}^{m} \alpha_j k(x_j, \cdot), \tag{7}$$

where $\alpha_j$ comes from the dual optimization problem

$$\min_{\alpha} \frac{1}{2} \sum_{j,j'=1}^{m} \alpha_j \alpha_{j'} k(x_j, x_{j'}), \tag{8}$$

subject to $0 \leq \alpha_j \leq \frac{1}{vm}, \sum_j \alpha_j = 1$.

The OCSVM decision function $f_\chi(x)$ is given as follows:

$$f_\chi(x) = \sum_{j}^{m} \alpha_j k(x_j, x) - \rho, \tag{9}$$

where the value of $\rho$ is calculated from knowing that $f_\chi(x_j) = 0$ for those $x_j \in \chi$ that verify both $\alpha_j \neq 0$ and $\alpha_j \neq \frac{1}{vm}$. Objects from $\chi$ that satisfies those conditions are located onto a decision boundary.

## III. Proposed approach

In this paper we propose to use a one-class classification approach to microarray analysis. Let us list the main features and advantages of the proposed approach:

1) We utilize one of the classes as the target concept $\omega_T$ and the remaining one as outliers. In case of imbalanced dataset the minority class is considered as the target concept, while in case of balanced distributions we chose the more numerous class. This is motivated by the fact, that while sacrificing the additional information about the second class, we gain a classifier that is able to adjust itself to the specificity of the given class and is more robust to difficulties that may be encountered, such as class imbalance or in-class noise.
2) The high dimensionality of the feature space is difficult to handle for one-class boundary classifiers. It significantly increases their complexity, the training and execution times and lead to a much more difficult optimization task (and hence to a degradation of the recognition quality). To cope with this problem we propose to use a Random Subspace ensemble to decompose the feature space into smaller competence areas and build an ensemble of simpler one-class models.
3) As Random Subspace may lead to creation of similar classifiers, or classifiers with low discriminative power, a pruning procedure is beneficial, as it may discard irrelevant predictors. We use a diversity-based method, which uses a criterion optimized for OCC task.

### A. Dealing with the high dimensionality problem

The one-class boundary compute a distance between the object $x$ and the estimated boundary, which encloses the target class $\omega_T$. This allows to apply fusion methods, that are based on the discrete output (returned class label) of the individual classifiers - such as the voting methods. However, to apply more sophisticated fusion methods, which assume the continuous outputs of each of the individuals, the support of an object $x$ for a given class is required.

We propose to use the following heuristic support function produced on the basis of a distance:

$$F(x, \omega_T) = \frac{1}{c_1} exp(-d(x|\omega_T)/c_2), \tag{10}$$

which models a Gaussian distribution around the classifier, where $d(x|\omega_T)$ is a distance (Euclidean distance is used) from the evaluated object to the support vectors describing the target concept, $c_1$ is the normalization constant and $c_2$ is the scale parameter. Parameters $c_1$ and $c_2$ should be fitted to the target class distribution.

Estimating this mapping for high dimension is very complex and requires a significant computational power and time. To cope with this difficulty we propose to use a Random Subspace method to partition the dataset into many subspaces of smaller dimensionality. Each base classifier is trained on a new subset, which is highly smaller than the original feature space size. This boosts the training time, while applying ensemble principles makes sure that despite using weaker predictors, we still get a satisfying accuracy [21].

### B. Pruning the ensemble

As Random Subspace may produce classifiers of different level of individual quality and diversity, a classifier section step is most beneficial to forming an one-class ensemble. Multiple Classifier Systems, in order to work properly, must consist of predictors of at the same time high accuracy and diversity. Only mutually complementary classifiers may lead to an improvement over using a single-model approach. Diversity is one of the most popular measures for this task. It may be applied to one-class classifiers, but after modifications, that take into consideration the nature of the OCC problem [22]. For this application an one-class entropy measure [23] is used.

Let's assume that the highest ensemble diversity for a given object $x_j \in X$ is displayed by $[R/2]$ of the ensemble votes with the same value ($\omega_T$ or $\omega_O$) and remaining $R$ - $[R/2]$ with the other value. If all votes returned identical response the ensemble cannot be considered as a diverse one. Let us denote by $r(x_j)$ the number of one-class classifiers that correctly recognize the object $x_j$. Assuming there are $N$ objects in the training set, one may use entropy to measure the diversity using the presented concept:

$$E_{oc}(\Pi^r) = \frac{1}{N} \sum_{j=1}^{N} \frac{1}{(R - [R/2])} \min\{r(x_j), R - r(x_j)\}. \tag{11}$$

where $\Pi^r$ is the considered pool of classifiers.

TABLE I
STATISTICS OF THE DATASETS USED IN THE EXPERIMENTS.

| dataset | samples (class 1 / class 2 ) | features |
|---|---|---|
| Breast Cancer | 78 (34 / 44 ) | 24481 |
| Breast Cancer - noise | 78 (34 / 44 ) | 24481 |
| Central Nervous System | 60 (21 / 39) | 7129 |
| Colon Tumor | 62 (22 / 40) | 6500 |
| Lung Cancer | 181 (31 / 150) | 12533 |

This is a non-pairwise (global) diversity measure, which take values from [0,1]. 0 corresponds to identical ensemble and 1 corresponds to the highest possible diversity.

### C. Fusion method

As a fusion method we use a one-class mean vote, which combines binary output labels of one-class classifiers. It can be written as:

$$y_{mv}(x) = \frac{1}{L}\sum_{k}[(P_k(x|\omega_T) \geq \theta_k)], \qquad (12)$$

where $[(\cdot)]$ is the *Iverson brackets* and $\theta_k$ is threshold for the target class. When a threshold equal to 0.5 is applied this rule transforms into a majority vote for binary problems.

## IV. EXPERIMENTAL INVESTIGATIONS

In this section we evaluate the proposed one-class ensemble on the basis of datasets available at [1], whose details are given in Table I. Four different datasets were used and additional, fifth one, was generated. It was based on the Breast Cancer dataset. To test the performance of classifiers in difficult scenarios we have affected 25% of objects with Gaussian noise, thus creating in-class outliers in the data.

As base classifier we have used an OCSVM with RBF kernel [24].

To put the obtained results into context we have tested the performance of multi-class classifiers used for this task - single SVM (trained with RBF kernel and SMO procedure), Random Forest (consisting of 100 decision trees) and Rotation Forest (consisting of 100 decision trees). Additionally we show the performance of a single OCSVM and the proposed ensemble without the pruning step.

Results are based on leave-one-out cross-validation (LOOCV).

All experiments were carried out in the R environment [25], with classification algorithms taken from the dedicated packages, thus ensuring that the results achieved the best possible efficiency and that the performance was not decreased by a bad implementation. The Friedman ranking test [26] was done for comparison over multiple benchmark datasets.

Firstly the parameters for the proposed pruned one-class ensemble are examined. We test the correlations between the accuracy and size of the subspaces / number of classifiers in the pool. For analyzing the optimal number of the classifiers, a subspace size equal to 0.2 was used. Then, when the size was selected, the subspace size parameter was investigated.

[1] http://datam.i2r.a-star.edu.sg/datasets/krbd/

One should note that these results are prior to the pruning phase - which firther improves the accuracy while reducing the number of classifiers in the pool. Results are presented in Fig 1 - 5.

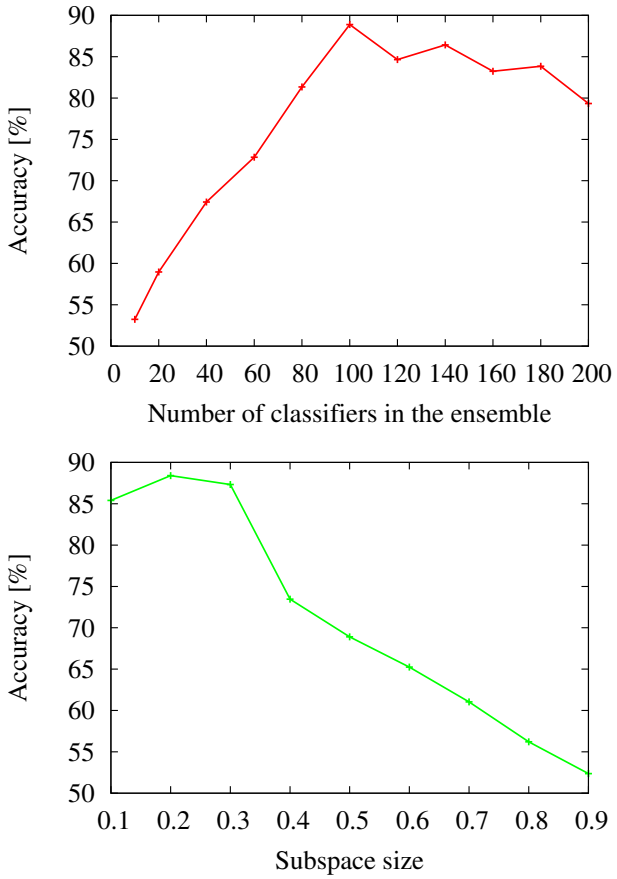


Fig. 1. Correlation between the accuracy and size of the pool of individual classifiers (*top*) and between the accuracy and size of the feature subspaces (*bottom*) for the Breast Cancer dataset.

The established optimal settings are then used for the second stage of the experimental investigation - comparison with other classification methods. Results with respect to sensitivity and specificity, are given in Tab. II.

Analyzing the results of parameter settings for one-class models shows us, that there are some common properties regardless the analyzed dataset. The optimal size of the ensemble was around 100-120 classifiers, built on a small subspaces (consisting of 10% - 20% of features). This allowed to maintain high diversity of the ensemble and allowed for a pruning procedure to select valuable classifiers with mutually complementary areas of competence. Additionally smaller size of the subspaces allowed for training less complex OCSVMs, which in turn prevented them for too overfitted decision boundary.

From the results one may clearly see, that in case of standard microarray datasets the proposed approach returns both specificity and sensitivity similar to those of the state-of-

TABLE II
RECOGNITION SENSITIVITY [%] AND SPECIFICITY [%] FOR EXAMINED METHODS. $RandF$ STANDS FOR RANDOM FOREST, $RotF$ FOR ROTATION FOREST, $OCCE$ FOR AN ONE-CLASS ENSEMBLE WITHOUT PRUNING AND $POCCE$ FOR THE PROPOSED PRUNED ONE-CLASS ENSEMBLE. AVERAGE RANK OF TESTED CLASSIFIERS, ACCORDING TO FRIEDMAN RANKING TEST, ARE GIVEN AT THE BOTTOM.

| Dataset | SVM | | RandF | | RotF | | OCSVM | | OCCE | | POCCE | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Sens [%] | Spec[%] | Sens [%] | Spec[%] | Sens [%] | Spec[%] | Sens [%] | Spec[%] | Sens [%] | Spec[%] | Sens [%] | Spec[%] |
| Breast Cancer | 90.23 | 91.46 | 92.32 | 93.65 | 92.32 | 93.65 | 87.85 | 90.07 | 88.11 | 90.86 | 92.89 | 92.70 |
| Breast Cancer - noise | 74.46 | 83.59 | 77.36 | 84.90 | 80.05 | 85.72 | 75.20 | 82.98 | 80.95 | 85.20 | 89.09 | 90.05 |
| Central Nervous System | 85.60 | 94.36 | 88.20 | 95.90 | 88.20 | 95.90 | 82.95 | 90.11 | 84.07 | 92.01 | 87.84 | 93.96 |
| Colon Tumor | 78.90 | 91.25 | 81.35 | 94.03 | 82.70 | 93.90 | 80.15 | 92.36 | 83.85 | 93.05 | 84.05 | 93.83 |
| Lung Cancer | 61.72 | 93.05 | 65.89 | 95.11 | 67.00 | 94.85 | 69.22 | 92.08 | 70.98 | 93.90 | 74.61 | 94.78 |
| Avg. score | 4.85 | | 2.90 | | 2.25 | | 5.21 | | 4.11 | | 1.68 | |

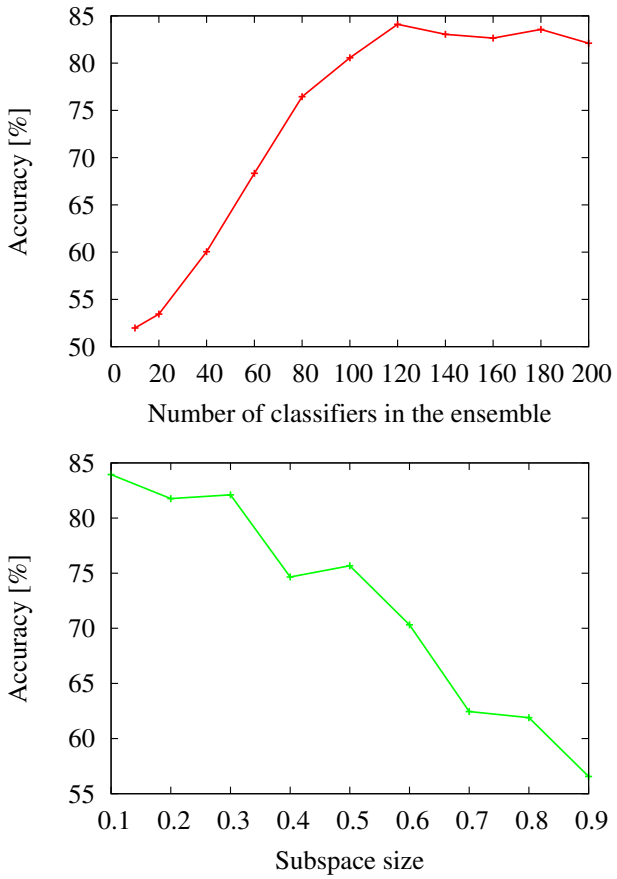


Fig. 2. Correlation between the accuracy and size of the pool of individual classifiers (*top*) and between the accuracy and size of the feature subspaces (*bottom*) for the Breast Cancer - noise dataset.

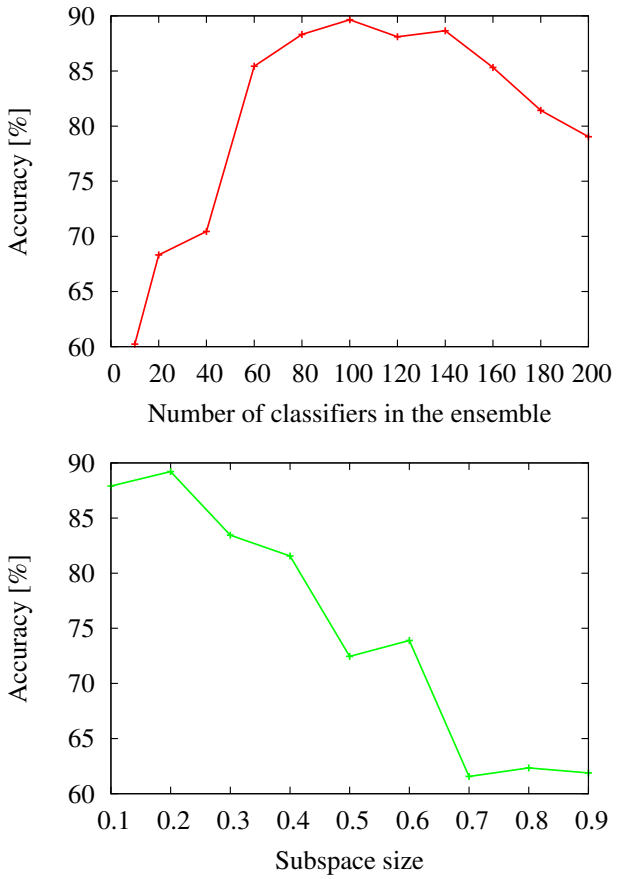


Fig. 3. Correlation between the accuracy and size of the pool of individual classifiers (*top*) and between the accuracy and size of the feature subspaces (*bottom*) for the Central Nervous System dataset.

the-art multi-class models. However in case of noisy (dataset no. 2) and imbalanced (datasets no. 4 and no. 5) our proposed approach is able to outperform significantly the standard classifiers. This happens due to the nature of OCC models - as they are able to learn the distinct properties of the target class, they are able to cope with in-class difficulties.

## V. CONCLUSIONS

In this paper a novel approach for microarray analysis, based on an ensemble of one-class support vector machines, was presented. To deal with the problem of high dimensionality, which may cause difficulties for one-class model, a Random Subspace method was applied. This, combined with a diversity-based pruning step, allowed for an effective classifier, returning similar performance as state-of-the-art multi-class methods. The strong points of the proposed method were revealed when dealing with noisy and imbalanced data. In such a case the proposed combined one-class classifier displayed superior quality over its competitors.

The proposed approach may be an attractive tool for bioinformatics decision support systems, in which we deal with uncertain, noisy data or data coming from uneven distributions.

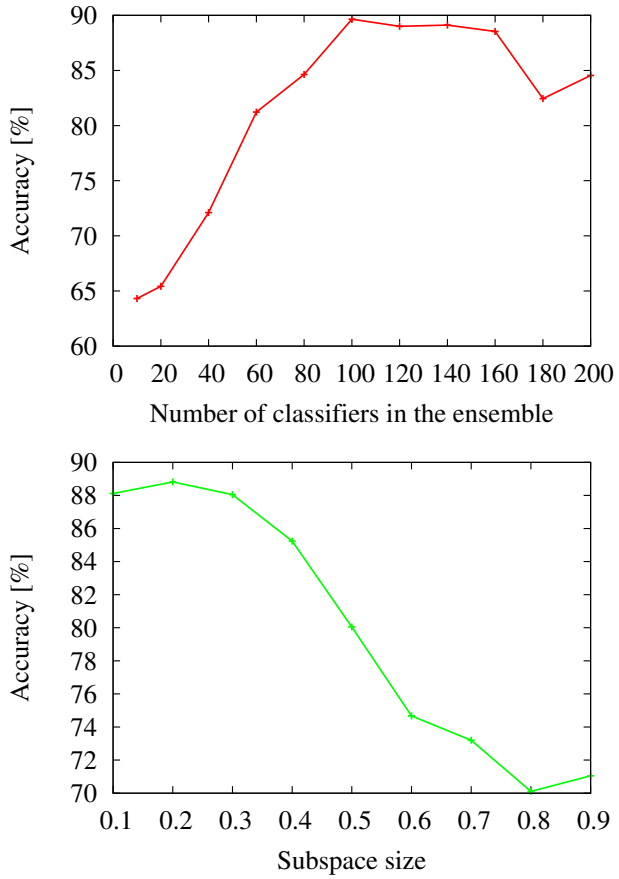


Fig. 4. Correlation between the accuracy and size of the pool of individual classifiers (*top*) and between the accuracy and size of the feature subspaces (*bottom*) for the Colon Tumor dataset.

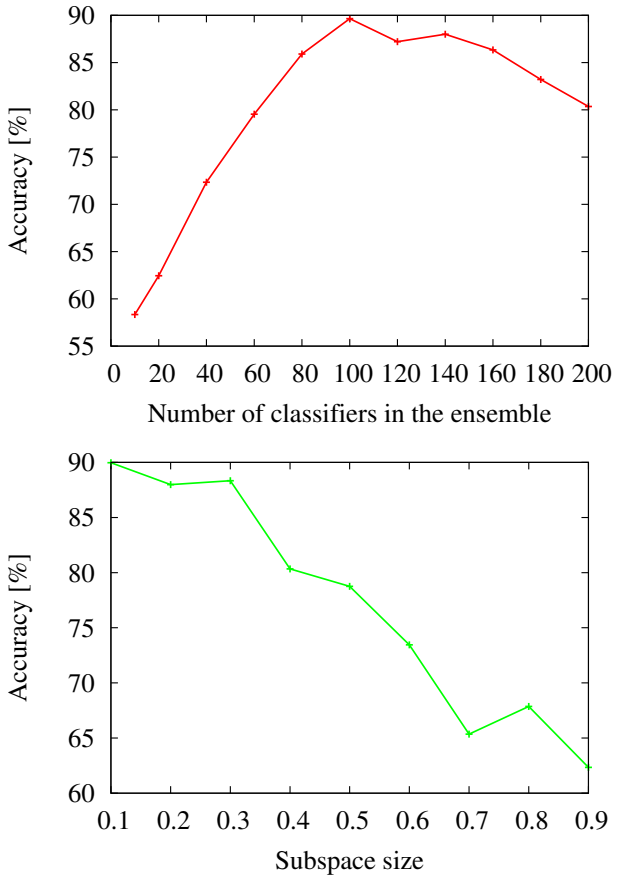


Fig. 5. Correlation between the accuracy and size of the pool of individual classifiers (*top*) and between the accuracy and size of the feature subspaces (*bottom*) for the Lung Cancer dataset.

## VI. Acknowledgment

The work was supported by the fellowship co-financed by European Union within European Social Fund.

## References

[1] A. V. Tinker, A. Boussioutas, and D. D. L. Bowtell, "The challenges of gene expression microarrays for the study of human cancer," *Cancer Cell*, vol. 9, no. 5, pp. 333–339, 2006.

[2] V. S. Silveira, C. A. Scrideli, D. A. Moreno, J. A. Yunes, R. G. P. Queiroz, S. C. Toledo, M. L. M. Lee, A. S. Petrilli, S. R. Brandalise, and L. G. Tone, "Gene expression pattern contributing to prognostic factors in childhood acute lymphoblastic leukemia," *Leukemia and Lymphoma*, vol. 54, no. 2, pp. 310–314, 2013.

[3] T. Schatton, G. F. Murphy, N. Y. Frank, K. Yamaura, A. M. Waaga-Gasser, M. Gasser, Q. Zhan, S. Jordan, L. M. Duncan, C. Weishaupt, R. C. Fuhlbrigge, T. S. Kupper, M. H. Sayegh, and M. H. Frank, "Identification of cells initiating human melanomas," *Nature*, vol. 451, no. 7176, pp. 345–349, 2008.

[4] G. Finak, N. Bertos, F. Pepin, S. Sadekova, M. Souleimanova, H. Zhao, H. Chen, G. Omeroglu, S. Meterissian, A. Omeroglu, M. Hallett, and M. Park, "Stromal gene expression predicts clinical outcome in breast cancer," *Nature medicine*, vol. 14, no. 5, pp. 518–527, 2008.

[5] C. C. Lynch, A. Hikosaka, H. B. Acuff, M. D. Martin, N. Kawai, R. K. Singh, T. C. Vargo-Gogola, J. L. Begtrup, T. E. Peterson, B. Fingleton, T. Shirai, L. M. Matrisian, and M. Futakuchi, "Mmp-7 promotes prostate cancer-induced osteolysis via the solubilization of rankl," *Cancer Cell*, vol. 7, no. 5, pp. 485–496, 2005.

[6] P. Larranaga, B. Calvo, R. Santana, C. Bielza, J. Galdiano, I. Inza, J. A. Lozano, R. Armananzas, G. Santaf, A. Perez, and V. Robles, "Machine learning in bioinformatics," *Briefings in Bioinformatics*, vol. 7, no. 1, pp. 86–112, 2006.

[7] Y. . Wang, Z. . Yu, and V. Anh, "Fuzzy c-means method with empirical mode decomposition for clustering microarray data," *International Journal of Data Mining and Bioinformatics*, vol. 7, no. 2, pp. 103–117, 2013.

[8] M. Ringner, C. Peterson, and J. Khan, "Analyzing array data using supervised methods," *Pharmacogenomics*, vol. 3, no. 3, pp. 403–415, 2002, cited By (since 1996):43. [Online]. Available: www.scopus.com

[9] D. Bariamis, D. Maroulis, and D. K. Iakovidis, "Unsupervised svm-based gridding for dna microarray images," *Computerized Medical Imaging and Graphics*, vol. 34, no. 6, pp. 418–425, 2010.

[10] M. Woźniak, M. Grana, and E. Corchado, "A survey of multiple classifier systems as hybrid systems," *Information Fusion*, 2013, article in Press.

[11] K. Moorthy and M. S. Mohamad, "Random forest for gene selection and microarray data classification," ser. Communications in Computer and Information Science, vol. 295 CCIS, 2012, pp. 174–183.

[12] K. Liu and D. Huang, "Cancer classification using rotation forest," *Computers in biology and medicine*, vol. 38, no. 5, pp. 601–610, 2008.

[13] I. Inza, P. Larraaga, R. Blanco, and A. J. Cerrolaza, "Filter versus wrapper gene selection approaches in dna microarray domains," *Artificial Intelligence in Medicine*, vol. 31, no. 2, pp. 91–103, 2004.

[14] T. K. Ho, "The random subspace method for constructing decision forests," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 20, pp. 832–844, August 1998.

[15] B. Cyganek, "Image segmentation with a hybrid ensemble of one-class support vector machines," ser. Lecture Notes in Computer Science

(including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 2010, vol. 6076 LNAI, no. PART 1, pp. 254–261.

[16] K. Noto, C. Brodley, and D. Slonim, "Frac: A feature-modeling approach for semi-supervised and unsupervised anomaly detection," *Data Mining and Knowledge Discovery*, vol. 25, no. 1, pp. 109–133, 2012.

[17] D. M. J. Tax, P. Juszczak, E. Pekalska, and R. P. W. Duin, "Outlier detection using ball descriptions with adjustable metric," in *Proceedings of the 2006 joint IAPR international conference on Structural, Syntactic, and Statistical Pattern Recognition*, ser. SSPR'06/SPR'06. Berlin, Heidelberg: Springer-Verlag, 2006, pp. 587–595.

[18] B. Krawczyk and M. Woźniak, "Experiments on distance measures for combining one-class classifiers," in *2012 Federated Conference on Computer Science and Information Systems, FedCSIS 2012*, 2012, pp. 89–92.

[19] B. Schölkopf and A. Smola, *Learning with kernels: support vector machines, regularization, optimization, and beyond*, ser. Adaptive computation and machine learning. MIT Press, 2002.

[20] D. M. J. Tax and R. P. W. Duin, "Support vector data description," *Machine Learning*, vol. 54, no. 1, pp. 45–66, 2004.

[21] T. Wilk and M. Woźniak, "Soft computing methods applied to combination of one-class classifiers," *Neurocomput.*, vol. 75, pp. 185–193, Jan. 2012.

[22] B. Krawczyk and M. Woźniak, "Accuracy and diversity in classifier selection for one-class classification ensembles," in *2013 IEEE Symposium Series on Computational Intelligence, Symposium on Computational Intelligence and Ensemble Learning CIEL*, 2013, pp. 46–51.

[23] B. Krawczyk, "Diversity in ensembles for one-class classification," in *New Trends in Databases and Information Systems*, ser. Advances in Intelligent Systems and Computing, M. Pechenizkiy and M. Wojciechowski, Eds. Springer Berlin Heidelberg, 2012, vol. 185, pp. 119–129.

[24] B. Cyganek, "One-class support vector ensembles for image segmentation and classification," *Journal of Mathematical Imaging and Vision*, vol. 42, no. 2-3, pp. 103–117, 2012.

[25] R. D. C. Team, *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria, 2008.

[26] J. Demšar, "Statistical comparisons of classifiers over multiple data sets," *J. Mach. Learn. Res.*, vol. 7, pp. 1–30, 2006.