

# COMPARATIVE STUDY OF MACHINE LEARNING TECHNIQUES FOR SUPERVISED CLASSIFICATION OF BIOMEDICAL DATA

Peter DROTÁR\*, Zdeněk SMÉKAL\*,

\*Department of Telecommunications, Faculty of Electrical Engineering and Communications, Brno University of Technology, Technická 12, Brno, Czech Republic, E-mail: {drotarp,smekal}@feec.vutbr.cz

## ABSTRACT

Several classification methods have been widely used in literature for identification of diseases or differential diagnosis of various types of disorders. Classification methods such as support vector machines, random forests, AdaBoost, deep belief networks, K- nearest neighbors, linear discriminant analysis or perceptron are probably the most popular ones. Even if these methods are frequently used there is a lack of comparison between them to find better framework for classification. In this study, we compared performance of the above mentioned classification methods. The 10-fold cross validation was used to calculate accuracy and Matthews correlation coefficient of the classifiers. In each case these methods were applied to eight binary biomedical datasets. The same evaluation was realized also in conjunction with feature selection technique that passed only hundred most relevant features. Even though there is no single classification method that dominates in terms of performance, we found that some methods provide more consistent performance than others.

**Keywords:** SVM, AdaBoost, Random Forests, Deep Belief Networks, bioinformatics, microarray

## 1. INTRODUCTION

With the expansion of computer methods in bioinformatics and other fields, researchers are more and more frequently faced with machine learning techniques. The machine learning (ML) technology has gained significant attention of biomedical community, mainly because of their potential to improve a process of the disease detection [24]. The most often task of any diagnostic system is a determining or attempting to determine disease or disorder based on observation if some signal. The machine learning techniques used for medical diagnosis have to provide high prediction performance, transparency of diagnostic knowledge and have to provide traceable decision [26].

Recent years witnessed development of wide variety of classification algorithms, beginning from relatively simple methods for classification of linearly separable data such as LDA, perceptron algorithm [23] or naive bayes classifier [13], [1] to more sophisticated and complex method such as support vector machines (SVM) [29] or deep learning [16]. Moreover, ongoing research produce further extension of existing techniques and methods that are better fitted to particular problem [6],[19]. Unfortunately, there are not many direct comparisons of the classification methods, let alone for biomedical applications. Exhaustive comparison of SVM and random forests is given in [25]. Studies including more types of classifiers can be found in [18] or [22].

Here we compare eight state-of-the art classifiers on eight biomedical datasets. Eight state-of-the art classifiers are considered: Support Vector Machines (SVM), AdaBoost classifier, Random Forests, Deep Belief Networks, K- Nearest Neighbours, Naive Bayes classifier, Linear Discriminant analysis and perceptron classifier. We investigate whether there is one or more techniques that clearly dominate in the terms of performance. Particularly, we focus on two evaluation scenarios. First, when all available features are fed to classifier and second, where only hundred features selected by ensemble FS technique are considered for classification.

This paper is organized as follows. Firstly dataset used in this study are described together with details on dataset preprocessing. We continue with brief description of classification algorithms that were applied for class prediction. Finally, classification results in terms of prediction accuracy and Matthews correlation coefficient are given. We conclude the paper with short discussion.

## 2. DATA

The datasets used in this study can be freely downloaded from internet or are available upon request from their author. We evaluated four high dimensional, small sample size dataset and four smaller biomedical datasets. All datasets are real-world datasets consisting of two classes. The basic overview of datasets is provided in Tab. 1.

The B2006 [5] dataset is used for molecular classification of Crohn's disease, ulcerative colitis based on microarray. We follow the approach taken in original paper and pool Crohn's disease and ulcerative colitis together resulting in one class.

C2006 [7] contains gene signatures of 104 subjects. The dataset is used for prediction of breast cancer.

The acute lymphoblastic leukemia (ALL)/ acute myeloid leukemia (AML) dataset G1999 [10] is one of the first datasets used for molecular classification of cancers based on microarray studies. The set consists of 72 patients (47 ALL + 25 AML) and 7129 genes. G1999 is considered as a two class dataset obtained by merging ALL-T and ALL-B together

Last high dimensional dataset G2002 [11] are data used for diagnosis of the lung cancer.

Datasets D2013[8] and T2014 [27] are new datasets for differential diagnosis of Parkinson's disease from handwriting and speech, respectively. The baseline classification accuracy of these two dataset is significantly lower than for four microarray datasets being below 80 % for D2013 and below 90 % for T2014.

Finally last two datasets Z2014[30] and K1988 [31]

contain in contrast to previous datasets more samples than features. The K1988 dataset is originally four class dataset, however third and fourth class are composed only from six subjects, that is too small to have any statistical power. Therefore we left these samples reducing size of the dataset from 148 to 142 samples. The Z2014 represents highly imbalanced dataset.

Our intention was to include various types of datasets that can be encountered in are of biomedical or bioinformatics research and examine the performance od classification algorithms.

### 2.1. Data Preprocessing and Feature Selection

The data were normalized before classification on a per-feature basis to have zero mean and a standard deviation of one.

The goal of the feature selection is to identify the most relevant features or to remove the noisy features in order to avoid some potential degradation in the predictive power [12]. Here, we implement filter features selection. it means that the FS process is applied prior to classification and only features that are evaluated as relevant are fed to the input of classifier.

Instead of implementing single FS algorithm, ensemble techniques apply several weak learners that contribute to final decision. Ensemble FS are especially recommended for small sample domains since they are quite robust to over-fitting and provide stable solutions [28]. The extremely randomized trees were used as base learners in this case [9]. The number of features selected with FS was  $N_{fs} = 100$ . The  $N_{fs} = 100$  was selected similar to other works [14],[21].

## 3. SUPERVISED CLASSIFICATION

Based on the types of training data that are available to learner one can distinguish between several learning or classification scenarios. We will consider supervised learning, when learner receives a set of labeled examples as training data and make prediction for all unseen points. However, in practice the amount of labeled data is relatively small and it is inconvenient to set aside validations sample since this would leave insufficient amount of training data. Instead approach known as  $n$ -fold cross-validation is used [20]. The  $n$ -fold cross-validation consist of randomly partitioning dataset into  $n$  subsamples (folds). Then,  $n - 1$  folds are used for training and the  $n$ -th fold is used as testing dataset.

We consider eight state-of-the art classifiers : Support Vector Machines (SVM), AdaBoost classifier, Random Forests, Deep Belief Networks, K- Nearest Neighbours, Naive Bayes classifier, Linear Discriminant analysis and perceptron classifier.

### 3.1. Support Vector Machines

The underlying idea of SVM classifiers is to calculate a maximal margin hyperplane separating two classes of the data. To learn non-linearly separable functions, the data are implicitly mapped to a higher dimensional space by means

of a kernel function. New samples are classified according to the side of the hyperplane they belong to. We used Radial Basis Function (RBF) kernel [29]. The RBF kernel is defined as

$$K(x, x_i) = e^{-\frac{\|x-x_i\|^2}{2\gamma^2}} \quad (1)$$

where  $\gamma$  controls the width of RBF function.

The parameters kernel gamma  $\gamma$  and penalty parameter  $C$  were optimized using grid search of possible values. Specifically, we searched over the grid  $(C, \gamma)$  defined by the product of the sets  $C = [2^{-10}, 2^{-9}, \dots, 2^6, 2^7]$ ,  $\gamma = [2^{-7}, 2^{-6}, \dots, 2^6, 2^7]$ .

### 3.2. AdaBoost

AdaBoost belongs to the important family of *ensemble methods* known as *boosting*. The key idea behind boosting techniques is to use ensemble methods to combine weak classifiers in order to build a strong learner. AdaBoost is an iterative boosting algorithm constructing a strong classifier as a linear combination of weak classifiers, each performing at least above chance level. As a week classifiers we used decision trees [4]. Similarly to SVM, we searched grid of possible classifier settings to find optimal performance. The grid was determined by the product of the sets  $n_e = [50, 100, 200]$ (maximum number of estimators at which boosting is terminated),  $n_{split} = [1, 2, 3, 5, 10]$  (the number of features to consider when looking for the best split) and  $n_{depth} = [1, 2, 3, 5, 10]$  (the maximum depth of the tree).

### 3.3. Random Forests

A drawback associated with decision trees classifiers is their high variance. In order to improve the stability proposed decision forest methodology [17] was proposed and later further improved by Breiman [3] to provide integrated form of random forest classifier. Random forest classifier is ensemble technique that uses an ensemble of unpruned decision trees, each of which is built on a bootstrap sample of the training data using randomly selected subset of variables. We considered different parameter configurations for the values of  $n_{tree} = [200, 500, 1000]$  (number of trees to build),  $m_{depth} = [1, 2, 3, 5, 10]$  (the maximum depth of the tree) and  $m_{split} = [1, 2, 3, 5, 10]$  (minimum number of samples required to split an internal node).

### 3.4. Deep Belief Network

It has been show that deep architectures have potential to better represent function than shallow ones [15],[16]. The deep belief network (DBN) are formed by stacking restricted boltzman machines (RBM) at the top of each other and train them in the greedy manner. The training strategy for DBNs may hold great promise as a principle to find solution for the problem of training deep networks. Upper layers of a DBN are represent more abstract concepts that explain the input observation, whereas lower layers extract low-level features from data. They learn simpler concepts

**Table 1** Datasets used in this study

Dataset name	source	classes	# samples	# features
B2006	Burczynski [5]	2	127	22,283
C2006	Chowdary [7]	2	104	22,283
G1999	Golub [10]	2	72	7129
G2002	Gordon [11]	2	181	12,533
D2013	Drotar [8]	2	75	204
T2014	Tsanas [27]	2	126	309
Z2014	Zieba [30]	2	470	16
K1988	Zwitter [31]	2	142	18

first and then use them to build and learn more abstract concepts. It is relatively difficult to finely train DBN and there are plenty of parameters to choose. We experimented only with limited number of parameters. The number of units in the input layer and the output layer was set automatically to number of features and number of classes respectively. The number of units in the middle layers was varied through  $n = [100, 300, 500, 1000]$  and  $m = [100, 300, 500]$  with number of epochs  $n_e = [50, 100]$ .

### 3.5. K-Nearest Neighbors

In the  $K$ -NN algorithm,  $K$  - nearest samples in references set are found, by taking a majority vote among the classes of these  $k$  samples. The goal is to determine true class of an undefined test pattern through finding of the nearest neighbors within a hyper-sphere of predefined radius. For the  $K$ -NN classifier, the optimal parameters were search through grid of values  $K = [3, 5, 10, 20, 50]$  and  $n_{leaf} = [10, 30, 50, 100]$ , where  $K$  is number of neighbors and  $n_{leaf}$  is the leaf size.

### 3.6. Naive Bayes Classifier

A naive Bayes classifier is relatively simple probabilistic classifier applying Bayes theorem with strong independence assumption. Basically, it assumes that the value of a particular feature is unrelated to the presence or absence of any other feature, given the class variable. Even though very simple it was frequently used and achieved satisfying results in many classification tasks [13]. We particularly used Gaussian Naive Bayes algorithm where the likelihood of the features is assumed to be Gaussian.

### 3.7. Linear Discriminant Analysis

Using linear discrimination we assume that samples of a class are linearly separable from instances of other classes. The LDA classifier is used frequently due to its simplicity. It does not have high computational requirements and linear model is easy to understand. The final output is a weighted sum of the input features. The magnitude of the feature weight shows the importance of particular feature and its sign indicates if the effect is positive or negative.

### 3.8. Perceptron Learning Algorithm

The perceptron is another algorithm for learning weights of features that tries to find linear decision boundary. In fact, if the data is linearly separable, i.e. there exist some hyperplane that puts all positive samples on one side and negative on the other side, then the perceptron will converge to weight vector separating the data [23].

## 4. EXPERIMENTAL RESULTS

Classifier validation was conducted using stratified ten-fold cross-validation. The process was repeated a total of five times, where in each repetition the original dataset was randomly permuted prior to splitting into training and testing subsets. As an evaluation metric we used conventional classification accuracy and Matthews correlation coefficient (MCC). We decided to use MCC since the classification accuracy alone as a measure is not suitable for imbalanced datasets. MCC takes into account true and false positives and negatives and is generally considered as a balanced measure [2]. MCC is defined as

$$MCC = \frac{(TP \times TN) - (FP \times FN)}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}} \quad (2)$$

where TP is the number of true positives, TN the number of true negatives, FP the number of false positives and FN the number of false negatives. Obviously, the scope of the MCC is within the range of  $< -1, 1 > \times 100\%$ . The larger the MCC value, the better the classifier performance.

Firstly, we consider case where all available features are taken into account during classification. Results for classification accuracy and MCC are presented in Tab. 2 and Tab. 3, respectively. One interesting observation is that SVM is clearly outperformed by other methods for classification of high dimensional datasets B2006, C2006, G1999 and G2002 and provide significantly worse results. When comparing classification accuracy of selected methods there is no single method that clearly dominates, however the highest accuracy is achieved using Ada, RF and DBN classifier. This is quite expected since Ada, RF and DBN are methods that can cope also with data that are hard to separate only by linear function. The results are different when considering MCC as performance measure. Even if the MCC of studied methods are not significantly differ-

**Table 2** Classification accuracy of different ML methods. All features.

Dataset	SVM	Ada	RF	DBN	KNN	NB	LDA	perceptron
B2006	66.92	<b>97.32</b>	94.37	94.62	92.88	86.95	95.79	77.65
C2006	73.22	97.31	96.49	97.13	95.00	91.51	96.20	<b>97.89</b>
G1999	65.00	97.71	<b>98.57</b>	94.46	84.61	98.36	90.50	84.89
G2002	73.71	<b>99.47</b>	99.32	99.33	90.14	97.29	98.47	95.95
D2013	72.64	<b>77.11</b>	76.25	72.61	70.96	68.04	60.21	71.43
T2014	86.81	86.35	85.12	<b>88.55</b>	80.33	52.36	67.95	82.55
Z2014	<b>85.19</b>	84.98	85.11	85.11	85.11	16.00	83.02	60.55
K1988	87.47	88.58	<b>88.64</b>	87.20	85.37	78.33	85.60	80.41
Avg.	76.41	91.10	90.48	89.87	85.55	73.60	84.71	81.41

**Table 3** MCC of different ML methods. All features.

Dataset	SVM	Ada	RF	DBN	KNN	NB	LDA	perceptron
B2006	0	<b>94.15</b>	87.88	88.36	86.17	72.77	91.08	64.39
C2006	42.47	<b>97.31</b>	93.06	94.34	90.50	83.25	92.46	96.01
G1999	0	<b>97.71</b>	96.79	88.74	64.39	96.82	76.41	73.34
G2002	0	<b>99.47</b>	98.27	98.28	70.85	93.38	96.67	91.67
D2013	46.85	<b>77.11</b>	54.60	47.63	44.09	37.94	20.97	44.25
T2014	2.29	<b>86.35</b>	66.92	75.84	59.13	27.00	35.07	64.46
Z2014	<b>18.40</b>	8.35	0	8.48	4.96	0.16	2.98	11.99
K1988	5.23	<b>78.01</b>	77.88	75.08	71.08	55.41	71.55	61.39
Avg.	14.40	<b>79.80</b>	71.92	72.09	61.39	58.34	60.89	63.43

ent Ada provides highest ACC in 7 out of 8 cases. Only dataset where SVM achieves highest MCC is highly imbalanced Z2014 dataset.

The feature selection was applied prior to classification in the second evaluation scenario. Only hundred most relevant features are included in classification process. Again classification accuracy and MCC are given in Tab 4 and Tab. 5, respectively. As can be seen, the best results are again provided by more complex methods such as Ada, RF or DBN. In contrast to previous case, where all features were fed to classifier, SVM performs much better and scores the best results in three datasets. This indicates that even if SVM is perceived as robust to overfitting in classification of high dimensional data the opposite is truth.

## 5. DISCUSSION

Eight classifiers were compared by the means of classification accuracy and MCC. When number of features is significantly higher than number of samples the best results are achieved by AdaBoost classifier. However the performance of RF and DBN is comparable to that of AdaBoost. Similarly, when the number of features are reduced to hundred, RF, SVM, Ada and DBN provide the most consistent performance.

Other studied techniques (KNN, NB, LDA, perceptron) that are relatively simpler can also achieve competitive performance. In fact perceptron or LDA had highest scores for some databases. However performance of these methods is not so consistent and for some databases are clearly outperformed by Ada or SVM.

When comparing MCC, that is more balanced measure, of classifiers with and without feature selection, there is no significant improvement when employing feature selection. However, it is still true that utilization of FS can reduce computational complexity of subsequent classifier and help to better interpret the results by choosing small number of relevant.

## ACKNOWLEDGEMENT

The described research was performed in laboratories supported by the SIX project; the registration number CZ.1.05/2.1.00/03.0072, the operational program Research and Development for Innovation. Peter Drotár was supported by the project CZ.1.07/2.3.00/30.0005 of Brno University of Technology.

## REFERENCES

- [1] S. A. – V. S.V.N. *Introduction to Machine Learning*. Cambridge University Press, Cambridge, UK, 2008.
- [2] BALDI, P. – BRUNAK, S. – CHAUVIN, Y. – ANDERSEN, C. A. F. – NIELSEN, H.: Assessing the accuracy of prediction algorithms for classification: an overview. *Bioinformatics*, 16(5):412–424, 2000.
- [3] BREIMAN, L.: Random forests. *MACHINE Learning*, 45:5–32, 2001.
- [4] BREIMAN, L. – FRIEDMAN, J. – OLSHEN, R. – STONE, C.: *Classification and Regression Trees*. Wadsworth, Belmont, California, 1984.

**Table 4** Classification accuracy of different ML methods. 100 best features according to ensemble FS.

Dataset	SVM	Ada	RF	DBN	KNN	NB	LDA	perceptron
B2006	<b>97.47</b>	96.87	96.86	96.73	95.96	94.42	81.35	96.15
C2006	98.07	96.89	96.18	98.09	<b>98.18</b>	98.00	77.64	98.09
G1999	97.32	97.25	<b>98.57</b>	97.14	95.71	95.89	95.89	94.64
G2002	<b>99.48</b>	98.50	99.17	99.17	98.33	98.33	93.33	99.17
D2013	73.54	<b>76.75</b>	70.18	70.18	76.43	64.82	62.50	69.46
T2014	88.58	86.35	86.47	<b>89.68</b>	85.19	59.74	61.99	86.54
Z2014	<b>85.19</b>	84.98	85.11	85.11	85.11	16.00	83.02	60.55
K1988	87.47	88.58	<b>88.64</b>	87.20	85.37	78.33	85.60	80.41
Avg.	90.89	90.89	90.14	90.41	90.03	75.69	80.16	85.62

**Table 5** MCC of different ML methods. 100 best features according to ensemble FS.

Dataset	SVM	Ada	RF	DBN	KNN	NB	LDA	perceptron
B2006	<b>94.64</b>	93.44	92.99	93.45	90.84	88.28	60.11	92.02
C2006	96.34	94.01	92.22	96.35	<b>96.61</b>	95.83	53.33	96.45
G1999	94.80	94.72	<b>97.30</b>	96.41	90.91	92.35	91.21	89.25
G2002	97.91	96.36	97.75	98.16	95.91	96.83	81.88	<b>98.16</b>
D2013	49.22	<b>56.30</b>	43.92	55.99	53.94	31.31	27.18	39.68
T2014	75.88	69.54	69.63	<b>76.93</b>	67.17	37.55	24.88	72.13
Z2014	<b>18.40</b>	8.35	0	8.48	4.96	0.16	2.98	11.99
K1988	5.23	<b>78.01</b>	77.88	75.08	71.08	55.41	71.55	61.39
Avg.	66.55	73.84	71.46	75.10				

- [5] BURCZYNSKI, M. E. – PETERSON, R. L. – TWINE, N. C. – ZUBEREK, K. A. – BRODEUR, B. J. – CASCIOTTI, L. – MAGANTI, V. – REDDY, P. S. – STRAHS, A. – IMMERMANN, F. – SPINELLI, W. – SCHWERTSCHLAG, U. – SLAGER, A. M. – COTREAU, M. M. – DORNER, A. J.: Molecular Classification of Crohn's Disease and Ulcerative Colitis Patients Using Transcriptional Profiles in Peripheral Blood Mononuclear Cells. *The Journal of Molecular Diagnostics*, 8(1):51–61, feb 2006.
- [6] CHENG, W. C. – JHAN, D. M.: Triaxial accelerometer-based fall detection method using a self-constructing cascade-adaboost-svm classifier. *Biomedical and Health Informatics, IEEE Journal of*, 17(2):411–419, March 2013.
- [7] CHOWDARY, D. – LATHROP, J. – SKELTON, J. – CURTIN, K. – BRIGGS, T. – ZHANG, Y. – YU, J. – WANG, Y. – MAZUMDER, A.: Prognostic Gene Expression Signatures Can Be Measured in Tissues Collected in RNAlater Preservative. *The Journal of Molecular Diagnostics*, 8(1):31–39, feb 2006.
- [8] DROTÁR, P. – MEKYSKA, J. – REKTOROVÁ, I. – MASAROVÁ, L. – SMÉKAL, Z. – FAUNDEZ-ZANUY, M.: Prediction potential of different handwriting tasks for diagnosis of Parkinsons. In *IEEE International Conference on e-Health and Bioengineering (EHB 2013)*, Iasi, Romania, November 2013.
- [9] GEURTS, P. – ERNST, D. – WEHENKEL, L.: Extremely randomized trees. *Machine Learning*, 63(1):3–42, 2006.
- [10] GOLUB, T. R. – SLONIM, D. K. – TAMAYO, P. – HUARD, C. – GAASENBEEK, M. – MESIROV, J. P. – COLLIER, 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, oct 1999.
- [11] GORDON, G. J. G. – JENSEN, R. V. R. – HSIAO, L. L. L. – GULLANS, S. R. S. – BLUMENSTOCK, J. E. J. – RAMASWAMY, S. S. – RICHARDS, W. G. W. – SUGARBAKER, D. J. D. – BUENO, R. R.: Translation of Microarray Data into Clinically Relevant Cancer Diagnostic Tests Using Gene Expression Ratios in Lung Cancer and Mesothelioma. *Cancer Research*, 62(17):4963–4967, sep 2002.
- [12] GUYON, I. – ELISSEEFF, A.: An introduction to variable and feature selection. *J. Mach. Learn. Res.*, 3:1157–1182, mar 2003.
- [13] HAND, D. J. – YU, K.: Idiots bayes not so stupid after all. *International Statistical Review*, 69(3):385–398, 2001.
- [14] HAURY, A. C. – GESTRAUD, P. – VERT, J. P.: The influence of feature selection methods on accuracy,

- stability and interpretability of molecular signatures. *PLoS ONE*, 6, dec 2011.
- [15] HINTON, G. E. – OSINDERO, S. – TEH, Y. W.: A fast learning algorithm for deep belief nets. *Neural Computation*, 18(7):1527–1554, 2006.
- [16] HINTON, G. E. – SALAKHUTDINOV, R.: Reducing the dimensionality of data with neural networks. *Science*, 313(5786):504–507, 2006.
- [17] HO, T.: Random decision forest. In *In Proc. International Conference on Document Analysis and Recognition*, pages 278–282, Montreal, Canada, 1995.
- [18] HU, H. – LI, J. – PLANK, A. – WANG, H. – DAGGARD, G.: A comparative study of classification methods for microarray data analysis. In *Proceedings of the Fifth Australasian Conference on Data Mining and Analytics - Volume 61*, AusDM '06, pages 33–37, Darlinghurst, Australia, Australia, 2006. Australian Computer Society, Inc.
- [19] JIANG, Z. – LIN, Z. – DAVIS, L.: Label consistent k-svd: Learning a discriminative dictionary for recognition. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 35(11):2651–2664, Nov 2013.
- [20] MOHRI, M. – ROSTAMIZADEH, A. – TALWALKAR, A.: *Foundations of Machine Learning*. The MIT Press, Cambridge, MA, 1st edition, 2012.
- [21] MOULOS, P. – KANARIS, I. – BONTEMPI, G.: Stability of feature selection algorithms for classification in high-throughput genomics datasets. In *Bioinformatics and Bioengineering (BIBE), 2013 IEEE 13th International Conference on*, pages 1–4, Nov 2013.
- [22] PIROOZANIA, M. – YANG, J. – YANG, M. – DENG, Y.: A comparative study of different machine learning methods on microarray gene expression data. *BMC Genomics*, 9(Suppl 1):1–13, 2008.
- [23] ROSENBLATT, F.: The perceptron - a perceiving and recognizing automaton. Technical Report 85-460-1, Cornell Aeronautical Laboratory, 1957.
- [24] SAJDA, P.: Machine learning for detection and diagnosis of disease. *Annual Review of Biomedical Engineering*, 8(1):537–565, 2006. PMID: 16834566.
- [25] STATNIKOV, A. – WANG, L. – ALIFERIS, C.: A comprehensive comparison of random forests and support vector machines for microarray-based cancer classification. *BMC Bioinformatics*, 9(1):1–10, 2008.
- [26] TRIPOLITI, E. – FOTIADIS, D. – MANIS, G.: Automated diagnosis of diseases based on classification: Dynamic determination of the number of trees in random forests algorithm. *Information Technology in Biomedicine, IEEE Transactions on*, 16(4):615–622, July 2012.
- [27] TSANAS, A. – LITTLE, M. – FOX, C. – RAMIG, L.: Objective automatic assessment of rehabilitative speech treatment in parkinsons disease. *IEEE Trans. Neural Syst. Rehabil. Eng.*, (22):181–190, 2014.
- [28] TUV, E. – BORISOV, A. – RUNGER, G. – TORKKOLA, K.: Feature selection with ensembles, artificial variables, and redundancy elimination. *J. Mach. Learn. Res.*, 10:1341–1366, Dec. 2009.
- [29] VAPNIK, V.: *Statistical Learning Theory*. John Wiley & Sons, London, 1st edition, 1998.
- [30] ZIEBA, M. – TOMCZAK, J. M. – LUBICZ, M. – SWIATEK, J.: Boosted {SVM} for extracting rules from imbalanced data in application to prediction of the post-operative life expectancy in the lung cancer patients. *Applied Soft Computing*, 14, Part A(0):99 – 108, 2014.
- [31] ZWITTER, M. – SOKLIC, M.: UCI machine learning depository. <http://archive.ics.uci.edu/ml/datasets/Lymphography>. Accessed: 2014-02-20.

Received August 13, 2014, accepted October 15, 2014

## BIOGRAPHY

**Peter Drotár** was born in 1984 in Košice, Slovakia. He received the M.Sc. in Electronics and Telecommunications from Technical University of Košice in 2007 and PhD from the same university in 2010. From 2010 to 2012 he worked as a scientist for Honeywell Aerospace - Advanced Technology Europe. Currently he is postdoctoral researcher at the Brno University of Technology, Czech Republic. His research interests include handwriting and biomedical signal processing. He also deals with machine learning and its application in decision support systems for biomedicine.

**Zdeněk Smékal** is a full professor at the Faculty of Electrical Engineering and Communications, Brno University of Technology. He received a professor degree in 2000 in the field of Theoretical Electrical Engineering. Since 1998 he is a member of IEEE (Signal Processing, Communications and Computer Societies), since 2006 as a Senior Member. He is also a member of Audio Engineering Society. He is a member of scientific councils (FEEC in BUT Brno and FEE in CTU Prague). He is an author of three monographs and more than 150 papers for international conferences and journals.