Results of an Adaboost Approach on Alzheimer's Disease Detection on MRI

Alexandre Savio[1], Maite García-Sebastián[1], Manuel Graña[1,*],
and Jorge Villanúa[2]

[1] Grupo de Inteligencia Computacional
www.ehu.es/ccwintco
[2] Osatek, Hospital Donostia Paseo Dr. Beguiristain 109, 20014 San Sebastián, Spain

Abstract. In this paper we explore the use of the Voxel-based Morphometry (VBM) detection clusters to guide the feature extraction processes for the detection of Alzheimer's disease on brain Magnetic Resonance Imaging (MRI). The voxel location detection clusters given by the VBM were applied to select the voxel values upon which the classification features were computed. We have evaluated feature vectors computed over the data from the original MRI volumes and from the GM segmentation volumes, using the VBM clusters as voxel selection masks. We use the Support Vector Machine (SVM) algorithm to perform classification of patients with mild Alzheimer's disease vs. control subjects. We have also considered combinations of isolated cluster based classifiers and an Adaboost strategy applied to the SVM built on the feature vectors. The study has been performed on MRI volumes of 98 females, after careful demographic selection from the Open Access Series of Imaging Studies (OASIS) database, which is a large number of subjects compared to current reported studies. Results are moderately encouraging, as we can obtain up to 85% accuracy with the Adaboost strategy in a 10-fold cross-validation.

1 Introduction

Alzheimer's disease (AD) is a neurodegenerative disorder, which is one of the most common cause of dementia in old people. Currently, due to the socioeconomic importance of the disease in occidental countries it is one of the most studied. The diagnosis of AD can be done after the exclusion of other forms of dementia but a definitive diagnosis can only be made after a post-mortem study of the brain tissue. This is one of the reasons why Magnetic Resonance Imaging (MRI) based early diagnosis is a current research hot topic in the neurosciences.

Morphometry analysis has become a common tool for computational brain anatomy studies. It allows a comprehensive measurement of structural differences within a group or across groups, not just in specific structures, but throughout the entire brain. Voxel-based morphometry (VBM) is a computational approach

to neuroanatomy that measures differences in local concentrations of brain tissue, through a voxel-wise comparison of multiple brain images [1]. For instance, VBM has been applied to study volumetric atrophy of the grey matter (GM) in areas of neocortex of AD patients vs. control subjects [3,16,9]. The procedure involves the spatial normalization of subject images into a standard space, segmentation of tissue classes using *a priori* probability maps, smoothing to correct noise and small variations, and voxel-wise statistical tests. Statistical analysis is based on the General Linear Model (GLM) to describe the data in terms of experimental and confounding effects, and residual variability. Classical statistical inference is used to test hypotheses that are expressed in terms of GLM estimated regression parameters. This computation of given contrast provides a Statistical Parametric Map (SPM), which is thresholded according to the Random Field theory.

Machine learning methods have become very popular to classify functional or structural brain images to discriminate them into normal or a specific neurodegenerative disorder. The Support Vector Machine (SVM) either with linear [10,15] or non-linear [6,11] kernels are the state of the art to build up classification and regression systems. Besides MRI, other medical imaging methods are being studied for AD diagnosis. There are different ways to extract features from MRI for SVM classification: based on morphometric methods [5,6], based on regions of interest (ROI) [13,11] or GM voxels in automated segmentation images [10]. Work has also been reported on the selection of a small set of the most informative features for classification, such as the SVM-Recursive Feature Elimination [6], the selection based on statistical tests [13,15], the wavelet decomposition of the RAVENS maps [11], among others.

Many of the classification studies on the detection of AD were done with both men and women. However, it has been demonstrated that brains of women are different from men's to the extent that it is possible to discriminate the gender via MRI analysis [11]. Moreover, it has been shown that VBM is sensitive to the gender differences. For these reasons, we have been very cautious in this study. We have selected a set of 98 MRI women's brain volumes. It must be noted that this is a large number of subjects compared with the other studies referred above.

Our approach is to use the VBM detected clusters as a mask on the MRI and Grey Matter (GM) segmentation images to select the potentially most discriminating voxels. Feature vectors for classification are either the voxel values or some summary statistics of each cluster. We both consider the feature vector computed from all the VBM clusters and the combination of the individual classifiers built from the clusters independently. We build our classification systems using the standard SVM, testing linear and non-linear (RBF) kernels. Best results are obtained with an Adaptive Boosting (AdaBoost) strategy tailored to the SVM [12]. Section 2 gives a description of the subjects selected for the study, the image processing, feature extraction details and the classifier system. Section 3 gives our classification performance results and section 4 gives the conclusions of this work and further research suggestions.

2   Materials and Methods

2.1   Subjects

Ninety eight right-handed women (aged 65-96 yr) were selected from the Open Access Series of Imaging Studies (OASIS) database (http://www.oasis-brains.org) [14]. OASIS data set has a cross-sectional collection of 416 subjects covering the adult life span aged 18 to 96 including individuals with early-stage Alzheimer's Disease. We have ruled out a set of 200 subjects whose demographic, clinical or derived anatomic volumes information was incomplete. For the present study there are 49 subjects who have been diagnosed with very mild to mild AD and 49 nondemented. A summary of subject demographics and dementia status is shown in table 1.

Table 1. Summary of subject demographics and dementia status. Education codes correspond to the following levels of education: 1 less than high school grad., 2: high school grad., 3: some college, 4: college grad., 5: beyond college. Categories of socioeconomic status: from 1 (biggest status) to 5 (lowest status). MMSE score ranges from 0 (worst) to 30 (best).

	Very mild to mild AD	Normal
No. of subjects	49	49
Age	78.08 (66-96)	77.77 (65-94)
Education	2.63 (1-5)	2.87 (1-5)
Socioeconomic status	2.94 (1-5)	2.88 (1-5)
CDR (0.5 / 1 / 2)	31 / 17 / 1	0
MMSE	24 (15-30)	28.96 (26-30)

2.2   Imaging Protocol

Multiple (three or four) high-resolution structural T1-weighted magnetization-prepared rapid gradient echo (MP-RAGE) images were acquired [7] on a 1.5-T Vision scanner (Siemens, Erlangen, Germany) in a single imaging session. Image parameters: TR= 9.7 msec., TE= 4.0 msec., Flip angle= 10, TI= 20 msec., TD= 200 msec., 128 sagittal 1.25 mm slices without gaps and pixels resolution of 256×256 (1×1mm).

2.3   Image Processing and VBM

We have used the average MRI volume for each subject, provided in the OASIS data set. These images are already registered and resampled into a 1-mm isotropic image in atlas space and the bias field has been already corrected [14]. The Statistical Parametric Mapping (SPM5) (http://www.fil.ion.ucl.ac.uk/spm/) was used to compute the VBM which gives us the spatial mask to obtain the classification features. Images were reoriented into a right-handed coordinate system to work

with SPM5. The tissue segmentation step does not need to perform bias correction. We performed the modulation normalization for grey matter, because we are interested in this tissue for this study. We performed a spatial smoothing before performing the voxel-wise statistics, setting the Full-Width at Half-Maximum (FWHM) of the Gaussian kernel to 10mm isotropic. A GM mask was created from the average of the GM segmentation volumes of the subjects under study. Thresholding the average GM segmentation, we obtain a binary mask that includes all voxels with probability greater than 0.1 in the average GM segmentation volume. This interpretation is not completely true, since the data are modulated, but it is close enough for the mask to be reasonable. We design the statistical analysis as a Two-sample t-test in which the first group corresponds with AD subjects. The general linear model contrast has been set as [-1 1], a right-tailed (groupN ¿ groupAD), correction FWE, p-value=0.05. The VBM detected clusters are used for the MRI feature extraction for the SVM classification.

2.4   Support Vector Machine Classification

The Support Vector Machine (SVM)[18] algorithm used for this study is included in the libSVM (`http://www.csie.ntu.edu.tw/~cjlin/libsvm/`) software package. The implementation is described in detail in [4]. Given training vectors $x_i \in R_n, i = 1, \ldots, l$ of the subject features of the two classes, and a vector $y \in R^l$ such that $y_i \in \{-1, 1\}$ labels each subject with its class, in our case, for example, patients were labeled as -1 and control subject as 1. To construct a classifier, the SVM algorithm solves the following optimization problem:

$$\min_{w,b,\xi} \frac{1}{2} w^T w + C \sum_{i=1}^{l} \xi_i$$

subject to $y_i(w^T \phi(x_i) + b) \geq (1 - \xi_i)$, $\xi_i \geq 0, i = 1, 2, \ldots, n$. The dual optimization problem is

$$\min_{\alpha} \frac{1}{2} \alpha^T Q \alpha - e^T \alpha$$

subject to $y^T \alpha = 0$, $0 \leq \alpha_i \leq C$, $i = 1, \ldots, l$. Where e is the vector of all ones, $C > 0$ is the upper bound on the error, Q is an l by l positive semi-definite matrix, $Q_{ij} \equiv y_i y_j K(x_i, x_j)$, and $K(x_i, x_j) \equiv \phi(x_i)^T \phi(x_j)$ is the kernel function that describes the behavior of the support vectors. Here, the training vectors x_i are mapped into a higher (maybe infinite) dimensional space by the function $\phi(x_i)$. The decision function is $sgn(\sum_{i=1}^{l} y_i \alpha_i K(x_i, x) + b)$. C is a regularization parameter used to balance the model complexity and the training error.

The kernel function chosen results in different kinds of SVM with different performance levels, and the choice of the appropriate kernel for a specific application is a difficult task. In this study two different kernels were tested: the linear and the radial basis function (RBF) kernel. The linear kernel function is defined as $K(x_i, x_j) = 1 + x_i^T x_j$, this kernel shows good performance for linearly separable data. The RBF kernel is defined as $K(x_i, x_j) = exp(-\frac{||x_i - x_j||^2}{2\sigma^2})$. This

kernel is basically suited best to deal with data that have a class-conditional probability distribution function approaching the Gaussian distribution [2]. One of the advantages of the RBF kernel is that given the kernel, the number of support vectors and the support vectors are all automatically obtained as part of the training procedure, i.e., they do not need to be specified by the training mechanism.

2.5   Feature Extraction

We have tested three different feature vector extraction processes, based on the voxel location clusters detection obtained from the VBM analysis.

1. The first feature extraction process computes the ratio of GM voxels to the total number of voxels of each voxel location cluster.
2. The second feature extraction process computes the mean and standard deviation of the GM voxel intensity values of each voxel location cluster.
3. The third feature feature extraction process computes a very high dimensional vector with all the GM segmentation values for the voxel locations included in each VBM detected cluster. The GM segmentation voxel values were ordered in this feature vector according to the coordinate lexicographic order.

First, we have considered all the VBM detected clusters together, so that each feature vector characterizes the whole MRI volume.

2.6   Combination of SVM

We have considered also the construction of independent SVM classifiers for each VBM detected cluster and the combination of their responses by a simple majority voting, and to use the cluster with greatest statistical significance to resolve ties. This can be viewed as a simplified combination of classifiers. Furthermore, we have defined a combination of classifiers weighted by the individual training errors, where the classifier weights are computed as in the AdaBoost-SVM algorithm in [12] (Algorithm 1), assuming an uniform weighting of the data samples.

2.7   Adaptive Boosting

Adaptive Boosting (AdaBoost)[17,8] is a meta-algorithm for machine learning that can be used in conjunction with many other learning algorithms to improve their performance. AdaBoost is adaptive in the sense that subsequent classifiers built are tweaked in favor of those instances misclassified by previous classifiers. AdaBoost is sensitive to noisy data and outliers. Otherwise, it is less susceptible to the over-fitting problem than most learning algorithms.

AdaBoost calls a weak classifier repeatedly in a series of rounds $t = 1, ..., T$. For each call a distribution of weights W_t is updated and indicates the importance of examples in the data set for the classification. On each round, the weights

---

Algorithm 1. Combining the independent SVM trained per cluster

---

1. **Input:** as many sets of training samples with labels as clusters in the statistical parametric map $T_k = \{(x_1, y_1), \ldots, (x_N, y_N)\}, k = 1..C$, where N is the number of samples of each cluster.
2. **Initialize:** the weights of training samples: $w_i^k = 1/N$, for all $i = 1, \ldots, N$
3. **For each k cluster do**

   (a) Search the best γ for the RBF kernel for the training set T_k, we denote it as γ_k.
   (b) Train the SVM with T_k and γ_k, we denote the classifier as h_k.
   (c) Classify the same training T_k set with h_k.
   (d) Calculate the training error of h_k: $\epsilon_k = \sum_{i=1}^{N} w_i^k$, $y_i \neq h_k(x_i)$.
   (e) Compute the weight of the cluster classifier h_k: $\alpha_k = \frac{1}{2} \ln(\frac{\epsilon_k}{1-\epsilon_k})$.

4. **Output:**   for   each   test   data x its   classification   is $f(x)$ $=$ $sign(\sum_{k=1}^{C} \alpha_k h_k(x))$.

---

of each incorrectly classified example are increased (or alternatively, the weights of each correctly classified example are decreased), so that the new classifier focuses more on those examples.

Following these ideas, we have also tested a combination of SVM classifiers along the ideas from the Diverse AdaBoost SVM [12], presented as Algorithm 2. In this approach we built a sequence of SVM classifiers of increasing variance parameter. The results of the classifiers are weighted according to their statistical error to obtain the response to the test inputs in the 10-fold validation process.

2.8   Classifier Performance Indices

We evaluated the performance of the classifiers built with the diverse strategy using 10 times the 10-fold cross-validation methodology. To quantify the results we measured the accuracy, the ratio of the number of test volumes correctly classified to the total of tested volumes. We also quantified the specificity and sensitivity of each test defined as $Specificity = \frac{TP}{TP+FP}$ and $Sensitivity = \frac{TN}{TN+FN}$, where TP is the number of true positives: number of AD patient volumes correctly classified; TN is the number of true negatives: number of control volumes correctly classified; FP is the number of false positives: number of AD patient volumes classified as control volume; FN is the number of false negatives: number of control volumes classified as patient. The regularization parameter C of all the SVM classifiers trained for this study was set to 1.

Algorithm 2. Diverse AdaBoostSVM

---

1. **Input:** a set of training samples with labels $\{(x_1, y_1), \ldots, (x_N, y_N)\}$; the initial σ, σ_{ini}; the minimal σ, σ_{min}; the step of σ, σ_{step}; the threshold on diversity DIV.
2. **Initialize:** the weights of training samples: $w_i^t = 1/N$, for all $i = 1, \ldots, N$
3. **Do while** $(\sigma > \sigma_{ini})$
   (a) Calculate gamma: $\gamma = (2\sigma^2)^{-1}$.
   (b) Use σ to train a component classifier h_t on the weighted training set.
   (c) Calculate the training error of h_t: $\epsilon_t = \sum_{i=1}^{N} w_i^t, \quad y_i \neq h_t(x_i)$.
   (d) Calculate the diversity of h_t: $D_t = \sum_{i=1}^{N} d_t(x_i)$, where $d_t(x_i) = \begin{cases} 0 & if\ h_t(x_i) = y_i \\ 1 & if\ h_t(x_i) \neq y_i \end{cases}$
   (e) Calculate the diversity of weighted component classifiers and the current classifier: $D = \sum_{t=1}^{T} \sum_{i=1}^{N} d_t(x_i)$.
   (f) If $\epsilon_t > 0.5$ or $D < DIV$: decrease σ by σ_{step} and go to (a).
   (g) Set weight of the component classifier h_t: $\alpha_t = \frac{1}{2} \ln(\frac{\epsilon_t}{1 - \epsilon_t})$.
   (h) Update the weights of training samples: $w_i^{t+1} = w_i^t exp(-\alpha y_i h_t(x_i))$.
   (i) Normalize the weights of training samples: $w_i^{t+1} = w_i^{t+1} (\sum_{i=1}^{N} w_i^{t+1})^{-1}$.
4. **Output:** $f(x) = sign(\sum_{k=1}^{C} \alpha_k h_k(x))$.

---

3   Results

In this section we present for each experiment the following data: the number of features, accuracy, specificity, which is related to AD patients and sensitivity, which is related to control subjects. We will give results on the global feature vectors, the simple voting of independent classifiers based on statistical significance of VBM, the weighted combination of individual cluster SVM based on training errors, and an adaptive boosting strategy for combining classifiers.

3.1   Global Feature Vectors

The VBM performed for this study was described in section 2. We present in table 2 the results of the three feature computation processes applied to the whole set of VBM clusters to obtain a single feature vector for the whole volume. Each table entry contains the SVM results using the linear (lk) and RBF (nlk) kernels upon the corresponding feature vector set. The table rows correspond to the feature extraction processes described in section 2.5. Table 2 best accuracy result is 80.6% with the RBF kernel, but this result is not too far from the results of the linear kernel SVM. This best accuracy result is obtained with a rather straightforward feature extraction method: the mean and standard deviation of the MRI voxel intensities. This means that MRI intensities may have discriminant value.

Table 2. Classification results with a linear kernel (lk) and a non-linear RBF kernel (nlk). The values of $\gamma = \left(2\sigma^2\right)^{-1}$ for non linear kernel were 0.5, 0.031, 0.0078 for each feature extraction process, respectively.

Feature extracted	#Features	Accuracy (lk/nlk)	Sensitivity (lk/nlk)	Specificity (lk/nlk)
GM proportion	12	69.39 / 68.36	0.63 / 0.61	0.88 / 0.90
Mean & StDev	24	78.57 / 80.61	0.72 / 0.75	0.88 / 0.89
Voxel intensities	3611	73.47 / 76.53	0.72 / 0.77	0.75 / 0.76

Overall the sensitivity results in table 2 is much lower than the specificity. We believe that the source of error is the confusion of mild demented AD patients with control subjects. Upon inspection, this hypothesis seems to be correct for this data.

3.2   Combination of Individual Cluster SVM

Table 3 presents the results of the combination of SVM classifiers built up over each cluster independently, searching for the best kernel parameter σ in each classifier independently. The voxel clusters are selected according to the VBM performed as described above. The results do not improve over the ones obtained with the whole image feature vector. We note that, contrary to the global feature vector, the results improve when considering the whole collection of MRI voxel intensities.

Table 4 presents the results of the combination of individual weighted SVM classifiers. Each SVM classifier was trained with one VBM cluster feature set and the weights were computed according to its training error. We obtain a

Table 3. Majority voting classification results with linear kernel (lk) and non-linear kernel (nlk) SVM built independently for each VBM cluster

Feature extracted	#Features	Accuracy (lk/nlk)	Sensitivity (lk/nlk)	Specificity (lk/nlk)
Mean & StDev	24	74% / 75%	0.51 / 0.56	0.97 / 0.95
Voxel intensities	3611	77% / 78%	0.74 / 0.76	0.80 / 0.82

Table 4. Weighted individual SVM per cluster classification results. The value of the RBF kernels for the nonlinear (nlk) classifiers were searched for the best fit to the training set.

Feature extracted	Features	Accuracy (lk/nlk)	Sensitivity (lk/nlk)	Specificity (lk/nlk)
Mean & StDev	24	71% / 79%	0.54 / 0.78	0.88 / 0.80
Voxel intensities	3611	73% / 86%	0.76 / 0.80	0.70 / 0.92

Table 5. Diverse AdaBoostSVM classification results

Feature extracted	Features	Accuracy	Sensitivity	Specificity
Mean & StDev	24	85%	0.78	0.92
Voxel intensities	3611	78%	0.71	0.85

significant improvement of the accuracy when considering the voxel intensities as features for the non-linear RBF SVM.

Table 5 shows the results of the Diverse . The σ_{min} is set as 0.1, the σ_{ini} is set as 100 and σ_{step} is set as 0.1. The DIV value is set as as 0.6.

4   Conclusions

In this work we have studied feature extraction processes based on VBM analysis, to classify MRI volumes of AD patients and normal subjects. We have analyzed different designs for the SPM of the VBM and we have found that the basic GLM design without covariates can detect subtle changes between AD patients and controls that lead to the construction of SVM classifiers with a discriminative accuracy of 86% in the best case. The weighted cluster SVM and the Diverse AdaBoostSVM methods improved remarkably the results, mainly the sensitivity of the classification models. In [5] they compare their results on a smaller population of controls and AD patients to the ones obtained with a standard VBM analysis, using a cluster and found a classification accuracy of 63.3% via cross-validation. Therefore, the results shown in this paper, along with the careful experimental methodology employed, can be of interest for the Neuroscience community researching on the AD diagnosis based on MRI. Further work may address the extraction of features based on other morphological measurement techniques, such as the Deformation-based Morphometry.

Acknowledgments

References

1. Ashburner, J., Friston, K.J.: Voxel-based morphometry: The methods. Neuroimage 11(6), 805–821 (2000)
2. Burges, C.: A tutorial on support vector machines for pattern recognition. Data Mining and Knowledge Discovery 2(2), 167 (1998)
3. Busatto, G.F., Garrido, G.E.J., Almeida, O.P., Castro, C.C., Camargo, C.H.P., Cid, C.G., Buchpiguel, C.A., Furuie, S., Bottino, C.M.: A voxel-based morphometry study of temporal lobe gray matter reductions in alzheimer's disease. Neurobiology of Aging 24(2), 221–231 (2003)
4. Chang, C.-C., Lin, C.-J.: LIBSVM: a library for support vector machines. Software (2001), `http://www.csie.ntu.edu.tw/~cjlin/libsvm`
5. Davatzikos, C., Fan, Y., Wu, X., Shen, D., Resnick, S.M.: Detection of prodromal alzheimer's disease via pattern classification of magnetic resonance imaging. Neurobiology of Aging 29(4), 514–523 (2008)
6. Fan, Y., Shen, D., Davatzikos, C.: Classification of Structural Images via High-Dimensional Image Warping, Robust Feature Extraction, and SVM, pp. 1–8 (2005)

7. Fotenos, A.F., Snyder, A.Z., Girton, L.E., Morris, J.C., Buckner, R.L.: Normative estimates of cross-sectional and longitudinal brain volume decline in aging and AD. Neurology 64(6), 1032–1039 (2005)
8. Freund, Y., Schapire, R.: A decision-theoretic generalization of on-line learning and an application to boosting. In: European Conference on Computational Learning Theory, pages 37, 23 (1995)
9. Frisoni, G.B., Testa, C., Zorzan, A., Sabattoli, F., Beltramello, A., Soininen, H., Laakso, M.P.: Detection of grey matter loss in mild alzheimer's disease with voxel based morphometry. Journal of Neurology, Neurosurgery & Psychiatry 73(6), 657–664 (2002)
10. Kloppel, S., Stonnington, C.M., Chu, C., Draganski, B., Scahill, R.I., Rohrer, J.D., Fox, N.C., Jack Jr., C.R., Ashburner, J., Frackowiak, R.S.J.: Automatic classification of MR scans in alzheimer's disease. Brain 131(3), 681 (2008)
11. Lao, Z., Shen, D., Xue, Z., Karacali, B., Resnick, S.M., Davatzikos, C.: Morphological classification of brains via high-dimensional shape transformations and machine learning methods. Neuroimage 21(1), 46–57 (2004)
12. Li, X., Wang, L., Sung, E.: A study of AdaBoost with SVM based weak learners. In: Proceedings of IEEE International Joint Conference on Neural Networks, IJCNN 2005, vol. 1, pp. 196–201 (2005)
13. Liu, Y., Teverovskiy, L., Carmichael, O., Kikinis, R., Shenton, M., Carter, C.S., Stenger, V.A., Davis, S., Aizenstein, H., Becker, J.T.: Discriminative MR image feature analysis for automatic schizophrenia and alzheimer's disease classification. In: Barillot, C., Haynor, D.R., Hellier, P. (eds.) MICCAI 2004. LNCS, vol. 3216, pp. 393–401. Springer, Heidelberg (2004)
14. Marcus, D.S., Wang, T.H., Parker, J., Csernansky, J.G., Morris, J.C., Buckner, R.L.: Open access series of imaging studies (OASIS): cross-sectional MRI data in young, middle aged, nondemented, and demented older adults. Journal of Cognitive Neuroscience 19(9), 1498–1507 (2007) PMID: 17714011
15. Ramirez, J., Gorriz, J.M., Lopez, M., Salas-Gonzalez, D., Alvarez, I., Segovia, F., Puntonet, C.G.: Early detection of the alzheimer disease combining feature selection and kernel machines. In: 15th International Conference on Neural Information Processing of the Asia-Pacific Neural Network Assembly (ICONIP 2008) (2008)
16. Scahill, R.I., Schott, J.M., Stevens, J.M., Rossor, M.N., Fox, N.C.: Mapping the evolution of regional atrophy in alzheimer's disease: Unbiased analysis of fluid-registered serial MRI. Proceedings of the National Academy of Sciences 99(7), 4703 (2002)
17. Schapire, R.E., Singer, Y.: Improved boosting algorithms using confidence-rated predictions. Machine Learning 37(3), 297–336 (1999)
18. Vapnik, V.N.: Statistical Learning Theory. Wiley-Interscience, Hoboken (1998)