

A voting ensemble method to assist the diagnosis of prostate cancer using multiparametric MRI

Patrick Riley¹, Ivan Olier¹, Marc Rea², Paulo Lisboa¹ and Sandra Ortega-Martorell¹

¹ Liverpool John Moores University – Department of Applied Mathematics
Byrom Street, Liverpool, L3 3AF – UK.

² Clatterbridge Cancer Centre – Department of Medical Imaging
Bebington, Wirral, CH63 4JY – UK.

P.J.Riley@2014.ljmu.ac.uk, I.A.OlierCaparroso@ljmu.ac.uk, M.Rea@nhs.net,
P.J.Lisboa@ljmu.ac.uk, S.OrtegaMartorell@ljmu.ac.uk

Abstract. Prostate cancer is the second most commonly occurring cancer in men. Diagnosis through Magnetic Resonance Imaging (MRI) is limited, yet current practice holds a relatively low specificity. This paper extends a previous SPIE ProstateX challenge study in three ways 1) to include healthy tissue analysis, creating a solution suitable for clinical practice, which has been requested and validated by collaborating clinicians; 2) by using a voting ensemble method to assist prostate cancer diagnosis through a supervised SVM approach; and 3) using the unsupervised GTM to provide interpretability to understand the supervised SVM classification results. Pairwise classifiers of clinically significant lesion, non-significant lesion, and healthy tissue, were developed. Results showed that when combining multiparametric MRI and patient level metadata, classification of significant lesions against healthy tissue attained an AUC of 0.869 (10-fold cross-validation).

Keywords: mpMRI, GTM, data mining, SVM, classification.

1 Introduction

Prostate cancer is the second most commonly occurring cancer in men worldwide [1]. Diagnosis through Magnetic Resonance Imaging (MRI) is the second line of diagnosis, usually after a trans-rectal ultrasound biopsy which is conducted after a positive blood screening test. However, this test has a relatively low specificity, leading to overdiagnosis and therefore overtreatment.

Although MRI diagnosis can overcome this, it requires specialist knowledge to review the prostate MRI, which is time consuming. Furthermore, the abundance of available MRI data can provide difficulty with where to begin. This study uses both Generative Topographic Mapping (GTM) and Support Vector Machines (SVMs) as assisting tools, in visualization and classification respectively, for the diagnosis of prostate cancer on multiparametric MRI (mpMRI) scans, through a voting ensemble technique.

Upon request from collaborating clinicians, the SPIE ProstateX challenge dataset [2] was extended in this study from its original use, whereby the contralateral of the lesion

location may be taken as healthy prostate tissue. This extends from the two classes available in the data set – the clinically significant lesions and non-significant lesions. The latter are denoted “non-significant” as prostate cancer treatment is not always needed for a lesion of a lower Gleason score [3].

The use of the contralateral to attain healthy tissue denotes a novel point for this study. Through pairwise tests of the three classes, the voting ensemble technique gives rise to a diagnosis aiding tool which is in demand from clinicians and radiologists.

Recent research from the ProstateX challenge tackles the two class problem – clinically significant against non-significant – with various machine learning methods such as deep learning [4], convolutional neural networks [5] and SVMs [6]; the efficiency of the latter inspiring its use in this study. The use of the GTM in this work provides insight into the structure of the data set, to allow for an explanation of the results of the ensemble voting method.

The Data & Classification Methods section details how the ProstateX challenge data set has been pre-processed and the setup of the SVM. The Model Evaluation section shows the ensemble voting systems results, with an Application to prostate lesion findings following. Using the GTM, an unsupervised explanation into the hidden data structure is presented, providing interpretability to the supervised SVM classifier.

2 Data

The ProstateX challenge training data used for this study is a collection of 330 lesion findings over 204 patients, each with mpMRI taken around the prostate. Metadata for each lesion finding, and the findings within each scan parameter were held in separate files. For each patient, one or more prostate lesions with their location were identified through its scanner coordinate position. Each lesion within the metadata holds a level denoting its clinical significance; where it is clinically significant if the Gleason score of the lesion is 7 or higher, or non-significant where the Gleason score is 6 or below. The numeric Gleason score of a given lesion is not available in the data set.

This work extends on the original ProstateX challenge data based on a request from collaborating clinicians. Through using the contralateral of the lesion location, healthy tissue can be extracted. The total number of significant cases in the dataset is 76, while the non-significant cases amount to 254. As these are imbalanced classes, we decided to subsample the class of non-significant cases to only 75 cases. In line with this, a total of 54 healthy tissue samples were extracted. Three of the various mpMRI scan parameters are analyzed in this study: T2-weighted, K^{trans} and Apparent Diffusion Coefficient (ADC). Only transverse images were used. These modalities are all shown to be related to lesion clinical significance [7].

T2-weighted imaging is a form of spin-echo pulse sequencing, showing fatty tissue and fluid brightly. ADC is a measure of the magnitude of diffusion (of water molecules) within tissue and is calculated from diffusion weighted imaging. K^{trans} , a type of perfusion imaging, represents a measure of capillary permeability, calculated from dynamic contrast-enhanced imaging.

3 Classification methods

For creating the classifiers we follow the steps outlined below:

- a) *Pre-processing*: The ProstateX dataset is a large data set, with each patient holding multiple scan parameters each – furthermore, the naming of the files was not consistent. Various rules were required to extract the relevant data. Metadata was collected from both the separate files, and the DICOM header data of the relevant image files; including the patient age which is used within the fitted models.
- b) *Patch Extraction and vectorization of the data*: For each lesion, a centered 5mm x 5mm patch is extracted at a resolution of 1px/mm for each of the three MRI scan parameters. The contralateral is defined as the opposite side of the MRI image to which the lesion is located. For example, if the lesion is located 25mm to the left of the center of the MRI image, then the contralateral is located 25mm to the right of the center of the MRI image, as validated by collaborating clinicians. They were used as the ‘healthy’ tissue data and is suitable for analysis as informed by collaborating clinicians. For the selection of those healthy samples, we identified the patient with only one lesion finding (as indicated in the metadata), which is of a certain distance away from the center of the image. In these cases, also a 5mm x 5mm patch was extracted at a resolution of 1px/mm for each of the three MRI scan parameters (as in the lesion patch extraction), following the method of a successful submission to the SPIE ProstateX challenge [6]. Images were flattened into 75 dimensional vectors. All processing was carried out with 32-bit floating point pixel values, preserving large dynamic range and subtle contrast differences. These vectors can be further augmented with patient level metadata, as described.
- c) *Augmenting the vectors with patient level metadata*: For the model presented in this study, the image data vector is further augmented with patient level metadata. The first is the zonal location of the region of interest (lesion or healthy tissue) within the prostate, denoted as a dummy variable. The second is the patient age.
- d) *Data standardization*: Each input dimension is scaled – for K^{trans} images, this is through the log transform. For all other input dimensions, this is through subtraction of the mean and division of the standard deviation. This ensures that the distribution of each dimension is approximately normal, again similar to [6].
- e) *Classification using Support Vector Machines*: For this classification study upon three class labels through pairwise testing, Support Vector Machines were utilized, as first described in [8]. SVM was successfully applied to this data in [6]. The linear kernel is not suitable for this data set – it was tested but discarded due to poor performance. The kernel selected for this work is the radial basis function (RBF-SVM), as defined in Equation 1. Through defining the kernel function there is no need to perform $\phi(\cdot)$ explicitly.

$$K(x^i, x^j) = \phi(x^i)^T \phi(x^j) = \exp\left(-\gamma \|x^i - x^j\|^2\right), \gamma > 0 \quad (1)$$

- f) *Initial cross-validation on all data for hyperparameter selection*: Two hyperparameters require tuning for an RBF-SVM: the soft-margin constant, C and the

kernel parameter, γ . A cross validation is used to evaluate each data combination and pairwise test, to find the best set of parameters. A grid search was carried out on $C \in \{0.1, 0.5, 1, 2, 5, 10, 20, 30, 50\}$ and $\gamma \in \{1.0, 0.2, 0.1, 0.05, 0.01, 0.001, 0.0001, 0.00001\}$. For the results presented in this paper, the 10-fold cross validation results across all the data are presented in Table 1.

4 Model Evaluation: Using a Voting Ensemble method for aiding prostate cancer diagnosis

In this study, a voting ensemble method across the three pairwise classifiers are proposed to give a final classification of each finding. This is to evaluate the system as a classifier. The majority vote across all classifiers is taken to be the final classification label. An undecided category was added for the cases where all three classifiers predict something different and a majority vote cannot be reached. In this case, as a tool for aiding diagnosis, a clinician would be alerted to this case for an expert to classify it themselves.

The voting ensemble method is evaluated through a 10-fold cross validation. Within each fold, 90% of the data is used for training and 10% of the data is partitioned according to the frequencies of the data, for testing the voting ensemble methodology. The results of each fold are used to calculate the classification across the data set used for this study. Each of the pairwise classifier is trained using the RBF-SVM with the parameters found through the cross-validation grid search (as explained in section 2.6). The results of applying the voting ensemble method (after cross validation) to the pairwise classifiers created using all image patch data and patient metadata is presented in Table 2.

For the significant cases, 72.4% of them were correctly classified as significant, with the majority of the misclassifications (23.6%) going to the non-significant lesions class. Only three cases (3.9%) were incorrectly classified as healthy tissue, and no cases were classified as undecided. In turn, the healthy tissue was correctly classified a 68.5% of the time, while again the majority of misclassification (25.9%) went to the non-significant class, and only 2 cases fell into the significant class.

The non-significant cases were correctly classified only 48.0% of the time, however, considering that this class is the intermediate between having a healthy prostate and a significant tissue, this is to some extent expected. Still, this is the less critical issue from the clinical viewpoint, as it is known that treatment is not always needed for non-significant prostate cancers [3] and techniques such as active surveillance may be employed within the patients care plan. Potentially, those non-significant cases predicted as significant (20%) will undergo more in-depth analysis, which may result in their final correct diagnosis as non-significant; and those predicted as healthy (32%) may hold a very low Gleason score and therefore closely resemble healthy prostate tissue. The latter group must probably would not require any treatment, therefore the risk to the patients is not considered high.

5 Classification results: Application to prostate lesion findings

An application of the ensemble voting system applied to given lesion findings and healthy tissue is presented in Figure 1. For cases A and B, a significant and non-significant lesion finding respectively, are classified through the three classifiers. For case C, the ensemble voting system is shown to classify healthy prostate tissue. In all three of these cases, the majority vote matches the original label and is therefore correctly classified. It is intended that this aids prostate lesion diagnosis.

However, for case D, which is known to be healthy prostate tissue, there is no majority vote; all three pairwise classifiers have predicted differently. In practice, this would be brought to the attention of a clinician, for their technical knowledge to make the final decision. Nevertheless, in all cases, lesion or not, the final decision would rest with them as this tool is designed to aid diagnosis.

Table 1: 10-fold cross validation results for hyperparameter grid search.

Classifier	AUC	Standard deviation	C	γ
Significant vs. Non-significant	0.722	0.262	50	0.001
Significant vs. Healthy	0.869	0.099	0.5	0.01
Non-significant vs. Healthy	0.713	0.134	2	0.01

Table 2: Results of the voting ensemble method applied to the dataset containing all image patch data and patient metadata.

Original label	No. cases	Final predicted label			
		Significant	Non-significant	Healthy	Undecided
Significant	76	55	18	3	0
Non-significant	75	15	36	24	0
Healthy	54	2	14	37	1

6 Discussion: Explanation of the classification results using Generative Topographic Mapping

Through visualizing the data in the latent space, it is possible to explain the supervised SVM classification results using the unsupervised GTM [8]. GTM is a nonlinear latent variable model of the manifold learning family, with sound foundations in probability theory. It performs simultaneous clustering and visualization of the observed data through a nonlinear and topology-preserving mapping from a visualization latent space in (with being usually 1 or 2 for visualization purposes) onto a manifold embedded in a multi-dimensional space, where the observed data reside. The mapping that generates the manifold is carried out through a generalized additive regression function:

$$y = W\varphi(u) \quad (2)$$

where $y \in D$, $u \in W$ is the matrix that generates the mapping, and ϕ is a vector with the images of S basis functions ϕ_s . To achieve computational tractability, the prior distribution of u in latent space is constrained to form a uniform discrete grid of M centers, analogous to the layout of the Self-Organizing Map (SOM, [9]) units, in the form of a sum of delta functions:

$$p(\mathbf{u}) = \frac{1}{K} \sum_{k=1}^K \delta(\mathbf{u} - \mathbf{u}_k) \quad (3)$$

Similar to how the RBF kernel has been used in this application through the SVM classifier, GTM typically uses a set of radial basis functions to map the results of the unsupervised analysis. By doing this, the interpretability problem of the SVM is somewhat alleviated. The similar kernel tricks allow for interpretation.

Figure 2 and Figure 3 show the mean projections for Significant against Healthy, and for all three classes respectively. Clustering can be seen within Figure 2 for both the Significant and the Healthy classes within the latent space. However, with the introduction of the Non-significant class, there is no clustering and it ‘invades’ both the Significant and the Healthy classes. To an extent this is to be expected. Where the Gleason score of a Significant lesion is 7 or higher, some Non-significant lesions at the higher end of their possible Gleason scores, such as 5 or 6, may overlap with the Significant class. Similarly, those with a lower Gleason score, such as 1 or 2, may overlap with the Healthy class.

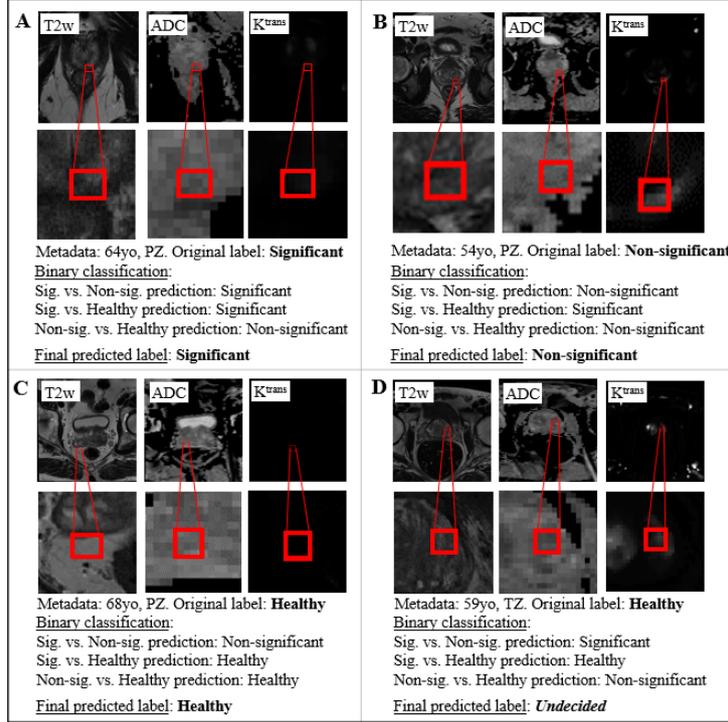


Figure 1: Application of the voting ensemble method to lesion findings. **A:** Classification of a Clinically Significant Lesion. **B:** Classification of a non-significant lesion. **C:** Classification of healthy tissue. **D:** Disagreement throughout the ensemble voting system. (yo – years old)

Figure 4 shows the boxplot of the distributions of the mean projections of the points. All classes overlap, which is to be expected as shown in Figure 3. The interquartile ranges of the Significant class and the Healthy class, showing separation (as shown in Figure 2). The interquartile range of the Nonsignificant class overlaps with both the Significant class and the Healthy class.

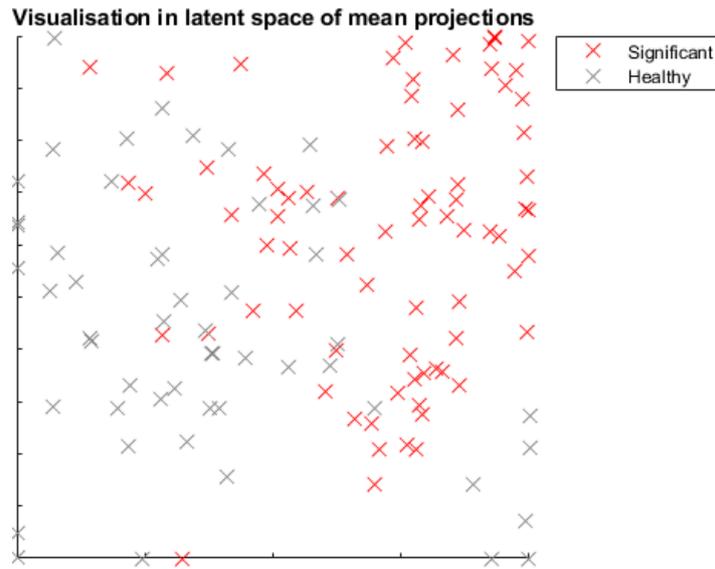


Figure 2: Mean projections in the latent space using GTM: Significant against Healthy. Clustering is present.

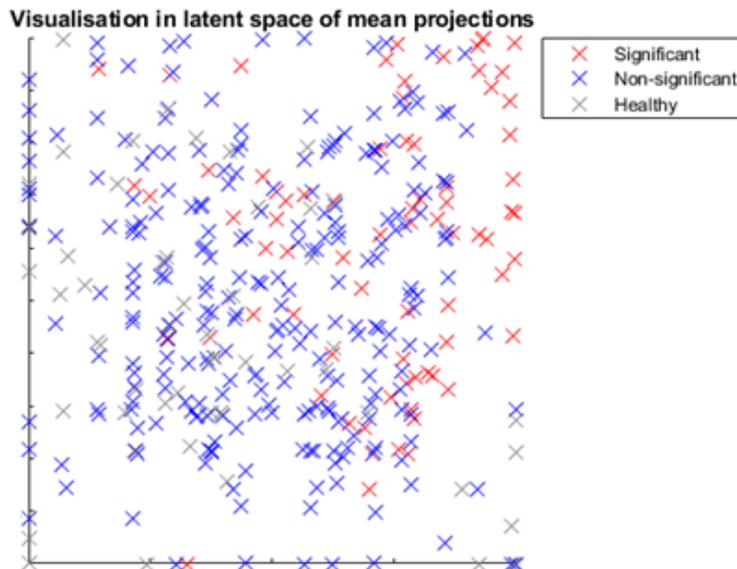


Figure 3: Mean projections in the latent space using GTM: all three classes. The clustering becomes distorted with the introduction of the non-significant (blue) class.

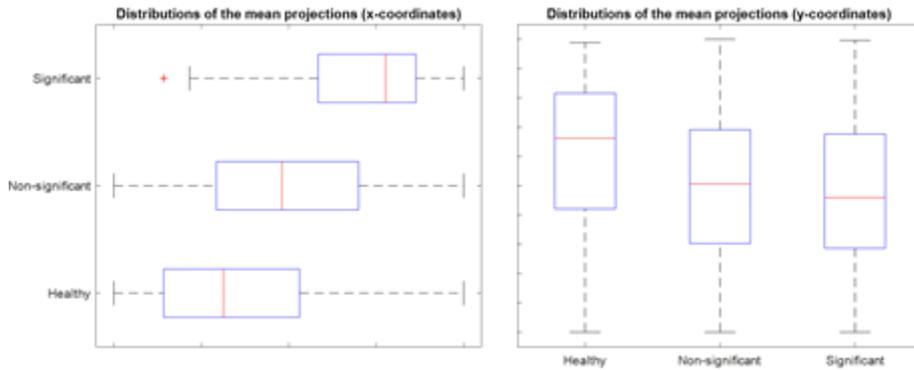


Figure 4: Boxplots of the distributions of the mean projections for each class label.

Through the addition of the healthy class, the initial 2-class classifier problem of the ProstateX challenge can be visualized, with much overlap of the Non-significant class within the Significant class. Submissions to the ProstateX challenge were for the 2-class classifier problem using AUC and hence comparisons are not presented alongside this study; enhanced interpretability using the GTM with the binary classifiers is provided.

6.1 Borderline cases

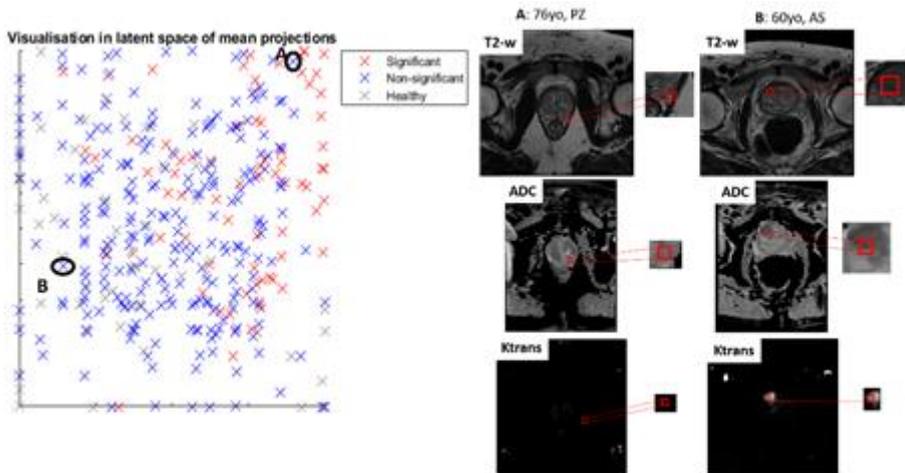


Figure 5: Examples of borderline cases - both are Non-significant lesions but appear within other clusters in the latent space, with the multiparametric profile displayed for each case.

Figure 5 shows examples of borderline cases of cases of Non-Significant lesions from the latent space, where case **A** is an example of a Non-significant lesion which is located within the Significant clustering, and case **B** is an example of a Non-significant lesion which is located within the Healthy clustering.

Reasons why cases **A** and **B** appear on different corners of the latent space visualization, in terms of the multiparametric images, may be because of the differences in brightness of the images. Although both Non-significant cases, case **A** for example may appear similar to the brightness of a Significant case and is therefore closer to that clustering. Indeed, the Gleason score is not available with the challenge data set and this would provide interesting insights.

7 Conclusions and further work

This study found that using an ensemble voting method across three classifiers, a majority vote system can assist clinicians in the diagnosis stage of a patient's cancer care. The use of healthy tissue has been included through demand and validation from clinicians. Through using a supervised SVM, the results can be explained, with enhanced interpretation, through an unsupervised visualization method, the GTM.

Further work will look at increasing the classifiers scope, leading to segmentation maps of the area to further aid diagnosis, particularly for non-invasive treatment techniques. Furthermore, utilizing the decision of a clinician with their classification of an undecided case through active machine learning techniques will be studied.

Acknowledgments. This work has been funded by the Liverpool John Moores University PhD Scholarship Fund. The authors would like to thank Andy Kitchen for his assistance in validating data extraction methods.

References

1. Bray, F., Ferlay, J., Soerjomataram, I., Siegel RL, Torre, L., Jemal, A.: Prostate cancer statistics | World Cancer Research Fund, <https://www.wcrf.org/dietandcancer/cancer-trends/prostate-cancer-statistics>
2. Litjens, G., Debats, O., Barentsz, J., Karssemeijer, N., Huisman, H.: Computer-Aided Detection of Prostate Cancer in MRI. *IEEE Trans. Med. Imaging.* 33, 1083–1092 (2014). doi:10.1109/TMI.2014.2303821
3. Gallagher, J.: Prostate cancer treatment “not always needed” - BBC News, <https://www.bbc.co.uk/news/health-37362572>
4. Liu, S., Zheng, H., Feng, Y., Li, W.: Prostate Cancer Diagnosis using Deep Learning with 3D Multiparametric MRI. Presented at the March 12 (2017)
5. Mehrtash, A., Sedghi, A., Ghafoorian, M., Taghipour, M., Tempany, C.M., Wells, W.M., Kapur, T., Mousavi, P., Abolmaesumi, P., Fedorov, A.: Classification of Clinical Significance of MRI Prostate Findings Using 3D Convolutional Neural Networks. In: Armato III, S.G. and Petrick, N.A. (eds.) *SPIE Medical Imaging 2017: Computer-Aided Diagnosis.* , Orlando (2017)
6. Kitchen, A., Seah, J.: Support vector machines for prostate lesion classification. In: Armato III, S.G. and Petrick, N.A. (eds.) *SPIE Medical Imaging 2017: Computer-Aided Diagnosis.* p. 1013427. International Society for Optics and Photonics, Orlando (2017)

7. Langer, D.L., van der Kwast, T.H., Evans, A.J., Plotkin, A., Trachtenberg, J., Wilson, B.C., Haider, M.A.: Prostate Tissue Composition and MR Measurements: Investigating the Relationships between ADC, T2, Ktrans, ve, and Corresponding Histologic Features. *Radiology*. 255, 485–494 (2010). doi:10.1148/radiol.10091343
8. Bishop, C.M., Svensén, M., Williams, C.K.I.: *The Generative Topographic Mapping*. (1998)
9. Kohonen, T.: *Self-Organizing Maps*. Springer Berlin Heidelberg, Berlin, Heidelberg (2001)