Macedonian Journal of Chemistry and Chemical Engineering, Vol. **33**, No. 1, pp. 65–71 (2014)

MJCCA9 – 638
Received: November 20, 2013
Accepted: February 5, 2014

ISSN 1857-5552
UDC: 543.545.2:[606:616.831.9-008.83
Original scientific paper

OPTIMIZATION OF SUPERVISED SELF-ORGANIZING MAPS WITH GENETIC ALGORITHMS FOR CLASSIFICATION ELECTROPHORETIC PROFILES

Natalija Tomovska, Igor Kuzmanovski, Kiro Stojanoski*

Institute of Chemistry, Faculty of Natural Sciences and Mathematics,
Ss. Cyril and Methodius University, Skopje, Republic of Macedonia

Standard electrophoresis methods were used to classify analyzed proteins in cerebrospinal fluid from patients with multiple sclerosis. Disc electrophoresis was carried out on polyacrylamide gels for the detection of oligoclonal IgG bands in cerebrospinal fluid, mainly from patients with multiple sclerosis and other central nervous system dysfunctions. ImageMaster 1D Elite and Gel-Pro specialized software packages were used for fast accurate image and gel analysis. The classification model was based on supervised self-organizing maps. In order to perform modeling in an automated manner, genetic algorithms were used. Using this approach and a data set composed of 69 samples, we developed models based on supervised self-organizing maps which were able to correctly classify 83% of the samples in the data set used for external validation.

Keywords*:* disc electrophoresis; cerebrospinal fluid; protein analysis; supervised self-organizing maps

ОПТИМИЗАЦИЈА НА САМООРГАНИЗИРАНИ МАПИ ТРЕНИРАНИ СО НАДГЛЕДУВАНО УЧЕЊЕ СО ГЕНЕТСКИ АЛГОРИТМИ ЗА КЛАСИФИКАЦИЈА НА ЕЛЕКТРОФОРЕТСКИ ПРОФИЛИ

Стандардна електрофоретска метода беше користена за анализа на протеините во цереброспиналниот ликвор кај пациенти претежно со мултипла склероза, но и со други заболувања на централниот нервен систем, со намера да се изврши класификација на експерименталните резултати. За оваа намена беше користена диск-електрофорезата за детекција на олигоклонални IgG ленти во електрофореграмите добиени со гел-електрофореза. Анализата на електрофореграмите беше направена со специјализираниот софтвер ImageMaster 1D Elite и софтверскиот пакет GelPro. За класификација на добиените експериментални резултати беа користени самоорганизирани мапи тренирани со надгледувано учење. За да се автоматизира моделирањето, беше користен генетскиот алгоритам. Користејќи го овој пристап и податоците од 69 примероци за анализа, добивме модели базирани на самоорганизирани мапи, со кои може коректно да се класифицираат 83 % од податоците користени за екстерна валидација.

Клучни зборови: диск-електрофореза; церебрален флуид; анализа на протеини;
самоорганизирани мапи тренирани со надгледувано учење

1. INTRODUCTION

Quantitative and qualitative analysis and the determination of different types of proteins, other biomolecules and their profiles play an increasing role in medical diagnosis. Standard electrophoresis methods and many emerging approaches such as lab-on-a-chip methods are well known for protein detection and analysis of cerebrospinal fluid (CSF) [1–5]. Proteomics technologies have been widely used in the investigation of neurodegenerative and psychiatric disorders. 2-D electrophoresis followed by mass spectrometry has been mainly applied, as this proteomics approach provides the possibility of convenient quantification of protein levels and detection of post-translational modifications. Many proteins with disrupted levels and modifications have been detected by proteomics approaches and related to neurodegeneration and psychiatric disorders [6]. However, cerebrospinal fluid analysis, coupled with other methods, remains the basis of the diagnosis of various neurological disorders, including multiple sclerosis and infectious diseases of the central nervous system (CNS) [3, 7].

In the routine procedure, electropherograms are classified into different groups according to the qualitative and quantitative composition of cerebrospinal fluid with regard to major protein fractions and the CSF/serum albumin quotient, coupled with McDonald diagnostic criteria [3]. In addition, for the detection of oligoclonal IgG bands in serum and in unconcentrated spinal fluid, some techniques have been used, such as the isoelectric focusing combined with polyethylene-enhanced gel immunofixation and silver staining, CSF/serum quotient diagrams, different body indices, etc. [7–9].

In addition, the automation and development of software has enabled the fast collection of huge amounts of electrophoretic data [10]. Image analysis software is used to extract much more information from the electropherogram for comparative analysis between gels generated in-house or available in web-based databases. Data acquisition, manipulation and computation for electrophoretic protein pattern recognition are performed using standard statistical signal analysis. Cluster analysis, along with other statistical methods such as principal component analysis (PCA), artificial neural networks (ANN) and fuzzy logic [10, 11] have been used in various areas of medicine.

Our previous results obtained using hierarchical cluster analysis, despite considerable similarities between electropherograms, have shown that different clustering approaches produced different dendrograms, and it was concluded that cluster analysis should be used cautiously [5]. Having the disadvantages of these methods in mind, here we decided to use self-organizing maps (SOM). This algorithm has become a valuable tool for data analysis purposes [12–22]. The most commonly used SOM algorithm is for clustering multidimensional data [12–19] and for process/reaction monitoring [22, 23], but also as a tool for variable selection [24]. The theoretical background of self-organizing maps [24] and their applications in chemistry are described in detail in the literature [25–26]. A variant of the SOM algorithm, called supervised self-organizing maps [24], has not been widely used in chemometrics. However, keeping in mind the fact that this version of the algorithm is suitable for classification purposes, we have used it for successfully developing classification models for different purposes [27, 28].

In this paper, we describe our efforts to develop classification models based on supervised self-organizing maps [24] in order to determine (1) whether the patients have multiple sclerosis or (2) other central nervous system dysfunctions (like polyradiculoneuritis, known as Guillain-Barré syndrome, encephalitis and paraproteinemia) or (3) whether the findings belong to patients without any disorder of the central nervous system.

2. EXPERIMENTAL

The electrophoregrams used here were analyzed in our previous work and the experimental details are described there [4, 5]. In addition to the results from 32 patients diagnosed with multiple sclerosis, we used the data from an additional 23 patients. The majority of these patients had a history of psychiatric disorders (polyradiculoneuritis, paraproteinemia and encephalitis) and no symptoms or signs of neurological disease, as shown by magnetic resonance imaging and electrophysiological investigations and routine biochemical analyses. Also, electrophoregrams were obtained from a control group of 14 healthy patients. Clinical investigations were performed according to the Regulations of the Macedonian Ethical Committee and the Ministry of Health of the Republic of Macedonia.

2.1. *Disc electrophoresis*

Disc electrophoresis was carried out on 7% polyacrylamide gels, using the Canalco (USA) electrophoresis system. CSF was used without preconcentration. Proteins were separated on polyacrylamide gels polymerized in glass tubes, approximately 5 mm in diameter and 15 cm in length. The experimental details are described by Spiroski et al. [5].

3. DATA PREPARATION AND MODELING

The collected data were digitized using ImageMaster 1D Elite and Gel-Pro software [29]. In all electropherograms, the dominant peak was that of albumin (Figure 1). In the preliminary data analysis, we noticed that using the entire electropherograms was not a good option because the albumin peak reduces the importance of the smaller peaks which bear in them the information we chose to model. Having this in mind, we removed the albumin peak from all electropherograms. As a result, the total number of data points was reduced from 450 to 270 in each of the electropherograms. In the next step, the remaining part of the electropherogram was normalized and further autoscaled.

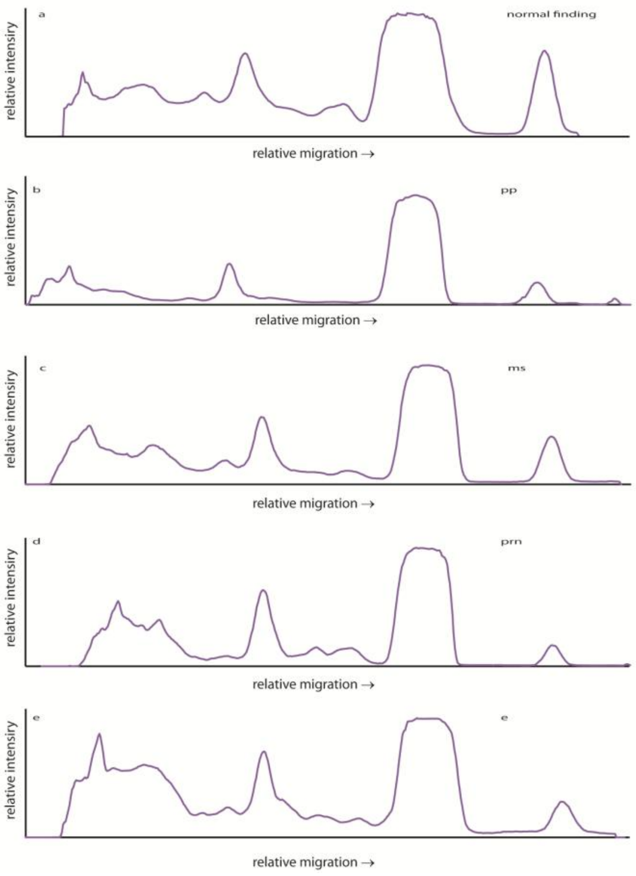

Fig. 1. Electrophoregrams of CSF from:
a) N – control subject, b) PRN – polyradiculoneuritis,
c) MS – multiple sclerosis, d) PP – paraproteinemia,
and e) E – encephalitis

3.1. *Supervised self-organizing maps*

Self-organizing maps have been developed as an algorithm for unsupervised learning. However, in cases where poor class separation is obtained, a modified version for supervised classification can be used. A slight modification of the algorithm can transform SOMs into a tool for supervised classification [24]. Namely, SOMs can be transformed into an algorithm for supervised classification if the input vectors for the training samples ($\boldsymbol{d}_s$) are augmented by a unit vector $\boldsymbol{d}_u$ (Figure 2a) with its components assigned into one of the four classes present in the training set. During the phase of prediction, the part of the weight vectors of SOM that correspond to the unit vector is excluded (Figure 2b). In other words, for each sample in the training set $\boldsymbol{d}_s$, the corresponding $\boldsymbol{d}_u$ must be used during training while during the recognition of an unknown sample $\boldsymbol{x}$, only the $\boldsymbol{x}_s$ part is compared with the corresponding part of the weight vectors of the trained SOM.

It is also important to mention that, in this work, instead of using autoscaled electropherograms for training the supervised SOMs, we used principal components extracted from the normalized and autoscaled electropherograms. The advantage of this approach is that PCA is able to extract most of the information (stored in the preprocessed data matrix composed of 270 data points) into a vector (principal component) composed of only a few data points.

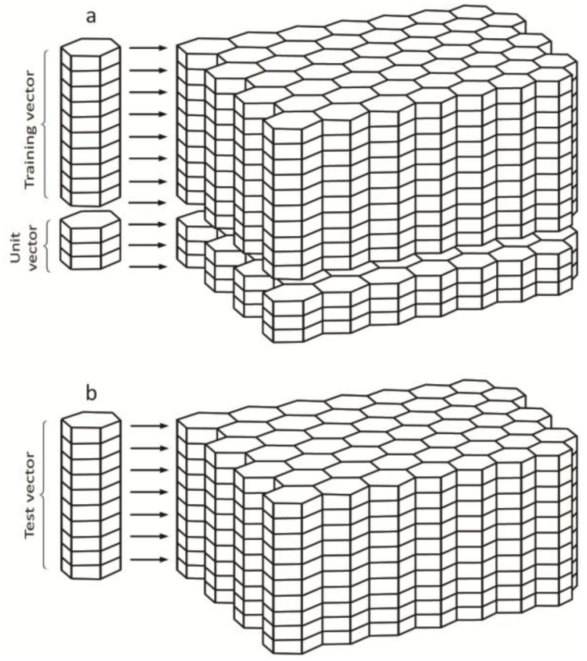

Fig. 2. Supervised SOM setup for classification of data set with four different classes. a) The training set vectors are augmented with the unit vector. The unit vector in this case has a length of three. Correspondingly, the number of additional weight levels of the supervised SOM which correspond to the unit vector is also four. b) After the training of the supervised SOM has finished, the developed model can be used (without the weight levels that correspond to the unit vector) for the classification of additional samples (test vectors)

In addition to this, the advantage of using autoscaling prior to extracting principal components is that it gives more importance to the signals which do not dominate the electropherograms [30]. However, the disadvantage of the same procedure is that it gives equal importance to the regions of the electropherograms which bear in themselves valuable information for the classification of our samples with those regions which are noisy or do not have any valuable information which could help in the correct classification of our samples [30]. In order to extract the data points in our electropherograms which are the most suitable for classification purposes, we used the genetic algorithm (GA) prior to extracting the principal components.

3.2. *Genetic algorithm*

In order to perform optimization of the supervised SOMs in an automated manner, but also to minimize the role of the analyst, the GA was used [31–33]. It has been shown that this algorithm is an effective tool for solving different optimization problems in chemistry [27, 28, 34–38]. GA is capable of performing relatively fast convergence toward a global minimum of the optimized function without the need to run every possible permutation of the optimized parameters [31–33].

Most commonly, GA is used for variable selection [31–33], but in our laboratory we use it not only for this purpose, but also for the determination of other important parameters which could influence the performance of the developed model [27, 28].

4. RESULTS AND DISCUSSION

As previously stated, in order to perform the optimization of the supervised SOMs in an automated manner, we used the GA. For these purposes, the search for the best possible model was performed on a population consisting of 100 chromosomes. The chromosomes were encoded, as presented in Figure 3. To find the width of the supervised SOM, three genes were used (in the interval 4–11 neurons). An additional three genes were used to adjust the length of the supervised SOM (in the interval 4–11 neurons); four genes were used to find the most suitable number of epochs in the rough training phase (in the interval: 10–25 neurons). Seven genes were used to adjust the number of training epochs in the fine training phase (the obtained number in the interval between 1 and 128 was increased by double the number of epochs in the rough training phase, in order to ensure that the number of epochs in this phase is larger than the one in the previous training phase). An additional 270 genes were used to select variables from electropherograms (normalized and autoscaled) and, at the end, four more genes were used to select the number of principal components, which was calculated from the preselected data points.

Initialization of the weights of the supervised SOMs was performed along the first two principal components obtained from the training data set. During training, we used the Gaussian neighborhood function and a linearly decreasing learning rate. The entire optimization using GA lasted 450 generations and, as previously stated, we used a population composed of 100 chromosomes. During optimization 20% of the chromosomes with the best performance were used as parents for the creation of the offspring chromosomes (80% of the population) for the following generation. Mating pairs were formed from the best (20%) chromosomes for formation of the new population using the roulette wheel selection rule [34].

In order to avoid fast convergence in a small area of the search, a mutation was applied during the optimization. Until generation 50, the probability of the occurrence of a mutation in the offspring chromosome was kept at 0.10. After that, until generation 150, the probability for the occurrence of a mutation linearly decreased down to 0.05. From there on, until the end of GA optimization, mutation was kept at the same level.

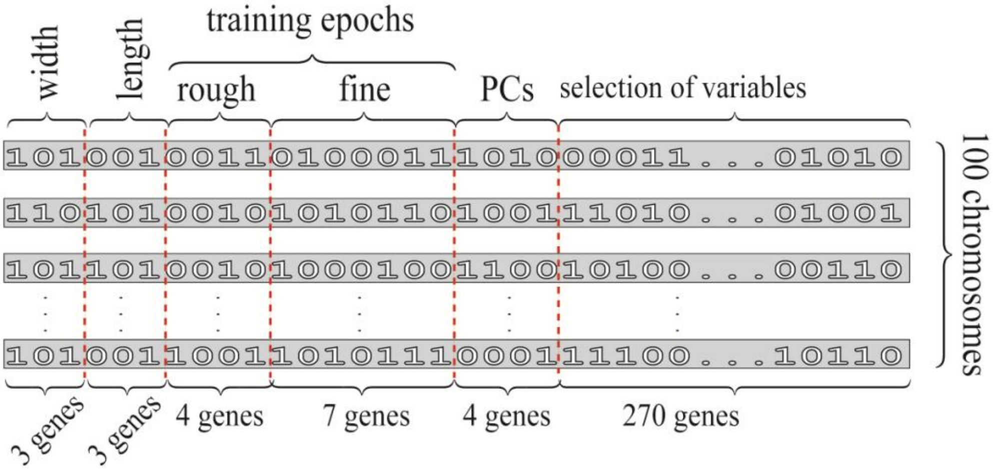

Fig. 3. Encoding the chromosomes used during the search for the best model performed using the genetic algorithm

The developed classification models were validated using both internal and external validation. Internal validation was performed during GA optimization using the cross-validation-leave-10%-out procedure with the training data set (composed of 52 samples). After the search for the best model ended, we used the test set for external validation of the best model. The test set was composed of 17 samples from the original data set. Prior to starting optimization (using GA), the data set was divided into the training set and the test set using the Kennard-Stone algorithm [39]. Some of the best models, the size of the SOM, the training epochs as well as the number of misclassified samples for these models are presented in Table 1.

Table 1

Parameters for some of the best models obtained using genetic algorithms

Model	No. of PCs	Training set errors	Cross-validation error	Test set errors	Size of the SOM		Training epochs	
					Width	Length	Rough phase	Fine phase
1	8	1	1	3	7	8	19	54
2	8	0	2	4	7	8	19	59
3	8	0	2	5	7	8	10	42

In the remaining part of this discussion, we will analyze model number 1 in more detail (Table 1). The trained supervised SOM for this model (in which different areas are properly labeled) is presented on Figure 4. One can notice here that the upper left corner of the supervised SOM is capable of *recognizing* samples from the healthy patients with normal protein status in the CSF. The central part of the supervised SOM is capable of recognizing the samples collected from patients with multiple sclerosis, while in the right part of the map, the samples from the remaining patients are grouped. Figure 5 also corresponds to model 1 (from Table 1). In this case (in Figure 5), SOM was labeled according to the results obtained using the training set. Here, we noticed that the only misclassified sample belongs to the *broad* class of samples, which were added to the data set in order to develop a more robust model. Without these additional 23 samples, our models would have been able to classify into only two classes (healthy individuals and patients with multiple sclerosis). If the model was developed using only these two types of samples, and if it had been used for the classification of additional samples (which do not belong to these two classes), it could have been "forced" to be mapped into a part which is capable of recognizing two classes of samples (in this hypotetical case, samples from healty patients and samples from patients with multiple sclerosis). So, having this explanation in mind, together with the fact that we had a relatively small number of samples in our data set, the additional 23 samples from the patients with three additional diagnoses were considered as a third class. So, we assume that the samples from this third class are misclassified because of the small number of samples with different diagnoses (polyradiculoneuritis, paraproteinemia and encephalitis) in it.

In Figure 6, we show the map that corresponds to the discussed model, but this time it is labeled with the labels from the samples which are part of the test set (used for external validation). In total, three samples from the test set were misclassified by this model. None of the misclassified samples belongs to a patient who was diagnosed with multiple sclerosis. (This is also the case with the other two models presented in Table 1.) In our oppinion, this is due to the fact that almost half of the samples belong to patients with multiple sclerosis, so the discussed model is capable of correctly classifying all samples of this type, not only the samples from the training set, i.e. the samples which are part of the test set. Two of the three misclassified samples for this model belong to healty patients, while the third sample is a sample from a patient with another psychological disorder.

In our opinion, the generalization performance of the developed models based on supervised SOM could be further improved if the data set were to be expanded with aditional experimental data. The number of misclassified samples from healthy patients, as well as from patients with psychological disorders, would be further reduced if the number of samples was at least comparable with the number of samples collected from patients with multiple sclerosis.

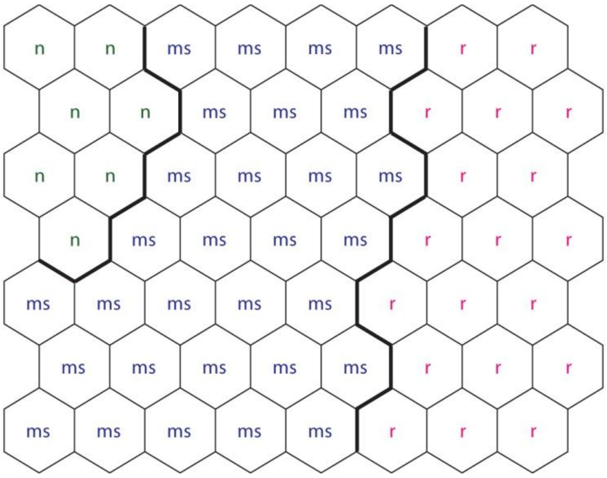

Fig. 4. Trained supervised SOM for model 1 (Table 1) labeled with the different types of classes
(n – healthy patients, ms – patients with multiple sclerosis, r – patients with other psychiatric disorders)

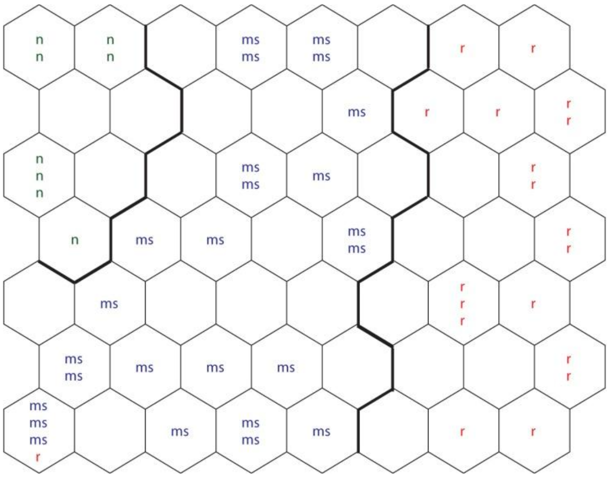

Fig. 5. Training samples with their labels mapped on the supervised SOM

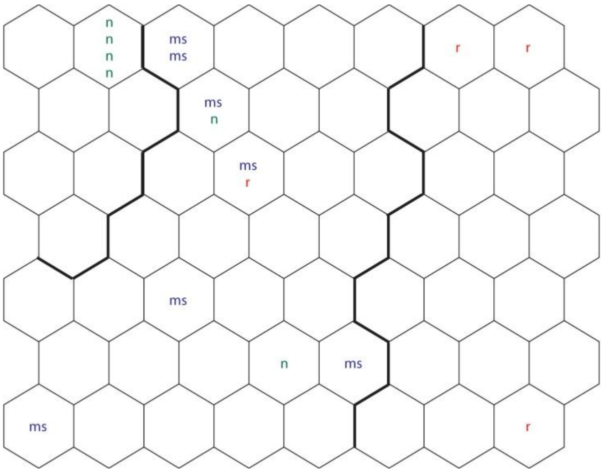

Fig. 6. Test samples with their labels mapped on the supervised SOM

5. CONCLUSIONS

We have shown that models based on supervised self-organizing maps coupled with genetic algorithms for the classification electrophoretic profiles is an attractive approach for automated diagnostics on samples from patients with multiple sclerosis. The results of our modeling show that it is possible to develop a model which can successfully classify samples from patients with multiple sclerosis. However, during the phase of external validation, for all three models examined in this work, we observed misclassifications of samples taken from healthy patients and from patients with other psychological disorders (not multiple sclerosis). We believe that, in future, the research on this subject in our laboratory will continue. With a larger number of samples, we will be able to present more detailed results from models which will be able to successfully recognize not only samples from patients with multiple sclerosis, but also samples from patients with other psychological disorders.

REFERENCES

[1] O. Trenchevska, V. Aleksovski, D. Nedelkov and K. Stojanoski, Developing Novel Methods for Protein Analysis and Their Potential, Implementation in Diagnosing Neurological Diseases, in: *Advanced Topics in Neurological Disorder*, chapter 7, pp. 129–158 (2012).

[2] H. Link, Y. Huang, Oligoclonal bands in multiple sclerosis cerebrospinal fluid: An update on methodology and clinical usefulness, *J. of Neuroimmunol.*, **180**, 17–28 (2006).

[3] McDonald diagnostic criteria: C. H. Polman, S. C. Reingold, B. Banwell, M. Clanet, J. A. Cohen, M. Filippi, K. Fujihara, E. Havrdova, M. Hutchinson, L. Kappos, F. D. Lublin, X. Montalban, P. O'Connor, M. Sandberg-Wollheim, A. J. Thompson, E. Waubant, B. Weinshenker, J. S. Wolinsky, *Annals of Neurology*, *Diagnostic criteria for multiple sclerosis*: 2010 Revisions to the McDonald criteria. **69**, 292–302 (2011).

[4] A. Mitrevski, K. Stojanoski and P. Korneti, Detection of oligoclonal IgG bands in cerebrospinal fluid on polyacrylamide support media: Comparison of isoelectric focusing and disc electrophoresis, *Acta Pharm.*, **3**, 163–171, (2001).

[5] F. Spiroski, K. Stojanoski, A. Mitrevski, Comparison of statistical cluster methods in electrophoretic protein pattern analysis, *Acta Pharm.*, **55**, 215–221 (2005).

[6] L. Liao, D. Cheng, J. Wang, T. Losik, D. Duong, M. Gearing, H. Rees, J.-J. Lah, AI Levey, J. Peng, Proteomic characterization of postmortem amyloid plaques isolated by laser capture microdissection, *J. Biol. Chem.*, **279**, 37061–37068 (2004).

[7] C. Sindic, M. Antwerpen and S. Goffette, The intrathecal humoral immune response: laboratory analysis and clinical relevance, *Clin. Chem. Lab. Med.*, **39,** 333–340 (2001).

[8] W. W. Tourtellotte, A. R. Potvin, J. O. Fleming, K. N. Murthy, J. Levy, K. Syndulko, J. H. Potvin, Multiple sclerosis: measurement and validation of central nervous system IgG synthesis rate, *Neurology*, **30,** 240–244 (1980).

[9] C. J. M. Sandic, P. Monteyne, G. Bigaignon and E. C. Laterre, Polyclonal and oligoclonal IgA synthesis in the cerebrospinal fluid of neurological patients. An immunoaffinity-mediated capillary blot study, *J. Neuroimmunol.* **94** 103–111(1994).

[10] J. Vohradsky, Adaptive classification of two-dimensional gel electrophoretic spot patterns by neural networks and cluster analysis, *Electrophoresis*, **18**, 2749–2754 (1997).

[11] M. Kratzer, B. Ivandic, A. Fateh-Moghadam, Neuronal network analysis of serum electrophoresis, *J. Clin. Pathol*., **45**, 612–615 (1992).

[12] P. K. Hopke, X. H. Song, Classification of single particles by neural networks based on the computer-controlled scanning electron microscopy data, *Anal. Chim. Acta*, **348**, 375–388 (1997).

[13] D. Wienke, Y. Xie, P. K. Hopke, Classification of airborne particles by analytical scanning electron microscopy imaging and a modified Kohonen neural network (3MAP), *Anal. Chim. Acta*, **310**, 1–14 (1995).

[14] R. Goodacre, J. Pygall, D. B. Kell, Plant seed classification using pyrolysis mass spectrometry with unsupervised learning: The application of auto-associative and Kohonen artificial neural networks, *Chemometr. Intell. Lab. Syst*., **38,** 69–83 (1997).

[15] J. Zupan, M. Novič, Classification of multicomponent analytical data of olive oils using different neural networks, *Anal. Chim. Acta*, **192**, 219–234 (1994).

[16] H. Yang, I. R. Lewis, P. R. Griffiths, Raman spectrometry and neural networks for the classification of wood types. 2. Kohonen self-organizing maps, *Spectrochim. Acta*, **55**, 2783–2791 (1999).

[17] Y. V. Heyden, P. Vankeerberghen, M. Novic, J. Zupan, D. L. Massart, The application of Kohonen neural networks to diagnose calibration problems in atomic absorption spectrometry, *Talanta*, **51,** 455–466 (2000).

[18] I. V. Pletnev, V. V. Zernov, Classification of metal ions according to their complexing properties: A data-driven approach, *Anal. Chim. Acta*, **455**, 131–142 (2002).

[19] F. Vandeerestraeten, C. Wojciechowski, N. Dupuy, J. P. Huvenne, Recognition of starch origin and modifications by chemometrics spectral data processing, *Analysis*, **26,** 57–62 (1998).

[20] V. Tanevska, I. Kuzmanovski, O. Grupče, Provenance determination of Vinica terra cotta icons using self-organising maps, *Annali di Chimica*, **97**, 541–552 (2007).

[21] M. Kolehmainen, P. Rönkkö, O. Raatikainen, Monitoring of yeast fermentation by ion mobility spectrometry measurement and data visualisation with Self-Organizing Maps, *Anal. Chim. Acta*, **484**, 93–100 (2003).

[22] C. Ruckebusch, L. Duponechel, J.-P. Huvenne, Degree of hydrolysis from mid-infrared spectra, *Anal. Chim. Acta*, **446,** 255–266 (2001).

[23] R. Todeschini, D. Galvagni, J. L. Vílchez, M. del Olmo, N. Navas, Kohonen artificial neural networks as a tool for wavelength selection in multicomponent spectrofluorimetric PLS modelling: Application to phenol, *o*-cresol, *m*-cresol and *p*-cresol mixtures, *Trends Anal. Chem*., **18,** 93–98 (1999).

[24] T. Kohonen, *Self-Organizing Maps*, 3rd Edition, Springer, Berlin, 2001.

[25] J. Zupan, M. Novič, I. Raisánchez, Kohonen and counterpropagation artificial neural networks in analytical chemistry, *Chemometr. Intell. Lab. Syst*., **38,** 1–23 (1997).

[26] J. Zupan, J. Gasteiger, *Neural Networks in Chemistry and Drug Design*, WCH, Weinheim, 1999.

[27] I. Kuzmanovski, M. Trpkovska, B. Šoptrajanov, Determination of the composition of human urinary calculi composed of whewellite, weddellite and carbonate apatite using artificial neural networks, *J. Mol. Struct*., **744–747**, 833–838 (2005).

[28] I. Kuzmanovski, S. Dimitrovska-Lazova, S. Aleksovska, Classification of perovskites with supervised self-organizing maps, *Anal. Chim. Acta*, **595**, 182–189 (2007).

[29] http://gel-pro-analyzer.software.informer.com/

[30] M. Otto, *Chemometrics*, 2nd Edition, Wiley, Weinheim, 2007.

[31] J. Holland, Outline for a Logical Theory of Adaptive Systems, *J. Comput. Machinery* **3,** 297-314 (1962).

[32] B. Kermani, S. Schiffman, H. T. Nagle, *IEEE Trans. Biomed. Eng*., **46,** 429–439 (1999).

[33] C. Henderson, W. Potter, R. McClendon, G. Hoogenboom, Predicting Aflatoxin Contamination in Peanuts: A Genetic Algorithm/Neural Network Approach, *Appl. Intell*., **12**, 183–192 (2000).

[34] K. Hasegawa, Y. Miyashita, K. Funatsu, GA Strategy for Variable Selection in QSAR Studies: GA-Based PLS Analysis of Calcium Channel Antagonists, *J. Chem. Inf. Comput. Sci*., **37,** 306–310 (1997).

[35] B. M. Smith, P. J. Gemperline, Wavelength selection and optimization of pattern recognition methods using the genetic algorithm, *Anal. Chim. Acta*, **423**, 167–177 (2000).

[36] S. S. So, M. Karplus, Genetic Neural Networks for Quantitative Structure−Activity Relationships: Improvements and Application of Benzodiazepine Affinity for Benzodiazepine/GABA$_A$ Receptors, *J. Med. Chem*., **39**, 5246–5256 (1996).

[37] R. Leardi, A. Lupiáñez González, Genetic algorithms applied to feature selection in PLS regression: How and when to use them, *Chemom. Intell. Lab. Syst*., **41**, 195–207 (1998).

[38] H. Yoshida, R. Leardi, K. Funatsu, K. Varmuza, Feature selection by genetic algorithms for mass spectral classifiers, *Anal. Chim. Acta*, **446,** 485–494 (2001).

[39] R. W. Kennard, L. A. Stone, Computer aided design of experiments, *Technometrics*, **11**, 137–148 (1969).