ELSEVIER

Artificial Intelligence in Medicine 28 (2003) 265–279

Artificial Intelligence in Medicine

www.elsevier.com/locate/artmed

Urinary nucleosides as potential tumor markers evaluated by learning vector quantization

Frank Dieterle[a,*], Silvia Müller-Hagedorn[b], Hartmut M. Liebich[b], Günter Gauglitz[a]

[a]*Institute of Physical and Theoretical Chemistry, Auf der Morgenstelle 8, D-72076 Tübingen, Germany*
[b]*Medizinische Universitätsklinik, Otfried-Müller-Strasse 10, D-72076 Tübingen, Germany*

Received 20 August 2002; received in revised form 3 April 2003; accepted 14 April 2003

Abstract

Modified nucleosides were recently presented as potential tumor markers for breast cancer. The patterns of the levels of urinary nucleosides are different for tumor bearing individuals and for healthy individuals. Thus, a powerful pattern recognition method is needed. Although backpropagation (BP) neural networks are becoming increasingly common in medical literature for pattern recognition, it has been shown that often-superior methods exist like learning vector quantization (LVQ) and support vector machines (SVM). The aim of this feasibility study is to get an indication of the performance of urinary nucleoside levels evaluated by LVQ in contrast to the evaluation the popular BP and SVM networks. Urine samples were collected from female breast cancer patients and from healthy females. Twelve different ribonucleosides were isolated and quantified by a high performance liquid chromatography (HPLC) procedure. LVQ, SVM and BP networks were trained and the performance was evaluated by the classification of the test sets into the categories ''cancer'' and ''healthy''. All methods showed a good classification with a sensitivity ranging from 58.8 to 70.6% at a specificity of 88.4–94.2% for the test patterns. Although the classification performance of all methods is comparable, the LVQ implementations are superior in terms of more qualitative features: the results of LVQ networks are more reproducible, as the initialization is deterministic. The LVQ networks can be trained by unbalanced sizes of the different classes. LVQ networks are fast during training, need only few parameters adjusted for training and can be retrained by patterns of ''local individuals''. As at least some of these features play an important role in an implementation into a medical decision support system, it is recommended to use LVQ for an extended study.

Keywords: Artificial neural networks; Learning vector quantization; Support vector machines; Breast cancer; Urinary nucleosides

* Corresponding author. Tel.: +49-7071-2978760; fax: +49-7071-295960.
E-mail address: frank.dieterle@ipc.uni-tuebingen.de (F. Dieterle).

0933-3657/$ – see front matter
doi:10.1016/S0933-3657(03)00058-7

1. Introduction

Breast cancer is the most frequent female cancer in Europe and in the USA [33]. Among many publications directed at the investigation of diagnosing and monitoring cancer, methods based on the use of objectively derived laboratory data for the diagnosis and prognosis of cancer show several benefits in theory [51]: the parameters can be contrasted for accuracy and precision and methods can be developed, which can be easily transferred from institution to institution. A final aim could be the implementation of these methods in a user-friendly automated medical decision support system.

Recently, modified nucleosides (MN) were presented as potential tumor markers by different groups [47,55,58]. They are also excreted in the urine of healthy subjects, but their levels show different patterns from tumor bearing organisms. The pathological process is often a gradual change and therefore it is difficult to assign numerical decision limits to single laboratory data. More often, these changes affect several laboratory data by only slight alterations compared to the differences between individuals, resulting in patterns of several changes to be discerned. There have been many attempts of exploiting computer-based models for the pattern recognition of pathological processes. Many studies are based on traditional statistical techniques like discriminant analysis (LDA) [58] or rule induction (CART) [16], but in numerous publications artificial neural networks (ANNs) provided more accurate predictions [22,30,37,48,58]. Thus, the use of ANN is becoming increasingly popular in several fields of medicine concerned with pattern recognition like decision-making in pathology [58,15], prediction of mortality of patients suffering from different diseases [7,12,15,20,34,49], analysis of EKG patterns [9], decision making in mammography [10,53] and many more [3,11]. Also a preliminary study showed that ANNs are outperforming common machine learning methods like CART, LDA and k-nearest neighbor (KNN) in terms of the classification performance.

Nevertheless, it astonishes that almost all studies published in clinical literature use practically only one class of neural networks, the backpropagation networks [11]. Backpropagation (BP) networks show a very broad range of applications like approximation of functions, prediction of output values, classifications of objects, recognition of patterns in multifactorial data and are therefore often considered to be the most universal tool among the different classes of neural networks. Nevertheless, several publications in different fields of science proved that vector-based neural networks, especially learning vector quantization (LVQ) [25–27,35], outperform BP in the field of supervised pattern recognition [18,21,32,36,37,40,48]. However, LVQ is much unknown in cancer research, indicated by references available on MEDLINE Database (1960 to July 2002): the search ‘‘neural networks’’ and ‘‘cancer’’ or ‘‘malignant’’ returned 342 hits with nearly all references using BP, whereas a search using ‘‘LVQ’’ and ‘‘cancer’’ or ‘‘malignant’’ returned only 11 hits. These 11 references utilize LVQ for image and morphometry analysis, but no publication was found on LVQ for pattern recognition of laboratory data like analyte concentrations.

In this study, the patterns of urinary nucleoside levels of healthy individuals and breast cancer patients are evaluated by two different LVQ implementations and by BP networks and support vector machines (SVM) as references. SVM have gained an increasing

popularity in the field of computer-based classifications in medicine and in decision support systems during the last few years [6,8,13,17,19,56]. The aims of this feasibility study are to find a reasonable indication of the classification performance of urinary nucleosides quantified by HPLC and evaluated by ANNs, to find an easier pattern recognition method than BP and to present advantages and drawbacks of LVQ compared with the more popular BP and SVM in respect to an implementation in a medical decision support system.

2. Methods

2.1. Collection of urine samples

Spot urine samples were collected from 85 female breast cancer patients 1 day before they underwent tumor removal at Tuebingen Frauenklinik. The diagnosis ''breast cancer'' was confirmed in each case by histopathological examination of the tumor tissue. All the patients were then staged concerning the TNM system. The patients included in the study were aged between 25 and 85 years.

The control group consisted of 121 healthy female volunteers in the same range of age. All these individuals did not take any form of medical treatment. In particular, anyone suffering from any infectious or malignant disease was excluded.

After collection, the urine samples were immediately frozen without any preservatives and stored at $-20\ ^{\circ}C$. Directly before analysis the samples were thawed at room temperature.

2.2. Determination of the different nucleosides

The concentrations of 12 different major and minor ribonucleosides (uridine (U), cytidine (C), adenosine (A), guanosine (G), inosine (I), xanthosine (X), pseudouridine (PseU), 1-methyladenosine (m1A), 3-methylguanosine (m3G), 1-methylinosine (m1I), 1-methylguanosine (m1G) and 2-methylguanosine (m2G)) in urine were assayed by a completely validated reversed phase HPLC procedure. The ribonucleosides were first isolated from urine by affinity chromatography on phenylboronate gel columns. This gel selectively binds ribonucleosides at alkaline pH, thus separating them from interfering substances in urine. The ribonucleosides were then eluted by lowering the pH of the columns' bed. Finally, the nucleosides were separated and quantitated by an internal standard method with a HPLC apparatus using a reversed phase column and UV detection. The HPLC analysis enabled the direct determination of nucleoside concentrations (μM) in urine; these were then transformed in nmol/μmol creatinine. The concentrations of creatinine in the urine samples were determined by using a modified Jaffé method [5]. A previous publication has described in detail the reagents, the nucleoside standard compounds, the instrumentation, the HPLC buffers, the chromatographic column and the phenylboronate affinity chromatography clean-up for nucleoside isolation [31]. Furthermore, the two classical tumor markers for breast cancer, CEA and CA15-3, were performed on an enzyme linked immunoassay.

2.3. Artificial neural networks

ANNs can learn and recognize patterns by imitating the learning process of the human brain. They consist of a number of simple neurons, which are also called nodes or units. These neurons are connected by links, which have strength or ''weights'' that are learned during a training step by the network. The training of an ANN used for classification is an iterative process in which the network learns by example to categorize data among different classes. Thereby, the patterns are presented to the input neurons of the network and the network estimates a category presented by an output neuron. If the category is not within an error tolerance of the true category of the corresponding pattern, the network's internal structure (mostly the weights) is adapted to reduce the error. The classification performance is finally measured in a test phase using separate test data. Thereby, the network predicts the category of patterns not encountered in training just on the basis of the patterns themselves.

This results in the need of two independent data sets, a test set and a training set. Due to the frequently limited sizes of the data in clinical research two methods are common in literature to divide one single data set into training and test sets: the leave-one-out technique (also called round-robin or full cross-validation) and the leave-*n*-out technique [4,15,54]. For this study the leave-one-out strategy was used simulating a confrontation of a network with a new object never encountered before.

The concentration levels of the MN, which represent the patterns, were mean-centered for each MN by subtracting the average from each MN level. Furthermore, dividing the data by the variance of the corresponding nucleoside level standardized the patterns. Hence, all MN levels are given the same weight to influence the estimation of the category. The category of the patterns was coded 0 for ''healthy pattern'' and 1 for ''cancer pattern''.

The main difference between BP, LVQ and SVM is the method by which knowledge is learned and stored. BP fits a function to the training data and the classification is performed according to the output, which is calculated by the network with the given unknown pattern. The LVQ partitions the feature space of the training data into clusters and models prototypes for these clusters. The classification of the LVQ is based on the similarity of the unknown data and these prototypes. SVM partitions the feature space of the training data by finding a maximum margin hyperplane after the projection of the data into an *n*-dimensional space. The classification is performed by projecting the test data into the same space and by assigning the data to the class of the corresponding side of the hyperplane.

2.4. Learning vector quantization

According to Fig. 1, the LVQ network can be described as a three-layer neural network. The input layer has as many neurons as the pattern variables. The second layer contains at least one ''prototype neuron'' for each class. These prototype neurons and the corresponding weighted links between the input neurons and the prototype neurons form prototype patterns, also called codebook vectors. For example, the two weights w_1 and w_2 form the black codebook vector belonging to the class ''cancer'' in Fig. 1. The output layer and the links to the output layer serve as static linear map transforming the class of the codebook vectors into the classes presented at the output neuron and do not have any adaptable parameters.

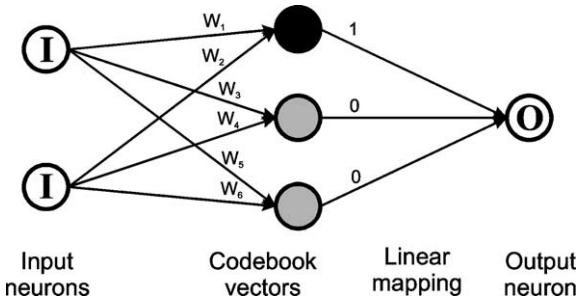

Fig. 1. LVQ network architecture. The network uses two input neurons for patterns consisting of two variables. Two codebook vectors for the patterns of healthy individuals (gray) and one codebook vector for cancer patterns are schematically shown.

The algorithm for the LVQ networks used in this study can be described by four steps:

1. Initializing the codebook vectors.
2. First training of the codebook vectors (450 cycles).
3. Balancing of the codebook vectors.
4. Second training of the codebook vectors (2500 cycles).

2.4.1. Initialization

The LVQ networks used in this study were initialized with five codebook vectors belonging to the class ''cancer'' and five codebook vectors belonging to the class ''healthy''. The initial vector values of the codebook vectors were picked up from the training data one at a time resulting in a rather high number of objects to be involved in the initialization process. Thereby, a KNN classification takes care that theses initial codebook vectors fall within the borders of the corresponding classes [24].

2.4.2. Training

LVQ networks are trained by iteratively optimizing the prototype codebook vectors. Thereby, an input pattern is compared with the codebook vectors by calculating the distance:

$$||x = m_i|| \tag{1}$$

with x as the presented input pattern and m_i as the codebook vectors. The codebook vector showing the smallest distance is most similar to the input pattern and therefore is the winner m_c. If the winner and the input pattern belong to the same class, the codebook vector is made more similar to the presented output pattern by moving the winner codebook vector towards the input vector:

$$m_c(t+1) = m_c(t) + \alpha_c(t)[x(t) - m_c(t)] \tag{2}$$

Each codebook vector has its individual learning rate $\alpha_c(t)$, which decrease over time and thus speeds up the convergence of the learning phase. The optimal learning rates are determined during the learning by the algorithm itself. This learning scheme is called OLVQ1 whereby more details can be found in [24].

Conversely, if the winning codebook belongs to the wrong class, it is made less similar to the presented input pattern by moving the winner away from the input pattern:

$$m_c(t+1) = m_c(t) - \alpha_c(t)[x(t) - m_c(t)] \quad (3)$$

The comparing step and the adaptation of the winning codebook form one training cycle. For this study, the number of training cycles of the first training (100–1000) and for the second training (500–5000), the initial learning rate (0.9–0.1) and the number of codebook vectors (2–20) were systematically varied using a full factorial design. The best parameters in terms of lowest classification errors corresponded to 450 training cycles during the first training phases, 2500 training cycles during the second training phase, an initial learning rate of 0.2 and 5 initial codebook vectors for each class. All calculations were performed by the program LVQ_PAK [28].

2.4.3. Balancing

The balancing step between the two training steps follows the strategy that for an optimal approximation of the class borders the average distances between adjacent codebook vectors should be the same in every class as the borders are represented piecewise linearly by segments of midplanes between codebook vectors of neighboring classes (borders of the so-called Voronoi tessellations). Therefore, the balance step calculates the medians of the shortest distances between the codebook vectors of each class. If the distances are very different new codebook vectors are added to the class with the longest distance. After that all codebook vectors are retrained in a second training phase.

In general the LVQ can be described as a prototype based method, which improves the Voronoi partition of the input space induced by the initial prototype codebook vectors by minimizing the empirical error using the training and balancing steps. Although LVQ can overfit the data similar to BP, the rather low number of codebook vectors used in this study helps to reduce the risk of overfitting.

2.5. Dynamic learning vector quantization

In contrast to LVQ, the dynamic learning vector quantization (DLVQ) generates the second layer dynamically during the training phase. The DLVQ algorithm works the following way:

1. For every class of the training data one codebook vector is calculated as the mean vector.
2. Each pattern of the training is presented to the network and the codebook vectors are adapted similar to LVQ. This cyclical process is repeated until the number of correctly classified vectors no longer increases.
3. A new codebook vector m_a is calculated from the vectors of a class a associated with a wrong class b. For every class one of the new mean vectors is added to the net. The algorithm returns to step 2.

For this study, the number of learning cycles of step 2 was limited to 50 and the maximal number of codebook vectors added to each class was set to 10. All calculations were performed by SNNS [57].

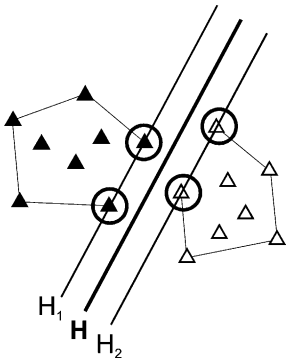

Fig. 2. Separation of two classes by SVM using two support vectors of each class and the maximal margin hyperplane H.

2.6. Support vector machines

SVM are an effective algorithm to find a special kind of linear model, the maximal margin hyperplane, to separate two classes of patterns. Thereby, a transformation of the data into a higher-dimensional space by a nonlinear mapping allows a linear separation of classes, which could not be linearly separated in the original space. The principles of SVM and the maximal margin hyperplane are illustrated in Fig. 2. The two classes of the black and the white triangles are linearly separable by any hyperplane between the hyperplanes H_1 and H_2. These two hyperplanes board the two classes. The objects that are located on these two hyperplanes are the so-called support vectors. The maximal margin hyperplane H, which is located in the middle of H_1 and H_2, gives the greatest separation between the classes. The maximal margin hyperplane is uniquely defined by the support vectors. The support vectors can be regarded as selected representatives out of the training data, which are most critical for the separation of the two classes. As usually only few support vectors are used there are only some parameters adjustable by the algorithm and thus overfitting is unlikely to occur. For polynomial kernel support vector machines, which were used in this study, the maximal margin hyperplane can be written as:

$$x = b + \sum \alpha_i y_i (a(i)a)^n \tag{4}$$

Thereby, y_i is the class value ($+1/-1$) of the training object $a(i)$, a the object to test, b and α_i the weights (parameters) determined by the algorithm and n the degree of polynomials determining the nonlinearity of the projection. Only training objects that serve as support vectors have a weight α_i different from zero. The support vectors for the training data and the weights can be determined by a constrained quadratic optimization, whereby many special algorithms have been developed to speed up learning. For data sets, which are not separable by SVM, the weights α_i have to be limited by an upper bound whereby this parameter has to be adjusted by the user.

For this study, the degree of polynomials (1–20) and the upper bound for the weights (1–10) were systematically varied using a full factorial design. The best parameters in terms of lowest classification errors corresponded to a degree of polynomials of 8 and an

upper bound for the weights of 5. A special implementation of the WEKA machine learning project was used for the SVM applied in this study [44,52].

2.7. Backpropagation

Due to the numerous excellent textbooks [23,41,42,59] describing in detail multilayer feedforward backpropagation (BP) networks only the differences between BP, SVM and LVQ will be outlined:

Most common in literature are BP networks consisting of three layers of neurons. In contrast to LVQ, there is no winner in the second layer, but all neurons will contribute to the output during training and prediction. The neurons of the third layer and the connections between the second and third layer are not static but have the same function as the preceding layer: each neuron sums up the inputs from the connected preceding neurons, applies a sigmoidal transfer function and feeds the output to the connected succeeding neurons. During the training phase, the error of prediction is sent backwards through the network and the weights of all connections are adjusted to reduce this error.

BP networks are affected to a greater extend by an overfitting [50] than SVM and LVQ networks. An overfitted ANN memorizes the often too small training set instead of generalizing the patterns and consequently performs badly on new data. The overfitting was anticipated by the so-called early stopping [46]. Early stopping means that the training is stopped when the error of cross-validation starts going up, as the net may start loosing its generalization ability at this moment.

For the BP networks implemented in this study a full factorial experimental design was used to find an optimal network topology. Thereby, the number of hidden neurons (1–20) and the number of hidden layers (1–2) were varied to minimize the classification errors evaluated by a full cross-validation. The best network topology consisted of 12 input neurons, 1 layer of 4 hidden neurons and 1 output neuron. The training included a maximum of 1000 learning cycles with an early stopping procedure. Scaled conjugate gradient (SCG) was used as training algorithm [38]. In contrast to standard backpropagation SCG converges faster, is more capable of dealing with local minima and is not sensitive to any parameters, which were all set to standard values suggested in [38]. All calculations were performed by a new implementation of the Stuttgart Neural Network Simulator [57,43] on a personal computer.

3. Results

For a valid comparison, only the results of the test data will be used, simulating a confrontation of the networks with new cases in a clinical setting. The three methods are compared by the sensitivity, specificity and the prediction rate. Clinical sensitivity is the percentage of true positives, clinical specificity is the percentage of true negatives and the prediction rate is the percent of true positives and true negatives in all individuals.

The results for all methods are shown in Table 1. LVQ classified 60 out of 85 cancer patients and 109 out of 121 healthy patients correctly, resulting in a sensitivity of 70.6%, a specificity of 90.1% and a prediction rate of 82.1%. A so-called sammon mapping [24]

Table 1
Results of the classification of the training and test objects

Method	Training patterns			Test patterns		
	True cancer (%)	True healthy (%)	Overall classification rate (%)	True cancer (%)	True healthy (%)	Overall classification rate (%)
LVQ	88.2	96.7	93.2	70.6	90.1	82.1
DLVQ	72.9	99.2	88.4	58.8	94.2	79.6
SVM	91.8	93.4	92.7	68.2	88.4	80.1
BP	94.1	95.0	94.7	70.6	91.7	83.0

Overview of the of the classifications by LVQ, DLVQ, SVM and BP for the training patterns and the test patterns.

can be used to reduce the 12-dimensional classification problem to 2-dimensions while trying to preserve the distances between all vectors. The classifications are visualized in Fig. 3. Thereby, the seven healthy codebook vectors and the five cancer codebook vectors, which represent the prototype training patterns vectors, are projected into this map as well as the 206 patterns of the data. It is visible that most correctly classified patterns are located in the outskirts of the cloud of patterns (for example the healthy patterns on the top left and bottom right side and the cancer patterns on the top right side). In the middle of the cloud the situation is more unclear with a number of misclassified cancer patterns surrounded by many correctly classified healthy patterns.

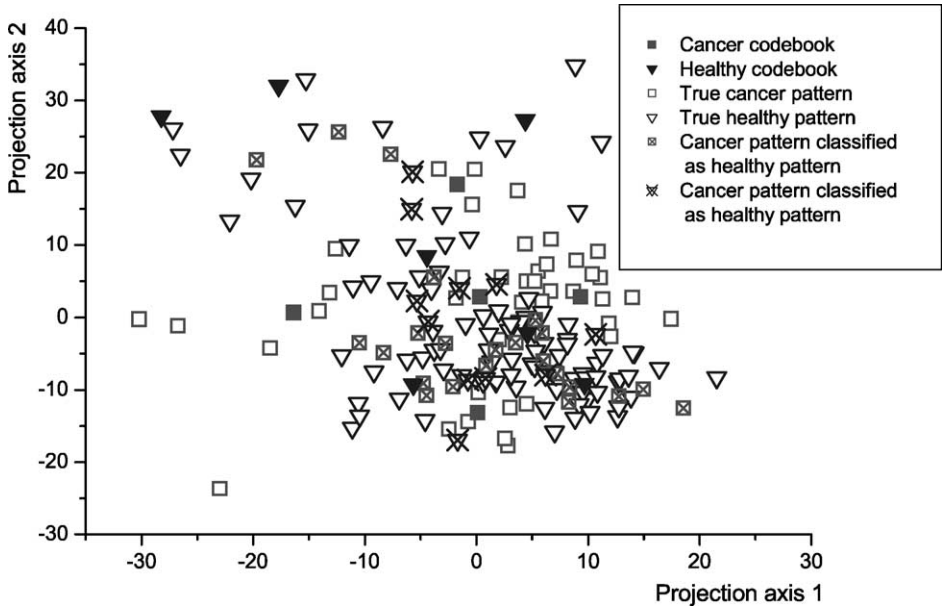

Fig. 3. Sammon mapping of the classification of the test data performed by LVQ.

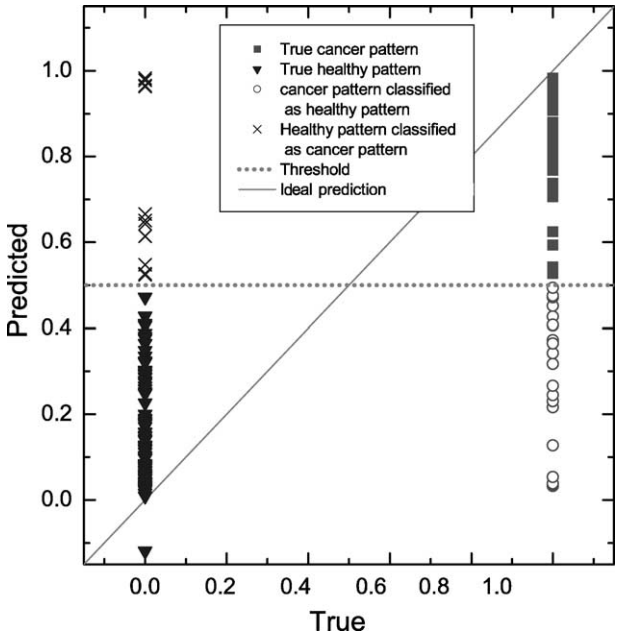

Fig. 4. True-predicted plots of the classification of the test data by BP.

This means that most misclassified cancer patterns are misclassified due to features, which are very similar to features of healthy patterns, and not due to outlying different features.

According to Table 1, DLVQ achieved a slightly lower prediction rate than LVQ with a higher sensitivity of 94.2% and a lower specificity of 58.8%. Thereby, DLVQ used (7–12) codebook vectors. The prediction rate of 82.1% by BP is practically identical to LVQ with exactly the same specificity (70.6%) and a slightly higher sensitivity (91.7%). As BP is not a vector network technology, a sammon mapping makes less sense. Yet, the outputs and thus the classifications can be visualized by true-predicted plots. Thereby, the calculated output of the ANN is plotted versus the true diagnosis. A diagonal represents the ideal output of the ANN and the threshold used for classification is plotted as horizontal line. In Fig. 4, the true-predicted plot is shown for the 206 patterns. It is obvious that most misclassified patterns (except of four misclassified healthy patterns) were not predicted as extreme outliers but rather as samples, which are continuously overlapping with the other class. This finding is similar to the sammon mapping of the classification by LVQ. The predictions by the SVM are also very similar to the predictions by LVQ with a slightly lower sensitivity (88.4%) and a slightly lower selectivity (68.2%). To test the significance of the prediction rates of LVQ, DLVQ, BP and SVM a Wilcoxon signed ranks test was performed for the predictions. The null hypothesis H_0 demonstrating that there is no significant difference between the methods was accepted at 5% error level for all combinations of the four methods.

In terms of the speed of operation, it is important to distinguish the training and the actual prediction phase. While all four algorithms are very fast classifying new cases

(practically not measurable for a modern PC), the four methods significantly differ concerning the time needed for the training: the calculation of the full cross-validation (206-folds) of the complete data set took 1050 s for BP, 178 s for SVM, 126 s for DLVQ and 73 s for LVQ whereby the time needed for the classification of test patterns can be neglected.

4. Discussion and conclusion

This study is part of the worldwide efforts to find objectively derived laboratory data for the diagnosis and prognosis of breast cancer. Recently, the patterns of MN levels were presented as potential tumor markers by different groups. In an early publication, it was indicated that an improved differentiation between healthy female individuals and breast cancer patients could be achieved by employing backpropagation neural networks as pattern recognition method for the concentrations of several modified nucleosides.

The interest in the use of BP is constantly growing in clinical applications. Yet, it has been shown in many other fields of science, that BP is not always the best choice out of the large family of neural networks. LVQ has been shown to be one of the most promising methods on pattern recognition in non-medical applications. Thus, in this study two implementations of LVQ were compared with the well-known BP and the SVM, which have become very popular during the last few years. Besides of the classification performance, the advantages and drawbacks of the different types of neural networks will be discussed. Especially the so-called qualitative criterions like complexity, retraining, outlier rejection, and speed will be discussed in respect to an implementation as medical decision support system.

All four classification methods performed very well on the calibration and prediction of the data set. The sensitivity of the prediction of the test data ranged from 88.4 to 91.7%, the specificity from 58.8 to 70.6% and the prediction rate for the test data resulted in 79.6–83.0% with LVQ and BP classifying most patterns correctly. The differences in the absolute numbers of correctly classifieds patterns for both classes together turned out not to be significant for all methods. The sensitivity and specificity of all methods are very promising compared with the most established markers for this type of malignancy CEA and CA15-3 showing only a sensitivity of 6%, respectively, 18% for the population of this study. The sensitivity of these two markers is generally below 20% at a specificity of 95% [2,14]. Nevertheless a definitive judgement of the usefulness of urinary nucleoside levels being evaluated by neural networks is only possible on the basis of a study incorporating the investigation of many more patterns. The study presented in this paper with a limited number of patterns is a feasibility study with two objectives: first, to find an alternative pattern recognition method which is easier to handle and to understand and more robust than BP showing at least a comparable performance. The second objective is to find a reasonable indication of the classification performance of urinary MN quantified by HPLC and evaluated by NN to justify an extensive and thus a cost intensive study. If in such an extended study other factors are included, which are also known to be predictive, the classification performance could even improve.

Although BP has become well established in several medical applications, many end users do not understand how neural networks work and often see artificial intelligence as a computerized oracle [29,45] due to two reasons: first, it is difficult to understand, how BP can classify patterns into different groups just by approximating a function during the training phase. In this issue, LVQ is a lot easier to understand with prototype patterns being formulated during training and with classifying by finding the most related prototype pattern. Also a graphical representation by a 2-dimensional mapping of LVQ classifications is more overt than true predicted plots of BP classifications. Although the classification by SVM can also be described by example patterns, which are most critical for a correct classification and which set up a boundary hyperplane, the high-dimensional projection makes this approach less insightful than LVQ. The second issue is the number of parameters used for training. While BP needs a lot of decisions to be made by the user for training, like choosing the number of hidden layers, the number of neurons, the type of activation function, parameters for cross-validation, several parameters for the learning function and several parameters for a possible pruning step, LVQ and SVM only need few parameters set by the user. This implies a more reliable classifier and an easier comparability between different studies. Further, the initialization of LVQ and SVM is deterministic in contrast to BP. The random initialization of the weights of BP results in different outputs for each run even when using the same datasets for training and prediction and thus complicates the reproducibility and comparability. In addition, the implementation of LVQ into software is easy, as the algorithms are very simple, whereas the algorithms for the SVM are more complicated and the algorithms for BP networks are often highly sophisticated. Another qualitative feature is the outlier rejection. LVQ, which is based on classifying by distance, can easily be configured to reject patterns that are not within a certain distance of the prototype patterns. For SVM the outlier detection is more difficult as the classification is based on the maximal margin hyperplane, which is set up only by the most critical training instances. The outlier detection of BP is also more difficult, which needs the notion of an object's classification margin introduced recently [1]. It has been further shown in this study that LVQ is the fastest method followed by SVM whereas BP needs significantly more time for training. In the case of studies using several hundred times more patterns, the time needed for training becomes an important issue, as the time increases nonlinearly with the number of patterns.

Two topics not investigated in this study due to the limited number of patterns are the unbalanced number of training patterns in different classes and the retraining. If a BP network is trained by patterns belonging to two groups of very different quantities, the network tends towards classifying all objects into the large group [21,39]. As LVQ uses prototypes of patterns for each group, this danger is not present. On the contrary, the number of patterns that are more frequent and easier to find can be increased to improve the overall classification performance of LVQ networks. Also the SVM should not be subject to tend to misclassifying patterns, which belong to the smaller class, as long as enough training patterns are available for selecting support vectors which build a reasonable maximal margin hyperplane. The second topic concerns the retraining, which is important if incremental learning is important. This could happen if further studies show that the classification of a local population is improved with a global model being retrained by a training set of local individuals. In this case, the model should be able to learn the new data

without forgetting learned knowledge. In [32], it is shown that the degradation of learned knowledge during retraining is less for LVQ than BP. From the theoretical point of view, the SVM should allow a retraining without a degradation of learned knowledge, whereas no study is known to the author investigating this topic.

While in respect of the classification performance all four algorithms perform well, LVQ outperforms BP in terms of the qualitative criterions like complexity, speed, retraining, unbalanced sizes of different classes and reproducibility. DLVQ is not mentioned extra in this discussion as it shows similar qualities as the original LVQ implementation in respect to most qualitative criterions, whereas the training time and the specificity are slightly inferior. The SVM can be considered as being between LVQ and BP for the ensemble of qualitative criterions. Thus, based on the results of this study, it is recommended to use the LVQ implementation for a further more extended study. Considering other studies inside and outside the medical literature in addition [18,21,32,36,37,40,48] it is advisable always considering LVQ when a powerful pattern recognition method is needed.

To prevent misunderstandings, some of the qualitative features discussed in this study like speed of the training, complexity and retraining capabilities are not always of primary importance in possible clinical applications. A common situation might be a network being trained with thousands of patterns on a fast computer by an expert on artificial intelligence outside the clinical application. Afterwards this neural network will be implemented in user-friendly software. In a clinical application, the end user will feed this software with the urinary nucleoside levels and in less than a second, the program will indicate the risk of the individual bearing breast cancer. One could think of implementing the trained network into the HPLC software, or even of assembling a dedicated device, which consists of an HPLC, automated HPLC software and a well trained LVQ network. This would be a user-friendly fully automated system supporting the oncologist in the decision making process without the need of exactly understanding what is inside the black box.

References

[1] Allwein EL, Schapire RE, Singer Y. Reducing multiclass to binary: a unifying approach for margin classifiers. J Machine Learn Res 2000;1:113–41.
[2] American Society of Clinical Oncology. Clinical practice guidelines for the use of tumor markers in breast and colorectal cancer. J Clin Oncol 1996;14:2843–77.
[3] Astion ML, Wilding P. The application of backpropagation neural networks to problems in pathology and laboratory medicine. Arch Pathol Med 1992;116:995–1001.
[4] Baker JA, Kornguth PJ, Lo JY. Artificial neural network for the prediction of breast cancer using BI-RADS standardized lexicon. Radiology 1995;196:817–22.
[5] Bartels H, Böhmer M, Heierli C. Serum Kreatininbestimmung ohne Enteiweissen. Clin Chim Acta 1972;37:193–7.
[6] Bredensteiner EJ, Bennett KP. Multicategory classification by support vector machines. Comput Optim Appl 1999;12:53–79.
[7] Burke HB, Goodman PH, Rosen DB, Henson DE, Weinstein JN, Harrell FE. Artificial neural networks improve the accuracy of cancer survival prediction. Cancer 1997;79:857–62.
[8] Cai YD, Liu XJ, Xu XB, Chou KC. Support vector machines for predicting HIV protease cleavage sites in protein. J Comput Chem 2002;23:267–74.
[9] Cios KJ, Chen K, Langenderfer RA. Use of neural networks in detecting cardiac diseases from echocardiographic images. IEEE Eng Med Biol 1990;9:58–60.

[10] Dawson AE, Austin RE, Weinberg RS. Nuclear grading of breast carcinoma by image analysis: classification by multivariate and neural network analysis. Am J Clin Pathol 1991;95:29–37.
[11] Dayhoff JE, DeLeo JM. Artificial neural networks. Cancer 2001;91:1615–35.
[12] Douglas TH, Connelly RR, McLeod DG. Neural network analysis of pre-operative and post-operative variable to predict pathologic stage and recurrence following radical prostatectomy. J Urol 1996;155:87.
[13] Dreiseitl S, Ohno-Machado L, Kittler H, Vinterbo S, Billhardt H, Binder M. A comparison of machine learning methods for the diagnosis of pigmented skin lesions. J Biomed Inform 2001;34:28–36.
[14] Ebeling FC, Schmitt UM, Untch M, Nagel D, Fateh-Moghadam A, Stieber P, et al. Tumour markers CEA and CA15-3 as prognostic factors in breast cancer—univariate and multivarite analysis. Anticancer Res 1999;19:2545–50.
[15] Floyd CE, Lo JY, Yun AJ, Sullivan DC, Kornguth PJ. Prediction of breast cancer malignancy using an artificial neural network. Cancer 1994;74:2944–8.
[16] Ford N. From information to knowledge management: the role of rule induction and neural net machine learning techniques in knowledge generation. J Inform Sci 1989;15:299–304.
[17] Furey TS, Cristianini N, Duffy N, Bednarsky DW, Schummer M, Haussler D. Support vector machine classification and validation of cancer tissue samples using microarray expression data. Bioinformatics 2000;16:906–14.
[18] Gardner JW, Shin HW, Hines EL, Dow CS. An electronic nose system for monitoring the quality of potable water. Sens Actuators B Chem 2000;69:336–41.
[19] Guyon I, Weston J, Barnhill S, Vapnik V. Gene selection for cancer classification using support vector machines. Machine Learn 2002;46:389–422.
[20] Hurst RE, Bonner RB, Ashenayi K, Veltri RW, Hemstreet GP. Neural net-based identification of cells expressing the p300 tumor-related antigen using fluorescence image analysis. Cytometry 1997;27: 39–42.
[21] Johnson SR, Sutter JM, Engelardt HL, Jurs PC. Identification of multiple analytes using an optical sensor array and pattern recognition neural networks. Anal Chem 1997;69:4641–8.
[22] Kattan MW, Beck JR. Artificial neural networks for medical classification decisions. Arch Pathol Lab Med 1995;119:672–7.
[23] Kaykin S. Neural networks a comprehensive foundation. Upper Saddle River: Prentice-Hall; 1999.
[24] Kohonen T, Hynninen J, Kangas J, Laaksonen J, Torkkola K. LVQ_Pak: the learning vector quantization program package, Report A30, Helsinki: Helsinki University of Technology; 1993.
[25] Kohonen T, Improved versions of learning vector quantization. In: Proceedings of the IEEE International Joint Conference on Neural Networks. New York: Institute of Electrical and Electronics Engineers; 1990. p. 545–50.
[26] Kohonen T, Kangas J, Laaksonen J, Torkkola K. LVQ_PAK: a program package for the correct application of learning vector quantization algorithms. In: Proceedings of the IEEE International Joint Conference on Neural Networks. New York: Institute of Electrical and Electronics Engineers; 1992. p. 725–30.
[27] Kohonen T. Self-organizing maps. Heidelberg: Springer-Verlag; 1995.
[28] Kohonen T. LVQ_PAK: the learning vector quantization program package. Helsinki: Helsinki University of Technology; 1993.
[29] Levin RF. Clinical problems, computational solutions. Cancer 2001;91:1595–602.
[30] Levin RF. Conference concluding remarks. Cancer 2001;91:1696–7.
[31] Liebich HM, Di Stefano C, Wixforth A, Schmid HR. Quantitation of urinary nucleosides by high-performance liquid chromatography. J Chromatogr A 1997;763:193–7.
[32] Llobet E, Hines EL, Gardner JW, Franco S. Nondestructive banana ripeness determination using a neural network-based electronic nose. Meas Sci Technol 1999;10:538–48.
[33] Luminachi F, Brandes AA, Boccagni P, Polistina F, Favia G, D'Amico DF. Long-term follow-up in breast cancer patients using serum tumor markers CEA and CA15-3. Anticancer Res 1999;19:4485–90.
[34] Maclin PS, Dempsey J, Brooks J, Rand J. Using neural networks to diagnose cancer. J Med Syst 1991;15:11–9.
[35] Makhoul J, Roucos S, Gish H. Vector quantization in speech coding. Proc IEEE 1985;73:1551–88.
[36] Maksiovic R, Popovic M. Classification of tetraplegics through automatic movement evaluation. Med Eng Phys 1999;21:313–27.

[37] Mattfeldt TM, Kestler HA, Hautmann R, Gottfried HW. Prediction of prostatic cancer progression after radical prostatectomy using artificial neural networks: a feasibility study. Br J Urol 1999;84:316–23.
[38] Moller MF. A scaled conjugate gradient algorithm for fast supervised learning. Neural Networks 1993;6:525–33.
[39] Namee BM, Cunningham P, Byrne S, Corrigan OI. The problem of bias in training data in regression problems in medical decision support. Artif Intell Med 2002;24:51–70.
[40] Pantazopoulos D, Karakitsos P, Pouliakis A, Iokim-Liossi A, Dimopoulos MA. Static cytometry and neural networks in the discrimination of lower urinary system lesions. Urology 1998;51:946–50.
[41] Patterson D. Artificial neural networks, theory and applications. Upper Saddle River: Prentice-Hall; 1996.
[42] Principe J, Euliano N, Lefebvre W. Neural and adaptive systems: fundamentals through simulations. New York: Wiley; 2000.
[43] Rapp FR. Kalibrierung und Validierung von Sensorarrays unter Einsatz Neuronaler Netze und genetischer Algorithmen. Master Thesis. Tuebingen: University of Tuebingen; 1996.
[44] Rapp FR. A data mining tool for search and optimization of lead structures, http://www.solves.org. Tuebingen: University of Tuebingen; 2002.
[45] Rogers E. AI and the changing face of health care. IEEE Intell Syst 1998;13:20–5.
[46] Sarle WS. Stopped training and other remedies for overfitting. In: SAS Institute Inc., editors. Proceedings of the 27th Symposium on the Interface of Computing Science and Statistics. Cary, NC: SAS Institute Inc.; 1995. p. 352–60.
[47] Sasco AJ, Rey F, Reynaud C, Bobin JY, Clavel M, Niveleau A. Breast cancer prognostic significance of some modified urinary nucleosides. Cancer Lett 1996;108:157–62.
[48] Shaffer RE, Rose-Pehrsson SL, McGill RA. A comparison study of chemical sensor array pattern recognition algorithms. Anal Chim Acta 1999;384:305–17.
[49] Snow P, Rawford ED, DeAntoni EP. Prostate cancer diagnosis from artificial neural networks using the prostate cancer awareness week (PCAW) database. J Urol 1997;157:365.
[50] Weigend A. On overfitting and the effective number of hidden units. In: Mozer A, Smolensky P, Touretzky DS, Elman JL, Weigend AS, editors. Proceedings of the 1993 Connectionist Models Summer School. Hillsdale, NJ: Lawrence Erlbaum Associates; 1994. p. 335–42.
[51] Wilding P, Morgan MA, Grygotis AE, Shoffner MA, Rosato EF. Application of backpropagation neural networks to diagnosis of breast and ovarian cancer. Cancer Lett 1994;77:145–53.
[52] Witten IH, Frank E. Data mining: practical machine learning tools and techniques with java implementations. San Francisco: Morgan Kaufmann; 1999.
[53] Wu Y, Doi K, Giger ML, Vyborny CJ, Schmidt RA, Nisikawa RM. Application of artificial neural networks in mammography for the diagnosis of breast cancer. Radiology 1991;181:143.
[54] Wu Y, Giger ML, Doi K. Artificial neural networks in mammography: application to decision making in the diagnosis of breast cancer. Radiology 1993;187:81–7.
[55] Xu G, Liebich HM, Lehmann R, Müller-Hagedorn S. Capillary electrophoresis of urinary normal and modified nucleosides of cancer patients. In: Mitchelson KR, Cheng J, editors. Methods in molecular biology, vol. 162. Capillary electrophoresis of nucleic acids, vol. 1. Introduction to the capillary electrophoresis of nucleic acids. Totowa, NJ: Humana; 2001. p. 459–75.
[56] Yeang CH, Ramaswamy S, Tamayo P, Mukherjee S, Rifkin RM, Angelo M. Molecular classification of multiple tumor types. Bioinformatics 2001;17:316–22.
[57] Zell A. Stuttgart Neural Network Simulator (SNNS), http://www-ra.informatik.uni-tuebingen.de/SNNS/. Tuebingen: University of Tuebingen; 2002.
[58] Zhao R, Xu G, Yue B, Liebich HM, Zhang Y. Artificial neural network classification based on capillary electrophoresis of urinary nucleosides for the clinical diagnosis of tumors. J Chromatogr A 1998;828:489–96.
[59] Zupan J, Gasteiger J. Neural networks in chemistry and drug design. Weinheim: Wiley; 1999.