

Hybrid neural network and statistical classification algorithms in computer assisted diagnosis

R. Stotzka

Forschungszentrum Karlsruhe, 76344 Eggenstein–Leopoldshafen, Germany

ABSTRACT

The development of computer assisted diagnosis systems for image-patterns is still in the early stages compared to the powerful image and object recognition capabilities of the human eye and visual cortex. Rules have to be defined and features have to be found manually in digital images to come to an automatic classification. The extraction of discriminating features is especially in medical applications a very time consuming process. The quality of the defined features influences directly the classification success. Artificial neural networks are in principle able to solve complex recognition and classification tasks, but their computational expenses restrict their use to small images. A new improved image object classification scheme consists of neural networks as feature extractors and common statistical discrimination algorithms. Applied to the recognition of different types of tumor nuclei images this system is able to find differences which are barely discernible by human eyes.

Keywords: computer assisted diagnosis, pattern recognition, neural networks, feature extraction, classification, image understanding

1. COMPUTER ASSISTED DIAGNOSIS

Computer assisted diagnosis systems are designed to support medical doctors by the analysis of medical data, e.g. tissue lesions or X-ray images, in their diagnostic decision.^{1,2} Automatic classification systems are essential components in computer assisted diagnosis systems.

Classification is the exclusive decision of assignment of patterns into one of several possible categories. Many automatic classification methods have been developed during the last decades.^{3–5} An observation consists of descriptions or measured numbers of the original pattern or the original pattern itself, e.g. pixels of an image. Statistical algorithms are based on estimation of membership probabilities of an observation. The Bayes test compares the probabilities and is so by construction the overall optimal rule to discriminate the patterns of several categories. To determine the membership probabilities the conditional density functions of the categories or classes have to be estimated. Normally in pattern or object recognition tasks the observation vector has many entries and spans a high dimensional space. Thus the accurate estimation of the true statistical properties is very expensive. If the shapes of the conditional density functions are known only the statistical parameters have to be estimated (parametric classification). Sometimes it can be assumed the patterns of each class are e.g. normal distributed within the observation space. This leads to quadratic decision boundaries between the classes. The *quadratic classifier* is very often successfully applied even if the patterns are not exact normally distributed. If two classes cannot be separated by a single decision boundary, the application of the quadratic classifier is senseless.³

In contrary to statistical classification the application of artificial neural network models in classification does not need the estimation of conditional density functions of the classes. Following the idea of learning in biological brains artificial neural networks are trained with a representative set of patterns of each class. A presented pattern produces a network output. A learning algorithm modifies the weights of the connections (synapses) between the processing units (neurons) iteratively to force the desired output. Multilayer perceptrons and error-backpropagation learning rules^{6,7} are frequently used in classification tasks. They are able to construct arbitrary decision boundaries by learning an internal representation of the presented patterns and their distribution and to extract important features by self-organization. Disadvantages of neural networks, especially multilayer perceptrons, are their slow convergence (many presentations of the patterns and iterations are needed until the network approximates a discriminating

Further author information:

phone: +49 7247 824738

E-mail: Rainer@Stotzka.de

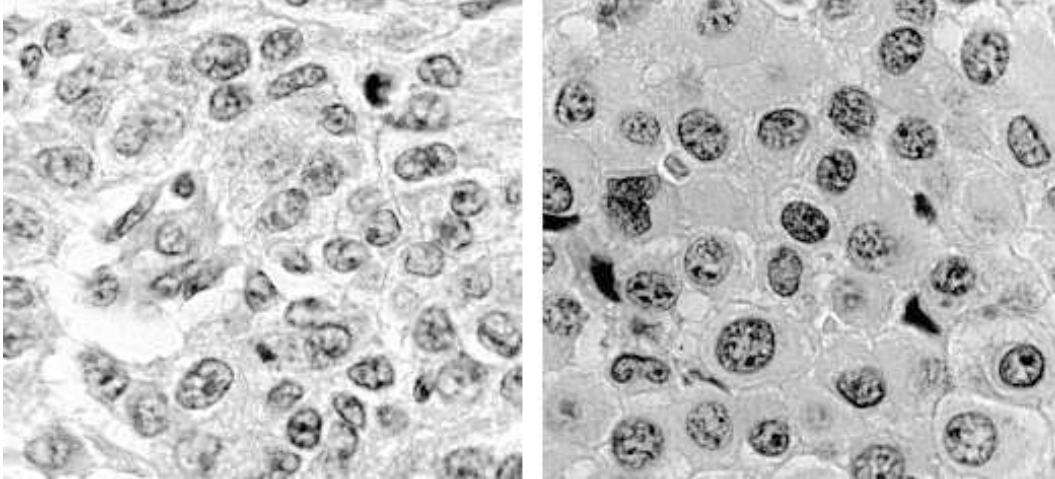


Figure 1. Examples of microscope scenes (recorded with an oil immersion objective at $630\times$) of pleura mesothelioma (left) and lung adenocarcinoma (right). Tumor nuclei vary very much in shape, size and chromatin structure. A decision based only on visual examination of the tumor nuclei often leads to no result.

surface) and the danger of converging to local optima instead of finding the optimal discrimination surface of the classification problem. The learning rule, e.g. backpropagation, tends to modify all connections globally. Even if “good” features already have been developed, the learning rule changes them due to the insufficient output caused by mal-developed output connections. Thus the convergence of the whole network is very slow.

A combination of multilayer perceptrons and statistical discrimination algorithms may result in improved classification schemes.^{8,9} Multilayer perceptrons are trained until feature detectors develop in the connections of the first layer. This takes approximately $\frac{1}{10} - \frac{1}{4}$ of the whole training time. The neural filters are extracted and used as linear feature filters. Their outputs are feed into a statistical classifier, e.g. a quadratic classification algorithm. The neural filters transform the distribution of the input patterns so that respective two classes are separable by a single decision boundary by reducing the observation space. The self-organized feature filters should pass only information important for discriminating the classes. The statistical classification algorithm builds the decision boundary. This method may lead sophisticated classification systems.

In the following a hybrid classification system consisting of multilayer perceptrons and quadratic classifiers is applied to separate lung adenocarcinoma and pleura mesothelioma tumor nuclei. In Section 2 the classification problem and used materials is described. Section 3 deals with the image preprocessing and Section 4 describes in detail the classification system. In Section 5 follow results and in Section 6 final conclusions.

2. MATERIALS

Knowledge about biological behavior of tumors is principally based on clinical experience. Showing similar morphological characteristics, different types of tumors are told apart by naming them after the tissue of their origin. Such a histogenetic classification is a difficult process and not always unambiguously practicable. However, even if the type of the tumor is established, many variations with different characteristics of the tumor are possible. From this follows a mapping according to criteria of the extension (staging) and the differentiation (grading) of the tumor. Characteristics of the disease, e.g. malignance of the tumors and life expectancy of the patient, can be derived from this diagnosis.¹⁰ However, the histogenetic classification the staging and grading of tumors are highly demanding tasks, whose success depends mostly on the experience of the examining expert. In contrast to grading based on histometric structures the structure of single cell nuclei, e.g. width, form and structure of chromatin, can be examined. These characteristics are widespreadly summoned to support histogenetic classification or grading. Another advantage of this method is that it requires only small tissue samples consisting of a few hundred cells of the tumor. The size of a probe obtained by a fine needle biopsy or a smear is sufficient to allow a diagnosis. Visual examination of different nuclei shows often only subtle differences which are hard to detect and are often misinterpreted by human observers. An example is the discrimination of two different types of tumors, pleura mesothelioma and lung adenocarcinoma.

The differential diagnosis of mesothelioma and adenocarcinoma is a common problem in lung pathology. The current methods in a routine laboratory include the following:

Histochemistry Several markers can be employed.¹¹ The main marker is Carcinoembryonic antigen, which is positive in adenocarcinoma and negative in mesothelioma. While there are other markers that can be used, at the present there is no positive marker for mesothelioma.

Histogeometry The inner structure and arrangement of nuclei within a tumor may vary in dependency of the tumor origin. Pathologists use histogeometric features to describe tumor lesions. Getting tumor tissue using fine needle biopsies, the arrangement of the nuclei is destroyed and histogeometric features are useless.

Each pathologist experienced in this field has his own criteria which includes knowledge of the patient, X-rays, operative findings and histological appearance. They use subjective criteria and often panels of pathologists have to decide in difficult cases. Visual examination of tumor nuclei to classify histogenetic tumor types or to grade tumor differentiation is a difficult problem. The classification is observer dependent and very often it is not executable reliably and reproducibly. The point in question is whether morphological features and chromatin textures of tumor nuclei are sufficient to solve given tumor type classification and grading problems.

The processed tumor lesions (see Figure 1) consisted of 17 adenocarcinoma and 31 mesothelioma patients. Each patient was diagnosed by 5 different pathologist independently. The images were recorded by a magnification of $630\times$ by Peter Bartels of the Optical Sciences Center, Tucson, AZ. An automatic image understanding¹² system segmented the tumor nuclei and produced ≈ 50 single nucleus images of each patient.

To determine the classification accuracy of a classification method the nuclei images are divided into training sets and test sets. The training sets consist of a similar number of segmented nuclei of adenocarcinoma and mesothelioma nuclei each. They are used as databases to build the classifier. The test sets consist of the remaining nuclei and are used to measure generalization abilities of the constructed classifier. It seems reasonable to divide the nuclei images into training sets and test sets by patients due to the aim diagnosing tumors. Anyway, the nuclei of a patient show similarities because of staining and microscopic imaging. Thereby the maximal size of the training set is limited to 17 adenocarcinoma and 17 mesothelioma patients which would leave no adenocarcinoma patient for testing.

3. IMAGE PREPROCESSING

Biological material can appear in many variations. Objects regarded under a microscope can be found in different locations, rotations or be viewed from different sides. A image recognition system has to be able to classify all objects regardless these variations.

The image preprocessing normalizes the positions of the single-nucleus-images by performing three steps:

- The plane of segmented nucleus is centered in a 128×128 image.
- The form of the nucleus is threatened as an ellipse and rotated to align the main axis of all nuclei uniformly from the upper left corner to the lower right.
- If necessary the nucleus is flipped to position the center of graylevel gravity at the upper left quarter of the image.

Now a classification system can be trained and tested to discriminate different types of nuclei objects without taking care of the position and the orientation.

However, images of the size 128×128 are still too large to be trained by neural networks or statistical discriminant analysis. On the other hand, standard reduction of the resolution to get smaller images causes a loss of information, small structures and variations of the chromatin within the nuclei will be lost. To reduce the amount of processed data and simultaneously preserve the fine structures a coarse-coding scheme is applied.^{13,14}

In Figure 2 overlaying fields of coarser pixels are generated using a 8-fold coarse-coding. A preprocessed nucleus image of the size 128×128 is coded in 8 images of lower resolution. The first 16×16 coarse image is generated by summing up 8×8 fields of the nucleus image. In the next image the 8×8 fields are shifted 1 pixel to the right and 1 pixel down to produce the coarse pixels. This lead to a different coding of the image. At the right and at the lower border zeros are appended to fill missing pixels. The other 6 coarse images are generated accordingly.

Each coarse image carries a slightly different version of the original nucleus by reduced resolution. Subtle structures are preserved in the differences of differently coarse images. Figure 2 shows a preprocessed nucleus by centering, rotating and flipping, and its corresponding 8 coarse images of lower resolution.

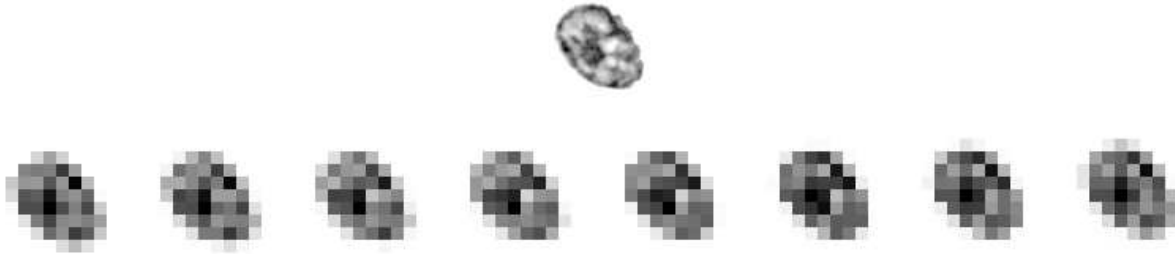


Figure 2. Nucleus after preprocessing. The image on the top shows a segmented nucleus found in the left microscope scene of Figure 1 in original resolution (128×128 pixels) after centering, rotation-normalization and straightening. The other 8 images show the corresponding coarse fields (16×16 pixels) using 8-fold coarse-coding.

4. HYBRID CLASSIFICATION

4.1. Neural Network Training

To perform image classification the coarse subimages of the nuclei are trained each by a neural network separately. A multilayer perceptron⁶ with one hidden layer, two output units to represent the two classes adenocarcinoma and mesothelioma, and RPROP⁷ as a learning function was chosen. Due to the smaller input images all networks together are able to learn their task much faster than a single network trained by the original nucleus images. For example using 8-fold coarse-coding described in the previous section, multilayer perceptrons with 5 hidden units 16×16 (coarse pixels) \times 8 (coarse fields) \times 5 (hidden units) = 10240 (synapses) have to be modified for a single training cycle of a nucleus instead of 128×128 (pixel) \times 5 (hidden units) = 81920 (synapses) of the original image. Furthermore networks with many input units need more training cycles to converge.

Multilayer perceptrons are in principle able to perform every classification tasks.¹⁵⁻¹⁷ Training until all training samples are sufficiently learned is computational very expensive. During learning multilayer perceptrons develop their synapses to produce the desired output and therefore to discriminate the two classes. Synapses of a hidden unit to the input layer are formed to measure a discriminating feature by the hidden unit. They represent a pattern or an image presented in the input layer which would result a maximum output of the corresponding hidden unit. These patterns or feature-filter are produced by internal self-organization. The synapses of the output units are developed to find an optimal decision boundary by given (hidden) features. These two layers of synapses are modified during learning concurrently which results in slow convergence of the whole network. We limit the training time to a fixed number of training set presentations (epochs) which should be sufficient for the networks to develop pattern sensitive feature-filters in the first synaptic layers but is not long enough to develop optimal feature-layer and decision-layer combinations within the networks.

4.2. Neural Feature Extraction

After training the self-organized feature-filters of the networks are extracted by selecting the synapses of the units of the hidden layer. A neural feature of a nucleus is calculated by summing up the products of the synapses of a hidden unit and the corresponding coarse pixels. This vector product is similar to the neural activation of a hidden unit. Using 8-fold coarse coding and 5 hidden units per neural network 40 neural features and additionally 8×2 neural outputs can be produced which are forwarded to the discriminant analysis.

4.3. Discriminant Analysis

Perceptrons are able to generate only linear discriminant surfaces. In multilayer perceptrons with one hidden layer the hidden units perform a nonlinear transformation of the input patterns and the output units generate a linear discriminating surface within the transformed space. A quadratic classification algorithm offers a more complex

decision boundary.³ An input pattern X consisting of neural features and neural outputs is mapped to class $\hat{\omega}(X) \in \{\omega_1, \omega_2\}$ ($\omega_1 \equiv$ adenocarcinoma, $\omega_2 \equiv$ mesothelioma) if

$$\hat{\omega}(X) = \begin{cases} \omega_1 & : & - \frac{1}{2}(X - M_2)^T \Sigma_2^{-1} (X - M_2) > \frac{1}{2}(X - M_1)^T \Sigma_1^{-1} (X - M_1) > 0 \\ \omega_2 & : & - \frac{1}{2}(X - M_2)^T \Sigma_2^{-1} (X - M_2) < \frac{1}{2}(X - M_1)^T \Sigma_1^{-1} (X - M_1) < 0 \end{cases}$$

where Σ_i denotes the covariance matrix and M_i the mean vector of class ω_i , $i \in \{1, 2\}$. To set up the discriminating algorithm only the parameters Σ_i and M_i have to be estimated using the neural features and neural outputs of the training set. Additionally the class membership probabilities of each pattern can be calculated based on these parameters.

5. RESULTS

5.1. Methodical measurement of classification accuracy

Apart from dividing the segmented nuclei into training set and test set and measuring the classification accuracy for the training set and the generalization accuracy for the test set, both values for the classification of single nuclei and for the average of all nuclei of a patient have been determined. Using these four values one can estimate the significance of the tested classification system and its parameters for pattern recognition tasks of single objects or its application in medical diagnosis as a decision averaged over all nuclei of a patient.

However, classification accuracies determined by single finite training and test sets are stochastic variables and their significance is modest in real world applications. Therefore different training and test sets by different patient assemblies are produced. To get most out of the given data the accuracy estimation method leave-one-out is suggested which uses all patients except one mesothelioma and one adenocarcinoma for training and the left out for testing, permuted over all possible combinations.¹⁸ By doing that 16 different classifiers have to be trained and tested which is computationally extremely expensive. In most of the experiments 2-fold crossvalidation with two different training and test sets to set up two classifiers is used. The results of both training and test sets were averaged to estimate the accuracies of the experiment.

5.2. Experimental results

5.2.1. Variation of neural network parameter

The success of multilayer perceptron learning depends mostly on the network parameters. The networks have not to be trained until all patterns are learned but until useful feature filter developed within the hidden unit. First tests have shown that after 300 RPROP epochs, e.g. 300 presentations of all training nuclei, hybrid classification do not further improve their classification and generalization accuracies. In contrast a network alone would need approximately 2000 epochs until it has learned the nuclei. Based on this result all further experiments are run with 400 training epochs.

Another significant parameter in multilayer perceptron training is the number of hidden units. In conventional training using too few hidden units causes underfitting of the data, e.g. pattern remain which can not be learned because not enough connections exist to store the patterns. Using too much hidden units makes it possible to store all patterns in memory (connections) but causes overfitting of the data which results normally in poor generalization.

In Figure 3 the number of hidden units is varied from 1 to 15. Nuclei images were centered, aligned and flipped like described in Section 3. An 8-fold coarse-coding produces 8 16×16 coarse images which were trained by 8 multilayer perceptrons concurrently with a similar number of hidden units. After 400 epochs of RPROP training the self-organized neural feature filters were extracted and neural features calculated by processing the input images. These features and the neural network outputs were fed into the quadratic discriminant analysis and the class membership probabilities and the classification and generalization errors estimated.

Classification of patients as an average of all their nuclei results in better classification on training and test sets. The probabilities of misclassifications of a assembly of nuclei of patients is smaller than those of single nuclei.

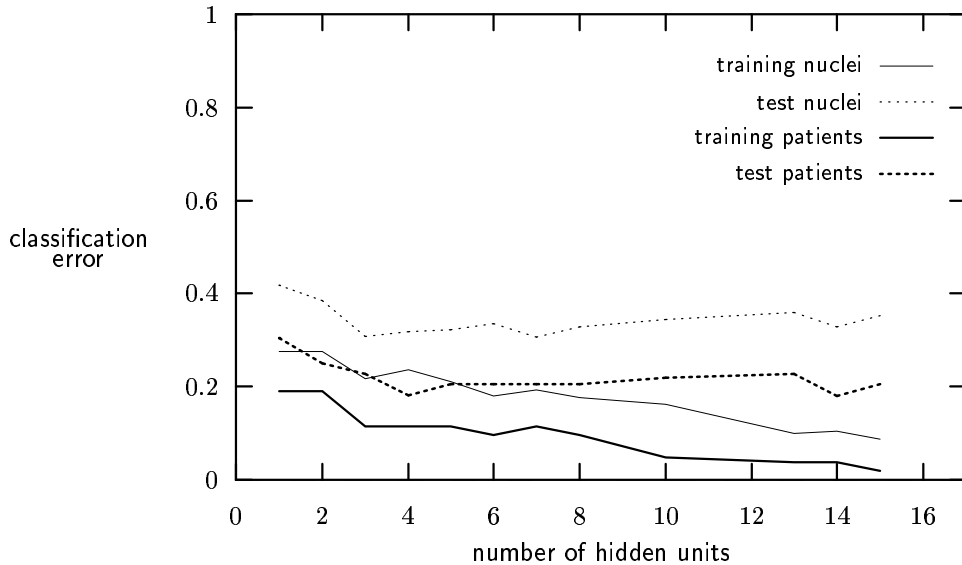


Figure 3. Influence of the number of hidden units. The images were coded using 8-fold coarse-coding, the network parameter were: multilayer perceptron, one hidden layer, RPROP, 400 epochs training. The classification success of the hybrid classification system of the training set nuclei, the average over the patients (13 mesothelioma and 13 adenocarcinoma), the test set nuclei and the corresponding average is shown.

Increasing the number of hidden units from 1 to 4 improves the classification and generalization abilities of the classifier significantly. Taking more hidden units only effects the classification error of the training set but does barely effect the test set errors. Figure 3 shows that the number of hidden units which is normally a critical parameter in multilayer perceptron learning between underfitting and overfitting has nearly no significance. The same behavior can be observed by varying the training time. The hybrid classification system is relatively independent of the neural network parameters.

Applying hybrid classification and training $2 \times 13 \times 50$ nuclei coded by 8-fold coarse coding requires in the experiment described above $1300 \text{ (nuclei)} \times 16 \times 16 \text{ (inputs)} \times 5 \text{ (hidden)} \times 8 \text{ (nets)} \times 400 \text{ (epochs)} = 5.3 \cdot 10^9$ (connection updates). SNNS¹⁹ performs $4.5 \cdot 10^5$ connection updates per second on a SUN SPARC-Station which results in approximately 3.5 hours computing time to train the multilayer perceptrons. If we used only a multilayer perceptron for classification we had to train 2000 epochs 128×128 pixel images which would require approximately 40 times more computing time.

5.2.2. Assembly of the training set

An important question in classification is whether the given dataset split up in training and test sets is large enough to perform good generalization by the classifier. A arbitrary chosen small training set of 'real world data' not containing the necessary information for discrimination will result in poor generalization on another arbitrary chosen test set. By varying size and assembly of training and test sets the possible generalization accuracies can be estimated.

Estimating the classification errors the training set size is varied from one adenocarcinoma and mesothelioma patient pair to 16 pairs. Training with one pair causes that all nuclei are perfectly learned but leads to bad generalization. Increasing the training set increases the training error and decreases the generalization (test set) error. Assuming the training sets could further increased the errors of nuclei classification would meet at a rough guess at 0.25 and the patient classification errors at 0.18. These errors should be viewed as lower limits. Generalization to classify unknown trained samples can never outdo the classification of already known samples. Therefore this limits give us a guess of the possible classification abilities of tumor nuclei using this hybrid classification system.

This experiment was repeated using nuclei patterns generated by 4-fold coarse-coding which produces 4 images of the resolution 32×32 of a original nucleus image. The results resemble to those described above using 8-fold coarse-coding. The errors of nuclei classification would meet approximately at 0.22 and the patient classification

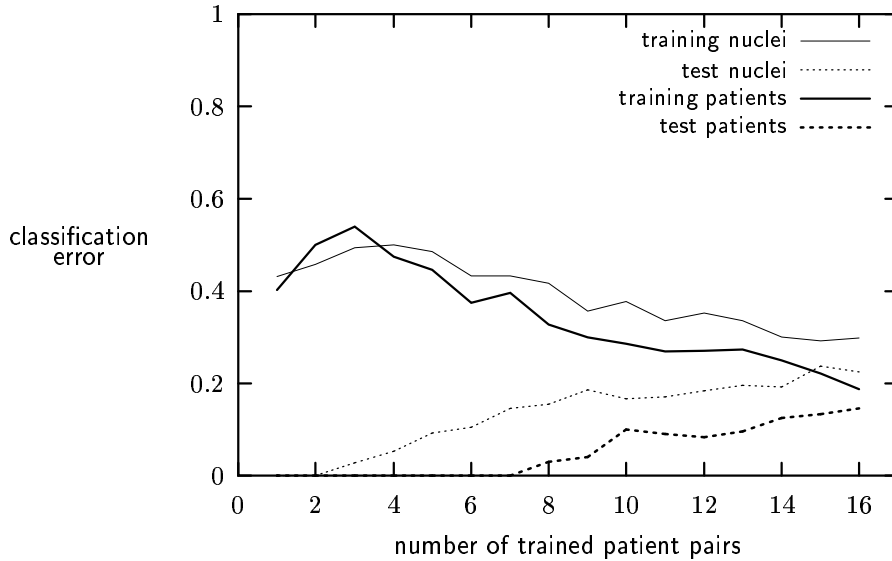


Figure 4. Variation of the size of the training set measured as adenocarcinoma and mesothelioma patient pairs. The images were coded using 8-fold coarse-coding, the network parameter were: RPROP, 5 hidden units per network, 400 epochs training.

errors at 0.16. These small improvements of the estimated classification accuracies do not justify the double training time needed because of the enlarged input patterns.

5.2.3. Discriminating feature filters

Using the hybrid classification system there is no need of feature extraction by using a priori information of experts. Nevertheless we are interested in which information the system uses for discriminating adenocarcinoma and mesothelioma patients.

All neural features of all hybrid classifiers with 1 to 15 hidden units were collected and a stepwise feature selection²⁰ performed. In Figure 5 the seven most significant feature filters are shown. These correspond to patterns activating the hidden units a lot. Interpreting Figure 5 from left to right the filters are sensitive for small nuclei, large nuclei, structure of the shells, form of the nuclei, internal chromatin structure and elongated nuclei. For the filter on the right side we could not imagine an interpretation. These filters give us hints how differences between adenocarcinoma and mesothelioma nuclei could be found in microscope images.

6. CONCLUSIONS

This new hybrid image classification system uses neural networks as feature extractors. By learning important features in multilayer perceptrons no expensive and time consuming feature extraction by experts using a priori information is necessary. Other features assembled by different methods can be easily added to the discriminant

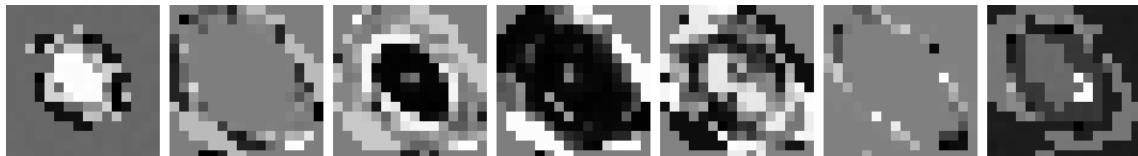


Figure 5. Self-organized filter during learning in the neural networks. The seven most discriminating feature filters found by stepwise feature selection are shown from left to right. A light pixel stands for a negative connection and a dark one for a positive one. Discriminating features found in the filters are size, shape and internal structure of the nuclei.

analysis to improve the classification accuracy. The training time of multilayer perceptrons could be decreased to a small fraction by preprocessing the images using coarse-coding and aborting training after feature filter emerge within the connections of the hidden units. Furthermore the classification success of the whole system is nearly independent of neural network parameter. Because of concurrent training of several multilayer perceptrons there is no danger for the entire classification success if a network steps into a local minimum during learning.

Still the training phase is the time consuming phase which can be done off-line. After training the extracted neural features can be fast calculated as a vector product of neural feature filters and coarse-coded images. The implementation of the quadratic classification algorithm is performed by vector products, too. These can be easily implemented in hardware²¹ to process many images in an on-line image classification process.

The methodology of hybrid classification using neural network feature extraction is applicable to many image and pattern recognition problems, e.g. detection of microcalcifications in X-ray mammograms. Neural networks are able to learn object shape as well as texture parameters.

Applied to images of adenocarcinoma and mesothelioma nuclei it could be shown that a visual discrimination of these images is possible. Shape and chromatin structure are important properties of different nuclei. This results may improve the certainty of diagnostic decisions combined with other histopathological and histochemical methods.

ACKNOWLEDGMENTS

Many thanks to Peter Bartels, Deborah Thompson, Hubert Bartels (University of Arizona, Optical Sciences Center, Tucson, AZ, USA) and Reinhard Männer (University of Mannheim, Computer Science 5, Mannheim, Germany) for provided material, scientific advice and fruitful discussions.

REFERENCES

1. R. Stotzka, T. Müller, W. Eppler, and H. Gemmeke, "A low-cost computer assisted mammography workstation," in *International Workshop on Digital Mammography 98*, 1998.
2. R. Stotzka, T. Müller, W. Eppler, and H. Gemmeke, "Three-dimensional reconstruction of clustered microcalcifications from two digitized mammograms," in *SPIE's Internl. Symposium Medical Imaging 1998*, 1998.
3. K. Fukunaga, *Statistical Pattern Recognition*, Academic Press, Harcourt Brace Jovanovich, Boston, 1990.
4. A. Fisher, *The Mathematical Theory of Probabilities*, vol. 1, McMillan, New York, 1923.
5. J. Schürmann, *Polynomklassifikatoren für die Zeichenerkennung*, R. Oldenbourg Verlag, München, Wien, 1977.
6. D. Rumelhard and J. McClelland, *Parallel Distributed Processing*, The MIT Press, Cambridge, MA, 1986.
7. M. Riedmiller and H. Braun, "RPROP – a fast adaptive learning algorithm," technical report, University of Karlsruhe, Karlsruhe, 1992.
8. R. Stotzka, R. Männer, and P. H. Bartels, "Neural network grading of prostate tumors," in *Neural Networks: From Biology to High Energy Physics*, O. Benhar, C. Bosio, P. Del Giudice, and E. Tabet, eds., pp. 323–333, 1991.
9. R. Stotzka, R. Männer, P. Bartels, and D. Thompson, "A hybrid neural and statistical classifier system for histopathologic grading of prostate lesions," *Analytical and Quantitative Cytology and Histology*, 1995.
10. U. Riede and H. Wehner, *Allgemeine und spezielle Pathologie*, Georg Thieme Verlag, 1986.
11. H. Battifora and M. Kopinski, "Distinction of mesothelioma from adenocarcinoma," *Cancer* **55**, pp. 1679–1685, 1985.
12. P. H. Bartels, M. Bibbo, A. Graham, S. Paplanus, R. Shoemaker, and D. Thompson, "Image understanding system for diagnostic histopathology," *Analyt. Cellular Pathol.* **1**, pp. 195–214, 1989.
13. J. Hinton, J. McClelland, and D. Rumelhart, "Distributed representations," in *Parallel Distributed Processing*, MIT Press, Cambridge, MA, 1986.
14. L. Spirkovska and M. Reid, "Coarse-coded higher-order neural networks for PSRI object recognition," *IEEE Transactions on Neural Networks* **4**, pp. 276–283, March 1993.
15. B. Ripley, "Statistical aspects of neural networks," in *Networks and Chaos – Statistical and Probabilistic Aspects*, E. Barndorff-Nielsen, D. Cox, J. Jensen, and W. Kendall, eds., Chapman and Hall, 1993.
16. J. Hertz, A. Krogh, and R. Palmer, *Introduction to the Theory of Neural Computing*, vol. 1 of *Santa Fe Institute, Studies in the Sciences of Complexity: Lecture Notes*, Addison-Westley Pub., Redwood City, 1991.

17. K. Hornik, M. Stinchcombe, and H. White, "Multilayer feedforward networks are universal approximators," *Neural Networks* **2**, pp. 359–366, 1989.
18. P. Lachenbruch and M. Mickey, "Estimation of error rates in discriminant analysis," *Technometrics* **10**, pp. 1–11, 1968.
19. A. Zell and et. al., *SNNS: Stuttgart Neural Network Simulator*. Institute for Parallel and Distributed High Performance Systems, University of Stuttgart, Stuttgart, 3.2 ed., 1994.
20. R. Jennrich, "Stepwise discriminant analysis," in *Statistical Methods for Digital Computers*, K. Enslein, A. Ralston, and H. Wilf, eds., pp. 76–96, John Wiley & Sons, Inc., 1977.
21. W. Eppler, T. Fischer, H. Gemmeke, R. Stotzka, and T. Köder, "Neural chip SAND/1 for real time pattern classification," in *IEEE Real Time '97*, (Beaune), 1997.