ELSEVIER Artificial Intelligence in Medicine 16 (1999) 283–297

Artificial Intelligence in Medicine

Feature selection for optimized skin tumor recognition using genetic algorithms

H. Handels [a,*], Th. Roß [a], J. Kreusch [b], H.H. Wolff [b], S.J. Pöppl [a]

[a] *Institute for Medical Informatics, Medical University of Lübeck, Ratzeburger Allee 160, 23538 Lübeck, Germany*

[b] *Department of Dermatology, Medical University of Lübeck, Ratzeburger Allee 160, 23538 Lübeck, Germany*

Received 4 April 1998; received in revised form 22 July 1998; accepted 4 September 1998

Abstract

In this paper, a new approach to computer supported diagnosis of skin tumors in dermatology is presented. High resolution skin surface profiles are analyzed to recognize malignant melanomas and nevocytic nevi (moles), automatically. In the first step, several types of features are extracted by 2D image analysis methods characterizing the structure of skin surface profiles: texture features based on cooccurrence matrices, Fourier features and fractal features. Then, feature selection algorithms are applied to determine suitable feature subsets for the recognition process. Feature selection is described as an optimization problem and several approaches including heuristic strategies, greedy and genetic algorithms are compared. As quality measure for feature subsets, the classification rate of the nearest neighbor classifier computed with the leaving-one-out method is used. Genetic algorithms show the best results. Finally, neural networks with error back-propagation as learning paradigm are trained using the selected feature sets. Different network topologies, learning parameters and pruning algorithms are investigated to optimize the classification performance of the neural classifiers. With the optimized recognition system a classification performance of 97.7% is achieved.

Keywords: Feature selection; Genetic algorithms; Melanoma; Artificial neural networks

* Corresponding author. Tel.: +49-451-5006621; fax.: +49-451-5006601.
E-mail address: handels@medinf.mu-luebeck.de (H. Handels)

0933-3657/99/$ - see front matter
PII: S0933-3657(99)00005-6

1. Introduction

The development of computer supported systems for melanoma diagnosis is of increasing importance: on the one hand, clinical accuracy of dermatologists in identifying malignant melanomas is only between 65 [2] and 85% [17]. On the other hand, one is faced with a world-wide increase of the incidence of malignant melanoma [10]. Several approaches to computer supported melanoma diagnosis are based on color images making use of image analysis methods to quantify visual features as described by the 'ABCD'-rule (Asymmetry, irregular Border, varying Color, Diameter) [14,16,39].

Laser profilometry opens up new possibilities to improve tumor diagnostics in dermatology [18,42]. With a laser profilometer, skin surfaces with an area of 4×4 mm^2 are taken at a resolution of 125 sample points per mm and a vertical resolution of 0.1 μm. Since surface sampling at this high resolution takes several hours, skin surface structures are preserved as silicone rubber imprints. The superior reflection characteristic of the silicon rubber is another reason to use imprints. Taking 125 samples per mm with a measuring interval of 8 μm results in a profile with 500×500 sample points, finally. Profiles are displayed using lambert shading technique [12] to visualize even fine surface structures of the lesions (Fig. 1).

The recognition task is to classify surface profiles of melanomas and nevi, also called moles. In a clinical study, surface profiles of 19 superficial spreading melanomas from 19 patients and 25 nevocytic nevi from 23 patients were analyzed. Most of the examined nevi were clinically atypical [5,27]. All diagnoses were confirmed histologically by at least two investigators. Due to the fact that all profiles contain regions with a structure similar to normal skin (like e.g. seen in Fig. 1, upper right), each profile is subdivided into 16 non overlapping quadratic sub-profiles and image analysis algorithms are applied to each sub-profile separately.

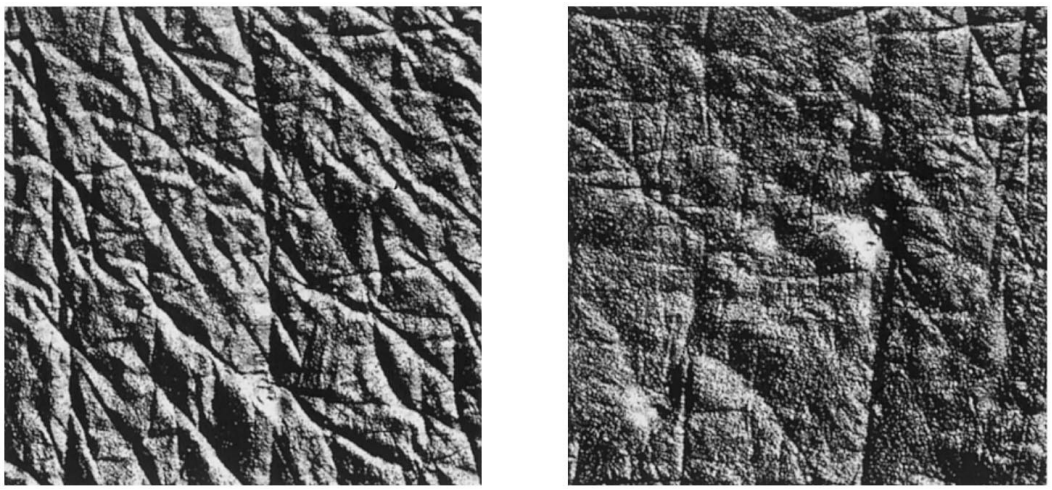

Fig. 1. Nevus profile (left) and melanoma profile (right) visualized using lambert shading.

In the next step, feature selection algorithms are applied to optimize the classification performance of the recognition system. This is motivated by the occurrence of a peaking phenomenon that is often observed in pattern recognition systems when finite sets of training samples are used to build classifiers [6,22,23]. Although theoretically the introduction of additional features can never decrease the classification rate (= 1-error rate) of the optimal Bayes classifier, adding new features to a feature set often leads to a decreasing classification rate of a classifier as trained on a limited set of samples. Genetic algorithms as well as heuristic strategies and greedy algorithms are used for feature selection. Finally, several architectures of feed forward neural networks trained with error back-propagation are evaluated for the recognition task.

2. Feature extraction

Three types of features are extracted to describe profile structures quantitatively: Texture parameters based on cooccurrence matrices as proposed in [19], features calculated from the Fourier power spectrum of the profiles as described in [3] and fractal features [11,16,31].

The application of texture analysis methods based on cooccurrence matrices [19] is motivated by the observation that textures can be characterized by the cooccurrence of its gray values. The cooccurrence matrix $S_{d,\theta}$ of a profile f is calculated as follows: A matrix element $S_{d,\theta}(i,j)$ is given as the relative frequency of cooccurrence of gray levels g_i and g_j in the spatial neighborhood described by displacement vector (d, Θ) with direction Θ and distance d and

$$f(x_1, y_1) = g_i \wedge f(x_2, y_2) = g_j \text{ with}$$
$$(x_2, y_2) = (x_1 + d\cos\theta, y_1 d\sin\theta) \quad (1)$$

Skin profile altitude values range between -869 μm and $+381.65$ μm with reference to a zero line. To calculate cooccurrence matrix entries, altitude values are normalized into 64 intervals with equal interval lengths. To assure rotation invariance of the texture features, cooccurrence matrices $S_{d,\theta k}$ for eight discrete angles $\theta_k = k\pi/4$, $k = 1,\ldots, 8$ are computed and a rotation invariant cooccurrence matrix is generated as follows [15]:

$$\bar{S}_d(i,j) = \frac{1}{8}\sum_{k=0}^{7} S_{(d,\pi/4k)}(i,j). \quad (2)$$

Due to the observation that texture patterns may vary strongly in scale, several rotation invariant cooccurrence matrices with $d = 1$, 3, 6, 10 and 16 are computed. From each matrix, 13 texture features h_i^d $(i = 1,\ldots, 13)$ as proposed in [19] are extracted.

The second approach quantifies textural information of skin surface profiles in the Fourier power spectrum [3]. Let $F(u, v)$ denote the Fourier transform of the two-dimensional signal function $f(x, y)$ and its power spectrum given by $|F(u, v)|^2$.

Spatially periodic or directional changes of profile values lead to peaks in the amplitude of corresponding frequencies in its Fourier power spectrum. Circular Fourier features are given by:

$$c_{r_i, r_{i+1}} = \int_{r_i^2 \le u^2 + v^2 \le r_{i+1}^2} \int |F(u, v)|^2 \, du \, dv \tag{3}$$

In a discrete power spectrum of a center ordered Fourier transform feature $c_{r_i, r_{i+1}}$ is obtained by adding up pixels within a ring defined by two concentric circles with radii r_i and r_{i+1}, respectively. Patterns varying on a large scale will appear as peaks in rings with small radii (low spatial frequencies), whereas a grainy texture will result in high values for rings with larger radii (high spatial frequencies). We extract features $\rho_0, \ldots, \rho_6$ with $\rho_i = c_{r_i, r_{i+1}}$, $r_0 = 0$ and $r_i = i/7X$, where X is the number of image rows/columns. Additionally, these features are normalized according to

$$\rho_i^n = \frac{\rho_i}{\sum_{i=0}^{6} \rho_i} \tag{4}$$

The standard deviations of circular Fourier features are computed as well.

Fourier features measuring directional changes of textures are given by

$$a_{\theta_i, \theta_{i+1}} = \int_{\theta_i \le \tan^{-1} u/v \le \theta_{i+1}} \int |F(u, v)|^2 \, du \, dv \tag{5}$$

The interval $[\theta_i, \theta_{i+1}]$ defines an angular sector emanating from the center of the power spectrum. High values for occur $a_{\theta_i, \theta_{i+1}}$, if there is a texture with a direction between $\theta_i + \pi/2$ and $\theta_{i+1} + \pi/2$, respectively. The rubber imprints were always scanned perpendicular to the main furrow direction. In total, ten features $\varphi_0, \ldots, \varphi_9$ with $\varphi_i = a_{\theta_i, \theta_{i+1}}$ (according to Eq. (5)), $\theta_0 = 0$ and $\theta_i = i\,\pi/10$ are computed to characterize directional changes in the textures of skin profiles.

The third feature extraction method is based on concepts from fractal geometry [11,31,16]. Apart from self similarity, one salient aspect of fractal structures is the increase in detail when magnified. The loss of detail in melanoma profiles motivates the extraction of fractal features. Here, a skin profile is interpreted as a Brownian surface. An ideal Brownian surface S_H can be characterized by its scaling parameter H $(0 < H < 1)$ as well as by its fractal dimension $D = 3 - H$ [11,16,31]. Furthermore, a horizontal cut $S'_H = \{(x, y) \mid S_H(x, y) = c_0\}$ of an ideal Brownian surface S_H shows (in most cases) the fractal dimension $D = 2 - H$ [11,31]. Analysis of more than 1000 digitized Brownian surfaces with 512×512 pixels has shown: calculated and given fractal dimensions coincide if threshold c_0 is chosen close to the histogram maximum [34].

The computation of fractal profile features starts with the generation of a profile's histogram. Then, the histogram maximum $h_{\max}$ is computed and three binary images $B_l = [b_l(i, j)]$ are created according to

$$b_l(i, l) = 1, \; h_{\max} \le f(i, j) \le h_{\max} + l - 1$$
$$b_l(i, l) = 0, \quad \text{otherwise} \quad l = 1, 2, 3 \tag{6}$$

where $f(i,j)$ denotes the profile's altitude value at position (i,j). Additional interval lengths ($l = 2,\ 3$) are considered to compensate for digitization effects. Finally, the fractal grid dimension D_l is computed from each binary image B_l as follows: Divide image B_l into a grid with squares of size δ completely covering the image and count the number of squares $N_g(B_l, \delta)$ containing marked pixels. This procedure is repeated with different grid sizes δ_i ($i = 1,\ldots,\ N$). Then, a linear regression based on the points $(-\log(\delta_i), \log(N_g(B_l, \delta_i)),\ i = 1,\ldots,\ N$ is performed and the fractal grid dimension D_l is computed as the slope of the regression line [11,16,31].

3. Feature selection

Feature selection is an essential step in order to optimize pattern recognition systems in practice [6,22–26,29,32,38]. The use of feature selection algorithms is mainly motivated by a peaking phenomenon often observed when classifiers are trained with a limited set of training samples: If the number of features is increased, the classification rate of the classifier decreases after a peak [6,22,23]. Furthermore, the need to reduce the number of features when designing a recognition system is motivated by computational reasons.

Feature selection can be modeled as an optimization problem: Select an optimal feature subset $F_{\text{opt}} \subseteq F = \{f_1,\ldots,f_n\}$ in order to maximize the classification performance of the recognition system. Since the number of feature subsets increases exponentially with the number of features (2^n-1 subsets), for many practical recognition problems it is impossible to consider all subsets with a brute force algorithm and select the best rated subset. Furthermore, $\Sigma_{i=1}^{m}\binom{n}{i}$ subsets with $i = 1,\ldots,\ m$ elements need to be examined, if the best feature subset with a maximal size m has to be selected out of n features. In our application, 94 features were extracted from each sub-profile by image analysis methods. Hence, for a maximal feature set size of, for example $m = 15$, approximately 32×10^{15} feature combinations have to be tested. Different optimization algorithms have been developed for feature selection in practical applications [24–26,29,32,38]. A survey of heuristic strategies determining good feature subsets is given in [29]. Furthermore, special approaches to feature evaluation and selection are described in [9,24–26,32,38].

Genetic algorithms, a heuristic feature selection strategy as well as greedy algorithms were implemented and compared to determine suitable feature subsets for our classification task. The quality of a feature subset is measured by the classification rate Q of the nearest neighbor classifier. Since the Euclidean distance used for the nearest neighbor classification is not invariant to changes in scale, each feature is normalized to a zero mean and unit variance in a pre-processing step. For a given feature subset, the classification rate is computed using the leaving-one-out method. After profile splitting leaving one profile out means: all 16 sub-profiles of a parent profile are removed from the training set. A parent profile is classified to one class, if more than 50% of its sub-profiles are assigned to this class (majority decision).

3.1. Genetic algorithms

Genetic algorithms are stochastic optimization methods which have been successfully applied in the field of discrete optimization [8,13,20]. They have been used for image segmentation [1], pattern recognition [4,40] as well as for the determination of sub-optimal solutions for NP-complete optimization problems [8,13,41].

In a genetic algorithm, an initial population of chromosomes (i.e. parameters of the function to optimize, feature subsets in our application) is iteratively altered by mechanisms inspired by natural evolution such as selection, recombination and mutation. A population represents a pool of possible solutions of the optimization problem. During the genetic optimization process, new chromosomes are iteratively derived from previous populations. The fitness of a chromosome describes how suitable the chromosome is for the solution of the problem considered. As in natural evolution, survival of the fittest is the main strategy. The aim of the genetic search is to find optimal or sub-optimal solutions corresponding to chromosomes with high fitness values.

In this paper, genetic algorithms are applied to feature selection. At first, basic mechanisms as well as two problem specific components in a genetic algorithm, the problem encoding and the fitness function, are described. A straight forward approach to represent subsets of a given set with n elements in a binary chromosome $C = c_1,c_2,\ldots,c_n$ is to index its elements and associate a position in the chromosome with the element's index. Thus, if feature f_i is included in set F_c, component c_i of the corresponding chromosome C is 1, otherwise $c_i = 0$. The fitness of chromosome C can be defined as the classification rate $Q(F_C)$ of its corresponding feature set $F_C \subseteq \{f_1,\ldots,f_n\}$.

After calculating the fitness of each chromosome in a population, a stochastic selection procedure determines the chromosomes that take part in the reproduction process. Two selection methods were investigated to optimize genetic algorithms for feature selection: stochastic universal sampling and ranking.

3.1.1. Stochastic universal sampling

Chromosome C_k is assigned a reproduction probability p_{select_k} according to

$$p_{\text{select}_k} = \frac{\text{fit}_k}{\sum_{i=1}^{p} \text{fit}_i} \tag{7}$$

p denotes the number of chromosomes in the population. It is well-known [13] that the use of this selection strategy may result in premature convergence to a local minimum. The ranking method is used to avoid this effect.

3.1.2. Ranking

In this method, chromosomes of a population are sorted according to their fitness values. Then, a reproduction probability p_{select} is computed as follows

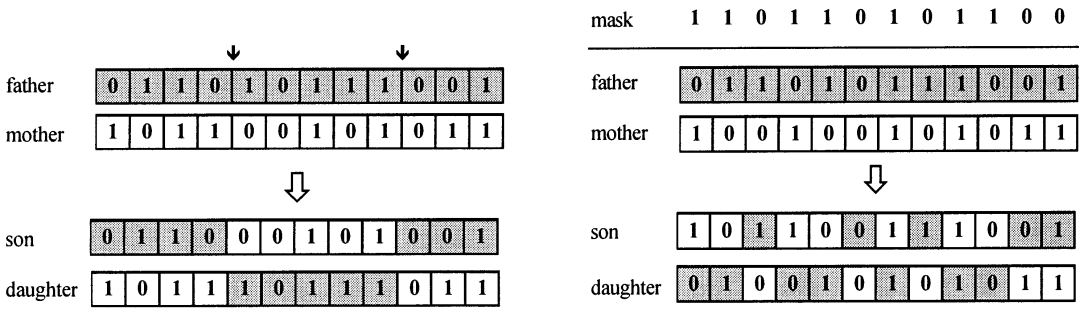

Fig. 2. Two-point crossover (left): The combination of the parents' chromosomes is directed by two crossover positions marked with arrows. The crossover positions are determined at random. Uniform crossover (right): The combination of the parents' chromosomes is directed by a binary mask. A bit in the mask is set with probability 0.5.

$$p_{\text{select}_k} = \frac{1}{p}\left(s - 2(s-1)\frac{rg(C_k)-1}{p-1}\right) \tag{8}$$

where p is the number of chromosomes in the population, $rg(C_k) \in \{1,\ldots,n\}$ denotes a chromosome's rank and $s \in [1,2]$ is a scaling parameter.

Recombination is the main mechanism to mix genetic information from different chromosomes. Hence in our application, chromosomes corresponding to feature subsets with high classification rates are combined with priority generating new feature subsets in the vicinity of suitable candidates in the search space. In our investigations, two-point crossover and uniform crossover are used as recombination mechanisms, alternatively (Fig. 2). While recombination of chromosomes by two-point crossover is guided by two crossover points, a binary mask is used with uniform crossover. Crossover points as well as entries of the binary masks are chosen at random. Crossover is a stochastic procedure that is performed with a predefined probability p_{cross}. After recombination, a new population is formed by substituting parents by their children.

Mutation, as in natural evolution, is a stochastic process where genes of a chromosome are altered with a small probability $p_{\text{mut}} \ll p_{\text{cross}}$. Bits of binary chromosomes are inverted randomly by the mutation procedure. Random changes of chromosomes (features sets) generate candidates for the solution of the optimization problem in different regions of the search space. Several runs with different values for the mutation probability $p_{\text{mut}} = 0.01,\ 0.001,\ 0.0001$ and the crossover probability $p_{\text{cross}} = 0.2,\ 0.5,\ 0.9$ were performed, each with a constant population size of 100 chromosomes. Parameters were chosen as suggested in [13].

3.2. Optimization of genetic algorithms for feature selection

Different strategies are applied in order to favor feature sets with a small number of features in the genetic optimization process. Firstly, the populations are initialized in a specific manner: The initial population is given by chromosomes representing all one element feature sets of the features considered in combination with randomly generated chromosomes. Secondly, the fitness definition is extended: If

$Q(F_C)$ denotes the classification rate assigned to feature set F_C, the fitness of chromosome C coding the feature subset F_C is defined as

$$\mathrm{fit}(C) = Q(F_C) + \frac{1}{k}\left(\frac{n - |F_C|}{n}\right) \tag{9}$$

where k is the sample size and n is the number of features. Thus it is guaranteed, that among feature sets with identical classification rates subsets with fewer elements receive higher fitness values. Furthermore, a chromosome with higher classification rate always receives a higher fitness value, because the smallest increment of the classification rate is $1/k$.

Genetic optimization processes were performed in four different populations with the ranking method and stochastic universal sampling as selection mechanisms as well as two-point and uniform crossover for recombination. The used elitist model guarantees that the chromosome with highest fitness value is always replicated in the next generation of chromosomes. Hence, the function of maximal fitness versus the number of generated chromosomes is a monotonous increasing function. In Fig. 3, the changes of mean and maximal fitness values in the populations are displayed during the optimization processes. Fig. 3 illustrates that the convergence of the genetic optimization processes is strongly influenced by the chosen selection mechanism. The optimization processes using stochastic universal sampling as selection procedure converge. This is shown by the mean fitness of a population which is near to the maximal fitness in a convergent state. By contrast, with the ranking method the mean fitness of the population decreases after a local maximum and the search process diverges. Table 1 summarizes the results obtained with different selection and recombination strategies.

Compared to the ranking procedure, stochastic universal sampling as selection mechanism leads to the generation of feature subsets with higher classification rates. Furthermore, these results are already obtained after 9000 or 9600 iterations, respectively, while more than 34 700 iterations are required if the ranking procedure is used. Since neighborhood relations of bits in a chromosome are not relevant to the feature selection problem, uniform crossover should be the preferred mechanism for recombination. This theoretical consideration is confirmed by simulation results (Table 1). In summary, genetic algorithms with stochastic universal sampling and uniform crossover show the best results for the genetic feature selection.

3.3. Comparison of genetic algorithms with heuristic strategies and greedy algorithms

Furthermore, heuristic strategies and greedy algorithms are used to select suitable feature subsets for the recognition system.

3.3.1. Heuristic strategy

Let $F = \{f_1, f_2, \ldots, f_n\}$ be the feature set with $Q(\{f_i\}) \geq Q(\{f_j\})$ for $i < j$. The heuristic algorithm selects the first m features with the highest values $Q(\{f_i\})$. The main disadvantage of this simple approach is that no combinations of features are considered in the selection process.

Greedy algorithms are fast algorithms to determine optimal or sub-optimal solutions of combinatorial optimization problems [21,30]. Two main variants of greedy algorithms can be distinguished for the feature selection problem.

3.3.2. Greedy algorithm I

In a first step, the best single rated feature f_1 is combined with all other features and the classification rate of these two element feature sets is calculated. The best rated two element subset is then combined with all other remaining features resulting in three element sets and so on. This search strategy is also called forward search. If there is more than one feature subset rated best, all these subsets were combined with all other features.

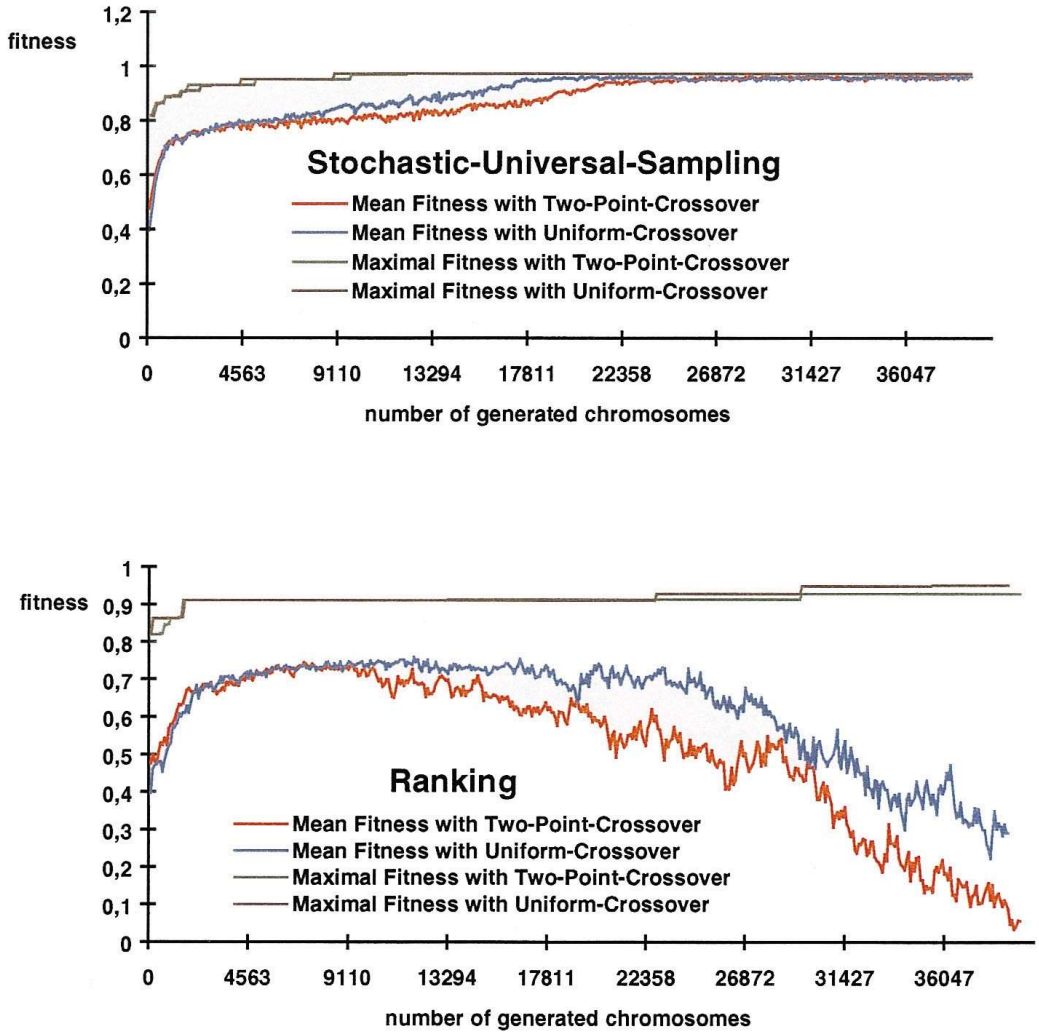

Fig. 3. Mean and maximal fitness in four populations with 100 chromosomes and different combinations of selection (stochastic universal sampling and ranking) and recombination mechanisms (two-point-crossover and uniform-crossover).

Table 1
Best classification rates and numbers of iterations needed to generate the fittest chromosome (feature set) during the genetic optimization process are given for different selection and recombination strategies: TPC: two-point crossover, UC: uniform crossover, SUS: stochastic universal sampling and ranking

Different variants of genetic algorithms				
	Ranking/TPC	Ranking/UC	SUS/TPC	SUS/UC
Best classification rate	0.93	0.95	0.97	0.97
Number of iterations	34 700	35 500	9600	9000

3.3.3. Greedy algorithm II

The second greedy algorithm starts with the complete feature set $F = \{f_1, f_2, ..., f_n\}$. In a first step, classification rates of all feature subsets with n-1 elements are computed and the feature subset with the highest classification rate is processed in the next iteration step. Based on the selected feature subset, all subsets with n-2 elements are generated and the feature subset with the highest classification rate is selected. The elimination process is repeated until the number of features in the subset is 1. Finally, all feature subsets generated during this process are considered and the subset with highest classification rate is selected. The search strategy implemented by greedy algorithm II is also denoted as backward search.

Compared to heuristic feature selection strategies and greedy algorithms, feature subsets determined by genetic algorithms show superior classification performance (Table 2). In total, 26 feature subsets with up to 12 features and a classification rate of 97.7% were determined by genetic algorithms. It was especially advantageous for the subsequent classification procedure that genetic algorithms could select high rated feature subsets with a small number of elements (Table 2). The selected set with the smallest number of features consisting of two Fourier and three texture features is given by $F_5 := \{\rho_1, \varphi_{10}, h_3^{16}, h_4^3, h_{13}^6\}$.

Table 2
Results of the feature selection process with different strategies. The table shows the best classification rates obtained with the selection strategy and the number of features in the selected subsets

Comparison of different feature selection algorithms		
	Best classification rate	Number of features
Heuristic algorithm	84.1%	5
Greedy algorithm I	95.5%	7
Greedy algorithm II	91.0%	12
Genetic algorithm	97.7%	5

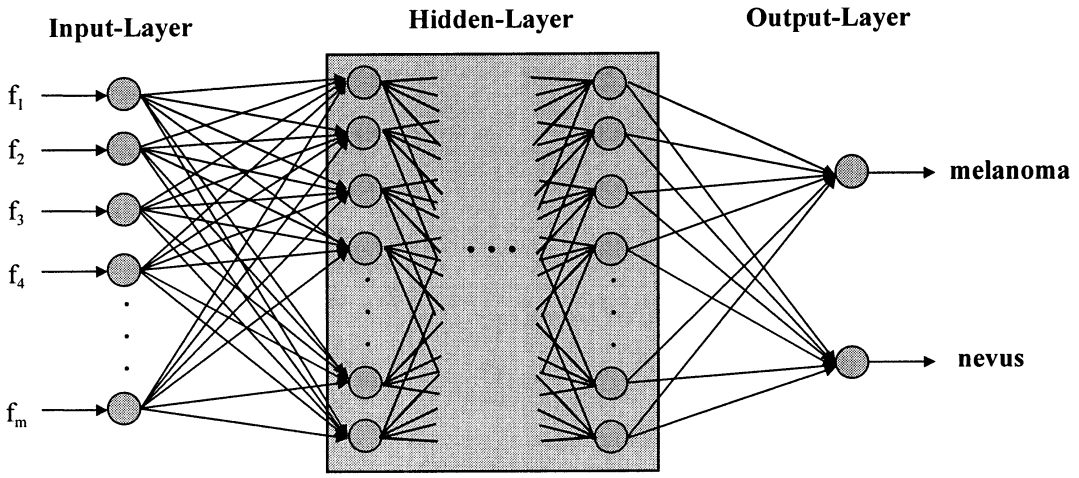

Fig. 4. Multilayer perceptron as a neural classifier: The number of input neurons is given by number features, the number of output neurons is defined by the number of classes. In contrast thereto, the number of hidden layers as well as the number of hidden neurons in each layer is not pre-defined by the classification task.

4. Back-propagation networks

Artificial neural networks are mathematical models inspired by the structure of natural neurons, the power and complexity of neural networks as well as the adaptive learning mechanisms in the human brain. They can be applied to function approximation, solution of optimization problems, simulation of biological processes as well as to pattern recognition problems. One important network type is the multilayer perceptron (MLP) with error back-propagation as learning paradigm [35,36]. Multilayer perceptrons have been employed successfully for numerous pattern recognition tasks, especially in the field of image processing [7,28,37].

In our application, back-propagation networks as illustrated in Fig. 4 are used to classify profiles of melanomas and nevi based on the selected feature subsets. After sub-profile analysis, a set of samples with 704 pre-classified feature vectors is available for network training. In a pre-processing step, each feature is normalized to a zero mean and unit variance2.

During the learning process, weights $w_{i,j} \in \mathrm{IR}$ of the connections between neurons are adjusted iteratively according to the back-propagation learning rule:

$$\begin{aligned} &\Delta w_{i,j} = \eta \delta_j o_i \quad \text{with} \\ &\delta_j = f'(\mathrm{net}_j)(t_j - o_j) \quad \text{if unit } j \text{ is an output unit} \\ &\delta_j = f'(\mathrm{net}_j)\textstyle\sum_k \delta_k w_{jk} \quad \text{if unit } j \text{ is a hidden unit} \end{aligned} \tag{10}$$

where f denotes the unit activation function, t_j the desired output and o_j the generated output of neuron j and η the learning rate. The learning algorithm realizes a parallel gradient descent method that minimizes the sum of squared

differences between actual and desired output values of the output neurons. Here, several networks were trained with learning rates $\eta = 0.1$, $\eta = 0.2$ and $\eta = 0.3$.

Classification of a sub-profile is performed using the winner-takes-all strategy: Let o_m and o_n denote the activation of output units corresponding to the classes melanoma and nevus (Fig. 4). The feature vector of a sub-profile is classified as melanoma, if $o_m > o_n$ and vice versa. A parent profile is assigned to one class, if more than 50% of its sub-profiles are assigned to this class (majority decision). All classification rates of neural classifiers were calculated with the leaving-one-out method. In this case, leaving one profile out means that its 16 sub-profiles were removed from the training set. The training of a network is continued until the number of correctly classified sub-profiles exceeds a given threshold value $n_{\min}$. Enforcing a correct classification of all feature vectors is avoided, because some sub-profiles of melanomas and nevi show similar structures. For a given network, the best value for the parameter $n_{\min}$ is determined experimentally.

Several fully connected networks with different topologies are investigated to optimize the network topology for the classification task. While the number of input neurons I is given by the number of features, the output units correspond to the classes melanoma and nevus (Fig. 4). The total number of weighted connections in a network is influenced by the sample size and chosen according to suggestions made in [22]. Hence, topologies with two hidden layers and 12 units per layer $\mathrm{MLP}_{I \times 12 \times 12 \times 2}$ as well as 15 units per layer $\mathrm{MLP}_{I \times 15 \times 15 \times 2}$ are applied, where the number of features I is limited to seven.

Alternatively, a weight pruning algorithm [33] is used to optimize the classification performance of the network resulting in dynamically changing network topologies during the learning process. Starting with a relatively large network, a normal learning process is performed. Then, the connection with the smallest weight in the network as well as isolated units without network input are deleted and the network is retrained. The pruning process is terminated, if one parent profile of the training set is classified wrongly after the removal of a connection and subsequent retraining.

5. Results and conclusions

Surface profiles of 19 superficial spreading melanomas and 25 nevocytic nevi with histologically confirmed diagnosis were analyzed. After division of each profile into 16 sub-profiles, 94 features were extracted from 704 sub-profiles using 2D-image analysis methods.

Feature selection was performed by adapting and optimizing genetic algorithms in order to obtain a classification system with high classification performance. Compared with heuristic feature selection strategies and greedy algorithms, feature subsets determined with genetic algorithms showed the best classification performance (Table 2). In total, 26 feature subsets with up to 12 features and a classification rate of 97.7% were determined by genetic algorithms. Furthermore, for the subsequent classification procedure it is of importance that genetic al-

gorithms selected high rated feature subsets with a small number of elements like $F_5 = \{\rho_1, \varphi_{10}, h_3^{16}, h_4^3, h_{13}^6\}$.

Several neural networks with varying parameters and topologies as described were trained to generate suitable neural classifiers. The classification rates of the neural classifiers were computed with the leaving-one-out method. Since the initialization of neural networks with random weights introduces an additional variability, several leaving-one-out runs were performed for each network topology and each feature set. With a static topology, network $\mathrm{MLP}_{5\times 15\times 15\times 2}$ and feature set F_5 performed best by achieving a classification rate of 90.9%. The fluctuations of classification rates in different runs are considerable: With the same neural network and the same feature set, the worst classification rate achieved in all runs was 77.3%. With the pruning algorithm, fluctuations are even more obvious: The best classification rate of 95.5% was achieved with feature set F_5, while in a second run with the same set and a different weight initialization a classification rate of 75% was obtained. A possible explanation for this effect could be that the neural learning processes converge in different local minima of the error function.

Nearest neighbor classification with the majority decision rule achieved a classification rate of 97.7%. The good classification performance of this comparatively simple classifier can be attributed to the feature selection pro cess, which optimizes the performance of the nearest neighbor classifier. The results show that the classification performance of the recognition system depends strongly on the features used in the classification process. Hence, feature selection is an essential step for the optimization of the developed pattern recognition system.

The presented approach to computer supported recognition of skin tumor profiles opens up new possibilities for the improvement of dermatological diagnosis in future. Further patients have to be examined to increase the number of cases in the classes melanoma and nevus. Furthermore, a reduction of measurement time for the generation of high resolution profiles is important for a wide application of the methods in clinical practice. Therefore, measurement techniques for the accelerated generation of high resolution surface profiles are considered in the next step to make surface profiles available in a fast and easy way.

References

[1] Andrey P, Tarroux P. Unsupervised image segmentation using a distributed genetic algorithm. Pattern Recogn 1994;27:659–73.
[2] Balch CM, Milton GW. Hautmelanome. Berlin: Springer, 1988.
[3] Ballard DH, Brown MB. Computer vision. Englewood Cliffs, NJ: Prentice-Hall, 1982.
[4] Bandyopadhyay S, Murthy CA, Pal KS. Pattern classification with genetic algorithms. Pattern Recogn Lett 1995;18.
[5] Busche H. Vergleichende Untersuchung der quantitativen Oberflächentopographie superfiziell spreitzender Melanome unnd nävozellulärer Nävi, master thesis, Medizinische Universität zu Lübeck, 1994.
[6] Campenhout JMv. On the peaking of the hughes mean recognition accuracy—the resolution of an apparent paradox. IEEE Trans Syst Man Cybern 1978;8:390–5.

[7] Carpenter GA, Grossberg S. Neural networks for vision and image processing. London: MIT Press, 1992 A Bradford Book.
[8] Davis L. Handbook of genetic algorithms. New York: Van Nostrand Reinhold, 1991.
[9] Egmont-Petersen M, Talmon JL, Hasman A, Ambergen AW. Assessing the importance of features for multi-layer perceptrons. Neural Net 1998;11:623–35.
[10] Elwood JM, Koh HK. Etiology, epidemiology, risk factors and public health issues of melanoma. Curr Opin Oncol 1994;6:179.
[11] Falconer KJ. Fractal geometry. Mathematical foundations and applications. Chichester: Wiley, 1990.
[12] Foley J, van Dam A, Feiner S, Hughes J. Computer graphics: principles and practice. Reading, MA: Addison Wesley, 1990.
[13] Goldberg DE. Genetic algorithms in search, optimization and machine learning. Reading MA: Addison Wesley, 1989.
[14] Golston JE, Stoecker WV, Moss RH, Dhillon IPS. Automatic detection of irregular borders in melanoma and other skin tumors. Comput Med Imag Graphics 1992;16:163–77.
[15] Gotlieb CC, Kreyszig HE. Texture descriptors based on co-occurrence matrices. Comput Vis Graphics Image Proc 1990;51:70–86.
[16] Green A, Martin N, Pfitzner J, O'Rourke M, Knight N. Computer image analysis in the diagnosis of melanoma. J Am Acad Dermatol 1994;31:958–64.
[17] Grin CM, Kopf AW, Welkovich B. Accuracy in the clinical diagnosis of malignant melanoma. Arch Dermatol 1990;126:763.
[18] Handels H, Roß T, Kreusch J, Wolf HH, Pöppl SJ. Image analysis and pattern recognition to support skin tumor diagnosis. In: Cesnik B, McCray AT, Scherrer JR, editors. Proceedings of the Ninth World Congress on Medical Informatics. Seoul (South Korea): IOS Press, Amsterdam, 1998:1056–1062.
[19] Haralick RM, Shanmugam K, Dinstein I. Textural features for image classification. IEEE Trans Syst Man Cybern 1973;3:610–21.
[20] Holland JH. Adaption in natural and artificial systems. Ann Arbor, MI: University of Michigan Press, 1975.
[21] Horowitz E, Sahni S. Algorithmen. Berlin: Springer, 1981.
[22] Jain AK. Advances in statistical pattern recognition. In: Devijer PA, Kittler J, editors. Pattern recognition, theory and applications. Berlin: Springer, 1986.
[23] Jain AK, Waller WG. On the optimal number of features in the classification of multivariate gaussian data. Pattern Recogn 1978;10:365–74.
[24] Jelonek J, Stefanowski J. Feature subset selection for classification of histological images. Artif Intell Med 1997;9:227–40.
[25] Kanal L, Chandrasekaran B. On dimensionality and sample size in statistical pattern recognition. Pattern Recogn 1971;3:225–34.
[26] Kohavi R, Sommerfield D. Feature subset selection using the wrapper method: overfitting and dynamic search space topology. Proceedings of the First International Conference Knowledge Discovery Data Mining, Montreal (Quebec, Canada), 1995.
[27] Kreusch J, Busche H, Connemann BJ, Roß T, Handels H, Wolff HH, Pöppl SJ. Differentiation between nevocellular nevi and malignant melanoma by skin surface parameters. In: Wilhelm KP, Elsner P, Beradesca E, Maibach H, editors. Bioengineering of the skin: skin surface imaging and analysis. Boca Raton, FL: CRC Press, 1997:289–300.
[28] Miller AS, Blott BH, Harmes TK. Review of neural network applications in medical imaging and signal processing. Med Biol Eng Comput 1992;30:449–64.
[29] Niemann H. Klassifikation von Mustern. Berlin: Springer, 1983.
[30] Papadimitrio CH, Steiglitz K. Combinatorial optimization. Englewood Cliffs, NJ: Prentice Hall, 1982.
[31] Peitgen H-O, Saupe D. The science of fractal images. Berlin: Springer, 1988.
[32] Raudys S, Jain AK. Small sample size problems in designing artificial nerual networks. In: Sethi IK, Jain AK, editors. Artificial neural networks and statistical pattern recognition. Amsterdam: North-Holland, 1991:33–50.

[33] Reed R. Pruning algorithms—a survey. IEEE Trans Neural Net 1993;4–5:740–7.
[34] Roß T. Analyse und Klassifikation zweidimensionaler Signale-Anwendung in der Erkennung von Hauttumoren anhand hochaufgelöster Oberflächenprofile, Ph.D. thesis, Shaker, Aachen, 1997.
[35] Rumelhart DE, Hinton GE, Williams RJ. Learning internal representations by error propagation. In: Rumelhart DE, McClelland JL, editors. Parallel distributed processing. Exploration in the microstructures of cognition. Cambridge, MA: MIT Press, 1986.
[36] Rumelhart DE, Hinton GE, Williams RJ. Learning representations by back-propagation errors. Nature 1986;323:533–6.
[37] Schürmann J. Pattern classification. New York: Wiley, 1996.
[38] Siedlecki W, Sklansky J. A note on genetic algorithms for large-scale feature selection. Pattern Recogn Lett 1989;10:335–47.
[39] Sober AJ, Burstein JM. Computerized digital image analysis: an aid for melanoma diagnosis—preliminary investigations and brief review. J Dermatol 1994;21:885–90.
[40] Srikanth R, George N, Warsi D, Prabhu D, Petry FE, Buckles BP. A variable-length genetic algorithm for clustering and classification. Pattern Recogn Lett 1995;16:789–800.
[41] Whitley D, Starkweather T, Shaner D. The traveling salesman and sequence scheduling: quality solutions using genetic edge recombination. In: Davis L, editor. Handbook of genetic algorithms. New York: Van Nostrand Reinhold, 1991.
[42] Wilhelm KP, Elsner P, Beradesca E, Maibach H. Bioengineering of the skin: skin surface imaging and analysis. Boca Raton, FL: CRC Press, 1997.