Detection of skin cancer by classification of Raman spectra

Sigurdur Sigurdsson, Peter Alshede Philipsen, Lars Kai Hansen
Jan Larsen, *Senior member, IEEE*, Monika Gniadecka and Hans Christian Wulf

Abstract— Skin lesion classification based on *in vitro* Raman spectroscopy is approached using a non-linear neural network classifier. The classification framework is probabilistic and highly automated. The framework includes feature extraction for Raman spectra and a fully adaptive and robust feed-forward neural network classifier. Moreover, classification rules learned by the neural network may be extracted and evaluated for reproducibility, making it possible to explain the class assignment. The classification performance for the present data set, involving 222 cases and 5 lesion types, was 80.5%±5.3% correct classification of malignant melanoma, which is similar to that of trained dermatologists based on visual inspection. The skin cancer basal cell carcinoma has a classification rate of 95.8%±2.7%, which is excellent. The overall classification rate of skin lesions is 94.8%±3.0%. Spectral regions, which are important for network classification, are demonstrated to reproduce. Small distinctive bands in the spectrum, corresponding to specific lipids and proteins, are shown to hold the discriminating information which the network uses to diagnose skin lesions.

I. Introduction

SKIN cancer is one of the most common cancers in humans and its incidence is increasing dramatically [1]. New incidences of the lethal skin cancer malignant melanoma in Denmark has increased 5- to 6-fold from 1942 to 1982 and the mortality rate has been doubled from 1955 to 1982 [2]. Currently, approximately 800 cases of malignant melanoma are reported in Denmark every year (approx. 15 per 100,000). In Germany $9000 - 10000$ new cases are expected every year (approx. 13 per 100,000) with an annual increase of $5 - 10\%$ [3]. Basal cell carcinoma (BCC) is the most common of skin tumors and is mainly considered to be provoked by ultraviolet radiation and does not metastasize. In contrast, malignant melanoma (MM) can metastasize rapidly. This cancer is also considered to be provoked by ultraviolet radiation, most probably by repeated high doses resulting in heavily burned skin.

Diagnosing malignant skin tumors is difficult as many common benign skin lesions resemble them visually, e.g. pigmented nevi and seborrhoeic keratosis. The diagnosis of BCC is approximately 65% for practicing dermatologists [4], while trained dermatologists diagnose MM clinically 63% [5]

to 75% [6] correctly. Dermatologists with less than one year experience diagnose only 31% of MM correctly [5].

Neural network diagnosis of skin tumors has previously been applied by classifying extracted features from digitized dermoscopy images of lesions [7], [8], [9]. The extracted features are based on geometry, colors and texture of the lesions, involving complex image processing techniques. With Raman spectroscopy the molecular structure of skin lesions can be exploited. Neural network analysis of Raman spectra for automatic classification of skin cancer is considered a promising tool for diagnosis of skin cancer, see e.g. [10], [11]. The network provides an objective method for skin lesion diagnosis, which could help both dermatologist and non-dermatologist to diagnose an early stage of skin cancer. However, difficulty in obtaining robustness and the lack of physical interpretation of neural network parameters have been considered a major drawback [12].

We suggest a framework for applying neural networks for skin cancer diagnosis from Raman spectra. The framework contains three main parts; a preprocessing stage for feature extraction from Raman spectra, a robust optimization scheme for training of feed-forward neural network classifiers, and finally a visualization method for interpretation of neural network results.

The main contributions of the paper are: The implementation of a highly automated artificial neural network framework, statistical evaluation of neural network visualization and a solution to an important medical application.

II. Raman Spectroscopy

Raman spectra are obtained by pointing a laser beam at a sample. The laser beam excites molecules in the sample and a scattering effect is observed. Inelastic scattering results in a frequency shift in the reflected Raman spectra. These frequency shifts are functions of the type of molecules in the sample, thus the Raman spectra holds useful information on the different chemical compounds.

Raman spectroscopy provides information about the molecular structure in the sample. The near-infrared Fourier transform (NIR-FT) Raman spectroscope has previously been investigated for medical diagnosis application in general [13], [14], as well as for the investigation of the skin [15], [16]. Previous studies on skin cancer [10], [17] indicate that this method has a potential for skin cancer diagnosis.

The Raman equipment used here for obtaining the spectra of the skin lesions is a NIR-FT Bruker FRA 106 Raman
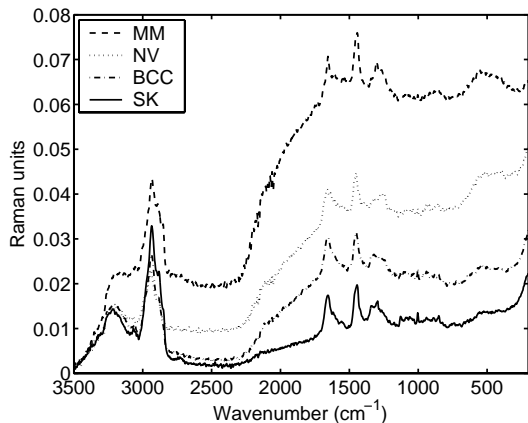
Fig. 1. Examples of the NIR-FT Raman spectra of benign and malignant skin lesions and tumors: basal cell carcinoma (BCC), malignant melanoma (MM), pigmented nevi (NV) and seborrhoeic keratosis (SK).

Fig. 2. The figure shows the original Raman spectrum, the background fit and the background suppressed spectrum of the (SK) lesion in Fig. 1.

spectroscope. The radiation source is a 1064 nm line Nd:YAG laser. The Raman spectra were obtained from skin samples *in vitro* at the skin surface of the punch biopsies or curetted lesions. Each spectrum is the average over 250 scans in the frequency range 200-3500 cm^{-1} with about 2 cm^{-1} frequency resolution, resulting in 1711 frequency components.

The data set comprises 5 different types of skin lesions; basal cell carcinoma (BCC), malignant melanoma (MM), normal skin (NOR), pigmented nevi (NV) and seborrhoeic keratosis (SK). BCC and MM are malignant, NV and SK are benign lesions. NV is a benign pigmented skin tumor and SK, also a benign skin lesion, is often seen in the older population and visually resembles MM. Class labels are obtained by histological analysis.

Fig. 1 shows representative examples of Raman spectra from the four classes of skin lesions and tumors.

III. PREPROCESSING OF DATA

The first step in the preprocessing is a visual screening for defective spectra. Medical data are noisy and contaminated by biological variability. One spectrum having very high amplitude and no visible distinctive narrow peaks was excluded from the data.

The Raman spectrum has a so-called background that originates from the skin fluorescence [18]. This background can clearly be seen in Fig. 1 as an amplitude elevation in the region 200-2800 cm^{-1}. The narrow peaks represent the vibration of chemical bonds. The background introduces both variance into the spectra and correlation between frequency components which makes modeling difficult. Also, the number of frequency components in the Raman spectrum is usually large $F \sim 10^3 - 10^4$; in the data set analyzed here $F = 1711$. The number of examples is $N = 222$, much less than the dimensions of the spectra, making the data *ill-posed*. Pattern recognition in such high dimensional input space with relatively few examples is a challenge as it suffers from the well-know *curse of dimensionality*. Hence, we want to create a mapping from the original high dimensional input space to a subspace, maintaining as much class discriminating information in the subspace as possible.
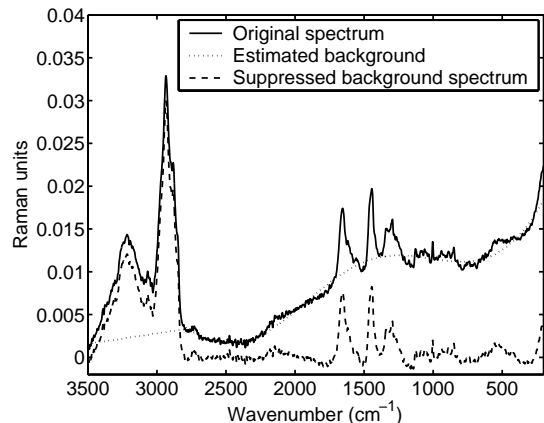
We suggest the following two-step preprocessing scheme of Raman spectra for skin lesion classification. Background is suppressed from the spectra by fitting a regression network to the background and subtracting it from the original spectrum. The input space is then reduced using *principal components analysis* (PCA). These steps are discussed in detail in the following sections.

A. Suppressing the background

One of the factors controlling the amount of background in skin lesions is the pigmentation in the skin [18]. Some MM lesions are often more pigmented than normal skin and other types of lesions, thus the amplitude of the background is higher as seen in Fig. 1. Raman spectra from skin lesions with a wide range of pigmentation, e.g. MM lesions, have higher variance, i.e. the frequency components are spread over a larger volume in the high dimensional input space, spanned by the Raman spectra. This makes predictions in those sparse areas difficult as training is based on few examples. By removing the background the variance of the input data for those classes is minimized, making the predictions more reliable. This will be shown with experiments, by comparing the classification results with and without background suppression, see section VI A.

The background spectra is estimated in two parts. The background below 2800 cm^{-1} is estimated with a regression network for each individual Raman spectrum. The spectrum 2800-3500 cm^{-1} is simply estimated as a straight line between the two end points. The Raman spectra split is based on the experience of medical experts. The estimated background spectra from these two parts are then subtracted from the original spectrum to form Raman spectra with suppressed background. The regression network models the background, by assuming the "noise" to be both the Raman peaks and small fluctuations in the spectrum. The peaks require a non-symmetric error function, as opposed to the usual symmetric quadratic error function used in regression. Here we assume the noise to be Gumbel distributed with zero location. For details on the neural regression network see [19].

An example of background suppression for a single Raman spectrum is shown in Fig. 2. The background noise has been suppressed while the peaks show little sign of being distorted.

B. Principal Components Analysis

Assuming that the presence of multiple signal classes in a skin lesion is the major source of variation, a natural choice for dimension reduction is PCA. PCA identifies an orthogonal basis on which projections are uncorrelated. Dimensional reduction is obtained by discarding transformed input dimensions with the low variance.

PCA is achieved using *singular value decomposition*. The background suppressed Raman spectrum data matrix $\mathbf{D}$ of size $F \times N$, where N is the number of data vectors and F is the dimension of the multivariate data vectors, are decomposed into

$$\mathbf{D} = \mathbf{U}SV^\top, \qquad (1)$$

where $\mathbf{U}$ is an $F \times N$ matrix, $\mathbf{S}$ is an $N \times N$ matrix and $\mathbf{V}$ is an $N \times N$ matrix by using an economy size decomposition. The matrix $\mathbf{S}$ is a diagonal matrix with nonnegative elements in descending order. These diagonal elements are the singular values that correspond to standard deviations of the input data, projected onto the given basis vector. The reduced input space is obtained by using only some fixed $I \leq N$ number of the largest principal components. The reduced data matrix is given by

$$\mathbf{X} = \widetilde{\mathbf{U}}^\top \mathbf{D}, \qquad (2)$$

where the transformation matrix from the original input space to the reduced inputs space is given by $\widetilde{\mathbf{U}}$, a $F \times I$ sub-matrix of $\mathbf{U}$. Note that the columns of $\mathbf{D}$, given by $\mathbf{d}^{(n)}$ where n is the column index, are the spectra after background suppression, and the columns of $\mathbf{X}$, given by $\mathbf{x}^{(n)}$, are the corresponding feature vectors.

The number of retained principal components I is selected by using a similar approach as in [20]. The likelihood is formulated for the PCA, based on a multivariate Gaussian distribution, and the Bayesian information criterion (BIC) is used for model selection [21]. The BIC adds a term to the likelihood which penalizes large models with many parameters. The number of parameters in the PCA increases with the number of principal components, thus models with many principal components are penalized harder. Hence, the BIC estimates an optimum number of principal components based on a tradeoff between too simple and too complex PCA models.

To ensure fully independent test data, the PCA is only applied to the training data to form the $\widetilde{\mathbf{U}}$ transformation matrix from the high dimensional input space to the low dimensional. This matrix is then used to transform the test data to the lower dimensional subspace.

The PCA is very sensitive to outliers in the data, i.e. data points far from the main mass cloud of the data. There are a few spectra in the skin lesion data set that have very high background amplitude, making them potential outliers. Removing the background makes the PCA less sensitive and the estimate of $\widetilde{\mathbf{U}}$ generalizes better. This may be shown by evaluating the difference for a single feature vector $\mathbf{x}^{(n)}$ when included in the evaluation of the PCA or not. The difference may be computed using the leave-one-out squared two norm, given by

$$e(n) = \| \left(\widetilde{\mathbf{U}}^\top - \widetilde{\mathbf{U}}_{\backslash n}^\top \right) \mathbf{d}^{(n)} \|^2, \qquad (3)$$

where $\widetilde{\mathbf{U}}_{\backslash n}^\top$ is evaluated with all examples except spectrum $\mathbf{d}^{(n)}$. This is done for all examples in a leave-one-out manner, with and without background suppression. To evaluate the influence of outliers, we compute the ratio of standard deviations of the error, given by $\sigma_e / \sigma_{e,0.95}$, where σ_e is the standard deviation using all examples, and $\sigma_{e,0.95}$ is the standard deviation where 5% of the examples with the largest error are removed. The results gave the ratio 5.5 and 27.2, with and without background suppression, respectively. Thus, by removing the background the 'tail to body' ratio $\sigma_e / \sigma_{e,0.95}$ is reduced by a factor 5.

IV. PROBABILISTIC CLASSIFICATION FRAMEWORK

We aim at modeling the posterior probability functions for multi-classification, given by $P(\mathcal{C}_k|\mathbf{x})$, $k = 1, 2, \ldots, c$, where $\mathbf{x}$ is the input feature vector with dimension I, $\mathcal{C}_k$ is the corresponding class label and $c = 5$ is the number of classes.

Multi-layer perceptron networks posses powerful approximation capabilities and when used for classification they can adapt to arbitrarily complex posterior probability functions. Such extreme flexibility calls for careful control of *overfit* and detection of *outliers*. Overfit control is aimed at regularization, typically using weight decay, i.e., controlling the roughness of decision surfaces, so that they do not get too rough in the face of noise in finite samples. Outlier detection, on the other hand, is aimed at modeling and controlling random label noise that can lead to wrong decision surface *topologies* by creating isolated "islands" of the wrong class.

Outliers are defined as an input pattern having the corresponding target class label erroneously "flipped" to another class. In skin lesion classification samples are labeled by histological examination. If a sample for some reason is erroneously registered, the label can have a random relation to the input pattern. Hence, we defined a probability ε of being assigned with random target label. The outlier probability $\varepsilon = [0, 1]$ is assumed to be independent of both "true" class label and input pattern value.

The posterior probability distribution has been previously formulated [22] as

$$P(\mathcal{C}_k|\mathbf{x}) = P_0(\mathcal{C}_k|\mathbf{x})(1-\varepsilon) + \frac{\varepsilon}{c-1} \sum_{l=1,l\neq k}^{c} P_0(\mathcal{C}_l|\mathbf{x}), \quad (4)$$

where $P_0(\mathcal{C}_k|\mathbf{x})$ is the posterior probability with zero outlier probability. The first term in (4) is the probability that the input pattern $\mathbf{x}$ is not an outlier, while the second term is the outlier contribution coming from classes other than $\mathcal{C}_k$. By defining a scaled outlier probability $\beta = \varepsilon/(c-1)$, (4) can be rewritten as

$$P(\mathcal{C}_k|\mathbf{x}) = P_0(\mathcal{C}_k|\mathbf{x})(1 - \beta c) + \beta, \qquad (5)$$

where $\beta = [0, 1/(c-1)]$.

A. Network architecture

The network architecture that represents the posterior probabilities is a two-layer feed-forward neural network with I inputs given by

$$h_j(\mathbf{x}) = \tanh\left(\sum_{i=1}^{I} w_{ji}x_i + w_{j0}\right), \qquad (6)$$

where w_{ji} are the input to hidden weights, w_{j0} is the input to hidden bias and $h_j(\mathbf{x})$ is the output of the jth sigmoidal activation function of the hidden layer. Network output k of the output layer is given by

$$y_k(\mathbf{x}) = \sum_{j=1}^{H} w_{kj}h_j(\mathbf{x}) + w_{k0}, \qquad (7)$$

where w_{kj} are the hidden to output weights, w_{kj} is the input to hidden bias and H is the number of units in the hidden layer. To be able to interpret the outputs as estimates of the posterior probabilities $\widehat{P}(\mathcal{C}_k|\mathbf{x})$ we used the normalized exponential or softmax [23]. The softmax is given by

$$\widehat{P}_0(\mathcal{C}_k|\mathbf{x}) = \frac{\exp(y_k(\mathbf{x}))}{\sum_{l=1}^{c} \exp(y_l(\mathbf{x}))} \qquad (8)$$

for output k. An estimate of the posterior probability incorporating the outlier probability is adapted from (5) giving $\widehat{P}(\mathcal{C}_k|\mathbf{x}) = \widehat{P}_0(\mathcal{C}_k|\mathbf{x})(1 - \hat{\beta}c) + \hat{\beta}$, where $\hat{\beta}$ is an estimate of the scaled outlier probability, see section IV C.

B. Inferring the weights

The data set for the supervised training of the model is given by the input-output pairs $\mathcal{D} = \{\mathbf{x}^{(n)}, \mathbf{t}^{(n)}\}$ $n = 1, 2, \ldots, N$, where $\mathbf{t}^{(n)}$ is the one-of-c coded target value vector given by

$$t_k^{(n)} = \begin{cases} 1 & \text{if } \mathbf{x}^{(n)} \in \mathcal{C}_k \\ 0 & \text{otherwise}, \end{cases} \qquad (9)$$

where $k = 1, 2, \ldots, c$. To simplify notation we define the network weight vector as $\mathbf{w}$, holding all weights.

To infer the weights we invoke the approach proposed by David MacKay [24], [25]. The posterior probability of the weights $\mathbf{w}$ can be written as

$$p(\mathbf{w}|\mathcal{D}, \alpha, \beta) = \frac{p(\mathcal{D}|\mathbf{w}, \beta)p(\mathbf{w}|\alpha)}{p(\mathcal{D}|\alpha, \beta)}, \qquad (10)$$

where $p(\mathcal{D}|\mathbf{w}, \beta)$ is the *likelihood*, $p(\mathbf{w}|\alpha)$ is the *prior* and $p(\mathcal{D}|\alpha, \beta)$ is the *evidence*. The α and β are hyperparameters, i.e., regularization parameter and scaled outlier probability respectively, both assumed to be known when inferring the weights.

For a classification problem with multiple classes the choice of likelihood is
$p(\mathcal{D}|\mathbf{w}, \beta) = \exp[-E_D(\mathbf{w}, \beta)]$, where

$$E_D(\mathbf{w}, \beta) = -\sum_{n=1}^{N} \sum_{k=1}^{c} t_k^{(n)} \ln(\widehat{P}(\mathcal{C}_k|\mathbf{x}^{(n)})) \qquad (11)$$

is the cross-entropy error function [26].

The prior over weights is a zero mean Gaussian distribution, better known as weight decay, given by

$$p(\mathbf{w}|\alpha) = \frac{\exp[-\alpha E_W(\mathbf{w})]}{Z_W(\alpha)}, \qquad (12)$$

where $E_W(\mathbf{w})$ is a regularization function given by

$$E_W(\mathbf{w}) = \frac{1}{2}\sum_{i=1}^{W} w_i^2, \qquad (13)$$

where W is the number of weights in the network and the normalization factor $Z_W(\alpha)$ is given by

$$\begin{align} Z_W(\alpha) &= \int \exp[-\alpha E_W(\mathbf{w})]d\mathbf{w} \qquad (14) \\ &= (2\pi/\alpha)^{W/2}. \qquad (15) \end{align}$$

The optimization of the weights is done by minimizing a cost function,
$S(\mathbf{w}) \propto -\ln p(\mathbf{w}|\mathcal{D}, \alpha, \beta)$, given by

$$S(\mathbf{w}) = E_D(\mathbf{w}, \beta) + \alpha E_W(\mathbf{w}), \qquad (16)$$

where weight independent terms have been omitted. The optimization algorithm uses a state-of-the-art Quasi-Newton scheme: A BFGS updating of the inverse Hessian, soft line search with quadratic fit and the Wolfe conditions for termination. Trust region type monitoring is applied to ensure appropriate relative changes in the weights at each iteration. See [27] for details.

C. Adapting the hyperparameters

The posterior distribution for the hyperparameters is given by

$$p(\alpha, \beta|\mathcal{D}) = \frac{p(\mathcal{D}|\alpha, \beta)p(\alpha, \beta)}{p(\mathcal{D})}. \qquad (17)$$

By using the normalized likelihood approach with an unspecified prior $p(\alpha, \beta) = 1$, we make the so-called evidence approximation by using the evidence $p(\mathcal{D}|\alpha, \beta)$ to evaluate $p(\alpha, \beta|\mathcal{D})$. For details on this approximation see [28]. The evidence can be written as

$$\begin{align} p(\mathcal{D}|\alpha, \beta) &= \int p(\mathcal{D}|\mathbf{w}, \beta)p(\mathbf{w}|\alpha)d\mathbf{w} \qquad (18) \\ &= \frac{1}{Z_W(\alpha)}\int \exp[-S(\mathbf{w})]d\mathbf{w}, \qquad (19) \end{align}$$

where the weights are marginalized. An analytic solution to the integral is intractable, so the evidence is evaluated using the Laplace approximation

$$p(\mathcal{D}|\alpha, \beta) \approx \frac{e^{-S(\mathbf{w}_{\text{MP}})}(2\pi)^{W/2}|\mathbf{A}(\mathbf{w})|^{-1/2}}{Z_W(\alpha)}, \qquad (20)$$

where $\mathbf{w}_{\text{MP}}$ maximizes the product $p(\mathcal{D}|\mathbf{w}, \beta)p(\mathbf{w}|\alpha)$ and $\mathbf{A}(\mathbf{w})$ is the Gauss-Newton approximation to the Hessian matrix. The approximation is given by

$$\mathbf{A}(\mathbf{w}) = \frac{1}{N}\sum_{n=1}^{N} \frac{\partial e_D^{(n)}(\mathbf{w})}{\partial \mathbf{w}} \frac{\partial e_D^{(n)}(\mathbf{w})}{\partial \mathbf{w}^\top} + \frac{\partial^2 E_W(\mathbf{w})}{\partial \mathbf{w}\partial \mathbf{w}^\top}, \qquad (21)$$

where $e_D^{(n)}(\mathbf{w}) = \sum_{k=1}^c t_k^{(n)} \ln(\widehat{P}(\mathcal{C}_k|\mathbf{x}^{(n)}))$, which is the cross-entropy error for a single example. The approximation ensures a symmetric matrix as $\frac{\partial^2 E_W(\mathbf{w})}{\partial \mathbf{w} \partial \mathbf{w}^\top} = \alpha \mathbf{I}$, where $\mathbf{I}$ is a $W \times W$ identity matrix. Note that $\alpha > 0$ in practice, so $\mathbf{A}(\mathbf{w})$ is positive definite, thus solving numerical problems when evaluating the determinant of $\mathbf{A}(\mathbf{w})$.

The scaled outlier probability β is estimated by minimizing

$$C(\beta) \quad \propto \quad -\ln p(\mathcal{D}|\alpha, \beta) \qquad (22)$$

$$= \quad S(\mathbf{w}_{\text{MP}}) + \frac{1}{2}|\mathbf{A}(\mathbf{w})|, \qquad (23)$$

where terms independent of β have been omitted. We suggest using Brent's minimization method [29], approximating $C(\beta)$ as a quadratic to find $\hat{\beta}$. This is possible as $C(\beta)$ is a smooth function and also as we have an upper and lower bound on β setting the range for the search of $\hat{\beta}$. As Brent's method does not use gradient information we avoid evaluating $\partial S(\mathbf{w})/\partial \beta$ which has a singularity at $\beta = 0$.

The α is computed as in [24], by maximizing $\ln p(\mathcal{D}|\alpha, \beta)$, evaluating $\partial \ln p(\mathcal{D}|\alpha, \beta)/\partial \alpha$ which gives the following update formula

$$\alpha^{new} = \frac{\gamma}{2E_W(\mathbf{w}_{\text{MP}})}, \qquad (24)$$

where $\gamma = W - \alpha \text{Trace} \mathbf{A}^{-1}(\mathbf{w})$ is the effective number of weights in the network.

A practical approach to adapting the hyperparameters would be to train the weights and update the α and β when the weights have converged. This is repeated in turn until the hyperparameters have converged. The neural network classifier is available as Matlab software [30].

V. MODEL VISUALIZATION BY SENSITIVITY MAPS

One of the disadvantages of using neural networks is the fact that they are non-parametric, i.e., black box methods. Finding the active inputs which the network uses to classify individual groups is interesting from a medical point of view. In our case we are interested in knowing which frequency components contribute most to the classification of the skin lesions, thus being able to interpret them as specific chemical substances.

A simple measurement of how much the network relies on each input is the *sensitivity map*, or sensitivity analysis, see e.g. [31], [32], [33]. The perturbation of each input is found by computing the gradient of each output with respect to all inputs. This means evaluating the derivative of the estimated posterior $\widehat{P}(\mathcal{C}_k|\mathbf{x})$ of each class with respect to the Raman spectrum. One way of computing a sensitivity map for a given network is the *absolute value average sensitivities* [32] given for class $\mathcal{C}_k$ by

$$\mathbf{s}^k = \frac{1}{N} \sum_{n=1}^N \left| \frac{\partial \widehat{P}(\mathcal{C}_k|\mathbf{x}^{(n)})}{\partial \mathbf{d}^{(n)}} \right|, \qquad (25)$$

where $\mathbf{d}^{(n)}$ is the nth column of $\mathbf{D}$, the training set of Raman spectra before PCA. The absolute operator operates on each element in the vector. Note that both $\mathbf{s}^k$ and $\mathbf{d}^{(n)}$ are vectors of length F, where the elements represent values for specific frequency components of the Raman spectrum. Computing the sensitivity with (25) is suggested when the values are disparate

[32]. Taking the absolute value of each training example's perturbation is necessary to avoid possible cancellation of positive and negative values because of multiple decision boundaries.

The derivative in (25), for a single training example and class k, is given by

$$\frac{\partial \widehat{P}(\mathcal{C}_k|\mathbf{x})}{\partial \mathbf{d}} = \widehat{P}(\mathcal{C}_k|\mathbf{x}) \left(\frac{\partial y_k(\mathbf{x})}{\partial \mathbf{d}} - \sum_{k'=1}^c \frac{\partial y_{k'}(\mathbf{x})}{\partial \mathbf{d}} \widehat{P}(\mathcal{C}_{k'}|\mathbf{x}) \right), \qquad (26)$$

where $y_k(\mathbf{x})$ is given in (7). For the network architecture used here the derivative in the previous equation is given by

$$\frac{\partial y_k(\mathbf{x})}{\partial \mathbf{d}} = \sum_{j=1}^H w_{kj} \left(1 - (h_j(\mathbf{x}))^2 \right) \sum_{i=1}^I w_{ji} \widetilde{\mathbf{u}}_i, \qquad (27)$$

where $h_j(\mathbf{x})$ is given in (6) and $\widetilde{\mathbf{u}}_i$ is the ith column vector of the PCA transformation matrix $\widetilde{\mathbf{U}}$ with length F, given in (2).

Some normalization should take place when comparing sensitivity maps from networks trained on different training data. Without normalization the sensitivity maps can be scaled, i.e., maps can be very similar except that they can have different amplitudes. This is due to the non-uniqueness of the hidden-to-output weights. For instance, multiplying the hidden-to-output weights with a constant does not change the predicted posterior of the network. To solve this problem we normalize the sensitivity maps for each class to unit vector length given by

$$\widetilde{\mathbf{s}}^k = \frac{\mathbf{s}^k}{\|\mathbf{s}^k\|}. \qquad (28)$$

A. Reproducibility of sensitivity maps

For the interpretation of sensitivity maps, reproducibility is essential. A large sensitivity which varies for sample to sample is not as important as a smaller sensitivity that is highly reproducible.

To measure the reproducibility of the sensitivity maps we adapt an approach very similar to NPAIRS (nonparametric prediction, activation, influence, and reproducibility resampling), which is used to validate statistical parametric maps (SPMs) in functional neuroimaging [34]. The method is based on a split-half resampling technique, where two networks are trained on each half of the data. The resulting scatter plot of the sensitivity maps from each network and class can now be directly compared without dealing with bias due to different class sizes, while at the same time the split-half resampling maximizes the power of each of the independent-class data.

To illustrate the method, Fig. 3 shows an example of two sensitivity maps for the SK group generated from single split-half resampling. By generating a scatter plot of the corresponding frequency components of the sensitivity maps, shown in Fig. 4, the correlation between them can be evaluated. A point in the scatter plot corresponds to a frequency component in the Raman spectra. The signal is the amplitude in the direction where the two sensitivity maps are equal, while the amplitude in the direction orthogonal to the signal is the noise. Note that the scatter plot is not symmetric around the signal axis,
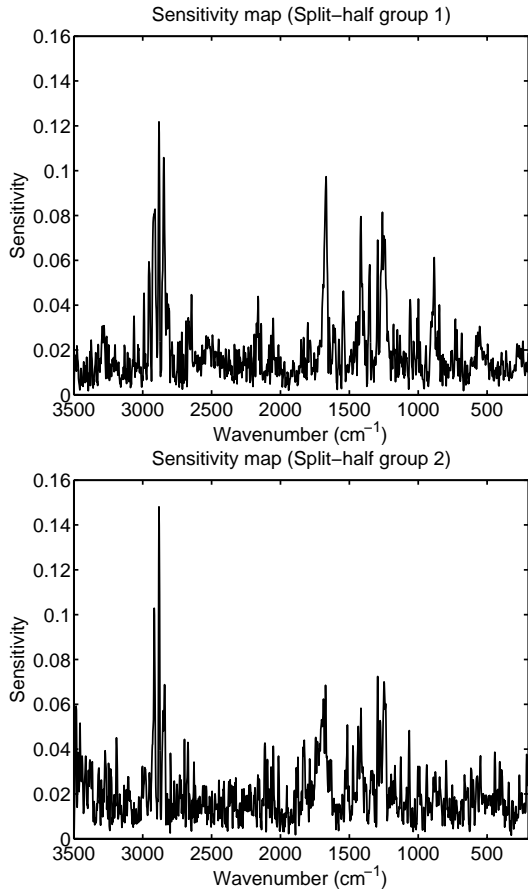
Fig. 3.   Sensitivity maps for the SK class generated from two networks trained on data using split-half resampling. Despite the noise the sensitivity maps are clearly correlated.
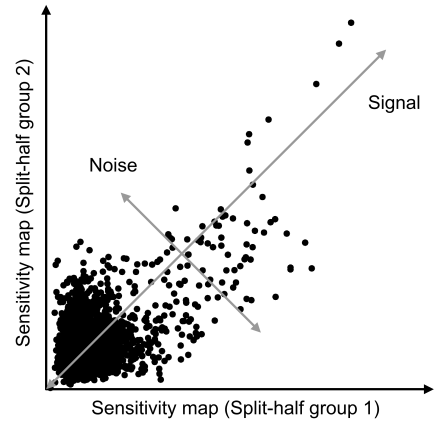

Fig. 4.   Scatter plot of the sensitivity maps in Fig. 3. The signal and noise arrows indicate the direction of dependence and independence, respectively, between the sensitivity maps.
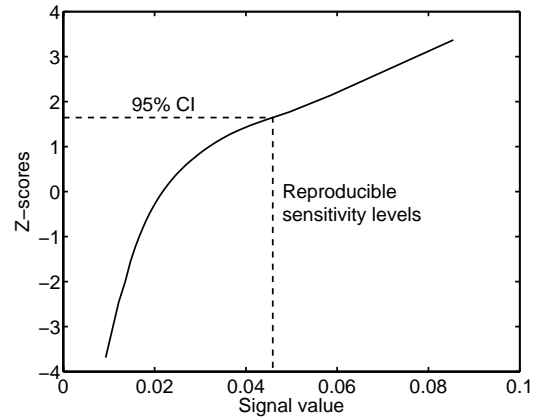

Fig. 5.   The Z-score as a function of signal level for the SK class. This is the average of 100 split-half resampled data sets. The dashed line indicates the 95% confidence interval of the Z-score which gives sensitivity levels above 0.045 as reproducible.

due to random variability of the two sensitivity maps in this single realization. Averaged over multiple split-half resampled sensitivity maps the scatter plot is symmetric around the signal axis.

The main difference between SPMs and sensitivity maps is the noise distribution in the scatter plot. The SPM assumes Gaussian distributed noise which makes a qualitative reproducibility measure with confidence intervals from Z-scores ideal. In [34] the reproducibility measure of the SPMs are based on the Pearson product correlation coefficient. As seen in Fig. 4 the noise of a single split-half scatter plot of the sensitivity maps is far from Gaussian distributed. Most of the density mass is a noise contribution concentrated close to the origin, while occasional peaks of signal radiate from the origin. Using the correlation coefficient here would result in an overestimated value.

When measuring reproducibility of sensitivity maps we suggest to divide the signal axis into bins and evaluate the Z-score for each bin. We assume that the noise is indeed Gaussian distributed, but having the variance dependent on the signal level. Given two column vector sensitivity maps $\mathbf{s}_1^k$ and $\mathbf{s}_2^k$, marked 1 and 2 from class k, generated from networks with split-half resampling of the data, the signal and noise contribution can be evaluated with a simple rotation given by $\mathbf{s}_{\text{signal}}^k = (\mathbf{s}_1^k + \mathbf{s}_2^k)/\sqrt{2}$ and $\mathbf{s}_{\text{noise}}^k = (\mathbf{s}_1^k - \mathbf{s}_2^k)/\sqrt{2}$,

respectively. Dividing the signal part into bins, each containing some fixed number of points, the mean signal μ_b and noise standard deviation σ_b for bin b is computed using the point belonging to the respective bin. The number of points in a bin is heuristically chosen 50. The Z-score is then easily evaluated for each bin with $Z_b = \mu_b/\sigma_b$. In Fig. 5 the average Z-score of 100 split-half resampling networks is plotted as a function of signal level. By referring to tables of the normal distribution, where a Z-score greater than 1.6449 corresponds to the 95% confidence interval, the sensitivity values above 0.045 are determined significant. The significant sensitivity level may be plotted with the sensitivity map, as shown in Fig. 13, thus indicating reproducible sensitivity in the map.

VI. EXPERIMENTS

In this section we will compare network results with and without the background suppression. We examined both the classification performance and visualization results. For this purpose the reproducibility of the generated sensitivity maps for visualization were examined.

	BCC	MM	NOR	NV	SK
BCC*	95.8	10.0	1.1	0.0	0.9
MM*	0.0	80.5	0.0	2.4	0.0
NOR*	0.0	4.8	97.8	5.4	0.0
NV*	2.1	4.8	1.1	92.2	0.0
SK*	2.1	0.0	0.0	0.0	99.1

Fig. 6. Confusion matrix showing classification rate in percentages on test data with background suppression. Stars indicate neural network prediction. Note that the classification rate of the malignant MM lesions is over 80%, which is similar to experts in dermatology. Most of the misclassified examples are classified as malignant BCC lesions.

	BCC	MM	NOR	NV	SK
BCC*	92.7	9.5	1.1	0.0	9.1
MM*	2.1	62.4	0.0	0.0	0.0
NOR*	0.0	0.0	97.8	7.3	5.2
NV*	3.1	21.0	1.1	92.7	3.9
SK*	2.1	7.1	0.0	0.0	81.7

Fig. 7. Confusion matrix showing classification rate in percentages on test data without background suppression. Stars indicate neural network prediction. Note that the classification rate of the malignant MM lesions is less than 63% and 21% of the examples are misclassified as benign NV lesions.

	Pos	Neg
Pos*	94.2	1.4
Neg*	5.8	98.6

	Pos	Neg
Pos*	87.8	2.0
Neg*	12.2	98.0

Fig. 8. Confusion matrix for cancer versus non-cancer skin lesions. The stars indicates network prediction. The left panel shows the results with background suppression and the right panel without background suppression.

The Raman data set of skin lesions consists of $N = 222$ examples, where the number of examples for each of the classes BCC, MM, NOR, NV and SK is 48, 21, 89, 41 and 23, respectively.

The data set was preprocessed as described earlier and $I = 25$ largest principal components were used as inputs to the network. The feed-forward neural network was initialized with $H = 20$ hidden units and $c = 5$ outputs to represent the posterior probabilities for the classes.

A. Classification

As the MM and SK classes have very few examples the classification performance of the network was evaluated using a leave-one-out cross-validation scheme. The results reported are the average over 10 runs of the data set. The computational time of a single optimization of the neural network classifier was approximately 30 minutes on an 800MHz Pentium III computer.

Error bars on classification were approximated with the normal approximation to the binomial distribution. The error bars are evaluated as two times the standard deviation, given by $2\sqrt{p(1-p)/N}$ where p is the classification rate and N is the number of examples.

The overall classification rate of skin lesions and tumors was 94.8%±3.0% and 90.7%±3.9% with and without background suppression, respectively. The hypothesis that the two classification rates were the same was examined with a McNemar's test [35] and rejected with 95.1% confidence. The classification performance of the neural network was also analyzed using a confusion matrix. The confusion matrices using data with and without suppressed background are shown in Fig. 6 and 7, respectively. The results showed that classes BCC, NOR, NV were well determined and had high classification rates with or without background suppression. When the background was suppressed both the SK and MM classification was improved significantly. When using no background suppression, the misclassified MM lesions where classified as benign NV lesions, which is a serious error. On the other hand when the background was suppressed, most of the misclassified MM lesions went to the malignant BCC group. The classification rate of MM was 80.5%±5.3%, which is similar to dermatology

experts. It should though be emphasized that the performance is based on limited amount of examples from the MM class and the error bars are rather large. The results for the BCC was excellent, 95.8%±2.7% of the lesions were classified correctly. Note that by suppressing the background a small percentage of MM lesions were classified as NOR.

From a clinical point of view it is interesting to look at the classification of cancer versus non-cancer lesions. The confusion matrices for cancer versus non-cancer lesions, with and without suppressed background are shown in Fig. 8. The performance using background suppression was much higher. 94.2%±3.1% of cancer lesions are classified correctly which is very good, compared to 87.8%±4.4% without background suppression.

It is appropriate to compare the classification results obtained here with other techniques for classification of skin cancer. A recent study, using a similar neural network architecture trained on features extracted from digitized dermatoscopic images, obtained 75% sensitivity for malignant melanoma [7]. Other recent studies [8], [36], [9], also using dermatoscopic features and neural networks, have shown an extremely good performance. The sensitivity was 92-94% for malignant melanoma, but these studies do not appear to apply independent test data. For comparison the neural network used here obtained 96% sensitivity on training data.

B. Visualization

The visualization is only based on networks trained on data with suppressed background as the classification performance of these networks is significantly higher. To examine the reproducibility of the sensitivity maps we used the previously
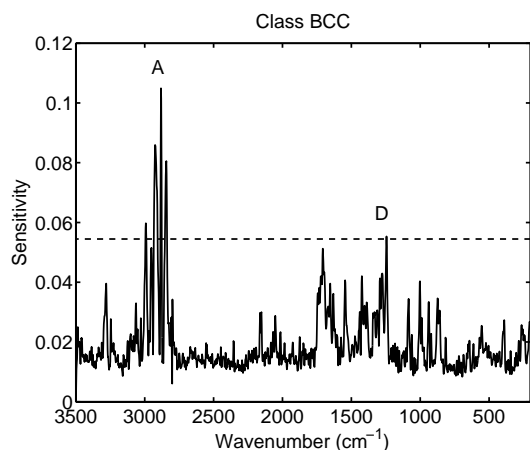
Fig. 9. Sensitivity maps for the BCC class. The dashed line indicates 95% confidence interval. The sensitivity is mainly in well determined peaks and most of the spectra does not contain any discriminative information. Region marked **A** represent the CH$^-$ vibrations in the lipids and proteins around 2940 cm^{-1} and region marked **D** are the amide III band in proteins around 1270 cm^{-1}.
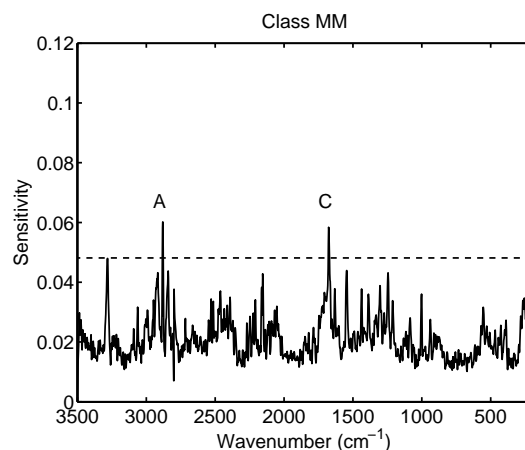

Fig. 10. Sensitivity maps for the MM class. The dashed line indicates 95% confidence interval. The sensitivity map seems more noisy than the BCC sensitivity map in Fig. 9. Region marked **A** represent the CH$^-$ vibrations in the lipids and proteins around 2940 cm^{-1} and region marked **C** reflects the amide I band of proteins 1600-1800 cm^{-1}.

described framework. Networks are trained on 100 random split-half of the data, in all 200 networks. Sensitivity maps for each class were computed by taking the average of the 200 generated sensitivity maps. The significance level of the average sensitivity maps were found by making Z-score curves for each class using 95% confidence intervals.

The sensitivity maps for BCC, MM, NOR, NV and SK lesions are shown in Fig. 9, 10, 11, 12 and 13, respectively. Classes BCC and SK reproduced well as sensitivity peaks lie well above the significance level. Classes MM, NOR and NV lie just above the significance level, but at the same time the confidence in those peaks is very high. The sensitivity maps show that the BCC and SK classes use a very small fraction of the Raman spectrum for classification and that they are well defined, giving very good classification results. On the other hand, the MM, NOR and NV classes use many parts of the spectrum as the features selected by the networks seem to be more complex. A probable reason for this could be the small sample size compared to the variability of the Raman spectrum for these classes, especially for the malignant MM lesions.

The spectral regions where the sensitivity maps give high importance, contain major protein and lipid bands. These are marked with letters **A**-**D** in Fig. 9 to 13. The spectral regions marked **A**, **C** and **D** have previously been visually identified by Raman spectroscopy trained experts in dermatology [37]. Region marked **A** represent the CH$^-$ vibrations in the lipids and proteins around 2940 cm^{-1}, the increase in the bandwidth was noted for BCC and SK classes. Region marked **C** reflects the amide I band of proteins 1600-1800 cm^{-1}, and the decrease in amplitude was found for the MM class. Region around 1270 cm^{-1} marked **D** shows the amide III band in proteins, for BCC and SK classes the decrease in amplitude was noted. Finally, by examining the Raman spectra, region marked **B** corresponds to the vibrations caused by skin fluorescence 2000-2350 cm^{-1}, which showed an increase in amplitude for NV class.
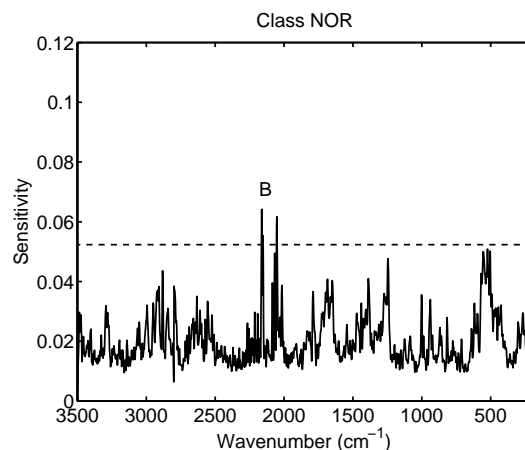

Fig. 11. Sensitivity maps for the NOR class. The dashed line indicates 95% confidence interval. The sensitivity map seems noisy like the MM class in Fig. 10. Region marked **B** corresponds to the vibrations caused by skin fluorescence 2000-2350 cm^{-1}.

VII. CONCLUSION

Skin tumor classification based on Raman spectroscopy was approached using a non-linear neural network classifier. A feature extraction scheme was suggested to reduce the dimension of the Raman spectra and to suppress background noise originating from the skin fluorescence. The proposed neural network classifier automatically avoids over-fitting by using adaptive regularization and outlier detection. The neural network framework is distributed for non-commercial use [30]. Finally, the classification rules are extracted from the neural networks with sensitivity maps. The reproducibility of the sensitivity maps was determined with an extension to the NPAIRS method, making it possible to assign confidence intervals to the sensitivity maps.

By applying this framework on the present data set, involving 222 cases and 5 classes, the classification rate was 80.5%±5.3% correct classification of malignant melanoma, similar rates as obtained with visual inspection by experts in
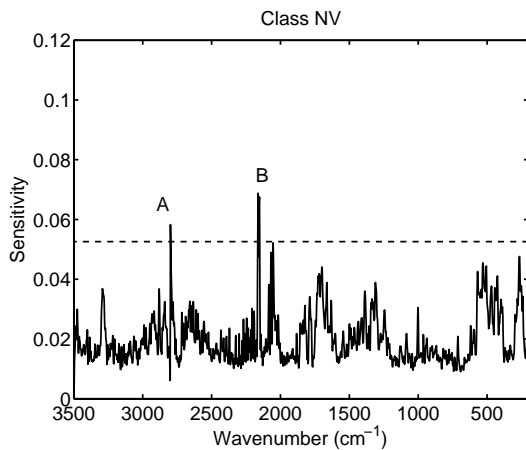
Fig. 12.   Sensitivity maps for the NV class. The dashed line indicates 95% confidence interval. The sensitivity map is noisy like the sensitivity maps for the MM and NOR class in Fig. 10 and 11, respectively. Region marked **A** represent the CH^- vibrations in the lipids and proteins around 2940 cm^{-1} and region marked **B** corresponds to the vibrations caused by skin fluorescence 2000-2350 cm^{-1}.
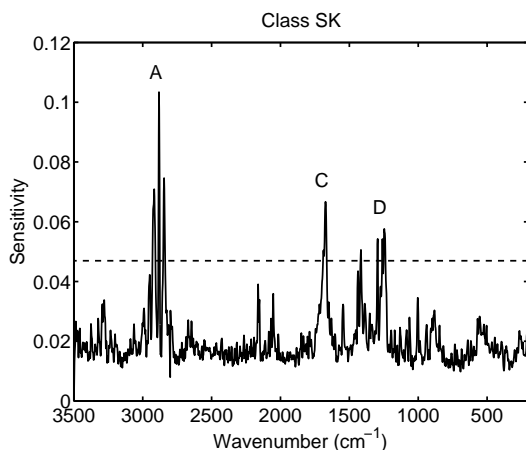

Fig. 13.   Sensitivity maps for the SK class. The dashed line indicates 95% confidence interval. Similar to the BCC class the sensitivity is mainly in very significant peaks and most of the spectra does not contain any discriminative information. Region marked **A** represent the CH^- vibrations in the lipids and proteins around 2940 cm^{-1}, region marked **C** reflects the amide I band of proteins 1600-1800 cm^{-1} and and region marked **D** are the amide III band in proteins around 1270 cm^{-1}.

dermatology [6]. The most common skin cancer, basal cell carcinoma, had a classification rate of 95.8%±2.7%, which is excellent. Taking both malignant lesion types as one group the networks classified 94.2%±3.1% of the cancer lesions correctly. The overall classification rate of skin lesions was 94.8%±3.0%. Sensitivity maps were shown to be a powerful tool for visualizing the learning of feed-forward network. They were shown to reproduce and important frequency bands were identified, corresponding to specific lipids and proteins.

REFERENCES

[1] R. Marks, "An Overview of Skin Cancers. Incidence and Causation," *Cancer*, vol. 75, pp. 607–612, 1995.

[2] A. Østerlind, "Malignant Melanoma in Denmark," Ph.D. dissertation, Danish Cancer Registry, Institute of Cancer Epidemiology, Denmark, 1990.

[3] G. Rassner, "Früherkennung des malignen Melanoms der Haut," *Hausartz*, vol. 39, pp. 396–401, 1988.

[4] S. Presser and J. Taylor, "Clinical Diagnostic Accuracy of Basal Cell Carcinoma," *Journal of the American Academy of Dermatology*, vol. 16, pp. 988–990, 1987.

[5] B. Lindelöf and M. Hedblad, "Accuracy in the Clinical Diagnosis and Pattern of Malignant Melanoma at a Dermatological Clinic," *The Journal of Dermatology*, vol. 21, no. 7, pp. 461–464, 1994.

[6] H. Koh, R. Lew, and M. Prout, "Screening for Melanoma/Skin Cancer," *Journal of American Academy of Dermatology*, vol. 20, no. 2, pp. 159–172, 1989.

[7] M. Hintz-Madsen, L. Hansen, J. Larsen, and K. Drzewiecki, "A Probabilistic Neural Network Framework for Detection of Malignant Melanoma," in *Artificial Neural Networks in Cancer Diagnosis, Prognosis and Patient Management*, R. Naguib and G. Sherbet, Eds.   CRC Press, 2001, pp. 141–183.

[8] P. Rubegni, M. Burroni, G. Cevenini, R. Perotti, G. Dell'Eva, P. Barbini, M. Fimiani, and L. Andreassi, "Digital Dermoscopy Analysis and Artificial Neural Network for the Differentiation of Clinically Atypical Pigmented Skin Lesions: A Retrospective Study," *Journal of Investigative Dermatology*, vol. 119, pp. 471–474, 2002.

[9] D. Piccolo, A. Ferrari, K. Peris, R. Daidone, B. Ruggeri, and S. Chimenti, "Dermoscopic Diagnosis by a Trained Clinician vs. a Clinician with Minimal Dermoscopy Training vs. Computer-aided Diagnosis of 341 Pigmented Skin Lesions: A Comparative Study," *British Journal of Dermatology*, vol. 147, pp. 481–486, 2002.

[10] M. Gniadecka, H. Wulf, N. Mortensen, O. Nielsen, and D. Christensen, "Diagnosis of Basal Cell Carcinoma by Raman Spectra," *Journal of Raman Spectroscopy*, vol. 28, pp. 125–129, 1997.

[11] P. Griffiths, I. Lewis, N. Chaffin, N. D. Jr., and J. Jegla, "Remote Characterization of Materials by Vibrational Spectrometry through Optical Fibers," *Journal of Molecular Structure*, vol. 347, pp. 169–185, 1995.

[12] E. Hanlon, R. Manoharan, T.-W. Koo, K. Shafer, J. Motz, M. Fitzmaurice, J. Kramer, I. Itzkan, R. Dasari, and M. Feld, "Prospects for in vivo Raman Spectroscopy," *Physics in Medicine and Biology*, vol. 45, pp. R1–R59, 2000.

[13] B. Schrader, S. Keller, T. Lochte, S. Fendel, D. Moore, A. Simon, and J. Sawatzki, "NIR FT Raman Spectroscopy in Medical Diagnosis," *Journal of Molecular Structure*, vol. 346, pp. 293–296, 1995.

[14] M. Shim and B. Wilson, "Development of an in vivo Raman Spectroscopy System for Diagnostic Applications," *Journal of Raman Spectroscopy*, vol. 28, pp. 131–142, 1997.

[15] B. Barry, H. Edwards, and A. Williams, "Fourier Transform Raman and Infrared Vibrational Study of Human Skin: Assignment of Spectral Bands," *Journal of Raman Spectroscopy*, vol. 23, pp. 641–645, 1992.

[16] H. Edwards, A. Williams, and B. Barry, "Potential Applications of FT-Raman Spectroscopy for Dermatological Diagnosis," *Journal of Molecular Structure*, vol. 347, pp. 379–388, 1995.

[17] M. Gniadecka, S. Wessel, O. Nielsen, D. Christensen, J. Hercogova, K. Rossen, and H. Wulf, "Potential of Raman Spectroscopy for in vitro and in vivo Diagnosis of Malignant Melanoma," in *Proceedings of the sixteenth International Conference on Raman Spectroscopy*, A. Heyns, Ed., vol. 16.   Chichester: John Wiley and Sons, 1998, pp. 764–765.

[18] L. Knudsen, C. Johansson, P. Philipsen, M. Gniadecka, and H. Wulf, "Natural Variations and Reproducibility of in vivo near-infrared Fourier Transform Raman Spectroscopy of Normal Human Skin," *Journal of Raman Spectroscopy*, vol. 33, pp. 574–579, 2002.

[19] S. Sigurdsson, J. Larsen, P. Philipsen, M. Gniadecka, H. Wulf and L. Hansen, "Estimating and Suppressing Background in Raman Spectra with an Artificial Neural Network," Informatics and Mathematical Modelling, Technical University of Denmark, Tech. Rep. 2003-20, 2003.

[20] L. Hansen, O. Paulson, J. Larsen, F. Nielsen, S. Strother, E. Rostrup, R. Savoy, N. Lange, J. Sidtis, and C. Svarer, "Generalizable Patterns in Neuroimaging: How Many Principal Components?" *NeuroImage*, vol. 9, pp. 534–544, 1999.

[21] G. Schwarz, "Estimating the Dimension of a Model," *The Annals of Statistics*, vol. 6, pp. 461–464, 1978.

[22] J. Larsen, L. Andersen, M. Hintz-Madsen, and L. Hansen, "Design of Robust Neural Network Classifiers," in *Proceedings of the 1998 International Conference on Acoustics, Speech and Signal Processing*, New York, New York, 1998, pp. 1205–1208.

[23] J. Bridle, "Probabilistic Interpretation of Feedforward Classification Network Outputs with Relationships to Statistical Pattern Recognition," in *Neurocomputing - Algorithms, Architectures and Applications*, F. Fougelman-Soulie and J. Herault, Eds.   Berlin: Springer-Verlag, 1990, vol. 6, pp. 227–236.

[24] D. MacKay, "A Practical Bayesian Framework for Backpropagation Networks," *Neural Computation*, vol. 4, no. 3, pp. 448–472, 1992.

[25] ——, "The Evidence Framework Applied to Classification Networks," *Neural Computation*, vol. 4, no. 5, pp. 720–736, 1992.

[26] C. Bishop, *Neural Networks for Pattern Recognition*. Oxford: Oxford University Press, 1995.

[27] H. Nielsen, "UCMINF - An Algorithm for Unconstrained Nonlinear Optimization," Department of Mathematical Modelling, Technical University of Denmark, Tech. Rep. IMM-REP-2000-19, 2000.

[28] D. MacKay, "Comparison of Approximate Methods for Handling Hyperparameters," *Neural Computation*, vol. 11, no. 5, pp. 1035–1068, 1999.

[29] W. Press, B. Flannery, S. Teukolsky, and W. Vettering, *Numerical Recipes in C: The Art of Scientic Computing*. Cambridge: Cambridge University Press, 1988.

[30] S. Sigurdsson, "The ANN:DTU Toolbox," 2002, *Available via:* `http://mole.imm.dtu.dk/toolbox/ann/index.html`.

[31] S. Hashem, "Sensitivity Analysis for Feedforward Artificial Neural Networks with Differentiable Activation Functions," *IJCNN International Joint Conference on Neural Networks*, vol. 1, pp. 419–424, 1992.

[32] J. Zurada, A. Malinowski, and I. Cloete, "Sensitivity Analysis for Minimization of Input Data Dimension for Feedforward Neural Network," in *Proceedings of IEEE Symposium on Circuits and Systems*, vol. 6, 1994, pp. 447–450.

[33] U. Kjems, L. Hansen, J. Anderson, S. Frutiger, S. Muley, J. Sidtis, D. Rottenberg, and S. Strother, "The Quantitative Evaluation of Functional Neuroimaging Experiments: Mutual Information Learning Curves," *NeuroImage*, vol. 15, no. 4, pp. 772–786, 2002.

[34] S. Strother, J. Anderson, L. Hansen, U. Kjems, R. Kustra, J. Sidtis, S. Frutiger, S. Muley, S. LaConte, and D. Rottenberg, "The Quantitative Evaluation of Functional Neuroimaging Experiments: The NPAIRS Data Analysis Framework," *NeuroImage*, vol. 15, no. 4, pp. 747–771, 2002.

[35] B. Ripley, *Pattern Recognition and Neural Networks*. Cambridge: Cambridge University Press, 1996.

[36] P. Rubegni, G. Cevenini, M. Burroni, R. Perotti, G. Dell'Eva, P. Sbano, C. Miracco, P. Luzi, P. Tosi, P. Barbini, and L. Andreassi, "Automated Diagnosis of Pigmented Skin Lesions," *International Journal of Cancer*, vol. 101, pp. 576–80, 2002.

[37] M. Gniadecka, H. Wulf, O. Nielsen, D. Christensen, and J. Hercogova, "Distinctive Molecular Abnormalities in Benign and Malignant Skin Lesions: Studies by Raman Spectroscopy," *Photochemistry and Photobiology*, vol. 66, pp. 418–423, 1997.

Lars Kai Hansen received the M.Sc. and Ph.D. degrees in physics from the University of Copenhagen, Denmark, in 1984 and 1986 respectively.

He was with Andrex Radiation Products A/S 1987-1990. Since 1990 he has been with the Technical University of Denmark, since 2000 as full professor of signal processing. He is the director of the THOR Center for Neuroinformatics and head of Copenhagen Image and Signal Processing graduate School (CISP). His current research projects concern modeling of adaptive systems, meta-analysis, and applications in neuroimaging.

Jan Larsen (S'89, M'93, SM'03) received the M.Sc. and Ph.D. degrees in electrical engineering from the Technical University of Denmark (DTU) in 1989 and 1994. Dr. Larsen is currently Associate Professor of Digital Signal Processing at Informatics and Mathematical Modelling, DTU. Jan Larsen has authored and co-authored around 80 papers and book chapters within the areas of nonlinear statistical signal processing, machine learning, neural networks and datamining with applications to biomedicine, monitoring systems, multimedia, and webmining.

He has participated in several national and international research programs and served as reviewer for many journals, conferences and publishing companies and took part in conference organizations, among others the IEEE Workshop on Neural Networks for Signal Processing (NNSP) 1999–2003, MLSP'2004. Currently he is chair elect of the IEEE Machine Learning for Signal Processing Technical Committee of the IEEE Signal Processing Society, and chair of IEEE Denmark Section's Signal Processing Chapter. He is member of the International Intelligent Knowledge Systems Society and member of the editorial board of International Journal of Computational Intelligence; International Journal of Signal Processing; International Journal of Intelligent Systems and Information Processing; and has been special issue guest editor of IEEE Transactions on Neural Networks; Journal of VLSI Signal Processing Systems; and Neurocomputing.

Sigurdur Sigurdsson received the M.Sc. and Ph.D. degrees from the Technical University of Denmark, Denmark, in 1998 and 2003, respectively.

He is currently working as Assistant Research Professor at Informatics and Mathematical Modelling, Technical University of Denmark. His research area is non-linear signal processing and pattern recognition for biomedical applications.

Monika Gniadecka M.D. is a dermatologist, graduated in 1991 from Warsaw Medical School, Poland. Ph.D. from Copenhagen University (1994) and since 2001 Assistant Professor.

Since 1995 she has been working with Raman spectroscopy for evaluation of skin diseases, especially skin cancer and connective tissue disorders. She extensively studied the structure of water in the skin which in 2001 culminated in the thesis *Studies on cutaneous water distribution and structure*.

Peter Alshede Philipsen received the M.S.E.E. and Ph.D. degrees from Technical University of Denmark, Denmark, in 1995 and 1999, respectively.

He is currently working as Research Fellow at the Department of Dermatology, Bispebjerg Hospital, University of Copenhagen. His research is focused on investigation and development of new techniques to diagnose and prevent skin cancer.

Hans Christian Wulf M.D. Dsc. graduated from the medical faculty Copenhagen University, Denmark.

He has a Ph.D. (Doctoral degree) in dermatology from University of Copenhagen and a Ph.D. in pharmacy from the Royal Danish School of Pharmacy. Until last year he was Professor and Head of Dermatology and still serves as Professor.

The main research interests are within photobiology and skin cancer. He has been author and co-author of 355 scientific papers.