

Journal of Biomedical Optics

SPIEDigitalLibrary.org/jbo

Statistical analysis of spectral data: a methodology for designing an intelligent monitoring system for the diabetic foot

Chanjuan Liu
Jaap J. van Netten
Marvin E. Klein
Jeff G. van Baal
Sicco A. Bus
Ferdinand van der Heijden

Statistical analysis of spectral data: a methodology for designing an intelligent monitoring system for the diabetic foot

Chanjuan Liu,^a Jaap J. van Netten,^b Marvin E. Klein,^c Jeff G. van Baal,^a Sicco A. Bus,^{b,d} and Ferdi van der Heijden^a

^aUniversity of Twente, Signals and Systems Group, Faculty Electrical Engineering, Mathematics and Computer Science, The Netherlands

^bHospital Group Twente, Diabetic Foot Unit, Department of Surgery, Almelo, The Netherlands

^cDEMCON Advanced Mechatronics BV, Enschede, The Netherlands

^dUniversity of Amsterdam, Department of Rehabilitation, Academic Medical Center, The Netherlands

Abstract. Early detection of (pre-)signs of ulceration on a diabetic foot is valuable for clinical practice. Hyperspectral imaging is a promising technique for detection and classification of such (pre-)signs. However, the number of the spectral bands should be limited to avoid overfitting, which is critical for pixel classification with hyperspectral image data. The goal was to design a detector/classifier based on spectral imaging (SI) with a small number of optical bandpass filters. The performance and stability of the design were also investigated. The selection of the bandpass filters boils down to a feature selection problem. A dataset was built, containing reflectance spectra of 227 skin spots from 64 patients, measured with a spectrometer. Each skin spot was annotated manually by clinicians as “healthy” or a specific (pre-)sign of ulceration. Statistical analysis on the data set showed the number of required filters is between 3 and 7, depending on additional constraints on the filter set. The stability analysis revealed that shot noise was the most critical factor affecting the classification performance. It indicated that this impact could be avoided in future SI systems with a camera sensor whose saturation level is higher than 10^6 , or by postimage processing. © 2013 Society of Photo-Optical Instrumentation Engineers (SPIE) [DOI: 10.1117/1.JBO.18.12.126004]

Keywords: diabetic foot ulcer; hyperspectral imaging; skin spectrum; feature selection; filter selection paper.

Paper 130655R received Sep. 9, 2013; revised manuscript received Nov. 3, 2013; accepted for publication Nov. 6, 2013; published online Dec. 10, 2013.

1 Introduction

Diabetes mellitus (DM) is one of the most common chronic diseases worldwide and continues to increase in population and significance.^{1,2} Vascular and neurological disorders in the lower extremities are common complications of DM. Approximately 15% to 25% of patients with DM eventually develop foot ulcers.³ If not adequately treated, these ulcers may ultimately lead to lower extremity amputation.⁴ The onset of diabetic foot ulcers is preventable by early detection and timely treatment of (pre-)signs of ulceration, such as callus formation, inflammation, and blisters. However, early detection depends on frequent assessment. This is not always possible, because self-examination is difficult or impossible due to the various health impairments caused by DM, and a sufficiently frequent examination by health care professionals would be too intrusive and too costly. Previous studies have shown the feasibility of such frequent assessment with help of telemedicine.⁵ However, digital photography, as applied in those studies, misses the opportunity for automatic detection, thereby still requiring help from health care professionals. The ultimate objective of our project is to develop an intelligent telemedicine monitoring system that can be deployed for frequent examination of the patients' feet to timely detect (pre-)signs of ulceration. That is, we want to automatically detect (pre-)signs of ulceration in the diabetic foot, and discriminate these from healthy skin. No such system is currently available.

Spectral imaging (SI) is a promising modality to achieve this objective. The SI is a hybrid modality for optical diagnosis, which obtains spectral data of the entire measured area and renders it in image form.⁶ Medical hyperspectral imaging (MHSI) provides a noncontact and noninvasive diagnostic tool that is able to quantify relevant features such as tissue oxygenation and epidermal thickness.⁷ Impairments in these features play an important role in the development of diabetic foot ulcerations and their subsequent failure to heal.⁸ Based on these features, studies have been conducted with the MHSI to assess the risk for ulcer formation,⁷ to monitor diabetic neuropathy,⁹ and to predict ulcer healing.¹⁰ Besides these three options, an intelligent monitoring system should also be capable of automatically recognizing ulcers and pre-signs of ulceration, such as abundant callus formation, inflammation and blisters, and discriminate these skin spots from healthy skin. This can be achieved with the help of statistical pattern classification techniques. The SI has not yet been used for automatic recognition of (pre-)signs of ulceration in the diabetic foot with statistical pattern classification techniques.

1.1 Statement of the Problem

The aim of the current article is to describe the methodology for designing an intelligent monitoring system involving SI for the diabetic foot, with spectral data acquired from spectrometer measurements on skin spots. Such a design is hampered by the lack of a physical model signifying the difference between spectral data coming from skin spots that are healthy or (pre-)

Address all correspondence to: Chanjuan Liu, University of Twente, Signals and Systems Group, Faculty Electrical Engineering, Mathematics and Computer Science, The Netherlands. Tel: +31 (0) 53 489 2897; Fax: +31 (0) 53 489 1060; E-mail: c.liu@utwente.nl

signs of ulceration (e.g., ulcers, abundant callus formation, inflammation, blisters, etc.). The alternative is an approach that exploits the statistical properties of a training set of spectral data, which has been annotated by an expert clinician. This immediately confronts us with the problem arising from the high-dimensionality of the spectral data acquired by current available MHSI systems. Exploiting the statistical properties by training a pixel classifier using high-dimensional MHSI data tends to overfit,¹¹ i.e., to adapt the classifier to accidental, statistically irrelevant peculiarities in the training set. This problem can be solved by a firm reduction of the dimensionality, by reducing the number of spectral bands (filters) employed in the SI system. Another advantage of such a reduction is that it will also increase the measurement and cost efficiency of the future system.

The methodology presented in this article will consist of determining suitable subsets of optical filters for the desired SI system by investigating their performance in discriminating different skin spots. Following, the stability of the selected filters in a simulated SI system will be analyzed with respect to shot noise and small changes of the experimental parameters.

1.2 Outline of the Article

Figure 1 presents an overview of the followed methodology and an outline of the article. In Sec. 2, the materials for collecting local skin spectral data on feet of diabetic patients with (pre-) signs of ulceration, the measurement procedure and resulting data set are described. In Sec. 3, the statistical analysis is performed on the acquired data to search for suitable subsets of commercial interference bandpass filters. With the selected filters, the stability of the virtual SI system with respect to shot noise and small changes of the experimental parameters is investigated with a Monte Carlo method in Sec. 4. Finally, conclusions and future plans are presented in Sec. 5.

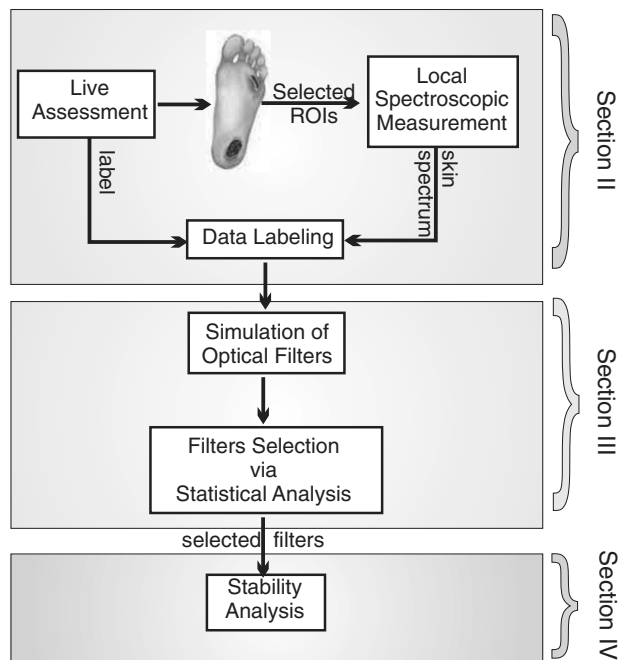


Fig. 1 The schematic representation of the methodology in this study.

2 Collection of Training Data

2.1 Data Acquisition

A spectroscopic system has been built to collect reflectance spectral data from the foot soles of patients with complications caused by diabetes. The experimental setup, as shown in Fig. 2, consists of (1) a spectrometer, (2) a broad-band illumination source, and (3) a reflection fiber optics probe. The reflection probe (6.5-mm diameter) contains 7 fibers: 6 illumination fibers (Fiber 3a) surrounding 1 reader fiber (Fiber 3b). The probe is mounted in a probe holder made from black anodized aluminum, which is placed on the target skin area. It fixes the distance between the probe end and target skin area at 5 mm. In this way, ambient light is blocked and a reproducible illumination under an angle of 45 deg is achieved. The coaxial arrangement of illumination and recording fibers minimizes the influence of superficial scattering and specular reflection on the spectra. The recorded skin spectra cover the wavelength range from 360 to 1000 nm and contain 1127 channels with a bandwidth ranging from 0.6 to 0.53 nm. The measurement area of the skin is around 60 mm², which is sufficient for small (pre-)signs of ulcerations or a small sample of bigger (pre-)signs of ulcerations.

Prior to the acquisition of *in vivo* skin spectra data for each patient, the dark current, $D(\lambda)$, of the spectrometer is recorded when the light source is off. In addition, a spectrum $S_{\text{ref}}(\lambda)$ of a white spectral standard (WS-2, Avantes, Apeldoorn, The Netherlands) is measured. This standard has a near-constant, high diffuse reflectance over the entire wavelength range, which is defined as 100%.

The diffuse reflectance spectrum, $R(\lambda)$, of the measured skin region of interest (ROI) is then calibrated according to:

$$R(\lambda) = \frac{S(\lambda) - D(\lambda)}{S_{\text{ref}}(\lambda) - D(\lambda)} \times 100\%, \quad (1)$$

where $S(\lambda)$ is the measured raw spectrum for the ROI.

2.2 Patients Recruitment and Live Assessments

Patients were recruited from the multidisciplinary diabetic foot clinic of the Hospital Group Twente, Almelo, the Netherlands. The patients included in this study were all diagnosed with diabetes and showed (pre-)signs of ulceration or had a history of

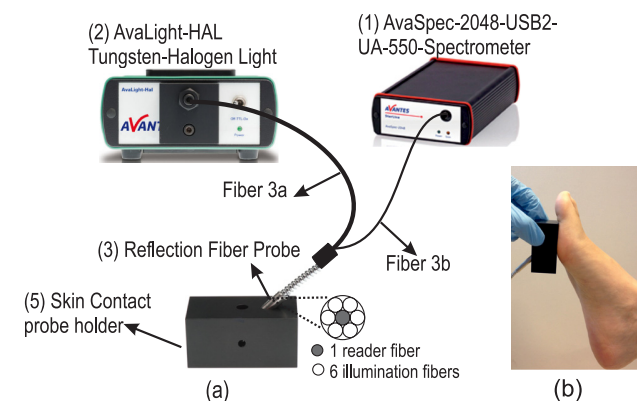


Fig. 2 Schematic of the experimental setup for local spectral measurements.

ulceration. The average age of the 64 included patients was 65.7 years (SD = 11.7; range 35 to 95).

The reflectance spectra were collected from skin regions selected by experienced clinicians classified with the following live assessment scores:

1. The ROIs on foot sole without (pre-)signs of ulceration. These were scored as “Healthy.”
2. The ROIs on foot sole with ulceration. These were scored as “Ulcer.”
3. The ROIs on foot sole with abundant callus. These were scored as “Callus.”
4. The ROIs on foot sole with other less frequent pre-signs of ulceration, such as redness, fissure, blister, and necrosis. These were scored as “Rest.”

During the measurement, the probe holder was placed on the skin in two directions, which were along the width of the foot and the orthogonal direction. Each ROI was measured two times in each direction. In some cases, when the areas of the ROIs were smaller than 60 mm² or on the edge of the foot, only two measurements along the first direction were performed. At least one “Healthy” ROI was measured on each foot of a patient. The final spectra of each ROI was obtained by taking the average of all the four or two measurements.

2.3 Summary of the Acquired Dataset

Local reflectance measurements were performed on 227 ROIs from 64 patients, among which there were 95 “Healthy,” 55 “Callus,” 38 “Ulcers,” and 39 “Rest.” The “Rest” category bundled several very diverse skin conditions, each of which occurred only infrequently in this study (e.g., 6 necrosis, 5 redness, 3 blister, etc.). In addition, the corresponding areas of some of these ROIs were much smaller than the 60 mm² measurement area, so that no pure spectra could be sampled from them. As a consequence, the within-class variation of the data in the “Rest” category was very large. For this reason, the initial statistical analysis, presented in Sec. 3, only focused on the data from the first three classes, which contained 188 of the measured 227 spectra. Examples of the measurements are given in Fig. 3.

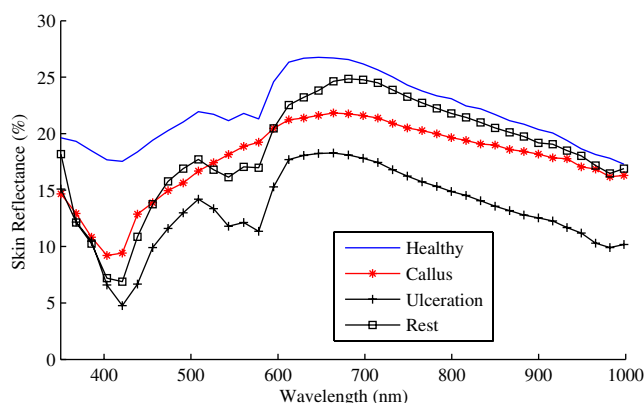


Fig. 3 Examples of spectrum from different classes.

3 Statistical Analysis for Selection of the Optical Filters

In this section, the statistical analysis will be performed on the acquired skin spectra to search for the most informative optical filters for the future SI system.

3.1 Methodology for Optical Filter Selection

The functional structure of the statistical analysis is shown in Fig. 4. The database, consisting of the measured and manually classified skin spectra, is fed into a simulator of a large bank of optical filters. The characteristics of these virtual filters match the specifications of real, commercially available bandpass filters. The output data of all simulated filters, as well as the original measured data, is also applied directly to “Classification 1” and “Classification 2” to find the baseline performances, which can help “Filters Selection” to make decisions over the performance of the selected filters. The “Filters Selection” put all the outputs of the simulated filters in a pool to find the (sub-) optimal subset of filters for classifying the measured spectral data. In the subsequent process step “Classification 3”, different classifiers are applied to find the most appropriate one that yields the best performance (i.e., accuracy of classification).

To make a comparison between the various selections of filters and classifiers, a relevant measure of the performance, such as the error rate (probability of misclassification), is essential. Since overfitting (resulting in a too optimistic estimate of the error rate) can easily occur if the same data is used both for training and for testing, k -fold cross validation and/or leave-one-out cross validation¹² is applied to avoid that.

The design of a classifier involves two aspects:¹² (1) selection of the type of classifier and (2) the training of the parameters of the classifiers. The type of classifier (linear, quadratic, Parzen, etc.) defines the structure of the decision function that maps the features to assigned classes. The more complex the mapping is, the more parameters are needed to be estimated to model the mapping, and the more the classifier is susceptible to overfitting. The complexity of a well designed classifier is a compromise between its ability to adapt its decision boundary to the requirements imposed by the statistical properties of the features at hand and its ability to withstand overfitting.

There are two common approaches to reduce complexity, which are feature extraction (FE) and feature selection (FS). The FE approach extracts a set of new features using a simple function that maps the original N -dimensional feature space to a lower dimensional space. The features extracted do not have clear physical meanings. The FS approach only selects a subset of existing original features without any mapping. Thus, the FS is the approach that is useful for finding the optimal subset of commercial available optical filters, as presented in Sec. 3.4, whereas the FE will be used in Sec. 3.3 as a reference.

3.2 Simulation of Optical Filtering

In total, 148 interference bandpass filters were simulated, where the central wavelength λ_c and the bandwidth W_d of these filters were provided by real filters from two manufacturers: Edmund Optics (York, UK)¹³ and Comar Optics (Cambridgeshire, UK).¹⁴ The λ_c 's changed from 400 to 1000 nm and the W_d 's varied between 10 and 75 nm. An overview of the parameters of all simulated filters is given in Fig. 5.

In addition to the parameters that are specified by the manufacturers, a third parameter W_t was assumed to model the

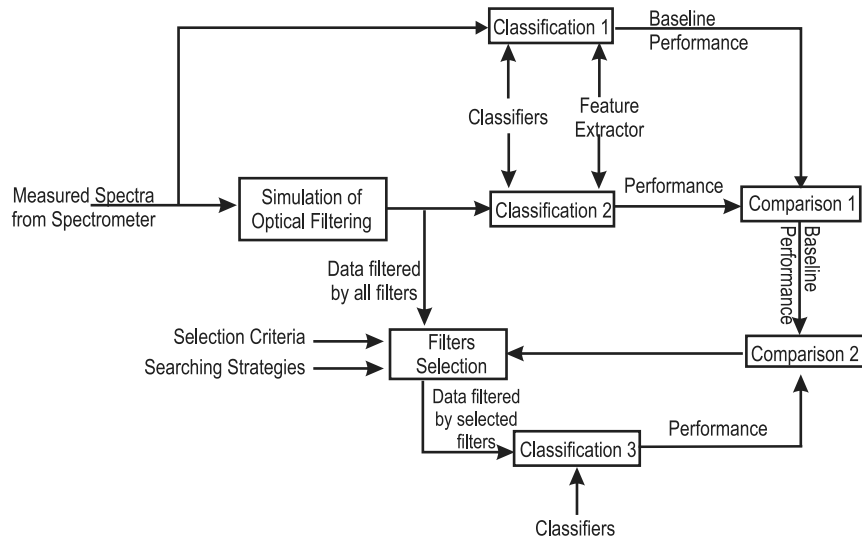


Fig. 4 The functional structure of the data analysis that is applied to find the optimal subset of optical filters.

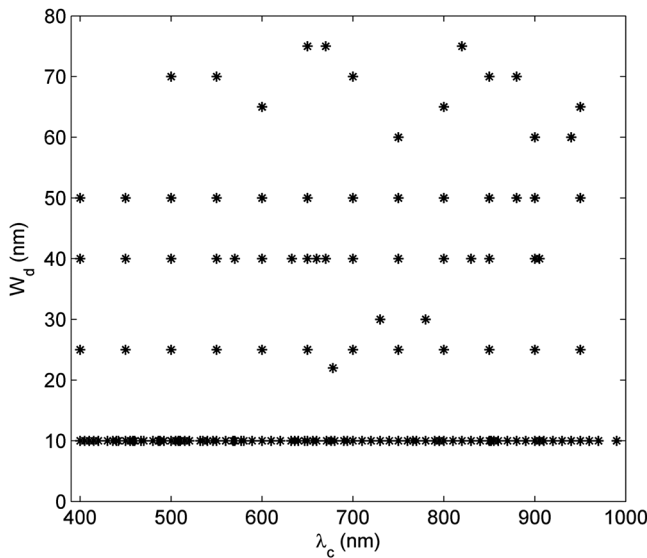


Fig. 5 The parameters of the simulated filters. Each point represents a filter.

steepness of the filter. W_t is the transition width between zero transmission and maximum transmission, which was set to $W_t = 10$ nm. Figure 6 shows the modelled shape of the transfer function $w(\lambda)$, where W_d represents the full width at half maximum and the transition zone is modelled by half a cosine function.

The $M_S = 188$ spot spectra in the data set were denoted by $R_m(\lambda)$, where m is the ROI number, $m = 1, \dots, M_S$. The $N = 148$ filter transfers are denoted by $w_n(\lambda)$. The filter simulation is accomplished by a numerical approximation of:

$$z_{nm} = \int w_n(\lambda) R_m(\lambda) d\lambda \quad \begin{cases} m = 1, \dots, M_S \\ n = 1, \dots, N \end{cases} \quad (2)$$

The filter outputs z_{nm} of each measured area are arranged in M_S N -dimensional vectors \mathbf{z}_m , called feature vectors. The true class for each ROI is available from the diagnosis by

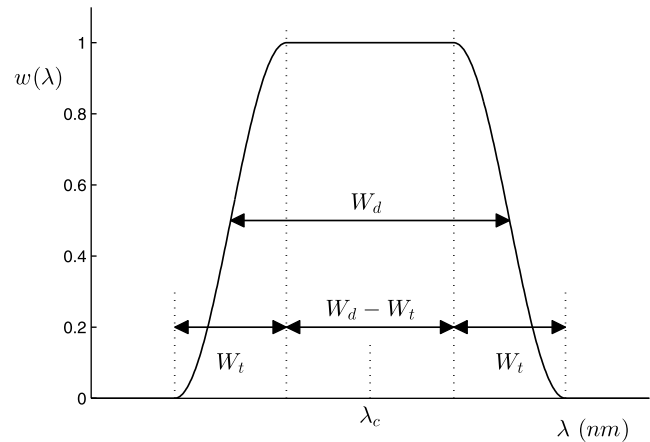


Fig. 6 The modelled filter transfer and the definition of parameters.

experienced clinicians. This yields a labelled training set of $M_S = 188$ ROIs, K classes ($K = 3$, “Healthy,” “Ulcer,” and “Callus”), and with each spot an associated N -dimensional feature vector ($N = 148$).

3.3 Baseline Performance

In order to find a reference for the effects of filter selection, first the performance of classification was assessed with the original measured data and when all filters were used. Direct application of a linear classifier to the full feature vectors ($N = 148$) yielded an error rate as high as 70%. Other classifiers yielded likewise performances. The explanation for this behaviour is overfitting. According to a simple rule of thumb,¹⁵ to avoid overfitting in a linear classifier, the number of samples should be at least about five times the number of features (= number of filters). The present case did not fulfill this, and the number of features should be reduced before applying classification. Linear FE was applied to accomplish that.

Linear FE is a linear mapping from \mathbb{R}^N to \mathbb{R}^D with $D < N$ implemented by a $D \times N$ matrix \mathbf{W} :

$$\mathbf{y}_m = \mathbf{W}\mathbf{z}_m \quad (m = 1, \dots, M_S). \quad (3)$$

Two methods are commonly used: linear principal component analysis (PCA) and linear discriminant analysis (LDA). The goal of the PCA is to find the D -dimensional linear subspace of \mathbf{z}_m that preserves as much variance in the data as possible. Often, D is chosen such that, for instance, 99% of the variance is still in the linear subspace. The PCA projection was always scaled by the mean class standard deviations in this article.

The LDA is a feature reduction method that takes class information into account. If there are K clusters in an N -dimensional space, the linear subspace that separates the clusters optimally is at most $D = K - 1$ dimension. Hence, for a data set with K classes, the LDA seeks to determine the $K - 1$ linear subspace to preserve as much class-discriminatory information as possible: it simultaneously tries to maximize the inter-class distance and to minimize the intra-class distance.

If the original feature space is high-dimensional, and the training set small (as in our case), the LDA must be preceded with PCA to prevent numerical instability.¹² The dimensionality preserved in the preceding PCA should be optimized by cross

validation. In our case, the optimal PCA subspace of the 148-D space is a 28-D space, which is further reduced to a 2-D space by LDA. The features extracted as such are illustrated in the scatter diagram in Fig. 7(a). In Fig. 7(b), the figure also shows the scattering of the two most important PCA features. These two features hold 98% of the variance.

When the PCA and/or the LDA is applied to the filtered data, as in Eq. (3), each extracted feature is a linear combination of the original filter outputs. Therefore, an extracted feature can be regarded as the output of a single virtual filter. If the elements of \mathbf{W} are denoted by W_{dn} , and the d 'th element from \mathbf{y}_m by y_{dm} , then Eq. (3) can be written as:

$$y_{dm} = \int \sum_{n=1}^N W_{dn} w_n(\lambda) R_m(\lambda) d\lambda. \quad (4)$$

This effectively implements an optical filter with transfer function:

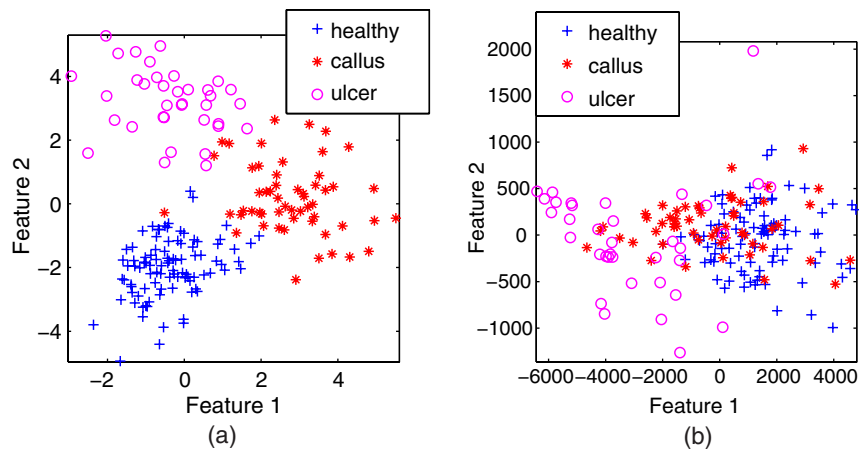


Fig. 7 Scatter diagrams in two different feature spaces. (a) Shows the feature space that is obtained by linear discriminant analysis (LDA) [preceded by principal component analysis (PCA)]. (b) Shows the space of PCA features solely.

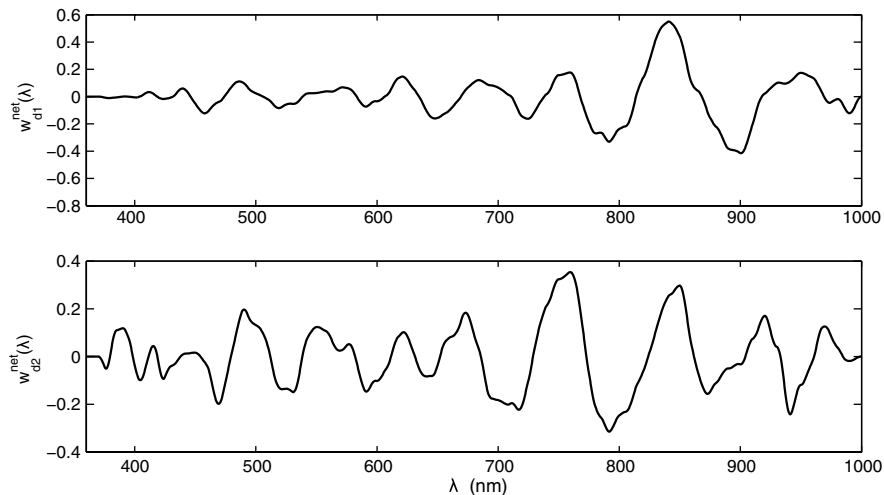


Fig. 8 The net optical filter transfers for the two LDA components normalized to [0,1].

$$w_d^{\text{net}}(\lambda) = \sum_{n=1}^N W_{dn} w_n(\lambda). \quad (5)$$

The net optical filter transfers for the two LDA components are illustrated in Fig. 8.

As shown in Fig. 7, the three classes are discriminated well by the two LDA features, indicating that a simple linear classifier suffices. We verified that by comparing its performance with the ones of a quadratic classifier, a Parzen classifier, a K-nearest neighbor classifier and support vector machine.¹² In all cases, the parameters were optimized using 10-fold cross validation. Besides, 10-fold cross validation was also used to assess the performance. Indeed none of them could out-perform the linear classifier, i.e., they all had similar performances, with a mean error rate 9%. Taking the time consumed for training and classification into consideration, the linear and quadratic classifiers were chosen for the filter selection.

With LDA as the FE and a linear classifier, the error rate of the full set of filtered data was $9\% \pm 1\%$ (obtained with 10-fold cross validation). Exactly the same result was obtained with 28 PCA extracted features directly classified by the linear classifier, instead of indirectly via LDA. Following the same procedure to extract 2 features using LDA (preceded by PCA), the original measurement data (1127 spectral measurements per ROI) achieved an error rate that was almost the same. The averaged error rate differed $<0.5\%$. We conclude that the simulated filtering did not cause much information loss/gain, and an error rate of $9\% \pm 1\%$ would be used as the baseline, i.e., the reference performance. For illustrative purposes, both of the error rates, from the original measurements (1127-D space) and the data after filtering (148-D space), will be presented in the following sections.

3.4 Optical Filter Selection

To select filters with the FS techniques, i.e., to find a subset of features corresponding to the full set of filters, two choices have to be made: the search strategy to select candidate subsets, and the search criterion to evaluate these candidates.

Among the available searching strategies,^{16,17} the optimal search methods are not applicable, because their searching complexity grows combinatorially. The suboptimal strategies build

up the subsets either incrementally (forward), or decrementally (backward). When searching for a relative small subset (<10 filters), backward selection is not adequate. Among the three forward strategies, which are sequential forward selection, plus- l -take-away- r selection, and sequential floating forward selection (SFFS), the most versatile approach, the SFFS, was chosen.¹⁷

The search criterion can be either a measure of the statistical distance of the class differences in the data, or it can be the estimated error rate of a classifier that is applied to the subset, the so-called wrapper. In this article, one measure, the Mahalanobis distance, and two wrappers, the linear discriminant classifier (LDC) and the quadratic classifier (QDC) are studied. Both wrappers are evaluated with leave-one-out cross validation to avoid overfitting during searching.

To make a comparison between the various selections of filters and classifiers, a relevant measure of the performance, such as the error rate (probability of misclassification), is essential.

The main light-absorbing factors are melanin in the epidermis, and oxy-haemoglobin (HbO₂) and deoxy-hemoglobin (Hb) in the dermis. The local concentration of each of these factors influences the measured skin spectrum in a specific way. In particular, HbO₂ absorption features a strong maximum around 412 nm and additional maxima at 542 and 577 nm, while Hb exhibits peaks at 430 and 555 nm.¹⁸ Thus, these five wavelengths could provide more information about physical parameters, e.g., the epidermal thickness and the oxygen saturation.^{19,20}

The filter selection was carried out in two different ways: (1) all features/filters had the same eligibility and (2) a constrained selection procedure was done in which center wavelengths, around the five wavelengths listed above, were evaluated and preselected with LDC. Once a subset of filters has been found, the subset must be evaluated empirically. This can be done with any type of classifier. Preliminary research showed that parametric-free classifiers, e.g., Parzen, did not perform better. Hence, we selected LDC and QDC once again for empirical evaluation. 10-fold cross validation was used to assess the error rate.

3.4.1 Filter selection without preselection

The evaluation results are shown in Fig. 9. The SFFS strategy may return with a smaller number of features than the number

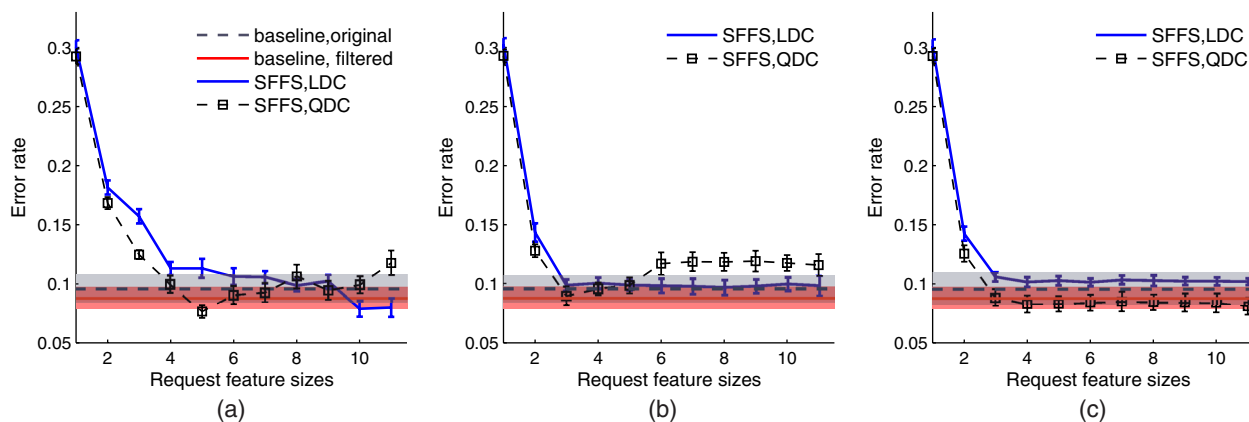


Fig. 9 Classification accuracy for sequential floating forward selection (SFFS) with different criteria, linear discriminant classifier (LDC) as classifier (solid lines) and quadratic classifier (QDC) as classifier (dashed lines) for empirical evaluation. The gray color bands represent baseline error rate achieved with features extracted from LDA with LDC as classifier with the original spectra data (1127-D), while the red colour bands represent the baseline achieved with the features extracted from LDA with LDC as classifier with the spectra data after filtering (148-D).

that is requested when it detects that addition of features does not improve the search criterion. Six features were returned when using LDC as the searching criterion, while four features were returned when using QDC as the searching criterion. When comparing the performances, the observations are:

- (1) With a set of appropriate chosen filters, the performance of all criteria compares well with the baseline performance.
- (2) The performance associated with the Mahalanobis search criterion converges slower than the other criteria. Apparently, the Mahalanobis criterion finds less strong features.
- (3) The Mahalanobis search criterion, evaluated with QDC, yields the lowest error rate at five filters.
- (4) The performances of LDC and QDC search criteria peak at three and four filters, respectively. Adding more filters does not improve the performance significantly, or could even worsen the performance.
- (5) For a smaller number of filters, the QDC classifier generally performs better, indicating that the classes are better separated by quadratic decision functions than linear ones.

The found subsets are given in Table 1. As can be seen in Fig. 10, each subset seeks information in the range of 550 to 600 nm, and the LDA features are formed by differentiating filter outputs. This results in a large impact on noise sensitivity, to be discussed in Sec. 4. The Mahalanobis distance seeks additional information in a range around 400 nm. The LDC and the QDC both find information in the range around 650 nm useful.

3.4.2 Filter Selection with pre-selection

In addition to the five filters predefined, this constrained FS procedure selected up to six other filters from the rest pool. The results are shown in Fig. 11. Ten-fold cross validation was used to assess the error rate.

Again, the QDC as an evaluation classifier often performed better than the LDC. The best results were obtained when QDC was used for the selection criterion as well as for empirical evaluation [Fig. 11(c)]. With 10 features, the number of ulcer ROIs (i.e., 38) is somewhat less than four times the dimension, whereas it should be at least five times to prevent overfitting, according to the rule of thumb.¹⁵ Apparently, it paid off to

Table 1 The best subsets of filters (unconstrained search). Each filter is defined by the parameter set " λ_c (W_d)" given in nm.

Search criterion	Filter ₁	Filter ₂	Filter ₃	Filter ₄	Filter ₅
Mahalanobis	570 (10)	570 (40)	589 (10)	415 (10)	400 (40)
Linear discriminant classifier (LDC)	570 (10)	550 (25)	647 (10)	—	—
Quadratic classifier (QDC)	632.8 (40)	570 (10)	550 (25)	690 (10)	—

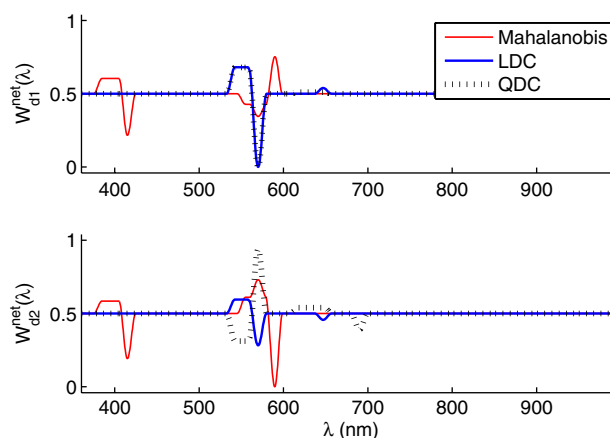


Fig. 10 The net optical filter transfers that produces the two LDA components of the 3-class problem. These transfers, normalized to [0,1], correspond to three feature subsets as given in Table 1.

allow quadratic decision boundaries (rather than linear), although the additional parameters for the quadratic boundaries resulted in some performance loss due to overfitting.

The QDC in Fig. 11(c) obtains a minimum with seven features. The performance improvement with the additional three features is not statistically significant. The filter parameters (λ_c and W_d , in nm) that are associated with these features are, where the first five are the predefined filters: 540 (10), 430 (10), 400 (50), 580 (10), 550 (40), 730 (30), 410 (10).

3.4.3 The "Rest" class

Up to now, the ROIs in the "Rest" category were ignored. As said before, the "Rest" consisted of a number of different subclasses that were often formed by tiny spots closely surrounded by healthy skin, callus, or ulcer. To see how the rest class behaves, Fig. 12 shows a scatter diagram of all spots in the posterior probability space that is spanned by $P(\text{callus}|\mathbf{y})$ and $P(\text{ulcer}|\mathbf{y})$. Here, \mathbf{y} represented the 7-D feature vector that was tabulated in Sec. 3.4.2.

The posterior probabilities were calculated under the assumption of class dependent, Gaussian likelihood functions, as the base of the quadratic classifier.¹² The quadratic decision boundaries in the feature space are mapped to linear boundaries in the probability space.

When "Callus" and "Ulcer," with or without "Rest," were conflated to one class "Unhealthy," the estimated performances of the subsets selected are listed in Table 2. When "Rest" is included, the sensitivities dropped from $\sim 96\%$ to $\sim 90\%$ and AUC decreased with 0 to 0.2, while the specificities all raised $\sim 1\%$. Based on the table, all subsets selected with different criteria did similar and sufficient good work in classification. It is still hard to make a choice among these selected subsets.

3.5 Discussion of the Optical Filters Selection

The four subsets of filters that are presented all operate in the wavelength range around 550 nm. Two selections also operate in the vicinity of 400 nm. These ranges are all in the biomedical window. None of the selections operate above 750 nm. There is some resemblance between the filters that are based on the LDC and the QDC criteria, but apart from that, the different solutions do not have much intersection. The selection of filters is not very

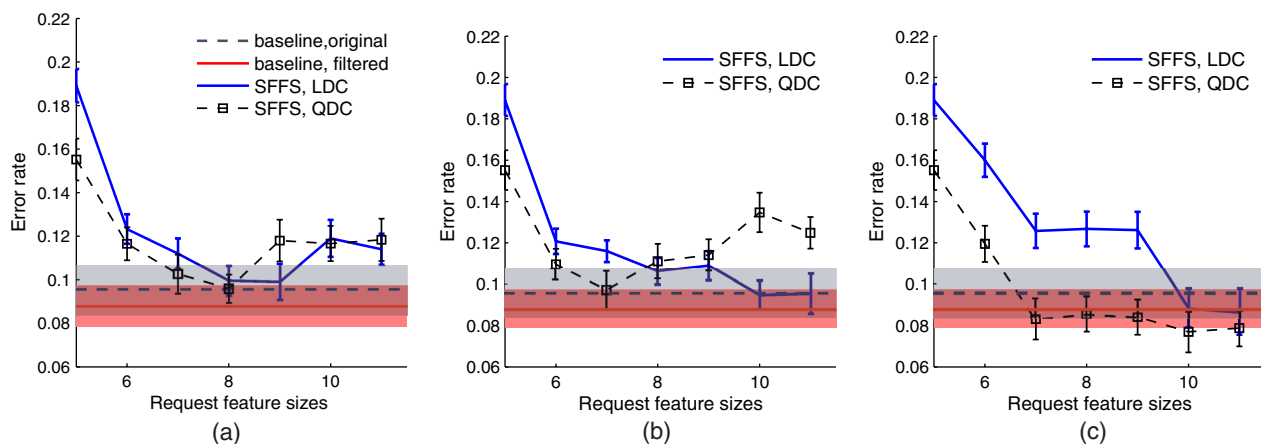


Fig. 11 Classification accuracy for the SFFS with different criteria, the LDC as classifier (solid lines) and the QDC as classifier (dashed lines). The feature selection started from 5 predetermined filters.

critical as different subsets perform equally well. In all subsets, the QDC outperforms the LDC. This suggests that the dispersion of the measurement data is class dependent. For class independent dispersion, the QDC degenerates into LDC. Since the dispersion of data is caused by the skin properties and not by measurement noise, a class dependent dispersion makes sense.

The “Rest” class is not included in the classifier design as yet. Figure 12 shows that the “Rest” ROIs are either situated at the three clusters that correspond with “Healthy,” “Ulcer,” and “Callus,” or situated in-between pairs of those clusters. The explanation is that the areas of “Rest” ROIs are smaller than the measurement area. As a consequence, the measured spectra represent a mixed response from the “Rest” ROIs and the surrounding areas belonging to one or several other classes. For instance, all six measurements for necrosis points in the “Rest” category were classified as “Ulcer,” as shown in

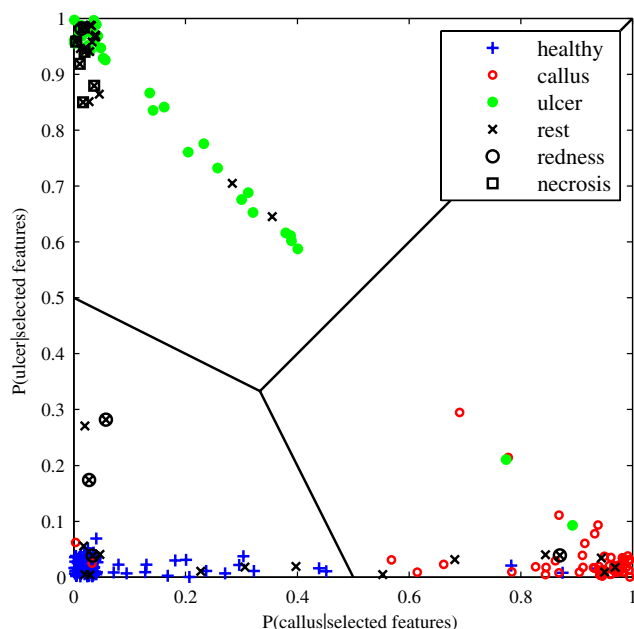


Fig. 12 The scattering of the data set in the posterior probability space and the decision boundaries, under assumption of equal prior probabilities and equal cost of misclassification.

Fig. 12. On the other hand, three out of five measurements from redness were classified as “Healthy,” one measurement was classified as “Callus” and the other as “Ulcer.” The same “mixed area” effect occurs in fact also for several ROIs labelled to the other classes. This also explains the phenomenon that the specificity rose when the “Rest” class was added, as shown in Table 2.

For a real spectral camera, the pixel size will be much smaller than 60 mm^2 , the spectrometer measurement area, and the problem will not occur. However, the classifier needs retraining then, because the rest class is not included as yet.

4 Stability Analysis

In the previous section, the randomness of the data set mainly stemmed from the variations in physical properties of the skin. All other factors were idealized, with transfer functions that were assumably exactly known, and without possible shot noise. This section addresses the stability of the design, i.e., its ability to withstand small deviations of the transfer functions, and the ability to withstand noise.

4.1 Monte Carlo Analysis

The stability analysis was carried out by means of a Monte Carlo analysis applied to three different configurations of filter subsets and classifiers:

Table 2 Performance of the filters selected.^a

Searching criterion	No. of filters	No “Rest” class			With “Rest” class		
		Sens ^b	Spec	AUC ^c	Sens	Spec	AUC
Mahalanobis	5	95%	96%	0.98	91%	97%	0.97
LDC	3	95%	92%	0.97	93%	93%	0.97
QDC	4	97%	94%	0.96	91%	95%	0.96
QDC ^d	7	97%	96%	0.99	89%	98%	0.97

^aResults obtained with 10-fold cross validation.

^bSens = Sensitivity and Spec = Specificity.

^cAUC = Area under Receiver operating characteristic curve.

^dWith preselection.

1. The subset of four filters that were selected without any constraints using SFFS/QDC as search strategy and with QDC as classifier [Fig. 9(c)].
2. The same subset as in 1, but with LDC as classifier.
3. The subset of seven filters that were selected with SFFS/QDC, but with the first five filters preselected in the biomedical window, and with QDC as classifier [Fig. 11(c)].

Configuration 1 was chosen since its performance and its parsimony were the best that could be obtained. Configuration 2 was added to see whether the quadratic classifier was less stable than a linear classifier. Configuration 3 was chosen to see whether a large number of filters (a less parsimonious design) would degrade the stability.

The Monte Carlo stability analysis was done as follows. After random sampling of the parameter of a configuration, the error rate of that configuration was estimated using 10-fold cross validation. This was repeated 100 times, and the average over these 100 estimates are shown in Fig. 13. The procedure is repeated for four parameters: the center wavelengths λ_c , the bandwidths W_d , the widths W_t of the transition band, and the maximum expected number of electrons. The latter parameter determines the signal-to-noise ratio (SNR) of the system.

The manufacturers specify tolerance regions for λ_c 's and W_d 's. For different W_d 's and different coating approaches, the tolerances may be different. For example, when $W_d = 50$ nm, the tolerance is ± 15 nm if traditionally coated, and

± 5 nm if hard coated. When $W_d = 10$ nm the tolerance is ± 2 nm. When $W_d = 25$ or 30 nm the tolerance is ± 3 nm.^{13,14}

Note that the transmission coefficient of the spectral filters, which varies from 30% to 90% for the types of filters considered here,^{13,14} need not be included in the stability analysis. This is because it can be assumed that the response of a given imaging system (including light source, image sensor, lens, and the filters) is calibrated for each filter using e.g., a white reference standard, and that it remains calibrated during the measurements. However, low transmission coefficients may influence the SNR.

The SNR of the imaging system may be another important aspect that affects the classification stability. The random count of electrons, which are generated at the image sensor during the exposure time, defines the image output and decides the shot noise.²¹ The probability of this count obeys Poisson distribution, and according to that property, the expectation of the count equals its variance, $E[\text{count}] = \text{Var}[\text{count}] = \Phi$. Then, the SNR can be expressed as

$$\text{SNR} = \frac{\Phi}{\sqrt{\Phi}} = \sqrt{\Phi}, \quad (6)$$

which grows with the expected number of electrons. So, Φ is another parameter of interest.

Each parameter was submitted to the Monte Carlo analysis by random sampling according to distributions of the parameters that modelled the tolerances:

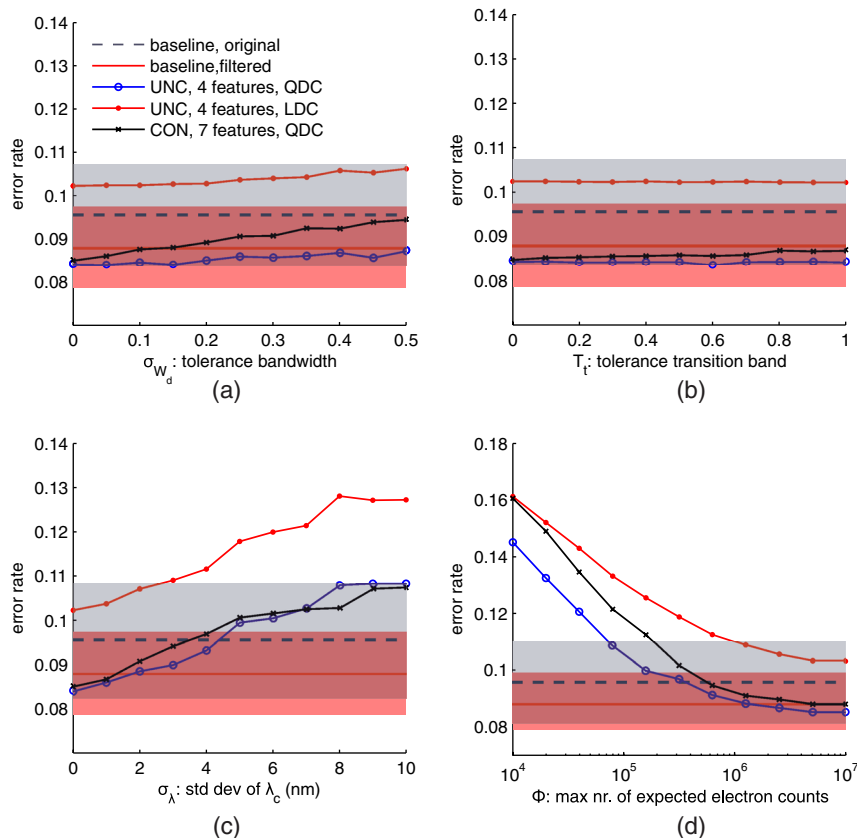


Fig. 13 The influence of tolerances of filter parameters and the influence of shot noise. The performance is given in classification error rate. “UNC” and “CON” present unconstrained and constrained selection, respectively.

- The central wavelengths were set to obey Gaussian distribution, with the nominal value as the mean, and a varying standard deviation σ_λ during the experiment.
- The bandwidth was modelled by $(1 + a)W_d$ in which a had a Gaussian distribution with zero mean, and varying standard deviation σ_{W_d} during the experiment.
- The variations of the transition bands were modelled by $(1 - a)W_t$ where the distribution of a was uniform between 0 and T_t . The parameter T_t was the tolerance of the transition band.
- The shot noise was modelled by the parameter Φ which was the the expectation of the electron count that corresponds to the maximum output y_{\max} (i.e., highest spectral reflectance value).

Then, all features are transformed to the expected electron count using the factor $K = \Phi/y_{\max}$. According to Eq. (6), Gaussian noise was added to each feature with a variance that equals Φ . Finally, the features were back-transformed by a factor $1/K$.

Figure 13 shows the results of the Monte Carlo analysis. The estimated errors all had standard deviations around 1%.

4.2 Discussion of Stability Analysis

When comparing the results with respect to the three configurations of filter subsets and classifiers in Fig. 13, no substantial differences in behavior can be observed. The linear classifier is not more stable than the quadratic classifier, and using only four filters instead of seven does not contribute much to stability either.

The shape of the filter transfer function is modelled by the bandwidth W_d and by the width W_t of the transition band. Figures 13(a) and 13(b) show that the classification performance is not sensitive to changes of these parameters. The results suggest that such an assumption is not very critical: with extreme values of W_t the modelled transfer functions vary between a rectangular window and a very smooth cosine window, but the performances are not affected.

Figure 13(c) shows that shifts of the center wavelengths of the filters may reduce the performance. If an increase of error rate of 1% is permitted (about equal to the uncertainty of the estimated error rate), the standard deviation of the shifts should not be more than about 4 nm. This requirement is equal to the specification of the low-cost commercial filters, so that it can be fulfilled easily when designing the SI system.

The most critical parameter appears to be the maximum expected electron count, Φ . The explanation is that the differences between the outputs of filters are important for classifiers to discriminate different classes (as illustrated in Fig. 10). This may drastically reduce the SNR, i.e., even a small amount of noise has a large impact on the difference. Figure 13(d) indicates that Φ should be above 10^6 to avoid a performance decline due to quantum limitations. In a hyperspectral camera, such a high density of electrons can be realized by (a) powerful illumination, (b) large exposure time, (c) large aperture of the optics, (d) large pixel size, (e) detective quantum efficiency close to 100%, and (f) transmission of the filters close to 100%. However, the ultimate limitation is defined by the saturation level of the image sensor. For a high-quality industrial camera module with 4 Mp resolution, a typical saturation level is 4×10^4 . The figure shows that at that level the increase of

error rate can be as high as 2%; the subset of seven filters a little higher than the one with four filters. If such a loss of performance cannot be tolerated, the effective electron count can be increased by using a sensor with a large pixel size, and by summing over neighboring pixels or subsequent frames. The price to pay for an increased pixel SNR is increased system costs, and reduction of spatial and temporal resolution.

5 Conclusion

The methodology for designing an intelligent monitoring system for the diabetic foot based on statistical pattern classification of spectral data has been proposed in this article. To avoid the possible overfitting problem arising from high-dimensional spectral data and to design a system with a limited number of optical filters, a FS procedure was applied. This design methodology can also increase the measurement and cost-efficiency of the future system. The number of needed filters in different subsets resulting from FS, ranged between 3 and 7. The sets were all able to automatically discriminate between (pre-) signs of ulceration and healthy skin spots with a specificity of 96% and a sensitivity of 97%. A stability analysis based on the Monte Carlo method revealed that the most critical design factor is the shot noise. The influence of the noise will be negligible if the saturation level of the image sensors is above about 10^6 electrons. If this cannot be achieved, temporal or spatial averaging in the image should be applied, at the cost of the resolution.

One of the limitations of this study was the small training data set in comparison to the abundant number of spectral bands. Such a set needs to be acquired during live assessment of patients using a spectrometer, and is expensive and therefore small. However, cross-validation during selection and evaluation helped in generalizing the analysis of the data set. Further, the ‘‘Rest’’ class could not be involved in the FS process, as this class was a repository of types of disorders that occur only infrequently. These spots were often so small that they fell below the resolution of the spot spectrometer. In a real SI device with a higher resolution this problem will not occur. Also, sensitivity and specificity were acceptable with levels around 90% with the rest class included.

The first future step to be taken is to actually build the SI system with the optical filters as determined in this article as an experimental setup first. Based on results obtained with such a setup, other relevant criteria concerning, for example, the ergonomics of the future system can be determined. These criteria will greatly influence the future system, and need to be investigated before issues such as hardware and costs can be extensively discussed.

Building an experimental setup containing the SI system with the optical filters described in this article will bring us one step closer to our ultimate goal: an intelligent telemedicine monitoring system that can be deployed for frequent examination of the patients’ feet to timely detect pre-signs of ulceration. Various options exist for application of such a system in daily clinical practice. For example, a system can be deployed at the home of high-risk patients who need close monitoring to prevent hospitalization. Another option may be placement of the system with the general practitioner, to facilitate weekly or monthly screening of both high- and low-risk patients in an automated manner. Future clinical studies are needed to investigate the most cost-effective application of an intelligent telemedicine monitoring system for the diabetic foot in daily clinical practice.

Acknowledgments

This project is supported by public funding from ZonMw, the Netherlands Organisation for Health Research and Development. The authors are very grateful for the patients who participated in the data collection, and the clinicians in Diabetic Foot Unit, Department of Surgery, Hospital Group Twente, Almelo, The Netherlands. We also appreciate Geert-Jan Laanstra of the Signals and Systems Group, Faculty of EEMCS, University of Twente, who gave us plenty support in building the experimental setup.

References

1. S. Wild et al., "Global prevalence of diabetes: estimates for the year 2000 and projections for 2030," *Diabetes Care* **27**(5), 1047–1053 (2004).
2. J. E. Shaw, R. A. Sicree, and P. Z. Zimmet, "Global estimates of the prevalence of diabetes for 2010 and 2030," *Diabetes Res. Clin. Pract.* **87**(1), 4–14 (2010).
3. A. J. Boulton et al., "The global burden of diabetic foot disease," *The Lancet* **366**(9498), 12–18 (2005).
4. J. Apelqvist et al., "International consensus and practical guidelines on the management and the prevention of the diabetic foot," *Diabetes/Metab. Res. Rev.* **16**(1), S84–S92 (2000).
5. C. E. Hazenberg et al., "Telemedical home-monitoring of diabetic foot disease using photographic foot imaging a feasibility study," *J. Telemed. Telecare* **18**(1), 32–36 (2012).
6. T. Vo-Dinh et al., "A hyperspectral imaging system for in vivo optical diagnosis," *IEEE Eng. Med. Biol. Mag.* **23**(5), 40–49 (2004).
7. D. Yudovsky et al., "Assessing diabetic foot ulcer development risk with hyperspectral tissue oximetry," *J. Biomed. Opt.* **16**(2), 026009 (2011).
8. A. Korzon-Burakowska and M. Edmonds, "Role of the microcirculation in diabetic foot ulceration," *Int. J. Lower Extremity Wounds* **5**(3), 144–148 (2006).
9. R. L. Greenman et al., "Early changes in the skin microcirculation and muscle metabolism of the diabetic foot," *The Lancet* **366**(9498), 1711–1717 (2005).
10. A. Nouvong et al., "Evaluation of diabetic foot ulcer healing with hyperspectral imaging of oxyhemoglobin and deoxyhemoglobin," *Diabetes Care* **32**(11), 2056–2061 (2009).
11. W. Li, "Locality-preserving dimensionality reduction and classification for hyperspectral image analysis," *IEEE Trans. Geosci. Remote Sens.* **50**(4), 1185–1198 (2012).
12. F. van der Heijden et al., *Classification, Parameter Estimation and State Estimation An Engineering Approach Using MATLAB*, John Wiley & Sons, Chichester, England (2004).
13. Edmund Optics Ltd, York, UK, "Edmund optics catalogue of the interference bandpass filters from 400 nm to 1000 nm," <http://www.edmundoptics.com/optics/optical-filters/> (25 November 2013).
14. Comar Optics Ltd-Photon House, Cambridgeshire, UK, "Comar optics catalogue of the interference bandpass filters from 400 nm to 1000 nm," <http://www.comaroptics.com/components/filters/interference-filters> (25 November 2013).
15. T. M. Cover, "Geometrical and statistical properties of systems of linear inequalities with applications in pattern recognition," *IEEE Trans. Electron. Comput.* **EC-14**(3), 326–334 (1965).
16. R. Kohavi and G. H. John, "Wrappers for feature subset selection," *Artif. Intell.* **97**(1–2), 273–324 (1997).
17. A. K. Jain, R. R. W. Duin, and J. Mao, "Statistical pattern recognition: a review," *IEEE Trans. Pattern Anal. Mach. Intell.* **22**(1), 4–37 (2000).
18. S. Prahl, "Optical absorption of hemoglobin," <http://omlc.ogi.edu/spectra/hemoglobin/index.html> (25 November 2013).
19. D. Yudovsky and L. Pilon, "Rapid and accurate estimation of blood saturation, melanin content, and epidermis thickness from spectral diffuse reflectance," *Appl. Opt.* **49**(10), 1707–1719 (2010).
20. D. Yudovsky et al., "Two-layer optical model of skin for early, non-invasive detection of wound development on the diabetic foot," *Proc. SPIE* **7555**, 755514 (2010).
21. P. Regtien et al., *Measurement science for engineers*, Chapter 10, pp. 277–307, Butterworth-Heinemann (2004).