

# Delineation of Malignant Skin Tumors by Hyperspectral Imaging Using Diffusion Maps Dimensionality Reduction

Valery Zheludev<sup>1,2</sup>, Ilkka Pölönen<sup>1</sup>, Noora Neittaanäki-Perttu<sup>3,4</sup>, Amir Averbuch<sup>2</sup>, Pekka Neittaanmäki<sup>1</sup>, Mari Grönroos<sup>3</sup>, Heikki Saari<sup>5</sup>

---

## Abstract

A new non-invasive method for delineation of lentigo maligna and lentigo maligna melanoma is demonstrated. The method is based on the analysis of the hyper-spectral images taken in-vivo before surgical excision of the lesions. For this, the characteristic features of the spectral signatures of *diseased* pixels and *healthy* pixels are extracted, which combines the intensities in a few selected wavebands with the coefficients of the wavelet frame transforms of the spectral curves. To reduce dimensionality and to reveal the internal structure of the datasets, the Diffusion Maps technique is applied. The averaged Nearest Neighbor and the Classification and Regression Tree (CART) classifiers are utilized as the decision units. To reduce false alarms by the CART classifier, the Aisles procedure is used.

*Keywords:* hyperspectral imaging, framelet, delineation, tumor, malignant

---

## 1. Introduction

In this paper, we introduce a non-invasive method for delineation of lentigo maligna (LM) and lentigo maligna-melanoma (LMM). Lentigo maligna is an early form of melanoma in which the malignant cells are confined to the upper part of the skin, the epidermis; hence it is often reported as *in-situ* melanoma.

---

<sup>1</sup>Department of Mathematical Information Technology, University of Jyväskylä

<sup>2</sup>School of Computer Science, Tel Aviv University, Israel

<sup>3</sup>Department of Dermatology and Allergology, Päijät-Häme Central Hospital, Finland.

<sup>4</sup>Skin and Allergy Hospital, Helsinki University Central Hospital, Finland

<sup>5</sup>VTT Technical Research Centre of Finland

Lentigo maligna melanoma is diagnosed when malignant melanoma cells have invaded the dermis and deeper layers of skin [1]. Both LM and LMM occur as a brown-to-black patch on a sun-damaged skin, typically in the facial area. Clinically LM cannot be differentiated from invasive LMM. Diagnosis is verified by histopathological evaluation of the skin biopsy samples. The lesion borders can be hard to define by eye, due to subclinical extension. Histopathological analyses of the removed lesion reveal the size of the lesion. Often a re-excision is required [2]. There is a need for a noninvasive method for tumor margin delineation before and during the surgery.

Eight lesions, 3 LMM and 5 LM, in 8 patients were included in the study. The study protocol followed the Declaration of Helsinki and was approved by the Ethics Committee of the Tampere University Hospital District, Finland. The study was conducted in Päijät-Häme Central Hospital, Department of Dermatology and Allergology during 2012-2013.

A hand-held hyperspectral imaging system was developed for the study. The hyperspectral sensor was produced by VTT Technical Research Centre of Finland [3]. This hyperspectral imager uses a wavelength area between 500 and 885 nm, which covers visible light and few wavebands in infrared.

The presented method is based on processing hyperspectral images. The research follows a manifold learning approach and supervised machine learning process. The process includes data preprocessing, definitions of training set, algorithm selection, training and evaluation with test set [4].

This article consists of five sections and two appendices. Section 2 provides a brief review of the related literature. In Section 3, we explain in details our hyperspectral imaging system and the data processing chain. Section 4 deals with the results from the executed clinical trials. Finally, in Section 5, we present our conclusions about how the presented method can affect the work of a clinician and about the desired further developments. Appendix A outlines the diffusion maps technique and embedding the new data into diffusion spaces by the Geometric Harmonics method. Outline of the CART algorithm is given in Appendix B.

## 2. Literature review

In the last years, medical applications of multispectral (MSI) and hyperspectral imaging (HSI) became a field of extensive research.

Different medical applications of HSI are reviewed in the recently published paper [5] by Lu and Fei. They outlined multiple medical applications, where HSI was utilized either for the disease diagnosis or for the surgical guidance. Two publications, [6] and [7], which were concentrated on in vivo human and histologies studies of the skin cancer, were cited in the review. In the paper by Dicker et al. [6], the differences between benign and malignant dermal tissue at different stages of treatment were studied. Hennesy et al [7] studied melanoma's segmentation from diffuse reflectance images. They utilized the principal component analysis (PCA) to reduce data's dimension and K-means algorithm for clustering. In their investigation, they used phantoms rather than the patients examination.

A melanoma screening system based on HSI was designed and tested by Nagaoka et al [8]. Their hardware consists of the traditional pushbroom scanner and fiberscope. In order to detect the difference between the healthy and melanoma affected skin, they derived an index related to the spectral angle mapper from the spectral library database.

Quinzán et. al. [9] studied band selection in non-invasive melanoma diagnosis. They utilized liquid crystal tunable filter for the spectral separation. For the band selection they used sequential forward floating selection procedure, which selects bands. The procedure is based on pattern recognition principles. They treat band selection problem as a feature selection problem in pattern recognition. Once the band selection is accomplished, the support vector machine is used for the classification. They indicated that band selection improve the classification sensitivity.

Diebele et al. [10] report the results of clinically evaluation their MSI system aimed to separate differences between melanoma and common nevi. They utilized Nuance EX multispectral imager to take spectral images. They use quite

simple parameter in order to distinguish between the melanoma and the nevus skin patches.

A number of studies were focused on modelling the skin's layered structure and utilizing the models for calculation of the spectral images inversion. For example, Tseng et. al. [11], Jolivot et. al. [12], and Galeano et. al. [13] studied different kinds of optical models for the skins reflectance. Basically, those models return thorough inversion maps for different parameters, such as the melanin concentration, the haemoglobin concentration, the water concentration and the dermis thickness.

A number of in vitro studies done with hyperspectral camera in dermatology are reported in Lue et al. [14], Akbari et al.[15], and Dicker et al. [16], to name a few.

*Specific features of the presented research:*. Our investigations differ of those reported in the above publications in a number of aspects.

1. We use the spectral imager based on Fabry-Perot interferometer, which takes the whole spatial plane image at once. This provides a significant advantage over traditional pushbroom scanning, during which either a patient or the camera (or both) can change its position.
2. Our analysis of data consists of blocks, which are conventional in the pattern analysis . These are band selection, features' extraction, dimensionality reduction and finally manifold learning approach. But methods we use are either novel or more sophisticated than those, which are reported in the above publications.
3. We implemented clinical trials with real patients under strict ethical conditions. Our results were subject to both clinical and histological evaluation, which is not the case in all the papers above.

### **3. Method**

A hyperspectral imaging system was developed, which consists of the hand-held hyperspectral imaging device and the data processing unit. Figure 1 illus-

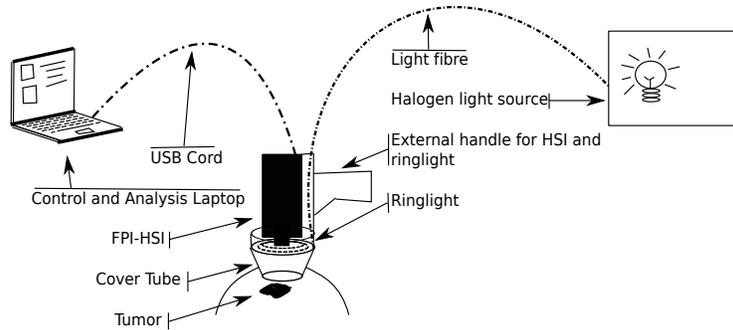


Figure 1: Hyperspectral imaging system setup.



Figure 2: Hand-held hyperspectral imager in holder. Ringlight is in front of the imager.

trates the system.

### 3.1. Imaging device

Unlike a regular camera, which captures pictures in three broad wavebands of red, green and blue, hyperspectral cameras capture pictures in dozens or even hundreds narrow wavebands. Our imaging device consists of a hyperspectral imager (FPI-HSI) based on the Fabry-Perot interferometer, a halogen light source, a ringlight and a 3D printed holder for the imager and the ringlight. The hand-held hyperspectral imager is presented in Figure 2. In front of the holder there is a cover tube, the purpose of which is to block background illumination.

Table 1: Specifications of Fabry-Perot hyperspectral imager

Parameter	Hand-held
Horizontal and vertical FOV (deg.)	>36,>26
Nominal focal length (mm)	$9.3 \pm 3$ (Custom lenses)
Wavelength range (nm)	500-885
Spectral resolution at FWHM (nm)	9-40
Adjustable spectral resolution step	<1
f-number	<6.7
Maximum spectral image size (pixels)	$2592 \times 1944$
Spectral image size with default binning (pixels)	$320 \times 240$
Camera dimensions (mm)	$62 \times 66 \times 219$
Weight (g)	<450

Because of the piezo-actualized Fabry-Perot interferometer, the hyperspectral imager is capable of capturing a whole plane at once. This enables fast imaging of patients, which is important because none patient can stay absolutely immobile even during a short time interval. A hyperspectral image itself consists of planes, which correspond with different wavebands. The imaging results in a stack of intensity images which are taken from the same location. If we trace a single spatially-identical pixel through the stack, we get a spectrum which represents the diffuse reflectance of this pixel.

As mentioned earlier, FPI-HSI captures images from 500 nm to 885 nm. The technical details of FPI-HSI are explained In Table 1 . In this research, we used  $N = 76$  different wavebands for image capture. The imaging resolution used ( $320 \times 240$  pixels) gave a spatial resolution of 0.125 mm/pixel.

### 3.2. Outline of the processing scheme

To start with, denote by  $\mathbf{x} = \{x_1, x_2, \dots, x_N\} \in \mathbb{R}^N$ , the spectrum of diffuse reflectance in different wavelngts from 1 to  $N$  of a single spatial pixel. From this, we derive the spectral signature of this specific pixel. The questions are

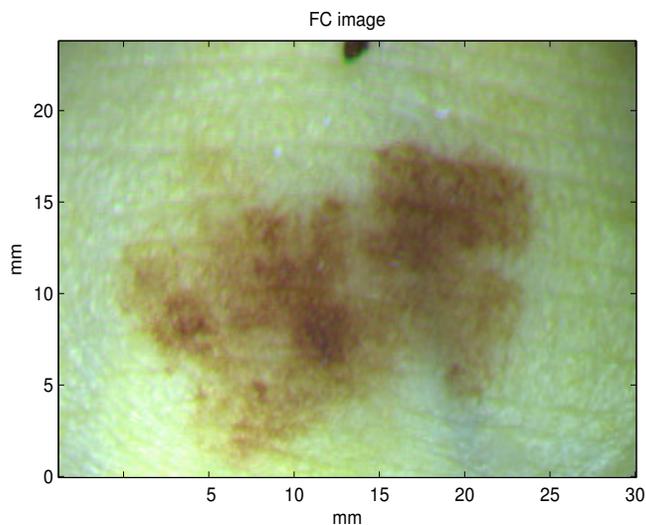


Figure 3: False color image composed from three different wavebands of hyperspectral image

how to correctly derive the signatures from the suspected *diseased* area and how to classify and interpret them.

In clinical trials, we imaged both LM and LMM tumors, which were indicated by physicians, and the surrounding *healthy* skin margins. Figure 3 displays a false-color composition of three different wavebands. In the image, there is a 2.5 cm wide LM tumor. The tumor-diseased part occupies the dark area in the image, while the *healthy* part is located in the lighter areas. The problem is that the tumor's edges are partly diffusing, which hampers delineation by the visual observation.

For training the algorithm, we collect spectra from the confirmed *healthy* and *diseased* areas of the hyperspectral image. The visual inspection of the spectra does not show a significant distinction of the spectra from the *diseased* regions from those belonging to the *healthy* regions. Figure 4, displays two spectra from either of the regions.

Because of the visual similarity of the *diseased* to the *healthy* spectra, we

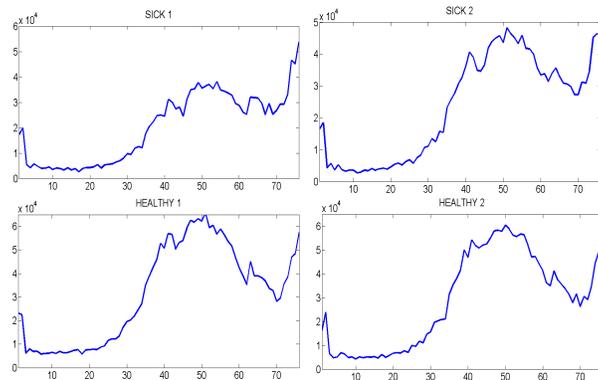


Figure 4: Top: two spectra from a *diseased* area. Bottom: two spectra from a *healthy* area

propose a scheme for the extraction of features that discriminate between these two classes. The features are derived from the combination of a number of selected wavebands in the raw data with the results from the framelet processing. For the dimensionality reduction revelation of the inherent structure of the data sets, the diffusion maps (DM) technique [17] is used.

The spectra from the small regions which definitely belong to the *diseased* and *healthy* areas were used for the construction of reference sets. In turn, the reference sets consisting of the extracted features were utilized for the design of the Classification and Regression Trees (CART's) [18] To reduce false alarms from the CART classification, we used the *Aisles* [19] procedure. In addition, the reference sets were used as a source for the *averaged Nearest Neighbors* (aNN) [20] decision unit.

Once an unknown spectrum is submitted for identification, the features from this spectrum are extracted and submitted to the three units (CART, Aisles, and aNN), whose decisions validate (cross-check) each other. The results of the identification of all the spectra from the given data cube are compared with the visual delineation and biopsies by a physician.

The scheme of the proposed algorithm is given in Fig. 5.

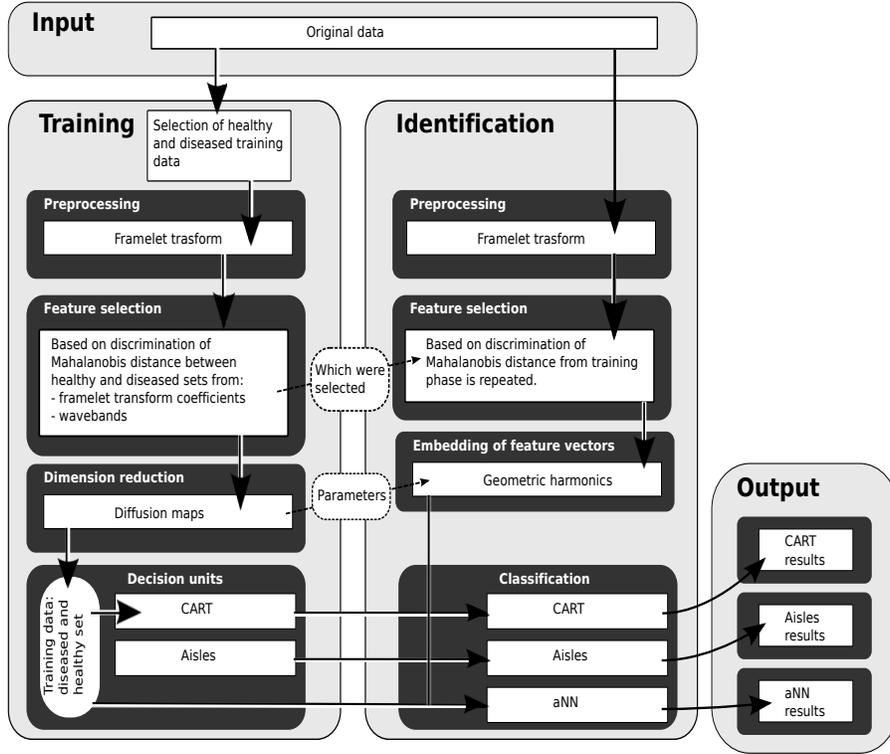


Figure 5: Processing pipeline

### 3.3. Selection of training data

The output from the imaging device is the spectral cube, denoted by  $\mathbf{X}^*$ , is  $\mathbf{X}^* = \{\mathbf{x}_{m,n,s} \in \mathbb{R} | m = 1, \dots, 240, n = 1, \dots, 320, s = 1, \dots, 76\}$ . Each horizontal plane  $\mathbf{X}_s^* \subset \mathbf{X}^*$  of the cube  $\mathbf{X}^*$  represents a "monochromatic" picture of size  $240 \times 320$  of the object, taken in the waveband  $\#s$ . Vertical lines of length 76 correspond to the spectra of pixels in the pictures.

The cube is rearranged into a data matrix of spectra, whose size is  $76800 \times 76$ , in the following way:

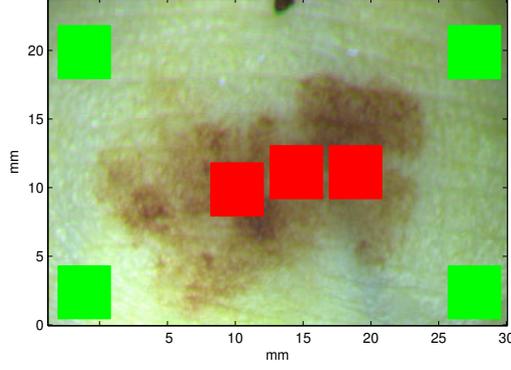


Figure 6: False-color image combined of three different wavebands. The regions are labeled for the training phase: the red squares in the middle mark *diseased* regions, while the green squares in the corners mark healthy regions. The remaining area is unlabeled.

$$\mathbf{X} = \begin{bmatrix} x_{1,1,1} & x_{1,1,2} & \dots & x_{1,1,76} \\ \vdots & \vdots & \vdots & \vdots \\ x_{1,320,1} & x_{1,320,2} & \dots & x_{1,320,76} \\ x_{2,1,1} & x_{2,1,2} & \dots & x_{2,1,76} \\ \vdots & \vdots & \vdots & \vdots \\ x_{240,320,1} & x_{240,320,2} & \dots & x_{240,320,76} \end{bmatrix}.$$

Each row of the matrix  $\mathbf{X}$  represents the spectrum of a spatial pixel.

In order to utilize the proposed methodology, which is based on supervised learning, we need to have labeled data sets for the training phase. For that, a clinician carried out a manual visual selection of the definitely *diseased* and the definitely *healthy* patches in the image. These patches are shown in Figure 6. We denote by  $\mathbf{S}^* \subset \mathbf{X}^*$  the dataset that contains all the labeled *diseased* spectra and by  $\mathbf{H}^* \subset \mathbf{X}^*$  the dataset that contains the *healthy* spectra where  $\mathbf{S}^* \cap \mathbf{H}^* = \emptyset$ . The labeled sets  $\mathbf{S}^*$  and  $\mathbf{H}^*$  of the spectra are arranged into the 76-columns matrices  $\mathbf{S}$  and  $\mathbf{H}$ , respectively.

### 3.4. Preprocessing the training data

*Refinement of the raw training data.* In order to enhance the representativeness of the data matrices  $\mathbf{S}$  and  $\mathbf{H}$ , we test their rows. For this, we calculate the Mahalanobis distances [21]  $d^{ss}$  and  $d^{sh}$  of each row in the matrix  $\mathbf{S}$  to the matrices  $\mathbf{S}$  and  $\mathbf{H}$ , respectively. If, for some row, the distance  $d^{ss} > d^{sh}$  then we discard that row from the matrix  $\mathbf{S}$ . The same is done for the the matrix  $\mathbf{H}$ . We denote the reduced matrices by  $\mathbf{S}_r$  and  $\mathbf{H}_r$ , respectively.

*Framelet transform.* The framelet transform is a generalization of the wavelet transform. Unlike the wavelet transform of a signal, which is one-to-one mapping of the signal into the set of coefficients, the framelet transform produces a larger number of coefficients than the number of samples in the original signal. Thus, the framelet transform provides a redundant representation of the signal, which results in better adaptivity compared to the performance of the wavelet transform. Similarly to the wavelet transform, the framelet transform is implemented in several resolution scales and the transform coefficients retain the spatial-frequency meaning which is inherent in the coefficients of the wavelet transforms.

Wavelet transform of a signal is implemented via the application of a critically sampled two-channel filter bank, which comprises one low-pass and one high-pass filter. Framelet transform, on the other hand, uses an oversampled filter bank, which typically comprises one low-pass, one high-pass and one or several band-pass filters. Once the filter bank has been applied to the signal, the next step of the transform is the application of the filter bank to the output from the low-pass filter, which produces the coefficients related to the coarser resolution scale. Then the procedure is iterated.

Each transform coefficient is associated with a certain small region in the spatial domain and a certain frequency band. Thus, the coefficient depends on the frequency content of this local region. On the other hand, the impulse responses of the high- and band-pass filters comprise finite differences of different orders and, consequently, the transform coefficients bear information on

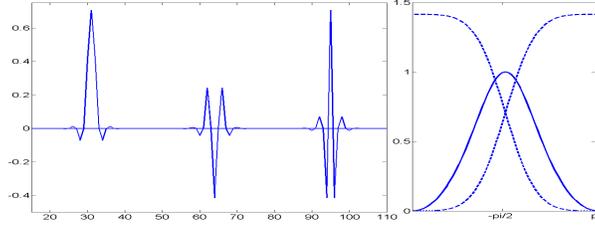


Figure 7: Left: Left-to-right impulse responses of the filters:  $\mathbf{h}^0 \rightarrow \mathbf{h}^1 \rightarrow \mathbf{h}^2$ . Right: magnitude responses of  $\mathbf{h}^0$  and  $\mathbf{h}^2$  (dashed lines) and  $\mathbf{h}^1$  (solid line)

geometrical properties of the respective fragments of the signal. Therefore, the coefficients from framelet transforms can serve as characteristic features for classes of spectra.

In this paper, we use the framelet transform designed in [22]. This transform is derived from the quadratic interpolating spline. Other framelet libraries can be found in [23]. This transform is implemented by recursive filtering the signal by a three-channel filter bank comprising the low-pass  $h^0$ , the band-pass  $h^1$  and the high-pass  $h^2$  filters. Filtering is followed by downsampling by factor of 2. Denote,  $D(z) = z^{-2} + 6 + z^2$  and  $N(z) = (z + 2 + z^{-1})^2$ . Then, the transfer functions of the filters are:

$$h^0(z) = \frac{N(z)}{\sqrt{2}D(z)}, \quad h^2(z) = \frac{N(-z)}{\sqrt{2}D(z)}, \quad h^1(z) = \frac{z^{-1}(z - z^{-1})^2}{D(z)}.$$

The impulse responses of the filters and their magnitude responses are displayed in Figure 7.

Two steps of the framelet transform were applied to each row in the matrices  $\mathbf{S}_r$  and  $\mathbf{H}_r$ , according to the diagram in Figure 8.

A spectrum and its framelet transform coefficients are displayed in Figure 9.

The transform coefficients of each row of  $\mathbf{S}_r$  and  $\mathbf{H}_r$  were gathered into 133-tap vectors, which formed the transform matrices  $\mathbf{T}_s$  and  $\mathbf{T}_h$  derived from  $\mathbf{S}_r$  and  $\mathbf{H}_r$ , respectively.

A detailed description of spline-based framelets and framelet transforms is given in the monograph [24], which is supplied with the MATLAB software.

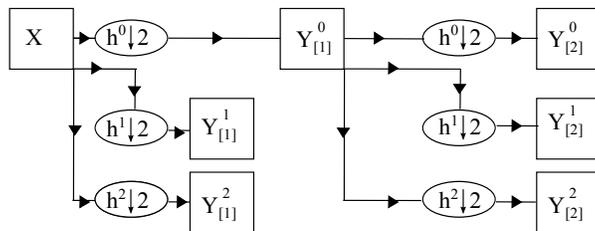


Figure 8: Diagram of 2 steps of framelet transforms, where  $\downarrow 2$  means downsampling by a factor of 2.

### 3.5. Selection of characteristic features

The dataset is prepared for the training phase. As a result of preprocessing, we have two matrices  $\mathbf{S}_r$  and  $\mathbf{H}_r$  whose rows comprise spectra of the pixels selected from the *diseased* and the *healthy* areas, respectively. In addition, we have matrices  $\mathbf{T}_s$  and  $\mathbf{T}_h$  whose rows are the vectors of the framelet transform coefficients of the corresponding spectra. The training procedure is to extract features that characterize the *diseased* and the *healthy* classes and to prepare a reference set for the decision units. The following sets of characterizing features were utilized:

1. A few selected wavebands that provide the best discrimination between the labeled training data.
2. A few selected coefficients of the framelet transforms.
3. A combination of framelet transform coefficients and wavebands.

#### 3.5.1. Refinement of the features' sets by cross-validation

We refine the features' sets by selection of features with the strongest discriminating power. This is done by the following cross-validation procedure.

*Selection of most discriminating wavebands.* Each of the 76 columns of the *diseased*  $\mathbf{S}_r$  and *healthy*  $\mathbf{H}_r$  matrices of size  $N_s \times 76$  and  $N_h \times 76$ , respectively, represent a certain waveband. To select a few wavebands that most discriminate

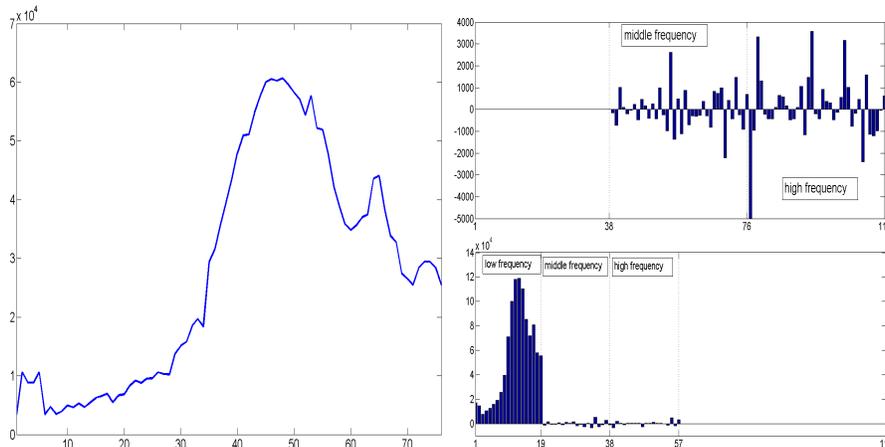


Figure 9: Left: a typical LM spectrum. Right: its framelet transform coefficients. Top: coefficients from the first decomposition level; bottom: coefficients from the second decomposition level

between the two classes, we implement cross-validation using the averaged Mahalanobis distance. For this, we evaluate, how removal of different wavebands from the features' sets affects the "distance" between the sets.

- We remove the first column from both matrices  $\mathbf{S}_r$  and  $\mathbf{H}_r$  and calculate the Mahalanobis distances ([21]  $D_{i,1}$ ,  $i = 1, \dots, N_s$ , from each row  $i$  of the squeezed matrix  $\mathbf{S}_{r,1}$  to the squeezed matrix  $\mathbf{H}_{r,1}$ . Then we calculate the averaged distance  $D_1 = (\sum_{i=1}^{N_s} D_{i,1})/N_s$ .
- After that, this row is returned back and the second row is removed. Similarly, the averaged distance  $D_2$  is calculated. This is repeated for each column, and we get the collection of distances  $D_k$  where  $k = 1, \dots, 76$ .
- This collection provides the knowledge about the contribution of different columns (wavebands) into the discrimination between two classes.
- The sorted distances  $D_k$  are displayed in Figure 10. We observe that the removal of some columns, which are depicted by initial bars in the figure, reduces substantially the distance between the two matrices, while removal

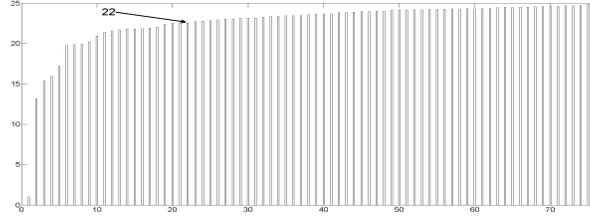


Figure 10: Sorted averaged Mahalanobis distances related to each waveband. The first 22 wavebands are chosen

of some others (bars in the right-hand side of the picture) does not affect it. This means that the first group of columns (wavebands) is essential for the discrimination because their removal makes the matrices to converge to each other.

- We select a number of initial bars. In our running example, we manually take the first 22 bars, which correspond to certain wavebands. This step could be automated in future. These selected wavebands discriminate best between the *diseased* and the *healthy* reference datasets. Figure 11 displays two spectra from the *diseased* region and two spectra from the healthy region. Six most discriminating out of 22 selected wavebands are indicated. We observe the difference in the behavior of "diseased" and the "healthy" spectra at these wavebands.
- We gather the selected columns of the matrices  $\mathbf{S}_r$  and  $\mathbf{H}_r$  into the 22-column features matrices  $\mathbf{S}_{fr}$  and  $\mathbf{H}_{fr}$  respectively.

*Selection of most discriminating framelet transform's coefficients.* To extend the set of characteristic features, we derive by selection of coefficients of the framelet transform of pixels' spectra. These coefficients are gathered into the transform matrices  $\mathbf{T}_s$  (*diseased*) and  $\mathbf{T}_h$  (*healthy*) of size  $N_s \times 133$  and  $N_h \times 133$ , respectively. The transform matrices  $\mathbf{T}_s$  and  $\mathbf{T}_h$  are subjected to the same cross-validation as the matrices  $\mathbf{S}_r$  and  $\mathbf{H}_r$  above. The sorted averaged

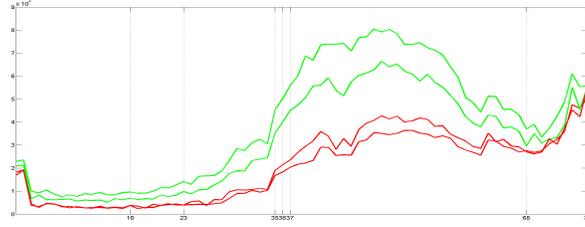


Figure 11: Two spectra from the *diseased* region (lower) and two spectra from the *healthy* region (upper).

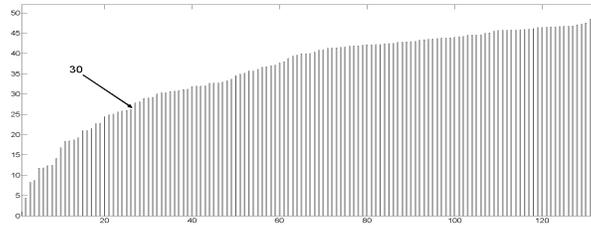


Figure 12: Sorted averaged Mahalanobis distances related to the framelet transform coefficients. The first 30 coefficients were selected.

Mahalanobis distances for these matrices are displayed in Figure 12. In this case, we select 30 coefficients, whose respective bars occupy the left-hand part of the picture.

We gather the selected columns of the matrices  $\mathbf{T}_s$  and  $\mathbf{T}_h$  into the features matrices  $\mathbf{T}_{f_s}$  and  $\mathbf{T}_{f_h}$  of size  $N_s \times 30$  and  $N_h \times 30$ , respectively. In order to enhance the discrimination power of the selected features sets, we combine them into a new features set. For this, we form two combined features matrices  $\mathbf{S}_{f_c} = \mathbf{S}_{f_r} \cup \mathbf{T}_{f_s}$  and  $\mathbf{H}_{f_c} = \mathbf{H}_{f_r} \cup \mathbf{T}_{f_h}$ . In our running example, each of the matrices  $\mathbf{S}_{f_c}$  and  $\mathbf{H}_{f_c}$ , has  $22+30=52$  columns.

As a result of the above procedures, we have three sets of characteristic features:

1. Wavebands features matrices  $\mathbf{S}_{f_w}$  and  $\mathbf{H}_{f_w}$ , where the data points are 22-tap vectors.
2. Transform coefficients features matrices  $\mathbf{T}_{f_s}$  and  $\mathbf{T}_{f_h}$ , where the data

points are 30-tap vectors.

3. Combined features matrices  $\mathbf{S}_{fc}$  and  $\mathbf{H}_{fc}$ , where the data points are 52-tap vectors.

### 3.6. Dimension reduction

The data points in all the three features' datasets are vectors from high-dimensional spaces. In order to reduce the dimensionality and to reveal the internal structure of the datasets, we project the data points onto a low-dimensional Euclidean space, using the DM technique [17]. The Euclidean space, which the data points are embedded into, is a linear hull of a few eigenvectors of the diffusion matrix, and it is called the diffusion space. Diffusion mapping is a non-linear mapping that retains the local (dis)similarity of the data points. Therefore, it provides an efficient tool for data clustering. The DM technique, which we utilize, is outlined in Appendix A.1.

For the classification purposes, we embed the features datasets into the diffusion spaces generated by the 7 – 10 strongest eigenvalues of the diffusion matrices. The projections on the three initial eigenvectors enable us to visualize the geometrical structure of the datasets.

Practically, we combine the features matrices  $\mathbf{S}_{fc}$  and  $\mathbf{H}_{fc}$  into one matrix

$$\mathbf{C} = \begin{pmatrix} \mathbf{S}_{fc} \\ \mathbf{H}_{fc} \end{pmatrix}$$

and form the diffusion matrix  $\mathbf{D}$  from the rows of the matrix  $\mathbf{C}$ . The eigenvectors of the matrix  $\mathbf{D}$  constitute a basis of the so-called *diffusion space*  $\Delta$ . The subspace generated by the 8 initial eigenvectors  $\mathbf{e}_k, k = 1, \dots, 8$ , is denoted by  $\Delta_8$ . The rows of the matrices  $\mathbf{S}_{fc}$  and  $\mathbf{H}_{fc}$  are projected onto the subspace  $\Delta_8$ . Thus, the matrices  $\mathbf{S}_{fc}$  and  $\mathbf{H}_{fc}$  are mapped onto the 8-column matrices  $\mathbf{S}_{ref}$  and  $\mathbf{H}_{ref}$ , respectively. Consequently, each spectrum from the training dataset is represented by an 8-tap vector, which consists of the coordinates of the spectrum's projection onto the space  $\Delta_8$ . Figure 13 visualizes three initial columns of the matrices  $\mathbf{S}_{ref}$  and  $\mathbf{H}_{ref}$ , which represent the *diseased* and the

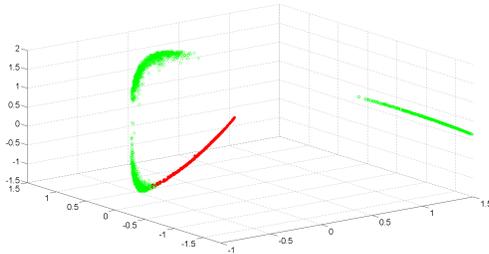


Figure 13: Projection of the training dataset onto three eigenvectors  $\mathbf{e}_k, k = 1, 2, 3$ , of the diffusion matrix. Green: projections of the *healthy* data points. Red: projections of the *diseased* data points.

*healthy* spectra, respectively. We observe that the projections of the *healthy* data points and the *diseased* data points are grouped in compact clusters which have a small overlap.

### 3.7. Decision units

The matrices  $\mathbf{S}_{ref}$  and  $\mathbf{H}_{ref}$  are used as the reference sets for the averaged Nearest Neighbors (aNN) [20] decision unit and also for the design of CART and *Aisles*.

*CART*. Prior to identification of unknown spectra, the Classification And Regression Tree (CART) [18] has to be trained using the reference matrices  $\mathbf{S}_{ref}$  and  $\mathbf{H}_{ref}$ . Rows of these matrices constitute the space of labeled input patterns  $\mathbf{X}$ . The tree is constructed by a binary split of the space  $\mathbf{X}$  in a way that the data in each descendant subset is “purer” than the data in the parent subset. A node that is not split further is declared *terminal* if it is (almost) completely populated by samples belonging to a certain class. The split is stopped when all the nodes become terminal. As a result, the space  $\mathbf{X}$  is decomposed into the union  $\mathbf{X} = \bigcup_{k=1}^K \mathbf{X}_m$  of subspaces (terminal nodes) each of which is labeled as belonging to either the *diseased* or the *healthy* class. This process is illustrated in Figure B.16.

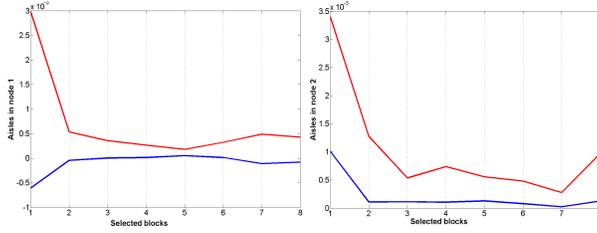


Figure 14: Examples of Aisles in two different CART nodes.  $X$ -axes: indices  $k = 1, \dots, 8$  of the coordinates of the row vectors. Upper lines:  $\bar{m}^k + 2\bar{s}^k$ . Lower lines:  $\bar{m}^k - 2\bar{s}^k$

Once a new vector is submitted to the designed tree, it is dispatched to one of the terminal nodes and its membership could be predicted with a sufficient reliability. More details about CART are provided in Appendix B.

*Aisles.* To reduce the false alarms, the CART classifier is refined in the following way. All the rows in the reference matrix  $\mathbf{S}_{ref}$  (8-tap vectors) are presented to the designed tree. Each of them is dispatched to one of the terminal nodes that are associated with the *diseased* class. We assume that the set  $\mathbf{R}^k$  of rows from the matrix  $\mathbf{S}_{ref}$  was dispatched to a terminal node  $\mathbf{X}_k$ . We calculate the 8-tap vectors  $\vec{m}^k = \{m_1^k, \dots, m_8^k\}$ , which is the mean value of the rows  $\mathbf{R}^k$ , and  $\vec{s}^k = \{s_1^k, \dots, s_8^k\}$ , which is the standard deviation (STD) of these rows. Then, we form the aisle  $\mathbf{A}^k = \bar{m}^k \pm 2\bar{s}^k$ . This is done for all the “*diseased*” terminal nodes.

Two examples of Aisles are displayed in Figure 14.

As a result from the training operations, we have

1. Two reference matrices  $\mathbf{S}_{ref}$  and  $\mathbf{H}_{ref}$ , which consist of coordinates of the *diseased* and the *healthy* spectra in the basis  $\mathbf{e}_k, k = 1, \dots, 8$ , of the eigenvectors.
2. The eigenvectors  $\mathbf{e}_k, k = 1, \dots, 8$ .
3. The CART classifier.
4. The *Aisles*, which reduce false alarms in the CART.
5. The lists  $L_w$  of the selected wavebands and  $L_f$  of the selected framelet

transform coefficients, which demonstrate the strongest discriminating property.

### 3.8. Identification phase

Once an unknown spectrum  $\mathbf{u}$  is submitted for identification, the following steps lead to a decision on its membership.

*Extraction of characteristic features.*

1. The wavebands listed in  $L_w$  are gathered into the features vector  $\mathbf{u}_w$ .
2. The framelet transform is applied.
3. The transform coefficients listed in  $L_f$  are gathered into the features vector  $\mathbf{u}_f$ .
4. The combined features vector  $\mathbf{u}_c = \mathbf{u}_w \cup \mathbf{u}_f$  is formed.

The feature vectors are embedded into the diffusion subspace. The diffusion space  $\Delta$  and its subspace  $\Delta_8$ , which the reference sets belong to, are derived from the features sets  $\mathbf{S}_{ref}$  and  $\mathbf{H}_{ref}$ . Therefore, the operator  $\mathbf{E}_8$ , which embeds into the subspace  $\Delta_8$ , is defined on the features vectors, which are the rows of the matrices  $\mathbf{S}_{ref}$  and  $\mathbf{H}_{ref}$ .

In order to embed the features vector  $\mathbf{u}_c$  to be identified into the subspace  $\Delta_8$ , the domain of the operator  $\mathbf{E}_8$  should be extended. That is done by the Geometric Harmonics (GH) technique [25, 26]. This technique is a generalization of the Nyström method [27], which is extensively used for an out-of-sample extension in dimensionality reduction schemes to find the coordinates of the unknown data points that have not participated in the training phase. The technique is outlined in Appendix A.2. As a result of embedding the features vector  $\mathbf{u}_c$  into  $\Delta_8$ , we get the 8-tap vector  $\mathbf{u}_\Delta \in \Delta_8$ , which represents the spectrum  $\mathbf{u}$ .

Decisions about the spectra membership are made as follows. The vector  $\mathbf{u}_\Delta \in \Delta_8$  is submitted to the CART tree, which dispatches it to a terminal node  $\mathbf{X}_k$ . The answer from the CART is  $\mathbf{R}_{cart} = 0$ , if the node  $\mathbf{X}_k$  is associated with the *diseased* class, and is  $\mathbf{R}_{cart} = 1$ , otherwise. We assume that the node  $\mathbf{X}_k$ ,

where the vector  $\mathbf{u}_\Delta = \{u_n\}, n = 1, \dots, 8$ , is dispatched to, is associated with the *diseased* class, thus,  $\mathbf{R}_{cart} = 0$ . Then, the vector  $\mathbf{u}_\Delta$  is subjected to an additional test. It is classified as *diseased* only if it passes through the Aisle  $\mathbf{A}^k$  associated with the node  $\mathbf{X}_k$ . This means that its coordinates should satisfy the inequalities  $m_n^k - s_n^k \leq u_n \leq m_n^k + s_n^k, n = 1, \dots, 8$ . If this is the case then the *diseased* classification is confirmed and the answer  $\mathbf{R}_{ais} = 0$ . Otherwise  $\mathbf{R}_{ais} = 1$ . Surely,  $\mathbf{R}_{ais} = 1$  in all cases when  $\mathbf{R}_{cart} = 1$ .

Independently, the decision on the membership of the spectrum  $\mathbf{u}$  is made by the averaged Nearest Neighbor (aNN) classifier. For this, the Euclidean distances  $\delta_m^s, m = 1, \dots, M$ , and  $\delta_m^h, m = 1, \dots, M$ , of the vector  $\mathbf{u}_\Delta$  to  $M$  nearest rows of the reference matrices  $\mathbf{S}_{ref}$  and  $\mathbf{H}_{ref}$ , respectively, are calculated (typically,  $M = 3 - 5$ ). Their averages are denoted as  $\bar{\delta}^s = M^{-1} \sum_{m=1}^M \delta_m^s$  and  $\bar{\delta}^h = M^{-1} \sum_{m=1}^M \delta_m^h$ , respectively. The ratio

$$p^s = \frac{\bar{\delta}^h}{\bar{\delta}^h + \bar{\delta}^s}$$

provides, in a sense, the probability for the spectrum  $\mathbf{u}$  to belong to the *diseased* class. We set the threshold  $T > 1/2$  and classify  $\mathbf{u}$  as *diseased* if  $p^s > T$ , then the answer is  $\mathbf{R}_{ann} = 0$ , otherwise  $\mathbf{R}_{ann} = 1$ . The above identification procedures are applied to all the spectra in the hyperspectral image.

#### 4. Experimental results

As mentioned previously, we imaged 8 lesions (3 LMM, 5 LM) in 8 patients before surgical removal of the lesions. The patients had poorly delineated tumors, which were delineated by using hyperspectral imaging before the removal of the lesions by wide-excision surgery. One of the lesions (lesion B) was a rare amelanotic lentigo maligna melanoma. We imaged the lesions with a developed hyperspectral imaging system. The data was processed as described in Section 2. The hyperspectral results were compared to lesion borders assessed by clinicians and to the histological results. Figure 15 shows results of our study. These results show the border delineation in hyperspectral images. Lesions B,

C and H represent LMM, lesions A and D-G represent LM without invasion. The analyses were not used for differentiation of the tumor types.

In histopathological sampling lesions, A, B, E and G were histologically wider than the clinically was detected. Lesions C, D, F and H were identical in size clinically and in histopathological evaluation.

A: The lesion continued histologically over visible borders to the right side and to the lower part of the image. All classifiers caught the extension at the bottom and showed minor extension to the right side.

B: The lesion continued histologically over visible borders both to the left and the right sides of the image. Both classifiers showed signs of extension.

C: Lesion borders were identical in size in visual inspection and histologically. ANN and Aisles give similar results. CART incorrectly catches a diffuse area in the lower left side of the image.

D: Lesion borders were identical clinically and histologically, as also delineated by all the classifiers used.

E: Histologically lesion was wider than seen in visual inspection in the lower part of the image and slightly wider than in the top part of the image. All classifiers caught the extension at the top of the image as well. The extension in the lower part was missed by all the classifiers.

F: Lesion borders were identical clinically and histologically. All the classifiers delineated the borders correctly.

G: The upper right corner of the lesion showed subclinical extension histologically. Only the ANN classifier caught the extension.

H: The lesion showed no subclinical extension. All the classifiers delineated the lesion borders correctly.

Our study indicates that in most of the cases the classifiers are useful in detection of the subclinical (not seen in the visual inspection) of the lesion

borders. By drawing surgery margins based on hyperspectral analysis so that each uniform *diseased* area is couple of millimeters away from the margin, we get better results than just by using traditional methods.

## 5. Conclusions and future work

The proposed features extraction scheme, which combines a number of selected wavebands with coefficients of framelet transforms, coupled with robust classifiers (CART, aNN, Aisles), proved to be efficient for the delineation of cancerous skin areas. A critical part of the algorithm is the dimensionality reduction scheme based on the DM.

The designed algorithm has showed potential in non-invasive delineation of tumor borders. By using hyperspectral imaging, lesion borders could be detected with more accuracy and re-excisions could be avoided.

In some cases, the classifiers misclassified the lesion borders due to the non-satisfactory illumination quality that affects the quality of the images captured by the hyperspectral camera. The illumination quality is to be improved in the next generation of the hyperspectral imager.

Currently, the training of the algorithm requires supervision and manual selection of features. Amount of the training data is essential for the analysis results.

In a future work, we plan to design a non-supervised automatic version of the algorithm. We will also focus on spectral separation of different tumors (LM, LMM and others), in order to provide a non-invasive diagnostics tool for clinicians.

## Appendix A. Diffusion Framework

### Appendix A.1. Diffusion Maps

Let  $\Gamma$  be a set of points in  $\mathbb{R}^n$ . The graph  $G = (\Gamma, K)$  with a kernel  $K \triangleq k(x, y)$ , which is a weight function, measures the pairwise similarity between the points is constructed. The weight function is: symmetric:  $k(x, y) = k(y, x)$ ;

non-negative:  $k(x, y) \geq 0$  for all  $x$  and  $y$  in  $\Gamma$ ; positive semi-definite: for all real-valued bounded function  $f$  defined on  $\Gamma$   $\sum_{x \in \Gamma} \sum_{y \in \Gamma} k(x, y) f(x) f(y) \geq 0$ .

The following normalization transforms the kernel into a Markov transition matrix  $P$ :  $P \triangleq p(x, y) = \frac{k(x, y)}{d(x)}$  where  $d(x) = \sum_{y \in \Gamma} k(x, y)$  is the degree of the node  $x$ . This normalization is known as the weighted Graph Laplacian normalization [28]. Since  $P$  consists of nonnegative real numbers, where each row is summed to 1, the matrix  $P$  can be viewed as the Markov transition matrix of a random walk between the points in  $\Gamma$ . The probability to move from  $x$  to  $y$  in one time step is given in  $P(x, y)$ . These probabilities measure the connectivity of the points within the graph.

The transition matrix  $P$  is conjugate to a symmetric matrix  $A$  given by  $a(x, y) = \sqrt{d(x)p(x, y)} \frac{1}{\sqrt{d(y)}}$ . Using a matrix notation,  $A \triangleq D^{\frac{1}{2}} P D^{-\frac{1}{2}}$ , where  $D$  is the diagonal matrix with the values  $\sum_y k(x, y)$  on its diagonal. The symmetric matrix  $A$  has  $n$  real eigenvalues  $\{\lambda_l\}_{l=0}^{n-1}$  and a set of orthonormal eigenvectors  $\{v_l\}$  in  $\mathbb{R}^n$ , thus, it has the following spectral decomposition:

$$a(x, y) = \sum_{k \geq 0} \lambda_k v_k(x) v_k(y). \quad (\text{A.1})$$

Since  $P$  is conjugate to  $A$ , the eigenvalues of both matrices are identical. In addition, if  $\{\phi_l\}$  and  $\{\psi_l\}$  are the corresponding left and right eigenvectors of  $P$ , then we have the following equalities:

$$\phi_l = D^{\frac{1}{2}} v_l, \quad \psi_l = D^{-\frac{1}{2}} v_l. \quad (\text{A.2})$$

From the orthonormality of  $\{v_i\}$  and Eq. A.2 it follows that  $\{\phi_l\}$  and  $\{\psi_l\}$  are biorthonormal i.e.  $\langle \phi_m, \psi_l \rangle = \delta_{ml}$ . Combing Eqs. A.1 and A.2 together with the biorthogonality of  $\{\phi_l\}$  and  $\{\psi_l\}$  leads to the following eigendecomposition of the transition matrix  $P$

$$p(x, y) = \sum_{l \geq 0} \lambda_l \psi_l(x) \phi_l(y). \quad (\text{A.3})$$

Because of the fast decay of the spectrum, only a few terms are required to achieve sufficient accuracy in the sum. The family of diffusion maps  $\{\Psi(x)\}$

defined by  $\Psi(x) = (\lambda_1\psi_1(x), \lambda_2\psi_2(x), \lambda_3\psi_3(x), \dots)$  embeds the dataset into a Euclidean space. This embedding constitutes a new parametrization of the data in a low-dimensional space. We recall the diffusion distance between two data points  $x$  and  $y$  as it was defined in [17]:

$$D^2(x, y) = \sum_{z \in \Gamma} \frac{(p(x, z) - p(z, y))^2}{\phi_0(z)}. \quad (\text{A.4})$$

This distance reflects the geometry of the dataset as the value of  $\frac{1}{\phi_0(x)}$  depends on the point's density. Two data points are close if there is a large number of paths connecting them. In this case, the diffusion distance is small. Substituting Eq. A.3 in Eq. A.4 together with the biorthogonality property allows to express the diffusion distance using the right eigenvectors of the transition matrix  $P$ :

$$D^2(x, y) = \sum_{l \geq 1} \lambda_l^2 (\psi_l(x) - \psi_l(y))^2. \quad (\text{A.5})$$

In these new coordinates, the Euclidean distance between two points in the embedded space represents the distances between the two high dimensional points as defined by a random walk. Moreover, this facilitates the embedding of the original points in a Euclidean space  $\mathbb{R}^\eta$  by:

$$\Xi_t : x_i \rightarrow (\lambda_2^t \psi_2(x_i), \lambda_3^t \psi_3(x_i), \dots, \lambda_\eta^t \psi_\eta(x_i)). \quad (\text{A.6})$$

which also provides coordinates on the set  $\Gamma$ . Essentially,  $\eta \ll n$  due to the fast decay of the eigenvalues of  $P$ . Furthermore,  $\eta$  depends only on the primary intrinsic variability of the data as captured by the random walk and not on the original dimensionality of the data.

### Appendix A.2. Geometric Harmonics

Geometric Harmonics (GH) is a method that facilitates the extension of any function  $f : \Gamma \rightarrow \mathbb{R}$  to a set of new points that are added to  $\Gamma$ . Specifically, every coordinate of the low-dimensional embedding constitutes such a function. Let  $\Gamma$  be a set of points in  $\mathbb{R}^n$  and  $\Psi$  be its diffusion map embedding. Let  $\bar{\Gamma}$  be a set in  $\mathbb{R}^n$  such that  $\Gamma \subseteq \bar{\Gamma}$ . The GH scheme extends  $\Psi_t$  into a new dataset  $\bar{\Gamma}$ . The GH generalize the *Nystrom extension* [29] method which we describe next.

*The Nyström extension.* The eigenvectors and eigenvalues of a Gaussian kernel on the training set  $\Gamma$  with width  $\epsilon$  are computed by

$$\lambda_l \varphi_l(x) = \sum_{y \in \Gamma} e^{-\frac{\|x-y\|^2}{2\epsilon}} \varphi_l(y), \quad x \in \Gamma. \quad (\text{A.7})$$

If  $\lambda_l \neq 0$ , the eigenvectors in Eq. A.7 can be extended to any  $x \in \mathbb{R}^n$  by

$$\bar{\varphi}_l(x) = \frac{1}{\lambda_l} \sum_{y \in \Gamma} e^{-\frac{\|x-y\|^2}{2\epsilon}} \varphi_l(y), \quad x \in \mathbb{R}^n. \quad (\text{A.8})$$

This is known as the Nyström extension, which is a common method for extension of functions to out-of-the-training-set points. This is common in online processes in which new samples arrive and we need to extrapolate a function  $f$  on the  $\Gamma$  to the new points. The extended distance of  $\bar{\varphi}_l$  from the training set is proportional to  $\epsilon$ . Let  $f$  be a function on the training set  $\Gamma$ . In our case, we are interested in extending each of the coordinates of the embedding function  $\Psi(x) = (\lambda_1 \psi_1(x), \lambda_2 \psi_2(x), \lambda_3 \psi_3(x), \dots)$ . The eigenfunctions  $\{\varphi_l\}$  are the outcome of the spectral decomposition of a symmetric positive matrix, thus, they form an orthonormal basis in  $\mathbb{R}^n$ . Consequently, any function  $f$  can be written as a linear combination of this basis  $f(x) = \sum_l \langle \varphi_l, f \rangle \varphi_l(x)$ ,  $x \in \Gamma$ . Using the Nyström extension, as given in Eq. A.8,  $f$  can be defined for any point in  $\mathbb{R}^n$  by

$$\bar{f}(x) = \sum_l \langle \varphi_l, f \rangle \bar{\varphi}_l(x), \quad x \in \mathbb{R}^n. \quad (\text{A.9})$$

The above extension allows to decompose each diffusion map coordinate  $\psi_i$  in the same way using  $\psi_i(x) = \sum_l \langle \varphi_l, \psi_i \rangle \varphi_l(x)$ ,  $x \in \Gamma$ . In addition, the embedding of a new point  $\bar{x} \in \bar{\Gamma} \setminus \Gamma$  can be evaluated in the embedding coordinate system by  $\bar{\psi}_i(\bar{x}) = \sum_l \langle \varphi_l, \psi_i \rangle \bar{\varphi}_l(\bar{x})$ .

*Geometric Harmonics.* The Nyström extension, which is given in Eq. A.8, has two major drawbacks:

1. The extension distance is proportional to the value of  $\epsilon$  used in the kernel. This extension numerically vanishes beyond this distance.

2. The scheme is ill conditioned since  $\lambda_l \rightarrow 0$  as  $l \rightarrow \infty$ .

The second issue can be solved by cutting off the sum in Eq. A.9 keeping the eigenvalues (and the corresponding eigenfunctions) satisfying  $\lambda_l \geq \delta\lambda_0$   $\bar{f}(x) = \sum_{\lambda_l \geq \delta\lambda_0} \langle \varphi_l, f \rangle \bar{\varphi}_l(x)$ ,  $x \in \mathbb{R}^n$ . The result is an extension scheme with a condition number  $\delta$ . In this new scheme,  $f$  and  $\bar{f}$  do not coincide on  $\Gamma$ , but they are relatively close. The value of  $\epsilon$  controls this error. An iterative method for modifying the value of  $\epsilon$  with respect to the function to be extended is introduced in [30]. The outline of the algorithm is as follows:

1. Determine an acceptable error for relatively big value of  $\epsilon$ . Denote them by  $err$  and  $\epsilon_0$ , respectively;
2. Build a Gaussian kernel using  $\epsilon_0$  such that  $k_{\epsilon_0}(x, y) = e^{-(\|x-y\|^2/2\epsilon_0)}$ ;
3. Compute the set of eigenvalues for this kernel. Denote this set by  $\varphi_l(x)$  and write  $f$  as a linear combination of this basis as  $f(x) = \sum_l \langle \varphi_l, f \rangle \varphi_l(x)$ ;
4. Compute the sum

$$r_{err} = \sum_{\lambda_l > \delta\lambda_0} \sqrt{|\langle \varphi_l, f \rangle|^2};$$

5. If  $r_{err} < err$  then expand  $f$  with this basis as explained in Eq. Appendix A.2. Otherwise, reduce the value of  $\epsilon_0$  and repeat the process.

The sum, which is computed in step 4 of the algorithm, consists of only large elements. The number of elements in the sum grows with respect to the number of oscillations the function has. This means that  $\epsilon$  will become smaller with respect to the behavior of  $f$ . We are interested in expanding the diffusion maps coordinates that correspond to the highest eigenvalues. These coordinates are usually smooth functions. Thus, the process in steps 1-5 is expected to find a large enough value for  $\epsilon$  that will allow to extend them to new data points.

## Appendix B. Outline of the CART algorithm

A comprehensive exposition of the CART scheme can be found in [18]. We consider a two-class classification problem.

*Building the tree* . The space  $X$  of input patterns from the reference set consists of two reference matrices  $V^l$ ,  $l = 1, 2$  of sizes  $\mu_l \times n$ , respectively. For simplicity, we assume that  $\mu_1 = \mu_2$ . The  $i$ -th row of the matrix  $V^l$  is a vector  $V^l(i, :)$  of length  $n$  representing the signal  $s_i^l$  which belongs to the class  $C^l$ . In our case,  $n$  is equal to the number of the selected eigenvectors of the diffusion matrix.

The tree structured classifier to be constructed has to divide our space  $\mathbf{X}$  into  $J$  disjoint subspaces  $\mathbf{X} = \bigcup_{\nu=1}^J \mathbf{X}_\nu^t$ . Each subspace  $\mathbf{X}_\nu^t$  must be “pure” in the sense that the percentage of vectors from one of the matrices  $V^l$ , must prevail the percentage of the vectors from the other matrix. (In the original space  $\mathbf{X}$  both are 50%.)

The construction of the binary tree is started by a split of  $\mathbf{X}$  into two descendant subspaces:  $\mathbf{X} = \mathbf{X}_1 = \mathbf{X}_2 \cup \mathbf{X}_3$ ,  $\mathbf{X}_2 \cap \mathbf{X}_3 = \emptyset$ .

To do so, CART chooses a split variable  $y_j$  and a split value  $z_j$  in a way to achieve minimal possible “impurity” of the subspaces  $\mathbf{X}_2$  and  $\mathbf{X}_3$ . The split rule for the space  $\mathbf{X}_1$  is as follows :

If a vector  $y = (y_1, \dots, y_n)$  satisfies the condition  $y_j \leq z_j$ , then it transferred to the subspace  $\mathbf{X}_2$ , otherwise it is transferred to the subspace  $\mathbf{X}_3$ .

Subsequently, we divide the subspace  $\mathbf{X}_2$  in a similar manner:  $\mathbf{X}_2 = \mathbf{X}_4 \cup \mathbf{X}_5$ ,  $\mathbf{X}_4 \cap \mathbf{X}_5 = \emptyset$ . The respective split variable  $y_k$  and split value  $z_k$  are selected so, that the data in each of the descendant subspaces were “purer” than the data in the parent subspace. Then, one of the subspaces  $\mathbf{X}_4$  or  $\mathbf{X}_5$  can be further divided recursively until we reach the so called terminal subspace  $\mathbf{X}_1^t$  which is not split further. The decisions whether a subspace is classified as terminal subspace depends on the predetermined minimal “impurity” and the minimal size of the subspace. The terminal subspace  $\mathbf{X}_1^t$  is assigned to the class  $C^l$ , with the probability

$$p_1^l = \frac{m_1^l}{m_1} = \frac{\#\{y \in \mathbf{X}_1^t \cap V^l\}}{\#\{y \in \mathbf{X}_1^t\}},$$

where  $m_1^l$  is the number of points in node  $\mathbf{X}_1^t$  that belongs to class  $C^l$  and  $m_i$  is the total number of points in the subspace  $\mathbf{X}_1^t$ . After termination is reached in the subspace  $\mathbf{X}_1^t$  we return to subspace  $\mathbf{X}_3$  which was not split.

Similarly, we reach the next terminal subspace  $\mathbf{X}_2^t$ . We do the same with one of yet non-splitted subspaces and finally the tree  $\mathbf{X} = \bigcup_{\nu=1}^J \mathbf{X}_\nu^t$  is constructed. In the terminology of graph theory, the space  $\mathbf{X}$  is called the root node, the nonterminal and terminal subspaces are the nonterminal and terminal nodes. This process is illustrated in Fig. B.16. The terminal nodes are marked as rectangles.

*Classification* . A vector  $x = (x_1, \dots, x_n)$  is fed to the tree in the first step is assigned to either node  $\mathbf{X}_2$  in the case where the coordinate  $x_j \leq z_j$  or to node  $\mathbf{X}_3$  otherwise. Finally, by checking subsequent split variables, the vector is forwarded into a terminal node  $\mathbf{X}_r^t$  which is labeled as class  $C^l$ , with probability  $p_r^l$ .

### Appendix C. Acknowledgement

This study is partly funded by Tekes – the Finnish Funding Agency for Innovation. Study involve human subjects and it complies with the principles laid down in the Declaration of Helsinki. The study protocol was approved by the Ethics Committee of the Tampere University, Finland (Pirkanmaa hospital district’s ethical commission, No. R12026). Authors do not have a conflict of interest.

- [1] J. McKenna, S. Florell, G. Goldman, G. Bowen, Lentigo maligna/lentigo maligna melanoma: Current state of diagnosis and treatment, *Dermatologic Surgery* 32 (2006) 493-504.
- [2] G. Clark, E. Pappas-Politis, B. Cherpelis, J. Messina, M. MÄüller, C. Cruse, L. Glass, Surgical management of melanoma in situ on chronically sun-damaged skin, *Cancer Control* 15 (2008) 216–224.
- [3] H. Saari, Spectrometer and interferometric method, US Patent (2012).
- [4] S. Kotsiantis, Supervised machine learning: A review of classification techniques, *Informatica* 31 (2007) 249–268.

- [5] G. Lu, B. Fei, Medical hyperspectral imaging: a review, *Journal of Biomedical Optics* 19 (2014) 010901.
- [6] D. T. Dicker, N. Kahn, K. T. Flaherty, J. Lerner, W. S. El-Deiry, Hyperspectral imaging: A non-invasive method of imaging melanoma lesions in a patient with stage iv melanoma, being treated with a raf inhibitor, *Cancer biology & therapy* 12 (2011) 326–334.
- [7] R. Hennessy, S. Bish, J. Tunnell, M. Markey, Segmentation of diffuse reflectance hyperspectral datasets with noise for detection of melanoma, in: *Engineering in Medicine and Biology Society (EMBC), 2012 Annual International Conference of the IEEE, 2012*, pp. 1482–1485. doi:10.1109/EMBC.2012.6346221.
- [8] T. Nagaoka, A. Nakamura, Y. Kiyohara, T. Sota, Melanoma screening system using hyperspectral imager attached to imaging fiberscope, in: *Engineering in Medicine and Biology Society (EMBC), 2012 Annual International Conference of the IEEE, 2012*, pp. 3728–3731. doi:10.1109/EMBC.2012.6346777.
- [9] I. Quinzán, J. M. Sotoca, P. Latorre-Carmona, F. Pla, P. García-Sevilla, E. Boldó, Band selection in spectral imaging for non-invasive melanoma diagnosis, *Biomedical optics express* 4 (2013) 514–519.
- [10] I. Diebele, I. Kuzmina, A. Lihachev, J. Kapostinsh, A. Derjabo, L. Valeine, J. Spigulis, Clinical evaluation of melanomas and common nevi by spectral imaging, *Biomedical optics express* 3 (2012) 467–472.
- [11] T.-Y. Tseng, C.-Y. Chen, Y.-S. Li, K.-B. Sung, Quantification of the optical properties of two-layered turbid media by simultaneously analyzing the spectral and spatial information of steady-state diffuse reflectance spectroscopy, *Biomedical optics express* 2 (2011) 901–914.
- [12] R. Jolivot, Y. Benezeth, F. Marzani, Skin parameter map retrieval

- from a dedicated multispectral imaging system applied to dermatology/cosmetology, *International journal of biomedical imaging* 2013 (2013).
- [13] J. Galeano, R. Jolivot, F. Marzani, Y. Benezeth, Unmixing of human skin optical reflectance maps by non-negative matrix factorization algorithm, *Biomedical Signal Processing and Control* 8 (2013) 169 – 175.
- [14] N. Lue, J. W. Kang, C.-C. Yu, I. Barman, N. C. Dingari, M. S. Feld, R. R. Dasari, M. Fitzmaurice, Portable optical fiber probe-based spectroscopic scanner for rapid cancer diagnosis: a new tool for intraoperative margin assessment, *PloS one* 7 (2012) e30887.
- [15] H. Akbari, L. V. Halig, H. Zhang, D. Wang, Z. G. Chen, B. Fei, Detection of cancer metastasis using a novel macroscopic hyperspectral method, in: *SPIE Medical Imaging, International Society for Optics and Photonics*, 2012, pp. 831711–831711.
- [16] D. T. Dicker, J. Lerner, P. Van Belle, S. F. Barth, M. Herlyn, D. E. Elder, W. S. El-Deiry, Research paper differentiation of normal skin and melanoma using high resolution hyperspectral imaging, *Cancer biology & therapy* 5 (2006) 1033–1038.
- [17] R. Coifman, S. Lafon, Diffusion maps, *Applied and Computational Harmonic Analysis* 21 (2006) 5–30.
- [18] L. Breiman, J. H. Friedman, R. A. Olshen, C. J. Stone, *Classification and Regression Trees*, Wadsworth, 1984.
- [19] A. Averbuch, E. Hulata, V. Zheludev, I. Kozlov, A wavelet packet algorithm for classification and detection of moving vehicles, *Multidimens. Systems Signal Process.* 12 (2001) 9–31.
- [20] D. Mount, A. Sunil, Ann: A library for approximate nearest neighbor searching, <http://www.cs.umd.edu/~simsmount/ANN/>, 2012.

- [21] P. C. Mahalanobis, On the generalised distance in statistics, in: Proceedings National Institute of Science, India, volume 2, 1936, pp. 49–55. URL: <http://ir.isical.ac.in/dspace/handle/1/1268>.
- [22] A. Averbuch, V. Zheludev, T. Cohen, Interpolatory frames in signal space, *Trans. Sig. Proc.* 54 (2006) 2126–2139.
- [23] V. Zheludev, P. Neittaanmäki, A. Averbuch, Spline-based frames for image restoration, in: 10th international conference on Sampling Theory and Applications (SampTA 2013), Bremen, Germany, 2013, pp. 464–467.
- [24] A. Z. Averbuch, P. Neittaanmäki, V. A. Zheludev, Spline and spline wavelet methods with applications to signal and image processing, Volume I: Periodic splines, Springer, 2014. (to appear).
- [25] R. Coifman, S. Lafon, Geometric harmonics: A novel tool for multiscale out-of-sample extension of empirical functions, *Applied and Computational Harmonic Analysis* 21 (2006) 31 – 52. Special Issue: Diffusion Maps and Wavelets.
- [26] A. Bermanis, A. Averbuch, R. R. Coifman, Multiscale data sampling and function extension, *Applied and Computational Harmonic Analysis* 34 (2013) 15 – 29.
- [27] W. H. Press, S. A. Teukolsky, W. T. Vetterling, B. P. Flannery, *Numerical recipes in C (2nd ed.): the art of scientific computing*, Cambridge University Press, New York, NY, USA, 1992.
- [28] F. R. K. Chung, *Spectral Graph Theory (CBMS Regional Conference Series in Mathematics, No. 92)*, American Mathematical Society, 1996. URL: <http://www.worldcat.org/isbn/0821803158>.
- [29] E. Nyström, Über die praktische auflösung von linearen integralgleichungen mit anwendungen auf randwertaufgaben der potentialtheorie, *Comment. Phys.-Math.* 4 (15) (1928) 1–52.

- [30] A. G. S. Lafon, Y. Keller, R. R. Coifman, Data fusion and multi-cue data matching by diffusion maps, *IEEE Transactions on Pattern Analysis and Machine Intelligence* 28 (11) (2006) 1784–1797.

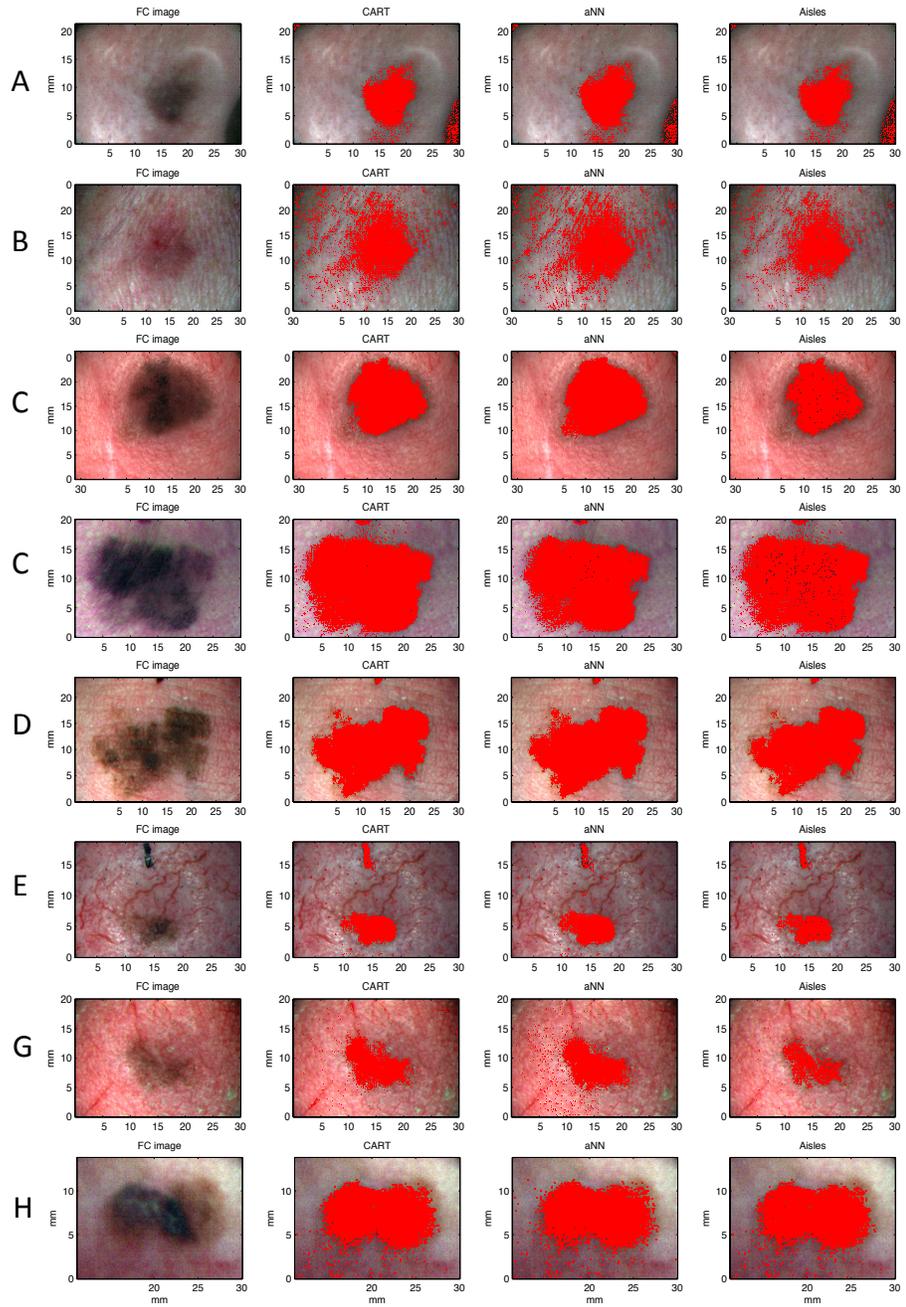


Figure 15: Analyzed results from 8 imaged lesions. On the left are the false-color compositions from tumors then are three different decision units. Regions marked with red indicate the *diseased area*

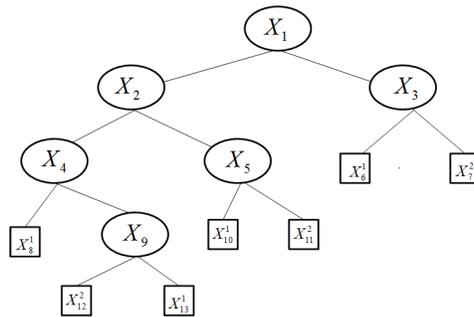


Figure B.16: Example of a tree for the two-class problem. The terminal nodes are indicated by rectangular boxes and are designated by a class label. Non-terminal nodes are indicated by circles.