Check for updates

On the spectral signature of melanoma: a non-parametric classification framework for cancer detection in hyperspectral imaging of melanocytic lesions

ARTURO PARDO,[1,*] JOSÉ A. GUTIÉRREZ-GUTIÉRREZ,[1] I. LIHACOVA,[2] JOSÉ M. LÓPEZ-HIGUERA, [1,3,4] AND OLGA M. CONDE [1,3,4]

[1]*Grupo de Ingeniería Fotónica, TEISA, Universidad de Cantabria, Avenida Los Castros S/N, 39006, Cantabria, Spain*
[2]*Biophotonics Laboratory, Institute of Atomic Physics and Spectroscopy, Raina Blvd. 19, Riga, LV-1586, Latvia*
[3]*Centro de Investigación Biomédica en Red – Bioingeniería, Biomateriales y Nanomedicina (CIBER-BBN), Cantabria, Spain*
[4]*Instituto de Investigación Sanitaria Valdecilla (IDIVAL), Calle Cardenal Herrera Oria S/N, 39011 Santander, Cantabria, Spain*
[*]*arturo.pardo@unican.es*

Abstract: Early detection and diagnosis is a must in secondary prevention of melanoma and other cancerous lesions of the skin. In this work, we present an online, reservoir-based, non-parametric estimation and classification model that allows for this functionality on pigmented lesions, such that detection thresholding can be tuned to maximize accuracy and/or minimize overall false negative rates. This system has been tested in a dataset consisting of 116 patients and a total of 124 hyperspectral images of nevi, raised nevi and melanomas, detecting up to 100% of the suspicious lesions at the expense of some false positives.

1.   Introduction

Although it only constitutes about 1–5% of total cancer cases –depending on the country–, melanoma is the deadliest form of skin cancer known to date. It accounts for at least two thirds of all skin cancer-related deaths in the US [1], and its incidence over time has also increased; for instance, it has duplicated since the 1990s (up to 128%) in the UK [2]. As a consequence of its incidence and mortality, there are numerous campaigns encouraging primary prevention, as well as increased costs in secondary prevention –also known as early detection– and treatment interventions. For example, yearly estimates of the cost of melanoma care currently hold at about $44.9 million dollars for already existing cases, and circa $932.5 million dollars for new cases across all age groups in the US Medicare system alone [3].

Current detection and early diagnosis methods are mainly clinical inspection, followed by histopatological diagnosis of the extracted lesion if the pathologist suspects malignancy. Human error has been reported to be up to 15% among experts when coming to diagnose a lesion as cancerous [4]. Therefore, it is key to devise reliable, automated screening procedures that minimize false negatives as a first priority, and –secondarily– minimize the number of false positives, if possible. It would also be ideal if early screening methods could be set in non-specialized clinics, therefore avoiding cluttering of dermatology wards with unsuspicious lesions, and providing experts with clinically difficult cases. In such a scenario, automated primary care screening could become just as reliable as skin self-examination (SSE) practices –which are still not prevalent among the general population– while also minimizing time spent by general practice physicians on whole-body skin examination (PSE) [5, 6].

The field of Applied Photonics (a *key enabling technology* (KET) as defined by the High-Level Expert Group of KETs of the European Commision [7]), has proven itself capable of minimizing both problematics. In particular, hyperspectral imaging (HSI) has shown promise for the detection of malignant lesions, given the rich chromophore-related information provided by radiation backscattered from tissues. Macroscopic HSI has been thoroughly employed in tissue classification [8–13], lesion segmentation [14, 15], and quantification of detectable chromophores [16,17], among others (for further discussion on the capabilities of HSI as a clinical technique, please refer to reviews [17], [18], [19] and [20]). In the field of dermatology, multi-spectral digital skin lesion analysis (MSDSLA) devices are capable of detecting inflammation [21], segmenting and differentiating between malignant and benign melanocytic and non-melanocytic lesions [22–29]. Chromophores present in melanocytic lesions have been evaluated in large spectroscopic studies as well [30]. Assisted diagnosis has already been tested for multispectral images in commercially available systems such as SiaScope or MelaFind, with improvements in sensitivity and specificity in the range of 5-10% [20, 31] and biopsy sensitivity and accuracy of up to 20% (from 64% to 86%) when used complementarily with standard clinical dermatology practice [32].

Most diagnostic methods in HSI imaging use well-known, standard machine learning algorithms, such as Support Vector Machines (SVMs), deep neural networks (both convolutional and feedforward), a combination of dimensionality reduction by Principal Component Analysis (PCA) and a clustering/classification algorithm (i.e. k-means, k-nearest neighbors), plain correlation [19], or a combination of these methods [33]. Unfortunately, algorithms that perform dimensionality reduction –such as PCA or the Singular Value Decomposition (SVD)– struggle with large datasets due to the computational complexity of the SVD. Neural networks and SVMs, on the other hand, rely on obtaining a hidden function that translates the problem domain into the solution domain by means of optimization, achieving classification. These two issues –i.e., the quadratic complexity of dimensionality reduction algorithms with respect to dataset size and being unable to easily determine the proper significance and clinical exploitability of any automated diagnosis– could complicate the usability of these methods in secondary prevention, and therefore should be avoided. Also, the advent of new MSDSLA systems in recent years call for a simple, concise detection protocol that allows the use of any molecular imaging system to assist in a clinical setting [32].

In this work, we describe and test an algorithm capable of dealing with large lesion datasets, reducing dimensionality while taking into account all incoming samples, and obtaining a metric or function that has a transparent, simple relationship between any pathology and its signature features, while also allowing for the classification of any additional lesions quickly and accurately, following a specified clinical policy. This combination of well-known algorithms is based on the Neyman-Pearson Lemma –the basis of the classical radar detection problem– and, given its constitutive elements, can be described as a *reservoir-based, online, multidimensional non-parametric probability density estimation and classification framework on large datasets*.

The proposed approach on the classification problem begins with finding a vector basis that can explain the dataset as a whole and reduce its dimensionality so that, through random population subsampling, it is possible to obtain statistically significant populations for each particular pathology under study, large enough to adequately interpolate the probability density function (PDF) of each tissue category in the low-dimensional, newly found feature space. Said PDFs are obtained and used to specify the most likely features or *spectral signatures* of each particular pathology. The interpolation –or, rather, estimate– of the PDF is calculated with *kernel density estimation* (KDE), and a Likelihood Ratio $\Lambda(x)$ and threshold parameter γ are established in such a way that detection is maximized. This PDF estimate is evaluated, then, on incoming new pixels, and the likelihood ratio is employed in their classification. Different applications have validated the feasibility of KDE for segmentation, in the fields of home range analysis [34], visual

surveillance systems [35] as well as hyperspectral remote sensing [36] and ship detection via synthetic aperture radar [37].

The combined action of population subsampling and density estimation of each population allows for the classification of hyperspectral images with tunable sensitivity and specificity, as the problem requires. In our case, we wish to either minimize false negatives at the expense of a few false positives (to minimize misdetection) or maximize overall accuracy.

2. Materials and methods

2.1. The melanoma HSI dataset

In order to test the proposed methodology, we require a multispectral dataset of pigmented lesions. The selected hyperspectral database was generated by Spigulis, Lihacova (formely Diebele), Kuzmina et al. during several clinical trials in a collaboration between the Biophotonics Laboratory of the Institute of Atomic Physics and Spectroscopy and several clinics in Riga, Latvia [22, 23, 38, 39]. This dataset consists of 116 patients (52 nevi patents, 33 raised nevi patients, 31 melanoma patients). In some nevi and raised nevi patients, more than one lesion has been imaged. The total number of samples is 124 (59 nevi, 35 raised nevi, and 31 melanomas). Table 1 describes the number of patients and number of lesions imaged per patient. There are no patients with more than one type of lesion (i.e. patients with benign and malignant lesions simultaneously).

For every patient, a concise protocol was followed, for which about 30 minutes were required per lesion, of which up to 2 minutes were used for long-exposure acquisition [22]. Nevi were inspected by dermatologists, whereas melanomas also underwent histological examination to prove their malignancy, as usual in these types of clinical studies [31, 40].

Table 1. Number of patients per pathology and total number of samples per category in the dataset.

Pathology	Patients with one lesion	Patients with two lesions	Total no. of samples
Nevus	45	7	59
Raised nevus	31	2	35
Melanoma	31	0	31
Total	116 patients		124 samples

The images vary in spatial resolution, but are consistent in spectral resolution per pixel. The device utilized was a 51-channel, 16-bit `Nuance EX` hyperspectral camera, which operates in the Vis-NIR range (450–950 nm), with a step of 10 nm. Optical density, OD, was directly calculated by `Nuance` software as a function of reflectance $R(\lambda)$ via the well-known equation

$$\text{OD} = -\log_{10}\left(\frac{I(\lambda)}{I_0(\lambda)}\right) = -\log_{10}\left(R(\lambda)\right), \tag{1}$$

with $I(\lambda)$ the received light intensity at the sample, and $I_0(\lambda)$ a reference spectrum, in particular a sheet of paper applied over the lesion [41]. The white reference image defines the exposure time required to maximize dynamic range in the hyperspectral system. That same exposure time is then used on the lesion itself [22].

Given the fact that there were no regions of interest (ROIs) in the dataset itself, two rectangular regions of interest were generated manually for each specimen, such that –for each of the samples– one ROI was placed within the melanocytic lesion, and the other in a surrounding area, where

there is visibly no lesion present. These ROIs (in-lesion and out-of-lesion) do not intend to provide clinical nor spatial information, as they just serve as locations for our method to extract spectra from within and outside each pigmented lesion. Special care was taken in order to keep the ROIs as clean and far away from fringe regions as possible. Also, in order to avoid any classification performance biases, we will evaluate the algorithm on a patient-by-patient basis, instead of analyzing samples separately, mimicking as realistically as possible what would happen in a real clinical setting.

2.2. Standard normal variate (SNV)

Interpatient –or intersubject– variability, i.e. the inconsistency in measurement characteristics amongst subjects that should *a posteriori* be classified as identical [42], is a prevalent phenomenon in fields such as hyperspectral imaging. Changes in position and illumination make non-contact diagnosis of material properties notably difficult in many cases [43]. In order to compensate for this variation, we chose to employ the *standard normal variate* (SNV), known for eliminating bias and trend in spectra [44]. In obtaining the SNV, each reflectance vector $x \in \mathbb{R}^n$ will be converted into a *normalized vector* x', by means of subtracting its average value $\mu = \frac{1}{N} \sum_{i=1}^{N} x_i$ and dividing by its variance:

$$x' = \frac{x - \mu}{\sigma}. \tag{2}$$

2.3. Sequential singular value decomposition (SVD)

Consider a corrected hyperspectral matrix as a tensor $\mathcal{A} \in \mathbb{R}^{m \times n \times l}$ with real elements $\{a_{mnl}\} = R(m\Delta x, n\Delta y, \lambda_0 + l\Delta\lambda)$ representing the diffuse surface reflectance of the material under test at discrete positions $(m\Delta x, n\Delta y)$ and wavelengths $\lambda_0 + l\Delta\lambda$ (with $\Delta x, \Delta y, \Delta\lambda$ the spatial and spectral resolutions established by the system, respectively), and $A = \mathcal{A}_{(3)} \in \mathbb{R}^{l \times mn}$ as its *mode-l matrization*, such that each column in A represents the spectrum of a single pixel. In other words, if we ignore pixel position and only pay attention to its spectral characteristics, the *singular value decomposition*

$$A = U\Sigma V^T \tag{3}$$

allows the representation of A as a sum of matrices of rank one

$$A = E_1 + E_2 + \cdots + E_r = \sum_{k=1}^{r} u_k \sigma_k v_k^T, \tag{4}$$

with $r = \text{rank}(A)$ the rank of our matrix, $u_1, \ldots, u_r$ the *left singular vectors* of A (columns of orthogonal matrix U), $\sigma_1, \ldots, \sigma_r$ its *singular values* (nonzero elements in diagonal matrix Σ), and $v_1, \ldots, v_r$ its *right singular vectors* (columns of orthogonal matrix V). This means that the SVD allows for lossy compression of spectra (i.e. dimensionality reduction) by means of an *L-rank approximation*, accomplished by truncating the sum in Equation (4). The error of this approximation is known; it is given by the next singular value, such that if $\tilde{A}_L = E_1 + \cdots + E_L$, $L \leq r$, then

$$\|A - \tilde{A}_L\|_F \leq \sigma_{L+1}$$

will hold true for any A [45]. Thus, if there is notable redundancy in the columns of A, it will be possible to represent each column a_t as a weighted sum of the first $L \ll r$ columns of U, namely

$$a_t \approx \tilde{a}_{L,t} = \sum_{i=i}^{L} u_i \alpha_{t,i} = \sum_{i=1}^{L} u_i \langle u_i, a_t \rangle, \tag{5}$$

where $\langle ., . \rangle$ represents the inner product of any two column vectors. The coefficients $\alpha_{t,1}, \ldots, \alpha_{t,L}$ are the coordinates of the t-th column vector of A in the space described by the first L vectors of U in Equation (3).

Finding the basis for a single image is not a problem from a practical perspective; the total number of operations required for executing the SVD on a matrix $A \in \mathbb{R}^{M \times N}$ is $O(MN^2)$, while using $O(MN)$ memory. These algorithms, though, become problematic as the number of column vectors N increases, which is something that is expected to happen when dealing with large datasets. That, and the fact that we are looking for an L-rank approximation with a very small L, calls for an algorithm that efficiently calculates the first few singular vectors of a matrix composed by concatenated matrices $A = (A_1|A_2|\dots)$ by analyzing each matrix separately.

The *sequential Karhunen-Loeve* (SKL) algorithm, which is based on the R-SVD algorithm, can be of good use for this purpose. Given a matrix $A \in \mathbb{R}^{M \times N}$, the SKL algorithm obtains –by taking the columns of A in batches of P columns– the first K columns of U in Equation (3) with negligible errors, in $O(MNK)$ operations and using $O(MK)$ space [46]. Thus, using an implementation of the SKL algorithm, and by specifying a value of K sufficiently large such that L fulfills $K > L$, but also $K, P \ll N_1 + N_2 + \dots$, it is possible to find the first K columns of U that can represent a complete dataset in a feasible manner.

Once we have a vector basis of size K, we may choose L by looking at the relative contribution (singular-value-wise) of each additional dimension to the trace of Σ (and, thus, the quality of the approximation). Instead of selecting a fixed value of L, once the SKL algorithm has provided us with the K-rank approximation $\tilde{A}_K = U_K \Sigma_K V_K^T$, we may look for the value where

$$\sigma_{contrib}(k) = \frac{\sum_{i=1}^{k} \sigma_i^2 - \sum_{i=1}^{k-1} \sigma_i^2}{\sum_{i=1}^{k-1} \sigma_i^2} \tag{6}$$

becomes negligible, e.g. where $\sigma_{contrib} \leq 10^{-3}$, which is a common dynamic threshold for dimensionality reduction [10, 14].

A projection of each spectrum a_t in the newly found feature space given by $u_1, \dots, u_L$ as in Equation (5) provides us with $x_t = (\alpha_{t,1}, \dots, \alpha_{t,L})^T$ which will correspond to the coordinates of spectrum a_t in a feature vector space of reduced dimensionality that explains with high fidelity most of the properties of all tissue samples in the dataset. This vector basis will be obtained from the ROIs specified in the previous section.

2.4. Kernel density estimation

After each spectrum is located in a vector space common to all subjects, the next step is to assume that each pixel x_t is a sample from an n-dimensional random variable X, which is assumed to have a different response to each given diagnostic hypothesis $X|H_0, \dots, X|H_m$, where m is the number of hypotheses. Specifically, this work evaluates 3 different hypotheses: H_0 for nevus; H_1 for melanoma and H_2 for healthy skin.

Ideally, we seek to obtain the probability density function of this variable given each hypothesis, namely $f(x|H_0), \dots, f(x|H_m)$. In other words, $f(x|H_1)$ represents the value of the likelihood of $x \in \mathbb{R}^L$ belonging to a specific hypothesis or tissue type, in this case H_1, melanoma. If we are given a sample set of n pixels given the k-th hypothesis, $X|H_k$, i.e. $X_{k,1}, \dots, X_{k,n}$, the probability density function (PDF) $f(x|H_k)$ at position x can be estimated using a *multivariate kernel density estimator* [47]:

$$\tilde{f}(x|H_k) = \frac{1}{n} \sum_{i=1}^{n} K(x - X_{k,i}), \tag{7}$$

and here $K(.)$ is referred to as the *kernel function*. Since the reference vectors are more determinant than the kernel function in terms of accuracy, we use the *standard multivariate normal kernel*

$$K(x) = (2\pi)^{-d/2} |H|^{-1/2} e^{-\frac{1}{2} x^T H^{-1} x}, \tag{8}$$

and here d is the dimension of the vector space and H is the *bandwidth matrix*. For our system, Silverman's rule-of-thumb estimator was used [47], which gives the optimal window with for

the smoothing of normally distributed data with variance σ_i on each dimension, and so H is a diagonal matrix whose nonzero elements correspond to

$$
h_{ij} = \begin{cases} \sigma_i \left(\frac{4}{d+2} \right)^{\frac{1}{d+4}} n^{\frac{-1}{d+4}}, & i = j, \\ 0, & i \neq j, \end{cases} \tag{9}
$$

and, since our dimension is given by the SKL algorithm, $d = L$.

2.5. C++ KDE implementation

Since kernel density estimation requires a total of n operations per reservoir (given by Equation (7)), KDE was implemented in C++ with hardware acceleration libraries, in order to reduce overall execution runtime. For this implementation, both the GSL (GNU Scientific Library) and the OpenCL (Open Computing Language) GPU (Graphics Processing Unit) APIs (Application Programming Interfaces) were used.

Verification of estimation accuracy was performed by comparing the numerical differences between the results given by the hardware-accelerated implementation and other high-level language implementations (Python 3 and MATLAB) on random RGB images, using a computer with an `Intel i7 6700 processor` (8-core, 3.4 GHz GPU) and an `nVidia GeForce GTX 1060` (3GB RAM) graphics card.

As a result of hardware acceleration, it was possible to obtain an average speedup factor of $45\times$, and a negligible average relative numerical error in the order of 10^{-7} percent. A secondary test was carried out where CPU (Central Processing Unit) multithreading was used, obtaining a $5\times$ speedup on average. The GPU-accelerated method was selected as preferable and wrapped accordingly for its use with the other Python3-implemented functions, which allowed us to generate PDF estimates of whole images in the order of milliseconds.

2.6. Reservoir sampling and Algorithm R

For a set of q pixels under analysis, and n reference pixels for each hypothesis, the number of operations specified in Equation (7) are in the order of qn products of vectors of dimension d. As the resolution of the image and/or the number of reference pixels increases, the operation becomes unfeasibly large, and so we must reduce either quantity, even when recurring to parallelization.

Random sampling –the selection without replacement of a random sample of size n from a pool of $N \gg n$ entries– is useful in this context. A *reservoir* is just the result of performing random sampling: given a complete set of samples $S, |S| = N$, a reservoir is a subset $R \subset S$ with size $|R| = n$. The main property of an adequate reservoir is that the probability of any vector in the pool to belonging to the reservoir is a constant, namely $P(s \in R) = n/N, \forall s \in S$.

There are several algorithms that allow this with more or less optimality, but for our case the simplest (*Algorithm R*) was employed, due to current processor speeds and random number generation not being the bottleneck of the system. In Algorithm R, the reservoir R is initialized with the first n entries of S (i.e. $R = \{s_1, \ldots, s_n\}$) and then the next $n + 1, \ldots, N$ entries are read consecutively. For each entry $s_t, t \in [n + 1, N]$ in S, we obtain a single sample of i, an integer, uniformly distributed random variable in the interval $[1, t]$. The i-th entry in R is substituted by s_t if $i < n$. By following this procedure, all the elements in S are read only once, enabling us to continue with an online approach [48].

2.7. Naïve Bayesian and weighted one-versus-all (OVA) classification

Given a newly acquired spectrum $x_t \in \mathbb{R}^L$ for which probability density function estimates for each tissue type $f(x_t|H_1), f(x_t|H_2), \ldots, f(x_t|H_m)$ have been calculated, we may proceed by

classifying such position by a *maximum a posteriori* (MAP) classifier:

$$f(x_t|H_k)P(H_k) \overset{H_k}{>} f(x_t|H_i)P(H_i), \forall i \neq k. \tag{10}$$

Thus, a pixel x_t is most likely to belong to class H_k if the estimated value of the likelihood function $\tilde{f}(x_t|H_k)$ at such position (times the *a priori* likelihood of the k-th hypothesis to be true) takes the largest value. If we assume that all hypotheses are equally probable, then Equation (10) becomes

$$f(x_t|H_k) \overset{H_k}{>} f(x_t|H_i), \forall i \neq k, \tag{11}$$

which is the equation corresponding to the Maximum Likelihood (ML) classifier [49]. We can therefore use the *one-versus-all* (OVA) criterion for pixel classification. A pixel x_t will be assiged a class $b_t \in \{H_0, \ldots, H_m\}$ by following the OVA equation

$$\hat{b}_t = \arg \max_i \{f(x_t|H_i)\}, \tag{12}$$

which can be weighted in order to change the probability of classification for any particular class, by including real-valued weights $\gamma_0, \ldots, \gamma_N \in \mathbb{R}$ to the classifier:

$$\hat{b}_t = \arg \max_i \{f(x_t|H_i)\gamma_i\}. \tag{13}$$

For the problem at hand, $\gamma_i = 1$ except for the melanoma class, which will be the varying parameter that sets, in practice, true positive and false positive rates.

2.8. Online interpatient-invariant property learning

The implementation of the reservoir-based classifier was achieved by the interconnection of the five different functions described so far, as shown in Fig. 1. The system has two modes of operation: *sample learning* and *sample evaluation*. We will define m as the number of classes –hypotheses– pertaining to the problem (in our case, categories for *nevi*, *melanoma* and *healthy skin*.)

During sample learning –depicted by the blue arrows in Fig. 1–, spectra-label pairs (a'_t, l'_t) are acquired in batches of size P. The SKL algorithm is then applied on the P-sized batch of spectra, ignoring their labels. Low-dimensional feature space basis vectors $u_1, \ldots, u_K$ are updated after every batch. After this calculation, each pixel in the batch is sent, without reducing its dimensionality, into its corresponding size-n reservoir $R_0, \ldots, R_m$ (depending on its label), where *algorithm R* selects a subset of them randomly, generating a statistically significant population of each class, $\{r_{0,i}\}, \ldots, \{r_{m,i}\}$. Pixels that have not been randomly selected are discarded in each reservoir.

Secondly, during sample evaluation (arrows in magenta), $L < K$ is recalculated dynamically by means of Equation (6). Pixels in each reservoir are projected onto the final vector basis $u_1, \ldots, u_L$ calculated with SKL, obtaining the *reference points* for each class in the L-dimensional vector space, $\{X_{0,i}\}, \ldots, \{X_{m,i}\}$. Once the n-sized reservoirs are represented in the updated feature space, a new incoming pixel a_t can be processed. This pixel is also projected onto the low-dimensional vector space, and its likelihood of belonging to each class is estimated $(f(x_t|H_0), \ldots, f(x_t|H_m))$. Classification follows suit, using Equation (13) (or, also, the MAP/ML classifier), obtaining its estimated tissue label $\hat{l}_t$.

These two steps describe, therefore, an on-line system that obtains a low-dimensional feature space that represents features in the whole dataset and, at the same time, the most frequent locations of these point clouds in this feature space are determined by the PDF estimates. A new pixel is most likely to belong to the class with the highest PDF value at its position in feature space, and therefore its assigned class will follow that rule.
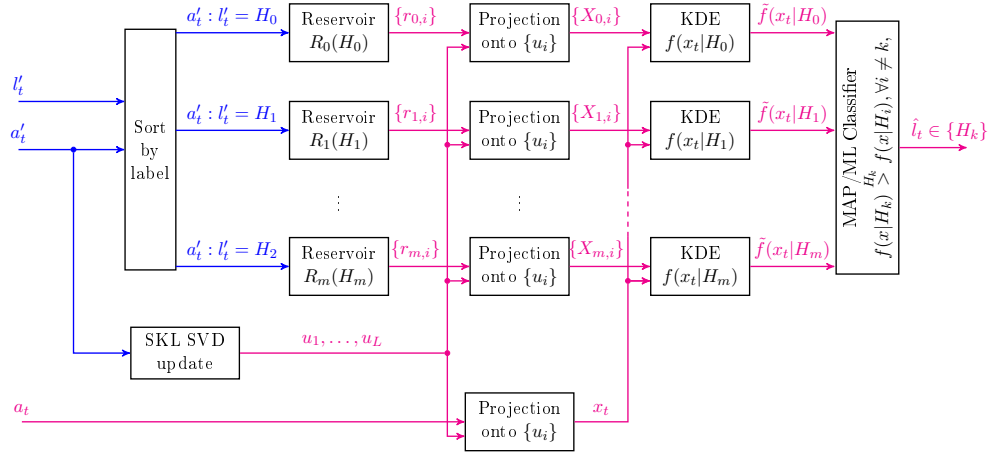
Fig. 1. General schematic description of the workflow and modes of operation implemented for the reservoir-based kernel density classifier. During learning (in blue) spectra-label pairs (a'_t, l'_t) are used to update a feature space vector basis $B = (u_1, \ldots, u_L)$ via the SKL algorithm and Equation (6), sorted by label and introduced into random reservoirs, where they have a chance of staying as reference spectra $\{r_{k,i}\}$. During evaluation (in magenta), a new incoming spectrum a_t is projected onto feature space, where it is compared with reference spectra. For that to be possible, reference spectra must also be projected onto B. The likelihood of a_t belonging to each tissue type is evaluated through KDE, and a final diagnosis $\hat{l}_t$ is calculated.

For the results section, in order to evaluate the generalization capabilities of this method, we have enforced different approaches. In particular, we have evaluated our algorithm with 10-fold and 3-fold cross-validation, as well as leave-one-out cross-validation (LOOCV). In all cases, we seek to learn from patients instead of separate samples, in order to avoid biasing our algorithm with *a priori* information. Diagnoses are accounted in a per-sample basis in order to simulate a realistic clinical setting.

In leave-one-out cross-validation, for each of the 116 patients, reservoirs were emptied, SKL was reset to its initial condition, and sample acquisition was performed with the other 115 patients. Then, evaluation was performed on the patient under test, as if it were a new incoming patient that the system had not seen before. Its samples are evaluated separately and their diagnosis are tallied on the confusion matrix. Threshold parameters $\gamma_0, \ldots, \gamma m$ are to be evaluated as a function of its effect on the Receiver Operating Characteristic (ROC). A similar approach is enforced for 3-fold and 10-fold cross-validation.

3. Results and discussion

3.1. Dimensionality reduction and vector basis

The first step in our process is the on-line extraction of a low-dimensional feature space. We must study the characteristics of the basis that describes such space, as well as how many dimensions are necessary to explain most of the features in it. The first five left-singular vectors that conform the space for the HSI melanoma dataset, as well as the cumulative sum of the square of the singular values of Σ in the SVD, are shown in Fig. 2. The first three basis vectors in feature space (those that represent an absorption peak around 550 nm and a slope in the red and infrared wavelengths) are the best feature descriptors for the problem at hand.

Although these feature basis vectors do not represent any isolated physical characteristics for
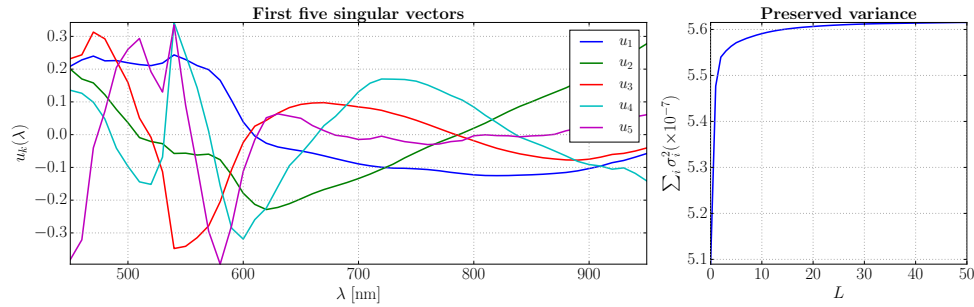
Fig. 2. First five left singular vectors, namely $u_1, \ldots, u_5$, of the HSI melanoma feature space (left) and cumulative sum of explained variance (square of singular values).

the dataset, a change in slope in the 600-900 nm range characteristic of melanin, and the typical absorption peaks of hemoglobin around 530-600 nm can be seen [17, 22]. It must also be noted that, since the SVD is a deterministic operation, the results and ordering of these feature vectors will be a result of the amount of data processed so far. These events are, therefore, the most repeated occurrences in all observed tissue spectra [45].

The low dimensionality required to represent melanocytic lesion spectra, as shown in the right subplot of Fig. 2, must be commented as well. Applying Equation (6) provides us with an L in the range of 3 to 10, depending on the degree of explained variance desired. For the aforementioned $\sigma_{contrib} \leq 10^{-3}$ criterion, L stays in the range $L < 11$. This confirms the fact that most of our data lies in a feature space of low dimensionality, as is common in hyperspectral images of biological tissues, and that using more dimensions will not benefit estimation nor classification significantly [9, 14]. Figure 3 represents the positions in feature space of 4000 points per pathology, selected via *Algorithm R* in three distinct reservoirs, and pre-processed by the Standard Normal Variate algorithm (hence, laying on an L-dimensional sphere).
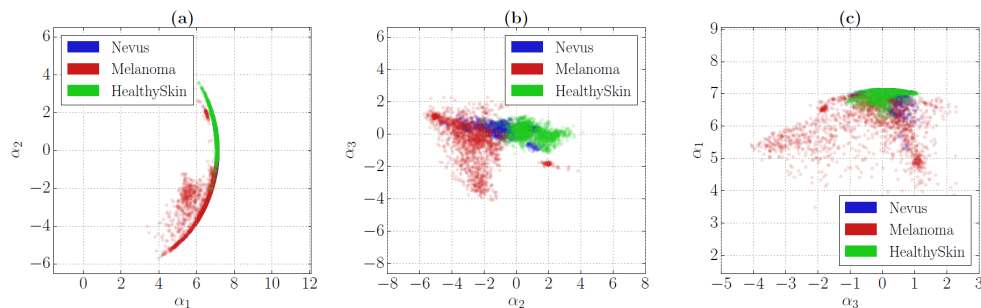

Fig. 3. Data clouds of the dataset in the basis defined by u_1, u_2 and u_3 (4000-point reservoir per category). The scatter plots are 2D and they display (a) the first two dimensions in feature space, (b) the second and third dimensions in feature space, and (c) the third and first dimensions in feature space, as given by their coordinates ($\alpha_1, \alpha_2, \alpha_3$). The data lay in an L-dimensional sphere due to the Standard Normal Variate.

In order to test the validity of our ROIs, we should at least compare the average spectra of each pathology with other state-of-the-art studies. Figure 4 displays the average spectra and standard deviation of nevi, melanoma and healthy skin, respectively. Average reflectance as well as standard deviations for each of the pathologies are within range of what has been shown in larger

clinical trials [28, 30] and in the articles that explain how the dataset was obtained [22, 38, 41].

A few things can be noted about the spectra. Melanoma spectra are notably different in both shape and absorption in certain wavelengths. The presence of melanin can be seen in the 600–700 nm range, whereas healthy skin shows lower absorption and variance. Note also the lack of absorption in the 400–500 nanometer range in melanoma, perhaps due to blue-white veil present in the images, as well as different concentrations of oxy- and deoxyhemoglobin [17]. Note, though, that the variability of melanoma –as shown by the scatter plots of Fig. 3– is more significant, and thus evaluating spectra with every spectrum in the dataset or with the average spectrum of melanoma turns practically unfeasible and prone to misdiagnosis. This can be explained by the fact that there will be non-cancerous spectra in feature space closer to the average spectrum of melanoma than to the average spectrum of nevi, and viceversa. Such a result suggests that estimating likelihood and class probability metrics in feature space will certainly provide a similarity function that holds true across all cases consequence of interpatient variability.
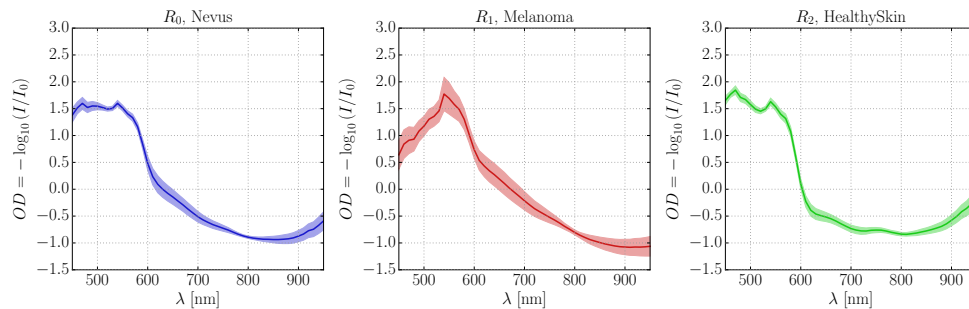

Fig. 4. Average values (solid line) and standard deviation (shaded regions) of spectra within reservoirs of size 4000 for each tissue class considered: nevi (left), melanoma (center) and healthy skin (right).

3.2. Segmentation and qualitative results

If we wish to diagnose a lesion by tissue type ratios, we must ensure lesion segmentation works in all the samples. We have chosen leave-one-out cross-validation as a first approach. For a given patient, every sample in the dataset except that patient was used to estimate the PDF function of each tissue type in feature space. Every pixel in each image will have a location in feature space, and therefore will be assigned a tissue class with weighted OVA (depending on the weights shown in Equation (13)).

From now on, we will consider three tissue types or hypotheses: the *Nevus*, *Melanoma* and *Skin* classes will be referred to as H_0, H_1, and H_2, respectively (skin-lesion segmentation must be performed automatically, hence why a 'Healthy Skin' reservoir is also considered). For a given sample, the PDF or likelihood values $f(x|H_0)$, $f(x|H_1)$ and $f(x|H_2)$ can be calculated for each pixel. As shown in Equation (13), for classification there is an associated weight for each tissue class, namely γ_0 for Nevus, γ_1 for Melanoma and γ_2 for Skin spectra. This constitutes a three-parameter system with two degrees of freedom: that given by γ_1 in relation to γ_0 will allow us to compare benign with malignant melanocytic spectra, and that given by γ_2 with respect to any pigmented spectra (γ_1 and γ_0) will provide segmentation.

Since skin pigmentation will be dependent on patient skin fairness, we have chosen a dynamic version of the *elbow method* [50]. We shall begin with the maximum likelihood scenario, where

$\gamma_0 = \gamma_1 = 1$. For any value of γ_2, we can calculate the ratio

$$\rho(\gamma_2) = \frac{N_{les}(\gamma_2)}{N_{total}} \in [0, 1], \tag{14}$$

where N_{les} is the number of pixels classified as pigmented spectra (Nevus or Melanoma) that are obtained with Equation (13), and N_{total} is the total number of pixels in the image. This ratio can take any value between 0 and 1, with $\rho = 1$ meaning all pixels are considered pigmented and $\rho = 0$ meaning no pixels in the image are pigmented. As γ_2 increases, more skin spectra will be classified as skin. The ratio will decrease up to a point of saturation, where ρ stabilizes for a finite interval of γ_2 and the lesion remains classified as such, and afterwards more and more pixels in the lesion are classified as skin until $\rho = 0$ and the lesion ROI vanishes. Similarly to the dynamic thresholding in Equation (6), we can calculate the finite-difference approximation of the second derivative of ρ with respect to γ_2. As shown in Fig. 5, the second derivative of $\rho(\gamma_2)$, $\rho''(\gamma_2)$, relates with the variation of its curvature. If the maximum value of $\rho''(\gamma_2)$ indicates where the elbow is, then stabilization will take place after the elbow, once ρ'' tends to zero again. Therefore, we can choose a relative dynamic threshold, such that the optimal segmentation value, $\gamma_{2,opt}$, is chosen by following

$$\gamma_{2,opt} = \left\{ \gamma_2 : \frac{\rho''(\gamma_2)}{\rho''_{max}} \le 0.01 \right\}, \tag{15}$$

and thus finding the end of the elbow in $\rho(\gamma_2)$. Here, ρ''_{max} is the maximum of the second derivative. In other words, we choose a value for γ_2 where the curvature of the size of the lesion stabilizes up to 1/100 of its maximum variation. Since numerical variables usually have noisy values, $\rho''(\gamma_2)$ has been stabilized with a Savitzky-Golay filter (window size 11, order 7). This procedure is followed analogously in n-fold cross-validation.
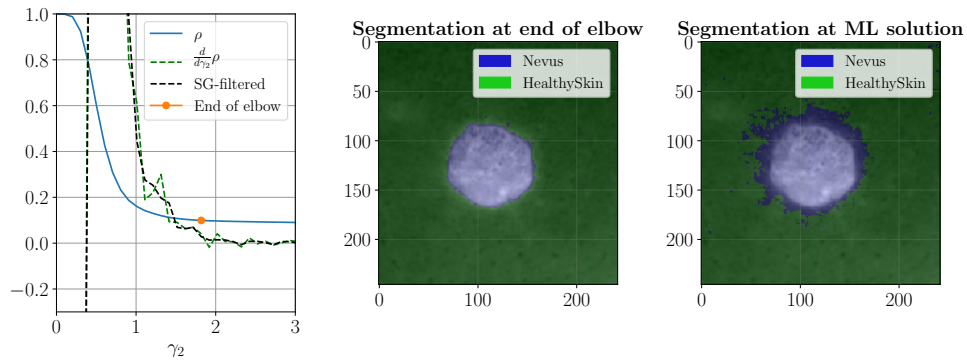

Fig. 5. Elbow method segmentation example for Patient 2, Sample 2. On the left plot, values of ρ, (in blue), its derivative (in green and black) and the elbow point chosen by the criterion in Equation (15). The center and right subplots show the segmentation given by the elbow method and standard maximum likelihood classification (i.e. $\gamma_0 = \gamma_1 = \gamma_2$), respectively.

Once segmentation is achieved, in-lesion classification just requires separating pigmented pixels with their PDF values. Figures 6 and 7 are two examples of the results provided by the system. The SKL subroutine was defined with batch size $P = 200$, forget factor $f = 1$, and $\epsilon = 10^{-5}$. Reservoir size was 4000 per spectral class. Each figure contains an RGB reconstruction of the absorption spectra, using the CIE 1931 color matching functions (CMFs) and considering a D65 illuminant [51]. The final classification/segmentation results are shown in the right-side

subplots of the first row of each example. Segmentation behaves generally well, for as long as there are no hairs in the image (hair has some degree of brown pigmentation, and is sometimes indistinguishable from with melanocytic tissue spectra). Note that this segmentation is obtained by directly comparing the three subplots in the bottom row. In most nevi cases, the melanoma hypothesis is illuminated within the melanocytic lesion, but its likelihood function is usually ten times lower than the nevus function – something which is not the case for malignant lesions, where the 'nevus' and 'skin' likelihood functions are zero and the melanoma likelihood function is larger.
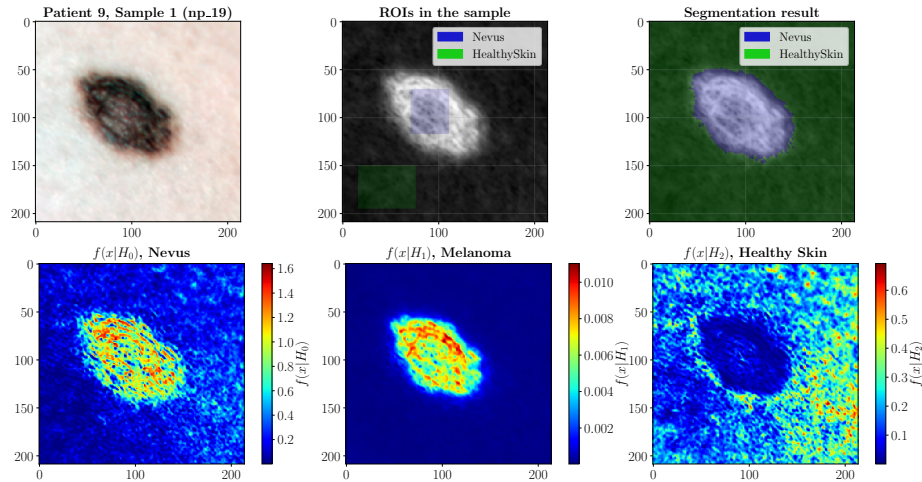

Fig. 6. **Top row.** CIE 1931 RGB reconstruction of Sample 24, a nevus (left), as well as the in-lesion and out-of-lesion ROIs (center); these ROIs aren't used in the classification result (shown as a semitransparent overlay on the right subplot). **Bottom row.** False-color likelihood maps for the three hypotheses (from left to right, H_0 : *Nevus*, H_1 : *Melanoma*, and H_2 : *Healthy Skin*); note that each figure is complemented with a color bar with a different scale, otherwise the likelihood map for melanoma would be too dark to see. Since the likelihood of being melanoma is lower than the likelihood of being nevi in each pixel, they are classified as non-malignant.

It must be noted that, given the higher variability of malignant spectra (as shown in Fig. 3), the PDF values for the Melanoma hypothesis are lower. This is an expected result, since KDE provides a PDF estimate and, by definition, the integral over all L-dimensional space must be equal to one. Nevertheless, the PDF values for the other hypotheses in the regions of feature space dominated by melanoma are zero or near zero, making the Melanoma class the dominant spectral signature. Another characteristic that is of notable interest is the fact that the algorithm tries –in some degree or another– to assign spectra of areas surrounding a melanoma as belonging to the nevi class. In such fringe areas the system simply assigns pixels to the class they are most likely to belong to, as constrained by the segmentation threshold. Given the lack of expert ROI information, is unknown if such regions would be pigmented, non-malignant areas; further research would be needed for a proper assesment of this response.

3.3. Diagnostic capability and ROC. Quantitative results

We will undertake now the study of the diagnostic capabilities of our algorithm, by evaluating its Receiver Operating Characteristic (ROC) and typical binary classification parameters. In order to describe a single sample as benign or malignant, a *diagnostic rule* must be established. For this scenario, given the fact that nevi pixels are rarely classified as malignant, the defined rule
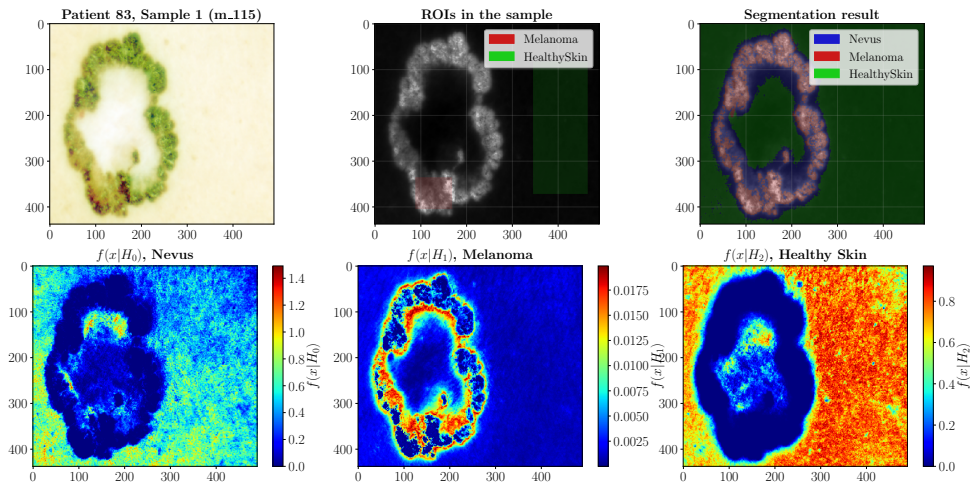
Fig. 7. **Top row.** CIE 1931 RGB reconstruction of Sample 67, a melanoma (left), its in-lesion and out-of-lesion ROIs (center) and its final classification/segmentation output as a semitransparent overlay (right). **Bottom row.** False-color likelihood maps for the three hypotheses (from left to right, H_0 : *Nevus*, H_1 : *Melanoma*, and H_2 : *Healthy Skin*) for this sample, with their accompanying color bars. The likelihood of pixels in the lesion for the nevi and skin hypotheses is zero, and therefore most pixels within the lesion are classified as more likely to be melanoma, while fringe regions in between skin and lesion are considered just pigmented, due to their likelihood values in feature space.

is a *diagnostic ratio*, calculated as follows: given a specific sample with a lesion of size N_{les} (in pixels) containing $N_{mal} \leq N_{les}$ pixels within the lesion classified as melanoma, it will be diagnosed as *malignant* if $N_{mal}/N_{les} \geq p$ holds. In summary, pixels are classified following the weighted *one-vs-all* method (Equation (13)), while whole-sample diagnosis is achieved with diagnostic ratio p.

In order to test the capabilities of our system, we have evaluated classification accuracy for 3-fold and 10-fold cross-validation, as well as leave-one-out. For leave-one-out, Fig. 8 depicts both the ROC as well as overall system accuracy as a function of γ_1, swept from $\gamma_1 = 0$ to $\gamma_1 = 30$, for several values of p: 0.01, 0.05, 0.1 and 0.2. As shown by this figure, the system can achieve stable accuracies up to 95%, depending on diagnostic criterion stringency, and it behaves like an almost-ideal radar detector. This implies, also, that lesion segmentation performs notably well for $\gamma_0 = 1$ and $\gamma_2 = 1$, given the fact that diagnosis is achieved by only counting pixels *within* the lesion itself. At peak performance for $p = 0.05$ ($\gamma = 6.1224$), the method is capable of reaching 96.8% sensitivity, 95.7% specificity, and 96.0% accuracy. For the same diagnostic criterion, a null false negative rate can be achieved at $\gamma = 7.9592$, with a false positive rate of about 10%. Such a tradeoff is to be expected, as false positives will increase in number as the false negative rate is minimized. Additional cross-validation tests can be found in Tables 2 and 3, with $p = 0.05$ and maximum accuracy and minimum false negative rate criteria, respectively. We must note that 3-fold cross-validation, due to small dataset size and reference spectra, has notably lower (16% less) accuracy and cannot achieve zero false negatives in the range $\gamma_1 \in [0, 30]$. This is to be expected, but it is kept in Table 3 for the sake of comparison. As is expected as well, a lower number of reference patients comes associated with lower accuracies, but still within range of state-of-the-art systems [28, 32].

This figure also exhibits the change in classification performance as the diagnostic ratio increases. As is expected, the larger the value that this ratio takes, the larger γ_1 must be in order
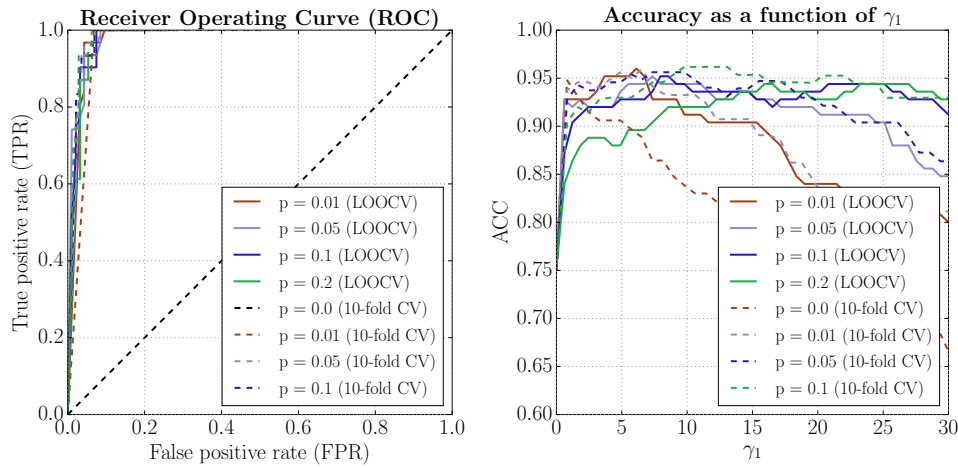
Fig. 8. System Receiver Operating Characteristic (ROC, left) and total accuracy as a function of γ_1 (right), for different diagnostic ratios in leave-one-out cross-validation, namely 0.01, 0.05, 0.1, and 0.2.
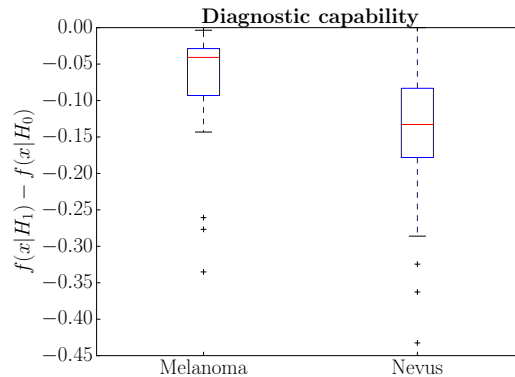

Fig. 9. Boxplot with the difference between the average f_1 and f_0 values inside each lesion, $f_1 - f_0$, organized by pathology class. The difference between average class likelihoods per sample is smaller in nevi, due to the fact that their likelihood for being melanoma $f(x|H_1)$ is lower. This variation in likelihood functions is the key to the diagnostic capability of our procedure.

to classify the same amount of pixels as malignant and thus reach the same diagnosis. This relationship between γ_1 and p is empirical and therefore a conservative criterion was chosen: if a given ratio p is selected such that it contains the peak accuracy point for a value of γ_1 in the order of the other values $\{\gamma_i\}$, then that p is adequate for classification, as long as the γ_1 that maximizes accuracy can be selected. Thus, $p = 0.05$ was selected as an adequate diagnostic ratio.

The diagnostic power of likelihood functions $f_1(x)$ and $f_0(x)$ can also be studied by obtaining their average values within each lesion. Then, relationships such as $d(x) = f_1(x) - f_0(x)$ or f_1/f_0 can be calculated for each lesion separately. The functions will be suitable for diagnostic classification if they present different values per category. The average values of $f_1(x)$ and

$f_0(x)$ per sample were obtained, their difference calculated and portrayed in a boxplot (Fig. 9). Although the algorithm tends to provide larger values to the nevi category, $d(x)$ presents clear separability per tissue class by at least one full standard deviation.

Table 2. Overall system performance for different cross-validation (CV) settings (leave-one-out, 10-fold and 3-fold) given a diagnostic ratio of $p = 0.05$ and maximizing for accuracy. For 3-fold and 10-fold CV, half the standard deviation of each parameter is also indicated.

Method	Optimal γ_1	Sensitivity at optimal γ_1	Specificity at optimal γ_1	Precision at optimal γ_1	Accuracy at optimal γ_1
LOOCV	6.122	0.968	0.957	0.882	0.960
10-fold CV	7.346	0.933 ± 0.067	0.964 ± 0.054	0.944 ± 0.083	0.956 ± 0.042
3-fold CV	8.163	0.799 ± 0.141	0.927 ± 0.051	0.870 ± 0.092	0.899 ± 0.035

Table 3. Overall system performance for different cross-validation (CV) settings (leave-one-out, 10-fold and 3-fold) given a diagnostic ratio of $p = 0.05$, and minimizing false negative rates (FNR). For 3-fold and 10-fold CV, half the standard deviation of each parameter is also indicated.

Method	Minimum FNR γ_1	Sensitivity at γ_1	Specificity at γ_1	Precision at γ_1	Accuracy at γ_1
LOOCV	7.9592	1.000	0.904	0.775	0.928
10-fold CV	15.5102	1.0 ± 0.0	0.931 ± 0.054	0.869 ± 0.09	0.947 ± 0.042
3-fold CV	>25.306	0.867 ± 0.094	0.871 ± 0.051	0.764 ± 0.088	0.873 ± 0.021

3.4. Performance as a function of reservoir size

It is clear, given Equation (7), that the presented methodology is highly dependent on how good the PDF approximations are. As with all interpolation procedures, the quality of such approximation will be a function of the number of pixels used in it. Although it has been shown in previous work that KDE can interpolate with few pixels with notable accuracy [14], we must study the influence of PDF estimation on overall classification performance.

For Figure 10, the effect of reservoir size on diagnostic criteria and PDF estimate fidelity was put to the test, in the following procedure. Leave-one-out cross validation was repeated on a fifth of the samples (i.e. Samples 1, 6, 11, 16, ..., up to Sample 116). First, ten simulations with a reservoir size per hypothesis of 200 (as depicted by Fig. 1) obtained the PDF estimates of each sample using leave-one-out on this reduced dataset, as in previous sections. This was then repeated for other reservoir sizes, namely 500, 1000, 2000 and 4000.

Each of the 10 simulations provide a different PDF estimate of each sample. For the leftmost subplot of Fig. 10, we calculated the variance in average PDF value among the 10 simulations in each pixel, then calculated and then averaged for all pixels and all 25 samples. Such procedure is then repeated for all the other reservoir sizes. This provides us a proof that the PDF estimates, although obtained by means of a random reservoir, will vary on average in the order of 10^{-4} for benign tissue and 10^{-8} for malignant tissue, depending on the reference pixels used, which is rather negligible ($< 1\%$) for PDF estimate values in the range [0.02, 2].

All 10 PDF estimates per sample are then classified following the accuracy-maximization criterion –weighted one-versus-all, $\gamma_1 = 6.114$, $p = 0.05$ (5%)–. Segmentation parameters were kept as specified in Section 3.2. For each simulation we classified, as indicated in Section 3.3, all

pixels in each sample, and obtained the number of lesion pixels N_{les} and the number of malignant pixels N_{mal} as well. A sample will be classified as malignant if $N_{mal}/N_{les} \geq p$, and therefore the variance of this ratio amongst simulations must be as close to zero as possible in order for this method to be valid. Variance in N_{mal}/N_{les} present among simulations was calculated and plotted, per sample, in the right and center subplots of Fig. 10. This variance is worse, as expected, between simulation runs with smaller reservoirs, although it remains well under 10^{-2} for reservoirs of size greater than 1000 in most samples. This, in turn, means that only a small number of pixels per lesion are differently classified in different simulations, which empirically demonstrates that both the PDF estimates provided by KDE and the defined diagnostic criterion are resilient to changes in the reference population, as long as such population is sufficiently large and accurately represents each pathology in feature space.
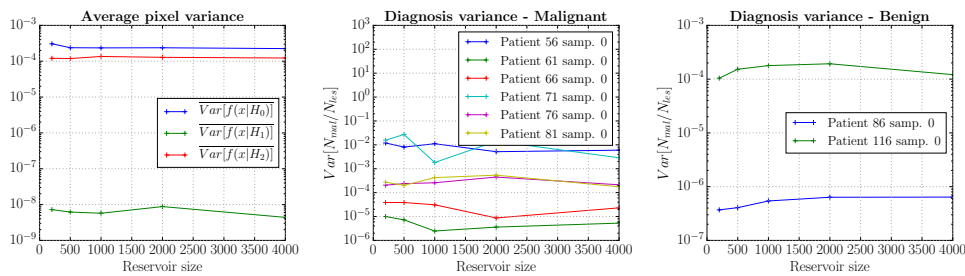

Fig. 10. Consistency in PDF estimates with random reservoirs. Average variance of pixel PDF values among machines of identical reservoir size (left subplot); per-sample variance in diagnostic ratio among machines of identical reservoir size for malignant (center) and benign lesions (right subplot). Probability density function and diagnosis variability are generally negligible, but they are stabilized once reservoir size is in the order of 2000-3000 reference samples, meaning that these sizes provide sufficient statistically and clinically significant information.

4. Conclusions

In this work, the ability of reservoir-based KDE to segment and classify hyperspectral images while still correcting for interpatient and intersample variability has been described and empirically tested, and a simple protocol inspired by the Neyman-Pearson lemma has been described. The system can be configured to maximize accuracy, obtaining a nominal sensitivity, specificity and accuracy of 96.8%, 95.7% and 96.0%, respectively, on a hyperspectral skin lesion dataset with 116 patients. It can also minimize false negative rates, obtaining a minimum false positive rate of 9.5%. These statistics are either comparable or an improvement over other state-of-the-art algorithms for spectral images of nevi [22, 25, 27, 31, 33, 38, 40], while its theoretical background simplifies its usage, as it has but one degree of freedom in its operation.

In order to avoid complicating the clinical translation of this methodology, we seek to adapt the method to the dermatologists' needs. Modifying classification parameter γ_1 provides for now two modes of operation, which allows the algorithm to work in different scenarios:

- If we let γ_1 take the optimal values of Table 2, that is, the value that maximizes accuracy for our diagnostic criterion, we would be operating at optimal performance, at the maximum true positive rate and minimum false negative rate that is achievable simultaneously.

- Finally, if we used γ_1 as shown in Table 3, we would be operating in a zero-false-negative-rate regime, which would perform in a greedier fashion, at the expense of not misdetecting any melanomas.

A notable difference is that, while classification results are well within those of state-of-the-art machine learning systems, the procedure by which it takes place does not require training *per se*. Instead, it estimates the probability distribution of each pathology in a feature space obtained by reference data. No hidden function is calculated and no parameter learning takes place. Segmentation threshold γ_2 is obtained dynamically in each lesion to achieve automatic segmentation and the malignant detection threshold parameter γ_1 could be either ignored ($\gamma_1 = 1$), or set to the aforementioned 'maximum accuracy' or 'minimum false negative rate' operation modes, which could be recalculated as new information arrives to the system.

It is fundamental to point also that this method could be used in other classification problems with high dimensionality, even those presenting noise. As long as the magnitude of the noise does not exceed significantly that of the features that differentiate classes among themselves –which would make separability impossible in any case–, KDE would provide a smooth PDF estimate, and classification by thresholding could be performed. Thus, applications of this algorithm include, but are not limited to: on-line tissue segmentation during surgery; detection of the presence of fluorescent markers and spurious artifacts; textural analysis and classification of superficial lesions.

In a clinical setting, imaging a single lesion takes about 20-30 minutes. Learning and diagnosis without methodology requires less than a minute (from milliseconds up to a second for the diagnostic step, if the algorithm has been prepared in advance), which is acceptable in the diagnosis of suspicious lesions [22]. Ideally, the rise of new modulated imaging systems that can provide depth-resolved absorption and scattering information, such as those utilized recently in melanocytic specimens [52], could also obtain with this method three-dimensional probability maps of the hyperspectral voxels they are capable of measuring. That could provide surgically relevant information in many subfields of oncological surgery, such as lumpectomy surgery –e.g. providing the depth of the surgical margins in any lesion– and, in the field of dermatology, provide an estimate of the Breslow depth of any skin lesion.

For now, given the exhibited high speed, accuracy, and segmentation capabilities of this algorithm, the time required for proper diagnosis will be only limited by the imaging system, and thus a first clinical trial could already take place, one in which the proposed methodology is employed alongside already-existing MSI/HSI cameras and systems such as MelaFind [20], providing a diagnostic assessment that minimizes false negatives and outputs the overall likelihood of malignancy of any melanocytic lesion. This information could be used in secondary prevention, and assist general practitioners and expert dermatologists in their mighty endeavor.

Funding

Disclosures

The authors declare that there are no conflicts of interest related to this article.

References

1. R. L. Siegel, K. D. Miller, and A. Jemal, "Cancer statistics, 2018," CA: A Cancer J. for Clin. **18**, 7–30 (2018).
2. C. R. UK, "Melanoma skin cancer statistics," cancerresearchuk.org (2015).
3. G. P. G. Jr, D. U. Ekwueme, F. K. Tankga, and L. C. Richardson, "Melanoma Treatment Costs: A Systematic Review of the Literature, 1990-2011," Am J Prev Med. **43**, 537–545 (2012).
4. G. Merlino, M. Herlyn, D. E. Fisher, B. C. Bastian, K. T. Flaherty, M. A. Davies, J. A. Wargo, C. Curiel-Lewandrowski, M. J. Weber, S. A. Leachman, M. S. Soengas, M. McMahon, J. W. Harbour, S. M. Swetter, A. E. Aplin, M. B. Atkins, M. W. Bosenberg, R. Dummer, J. Gershenwald, A. C. Halpern, D. Herlyn, G. C. Karakousis, J. M. Kirkwood, M. Krauthammer, roger S. Lo, G. V. Long, G. McArthur, A. Ribas, L. Schuchter, J. A. Sosman, K. S. Smalley, P. Steeg,

N. E. Thomas, H. Tsao, T. Tueting, A. Weeraratna, G. Xu, R. Lomax, A. Martin, S. Silverstein, T. Turnham, and Z. A. Ronai, "The State of Melanoma: Challenges and Opportunities," Pigment. Cell Melanoma Res. **29**, 404–416 (2016).

5. C. Curiel-Lewandrowski, S. C. Chen, and S. M. Swetter, "Screening and Prevention Measures for Melanoma: Is There a Survival Advantage?" Curr Oncol Rep **14**, 458–467 (2012).

6. M. K. Tripp, M. Watson, S. J. Balk, S. M. Swetter, and J. E. Gershenwald, "State of the Science on Prevention and Screening to Reduce Melanoma Incidence and Mortality: The Time is Now," CA Cancer J Clin. **66**, 460–480 (2017).

7. European Commision, "Key enabling technologies: Time to act (final report)," (2015).

8. H. Akbari, K. Uto, Y. Kosugi, K. Kojima, and N. Tanaka, "Cancer detection using infrared hyperspectral imaging," Cancer Sci **102**, 852–857 (2011).

9. P. B. García-Allende, V. Krishnaswamy, P. J. Hoopes, K. S. Samkoe, O. M. Conde, and B. W. Pogue, "Automated identification of tumor microscopic morphology based on macroscopically measured scatter signatures," J. Biomed. Opt. **14**, 034034 (2009).

10. A. Eguizábal, A. M. Laughney, P. B. García-Allende, V. Krishnaswamy, W. A. Wells, K. D. Paulsen, B. W. Pogue, J. M. López-Higuera, and O. M. Conde, "Direct identification of breast cancer pathologies using blind separation of label-free localized reflectance measurements," Biomed. Opt. Express **4**, 1104–1118 (2013).

11. A. M.Laughney, V. Krishnaswamy, P. B. García-Allende, O. M. Conde, W. A. Wells, K. D. Paulsen, and B. W. Pogue, "Automated classification of breast pathology using local measures of broadband reflectance," J. Biomed. Opt. **15** (2010).

12. A. M. Laughney, V. Krishnaswamy, E. J. Rizzo, M. C. Schwab, R. J. Barth, B. W. Pogue, K. D. Paulsen, and W. A. Wells, "Scatter spectroscopic imaging distinguishes between breast pathologies in tissues relevant to surgical margin assessment," Clin. Cancer Res. **18**, 6315–6325 (2012).

13. V. Krishnaswamy, P. J. Hoopes, K. S. Samkoe, J. A. O'Hara, T. Hasan, and B. W. Pogue, "Quantitative imaging of scattering changes associated with epithelial proliferation, necrosis, and fibrosis using microsampling reflectance spectroscopy," J. Biomed. Opt. **14**, 014004 (2009).

14. A. Pardo, E. Real, V. Krishnaswamy, J. M. L. Higuera, B. W. Pogue, and O. M. Conde, "Directional Kernel Density Estimation for Classification fo Breast Tissue Spectra," IEEE Transactions on Med. Imaging **36**, 64–73 (2017).

15. R. Hennessy, S. Bish, J. W. Tunnell, and M. K. Markey, "Segmentation of diffuse reflectance hyperspectral datasets with noise for detection of melanoma," in *34th Annual International Conference of the IEEE EMBS, San Diego, California, USA, 28 August – 1 September,* (2012).

16. J. McMudry, G. Jay, S. Suner, and G. Crawford, "Photonics-based *in vivo* total hemoglobin monitoring and clinical relevance," J. Biophotonics **2**, 277–287 (2009).

17. S. L. Jacques, "Optical properties of biological tissues: a review," Phys. Medicine Biol. **58**, R37–R61 (2013).

18. S. A.Leachman, P. B. Cassidy, S. C. Chen, C. Curiel, A. Geller, D. Gareau, G. Pellacani, J. M. Grichnik, J. Malvehy, J. North, S. L. Jacques, T. Petrie, S. Puig, S. M. Swetter, S. Tofte, and M. A. Weinstock, "Methods of melanoma detection," in *Melanoma (Cancer Treatment and Research, volume 167),* H. L. Kaufman and J. M. Menhert, eds. (Springer, 2015), chap. 3, pp. 51–105.

19. G. Lu and B. Fei, "Medical hyperspectral imaging: a review," J. Biomed. Opt. **19**, 010901_1–010901_23 (2014).

20. L. K. Ferris and R. J. Harris, "New diagnostic aides for melanoma," Dermatol. Clin. **30**, 535–545 (2012).

21. T. Kim, M. A. Visbal-Onufrak, R. L. Konger, and Y. L. Kim, "Data-driven imaging of tissue inflammation using rgb-based hyperspectral reconstruction toward personal monitoring of dermatologic health," Biomed. Opt. Express **8**, 5282–5296 (2017).

22. I. Diebele, I. Kuzmina, A. Lihachev, J. Kapostinsh, A. Derjabo, L. Valeine, and J. Spigulis, "Clinical evaluation of melanomas and common nevi by spectral imaging," Biomed. Opt. Express **3**, 467–472 (2012).

23. I. Kuzmina, I. Diebele, J. Jakovels, J. Spigulis, L. Valeine, J. Kapostinsh, and A. Berzina, "Towards noncontact skin melanoma selection by multispectral imaging analysis," J. Biomed. Opt. **16**, 060502_1–060502_3 (2011).

24. T. Nagaoka, A. Nakamura, H. Okutani, Y. Kiyohara, H. Koga, T. Saida, and T. Sota, "Hyperspectroscopic screening of melanoma on acral volar skin," Ski. Res. Technol. **19**, e290–e296 (2013).

25. N. Neittaanmäki-Perttu, M. Grönroos, T. Tani, I. Pölönen, A. Ranki, O. Saksela, and E. Snellman, "Detecting field cancerization using a hyperspectral imaging system," Lasers Surg. Medicine **45**, 410–417 (2013).

26. N. Neittaanmäki, M. Salmivuori, I. Pölönen, L. Jeskanen, A. Ranki, O. Saksela, e. Snellman, and M. Grönroos, "Hyperspectral imaging in detecting dermal invasion in lentigo maligna melanoma," Br. J. Dermatol. **177**, 1742–1744 (2017).

27. N. Neittaanmäki-Perttu, M. Grönroos, L. Jeskanen, I. Pölönen, A. Ranki, O. Saksela, and E. Snellman, "Delineating margins of lentigo maligna using a hyperspectral imaging system," Acta Derm Venereol **95**, 549–52 (2015).

28. S. Tomatis, M. Carrara, A. Bono, C. Bartoli, M. Lualdi, G. Tragni, A. Colombo, and R. Marchesini, "Automated melanoma detection with a novel multispectral imaging system: results of a prospective study," Phys. Medicine Biol. **50**, 1675–1687 (2005).

29. M. Salmivuori, N. Neittaanmäki, I. Pölönen, L. Jeskanen, E. Snellman, and M. Grönroos, "Hyperspectral imaging system in the delineation of ill-defined basal cell carcinomas: a pilot study," J. Eur. Acad. Dermatol. Venereol. **0**, 0 (2018). Online Version of Record before inclusion in an issue.

30. R. Marchesini, A. Bono, and M. Carrara, "*In vivo* characterization of melanin in melanocytic lesions: spectroscopic study on 1671 pigmented skin lesions," J. Biomed. Opt. **14**, 014027 (2009).

31. J. D. Emery, J. Hunter, P. N. Hall, A. J. Watson, M. Moncrieff, and F. M. Walter, "Accuracy of siascopy for pigmented

skin lesions encountered in primary care: development and validation of a new diagnostic algorithm," BMC Dermatol. **10**, Article number 9, 9 pages (2010).

32. A. S. Farberg, R. R. Winkelmann, N. Tucker, R. White, and D. S. Rigel, "The impact of quantitative data provided by a multi-spectral digital skin lesion analysis device on dermatologists' decisions to biopsy pigmented lesions," J. Clin. Aesthetic Dermatol. **10**, 24–26 (2017).

33. Q. Wang, Q. Li, M. Zhou, L. Sun, S. Qiu, and Y. Wang, "Melanoma and melanocyte identification from hyperspectral pathology images using object-based multiscale analysis," Appl. Spectrosc. **72**, 1538–1547 (2018).

34. E. Seaman and R. A. Powell, "An Evaluation of the Accuracy of Kernel Density Estimators for Home Range Analysis," Ecology **77**, 2075–2085 (1996).

35. A. Elgammal, R. Duraiswami, D. Harwood, and L. S. Davis, "Background and Foreground Modeling Using Nonparametric Kernel Density Estimation for Visual Surveillance," Proc. IEEE **90**, 1151–1163 (2002).

36. A. Villa, J. Benediktsson, J. Chanussot, and C. Jutten, "Hyperspectral Image Classification with Independent Component Discriminant Analysis," IEEE Transactions on Geoscience and Remote Sensing **49**, 4865–4876 (2011).

37. Y. Cui, J. Yang, Y. Yamaguchi, G. Singh, S.-E. Park, and H. Kobayashi, "On Semiparametric Clutter Estimation for Ship Detection in Synthetic Aperture Radar Images," IEEE Transactions on Geoscience and Remote Sensing **51**, 3170–3180 (2013).

38. I. Diebele, I. Kuzmina, J. Kapostinsh, A. Derjabo, and J. Spigulis, "Melanoma-nevus differentiation by multispectral imaging," in *Clinical and Biomedical Spectroscopy and Imaging II,* vol. 8087 (SPIE-OSA, 2011).

39. I. Kuzmina, I. Diebele, L. Valeine, D. Jakovels, A. Kempele, J. Kapostinsh, and J. Spigulis, "Multi-spectral imaging analysis of pigmented and vascular skin lesions: results of a clinical trial," in *Photonic Therapeutics and Diagnostics VII,* vol. 7883 (2011).

40. A. M. Witkowski, J. Ludzik, F. Arginelli, S. Bassoli, E. Benati, A. Casari, N. D. Carvalho, B. D. Pace, F. Farnetani, A. Losi, M. Manfredini, C. Reggiani, J. Malvehy, and G. Pellacani, "Improving diagnostic sensitivity of combined dermoscopy and reflectance confocal microscopy imaging through doble reader concordance evaluation in telemedicine settings: A retrospective study of 1000 equivocal cases," PLoS One **12**, 1–14 (2017).

41. I. Diebele, A. Bekina, A. Derjabo, J. Kapostinsh, I. Kuzmina, and J. Spigulis, "Analysis of skin basalioma and melanoma by multispectral imaging," in *Proceedings of SPIE,* vol. 8427 (2012).

42. I. N. Bankman, *Handbook of Medical Imaging - Processing and Analysis* (Academic Press, 2000).

43. T. Skauli, I. Kåsen, V. Roy, and E. Puckrin, "Experimental study of spectral signature variability in hyperspectral remote sensing," in *Imaging and Applied Optics 2016,* (Optical Society of America, 2016). paper IW1E.4.

44. R. J. Barnes, M. S. Dhanoa, and S. J. Lister, "Standard normal variate transformation and de-trending of near-infrared diffuse reflectance spectra," Appl. Spectrosc. **43**, 772–777 (1989).

45. J. W. Demmel, *Applied numerical linear algebra* (Society for Industrial and Applied Mathematics (SIAM), Philadelphia, US, 1997).

46. A. Levy and M. Lindenbaum, "Sequential Karhunen-Loeve basis extraction and its application to images," IEEE Transactions on Image Process. **9**, 1371–1374 (2000).

47. B. W. Silverman, *Density Estimation for Statistics and Data Analysis* (Chapman & Hall / CRC, 1986).

48. J. S. Vitter, "Random sampling with a reservoir," ACM Transactions on Math. Softw. (TOMS) **11**, 37–57 (1985).

49. P. Z. P. Jr., *Probability, Random Variables and Random Signal Principles* (McGraw-Hill, 2001), 4th ed.

50. C. Goutte, P. Toft, E. Rostrup, F. A. Nielsen, and L. K. Hansen, "On clustering fmri time series," NeuroImage **9**, 298–310 (1999).

51. J. Schanda, *Colorimetry: Understanding the CIE System* (John Wiley & Sons (Wiley Interscience), 2007).

52. F. Vasefi, N. MacKinnon, N. Booth, A. J. Durkin, and D. Farkas, "Melanoma surveillance by multimode, hyperspectral dermoscopy and self-imaging using smartphone in high-risk patients," J. Am. Acad. Dermatol. **76**, AB168 (2017).