Hybrid feature selection and SVM-based classification for mouse skin precancerous stages diagnosis from bimodal spectroscopy

F. Abdat,[1] **M. Amouroux,**[1] **Y. Guermeur,**[2] **and W. Blondel**[1,*]

[1]*Centre de Recherche en Automatique de Nancy (CRAN), UMR 7039, Nancy University-CNRS, 2 avenue de la forêt de haye, 54500 Vandœuvre-lès-Nancy, France*
[2]*Laboratoire Lorrain de Recherche en Informatique et ses Applications (LORIA), UMR 7503, Nancy University-CNRS, Campus Scientifique, BP239, 54506 Vandœuvre-lès-Nancy Cedex, France*

[*]*Walter.Blondel@ensem.inpl-nancy.fr*

Abstract: This paper deals with multi-class classification of skin pre-cancerous stages based on bimodal spectroscopic features combining spatially resolved AutoFluorescence (AF) and Diffuse Reflectance (DR) measurements. A new hybrid method to extract and select features is presented. It is based on Discrete Cosine Transform (DCT) applied to AF spectra and on Mutual Information (MI) applied to DR spectra. The classification is performed by means of a multi-class SVM: the M-SVM^2. Its performance is compared with the one of the One-Versus-All (OVA) decomposition method involving bi-class SVMs as base classifiers. The results of this study show that bimodality and the choice of an adequate spatial resolution allow for a significant increase in diagnostic accuracy. This accuracy can get as high as 81.7% when combining different distances in the case of bimodality.

OCIS codes: (170.6510) Spectroscopy, tissue diagnostics; (070.5010) Pattern recognition; (070.2025) Discrete optical signal processing; (070.4790) Spectrum analysis.

References

1. C. Zhu, T. M. Breslin, J. Harter, and N. Ramanujam, "Model based and empirical spectral analysis for the diagnosis of breast cancer," Opt. Express. **16**, 14 961–14 978 (2008).
2. M. F. Mitchell, S. B. Cantor, N. Ramanujam, G. Tortolero-Luna, and R. Richards-Kortum, "Fluorescence spectroscopy for diagnosis of squamous intraepithelial lesions of the cervix," Obstet. Gynecol. **93**, 462–470 (1999).
3. N. Ramanujam, M. F. Mitchell, A. Mahadevan, S. Thomsen, A. Malpica, T. Wright, N. Atkinson, and R. Richards-Kortum, "Spectroscopic diagnosis of cervical intraepithelial neoplasia (cin) *in vivo* using laser-induced fluorescence spectra at multiple excitation wavelengths," Lasers Surg. Med. **19**, 63–74 (1996).
4. W. C. Lin, S. A. Toms, M. Johnson, E. D. Jansen, and A. Mahadevan-Jansen, "*In vivo* brain tumor demarcation using optical spectroscopy," Photochem Photobiol. **73**, 396–402 (2001).
5. W. C. Lin, S. A. Toms, M. Motamedi, E. D. Jansen, and A. Mahadevan-Jansen, "Brain tumor demarcation using optical spectroscopy; an in vitro study," J. Biomed. Opt. **5**, 214–220 (2000).
6. R. Gillies, G. Zonios, R. R. Anderson, and N. Kollias, "Fluorescence excitation spectroscopy provides information about human skin *in vivo*," J. Invest. Dermatol. **115**, 704–707 (2000).
7. A. M. Pena, M. Strupler, and T. Boulesteix, "Spectroscopic analysis of keratin endogenous signal for skin multiphoton microscopy," Opt. Express. **13**, 6268–6274 (2005).

8. E. Pery, W. Blondel, J. Didelon, A. Leroux, and F. Guillemin, "Simultaneous characterization of optical and rheological properties of carotid arteries via bimodal spectroscopy: Experimental and simulation results," IEEE Trans. Biomed. Eng. **56**, 1267–1276 (2009).

9. E. Widjaja, W. Zheng, and Z. Huang, "Classification of colonic tissues using near-infrared Raman spectroscopy and support vector machines," Int. J. Oncol. **32**, 653–662 (2008).

10. P. K. Gupta, S. K. Majumder, and A. Uppal, "Breast cancer diagnosis using N2 laser excited autofluorescence spectroscopy," Lasers Surg. Med. **21**, 417–422 (1997).

11. S. K. Majumder, P. K. Gupta, B. Jain, and A. Uppal, "UV excited autofluorescence spectroscopy of human breast tissues for discriminating cancerous tissue from benign tumor and normal tissue," Lasers in the Life Sciences **8**, 249–264 (1999).

12. A. Molckovsky, K. Wong, M. Shim, N. Marcon, and B. Wilson, "Diagnostic potential of nearinfrared Raman spectroscopy in the colon: differentiating adenomatous from hyperplastic polyps," Gastrointest. Endosc. **57**, 396–402 (2003).

13. J. Backhausa, R. Muellera, N. Formanskia, N. Szlamaa, H. G. Meerpohlb, M. Eidtb, and P. Bugertc, "Diagnosis of breast cancer with infrared spectroscopy from serum samples," Vibrat. Spect. **52**, 173–177 (2010).

14. N. Cristianini and J. Shawe-Taylor, *An Introduction to Support Vector Machines and other kernelbased learning methods* (Cambridge University Press, Cambridge, 2000).

15. W. Lin, X. Yuan, P. Yuen, W. I. Wei, J. Sham, P. C. Shi, and J. Qu, "Classification of *in vivo* autofluorescence spectra using support vector machines," J. Biomed. Opt. **9**, 180–186 (2004).

16. G. M. Palmer, C. Zhu, T. M. Breslin, F. Xu, K. W. Gilchrist, and N. Ramanujam, "Comparison of multiexcitation fluorescence and diffuse reflectance spectroscopy for the diagnosis of breast cancer," IEEE Trans. Biomed. Eng. **50**, 1233–1242 (2003).

17. S. Majumder, N. Ghosh, and P. Gupta, "Support vector machine for optical diagnosis of cancer," J. Biomed. Opt. **10**, 024034 (2005).

18. G. S. Nayak, S. Kamath, K. M. Pai, A. Sarkar, S. Ray, J. Kurien, L. D'Almeida, B. R. Krishnanand, C. Santhosh, V. B. Kartha, and K. K. Mahato, "Principal component analysis and artificial neural network analysis of oral tissue fluorescence spectra : classification of normal premalignant and malignant pathological conditions," Biopolymers **82**, 152–166 (2006).

19. M. Amouroux, G. Diaz-Ayil, W. Blondel, G. Bourg-Heckly, A. Leroux, and F. Guillemin, "Classification of ultra-violet irradiated mouse skin histological stages by bimodal spectroscopy (multiple excitation autofluorescence and diffuse reflectance)," J. Biomed. Opt. **14**, 14 011–14 024 (2009).

20. G. Diaz-Ayil, M. Amouroux, W. Blondel, G. Bourg-Heckly, A. Leroux, F. Guillemin, and Y. Granjon, "Bimodal spectroscopic evaluation of ultra violet-irradiated mouse skin inflammatory and precancerous stages: instrumentation, spectral feature extraction/selection and classification (k-NN, LDA and SVM)," Europ. Physic. J. App. Physic. **47**, 12 707–12 718 (2009).

21. A. M. Sarhan, "Iris recognition using discrete cosine transform and artificial neural networks," J. Comp. Scienc. **5**, 369–373 (2009).

22. W. B. Pennebaker and l. J. L. Mitchel, *Jpeg still image data compression standard* (Van Nostrand Reinhold, New York, NY, 1993).

23. G. Potamianos, H. P. Graf, and E. Cosatto, "An image transform approach for hmm based automatic lipreading," IEEE Int. Conf. Image. Process. (1998).

24. H. Chang and N. S. Kim, "Speech enhancement using warped discrete cosine transform," Speech Coding. 175–177 (2002).

25. A. Jain, *Fundamentals of Digital Image Processing* (Englewood Cliffs, NJ: Prentice-Hall 1989).

26. T. M. Cover and J. Thomas, *Elements of information theory* (Wiley Series in Telecommunications, New York, 1991).

27. R. Miranda-Luna, C. Daul, W. C. P. M. Blondel, Y. Hernandez-Mier, D. Wolf, and F. Guillemin, "Mosaicing of bladder endoscopic image sequences: Distortion calibration and registration algorithm," IEEE Trans. Biomed. Eng. **55**, 541–553 (2008).

28. H. Peng, F. Long, and C. Ding, "Feature selection based on mutual information: Criteria of maxdependency, max-relevance, and min-redundancy," Pat. Anal. Mach. Intel. **27**, 1226–1238 (2005).

29. M. Inaguma and K. Hashimoto, "Porphyrin-like fluorescence in oral cancer : *In vivo* fluorescence spectral characterization of lesions by use of a near-ultraviolet excited autofluorescence diagnosis system and separation of fluorescent extracts by capillary electrophoresis," Cancer. **86**, 2201–2211 (1999).

30. M. Anthony and P. Bartlett, *Neural Network Learning: Theoretical Foundations* (Cambridge University Press, Cambridge, 1999).

31. L. Devroye, L. Györfi, and G. Lugosi, *A Probabilistic Theory of Pattern Recognition* (Springer- Verlag, New York 1996).

32. F. Liang, "An effective bayesian neural network classifier with a comparison study to support vector machine," Neural Comput. **15**, 1959–1989 (2003).

33. B. Schölkopf, C. Burges, and V. Vapnik, "Extracting support data for a given task," Int. Conf. Knowledge Discov. Data. Mining. 252–257 (1995).

34. V. Vapnik, *The nature of statistical learning theory* (Springer-Verlag, New York, 1995).
35. C. Cortes and V. Vapnik, "Support-vector networks," Mach. Lear. **20**, 273–297 (1995).
36. J. Weston and C. Watkins, *Multi-class support vector machines*, (Technical Report CSD-TR- 98-04, Royal Holloway, University of London, Department of Computer Science, 1998).
37. Y. Guermeur and E. Monfrini., "A quadratic loss multi-class svm for which a radius-margin bound applies," Informatica **22**, 73-96 (2011).
38. F. Lauer and Y. Guermeur, "MSVMpack: a multi-class support vector machine package", J. Mach. Lear. Res. **12**, 2293-2296 (2011).
39. Y. Guermeur, "A generic model of multi-class support vector machine," Int. J. Int. Inf. Datab. Sys. (2011), (accepted).
40. M. Mendez, A. Bianchi, M. Matteucci, S. Cerutti, and T. Penzel, "Sleep apnea screening by autoregressive models from a single ecg lead," IEEE Trans. Biomed. Eng. **56**, 2838–2850 (2009).
41. M. D. Keller, *Optical spectroscopy for the evaluation of surgical margin status following breast cancer resection*, (Ph.D. dissertation, Nashville, Tennessee, 2009).
42. H. J. van Staveren, R. L. P. van Veen, O. C. Speelman, M. J. H. Witjes, W. M. Star, and J. L. N. Roodenburgb, "Classification of clinical autofluorescence spectra of oral leukoplakia using an artificial neural network: a pilot study," Oral Oncol. **36**, 286–293 (2000).
43. J. Dhingra, D. Perrault, K. McMillan, E. Rebeiz, S. Kabani, R. Manoharan, I. Itzkan, M. Feld, and S. M. Shapshay, "Early diagnosis of upper aerodigestive tract cancer by autofluorescence," Archives of Otolaryngology-Head and Neck Surgery. **122**, 1181–6 (1996).
44. M. Harries, S. Lam, C. Macaulay, J. Qu, and B. Palcic, "Diagnostic imaging of the larynx: autofluorescence of laryngeal tumours using the helium-cadmium laser," J. Laryngol. Otol. **109**, 108–110 (1995).
45. G. Wagnieres, W. Star, and B. Wilson, "*In vivo* fluorescence spectroscopy and imaging for oncological applications," Phot. Chem. Photobiol. **68**, 603–32 (1998).

1. Introduction

Optical spectroscopies such as autofluorescence and diffuse reflectance spectroscopies have been applied to cancer detection *in vivo* in many areas of the body, including breast [1], cervix [2, 3], brain [4, 5] and skin [6, 7]. The principle consists in exposing the tissue to an incident light flux which is then absorbed, elastically scattered or gives rise to intrinsic fluorescence. The spectral characteristics of the reflected out-coming light intensity measured can be related to the tissue state [8]. Fundamental mechanisms between optical radiation and biological specimens are absorption, reflection, elastic or inelastic scattering and luminescence. The latter is subdivided into fluorescence and phosphorescence. The current study focuses on fluorescence that corresponds to an allowed optical transition with rather high quantum yield and short (nanosecond) lifetime. Fluorescence arises upon light absorption and is related to an electronic transition from the excited state to the ground state of a molecule. Fluorescence of endogenous fluorophores is usually called autofluorescence. Biological tissues' autofluorescence depends on several parameters that are different whether the tissue under study is benign or malignant: concentration and spatial distribution of endogenous fluorophores, cell metabolism, vascularization (tumor's angiogenesis), etc. Elastic-scattering spectroscopy (ESS) is sensitive to sub-cellular architectural changes, such as nuclear grade or nuclear to cytoplasm ratio, that correlate with features used in histological assessment of cancer. Scattering intensity and spectral distribution of the signals detected could give information about the scatterers' size and distribution (cells, nuclei, etc.). Most of the tissue pathologies, including tumors, exhibit significant architecture changes on a cellular and sub-cellular level. Diffuse reflectance is the name given to the signal defined as the ratio of elastic scattering signal intensity and a standard spectrum acquired on a Lambertian surface. Diffuse reflectance signal is a superposition of diffuse scattering and absorption from tissue pigments, the resultant spectrum on the tissue surface reveals also information about the main absorbers in the biological tissues, such as hemoglobin and melanin (skin). To benefit fully from reflectance spectroscopy's advantages, one needs to relate the spectral features with the morphology and biochemical composition of the tissue investigated.

These noninvasive techniques show great promise for diagnosing early stages of cancer development *in vivo* and are particularly well adapted to everyday clinical practice. Pre-cancerous and cancerous evolutions of biological tissues are characterized by morphological and functional modifications of the tissue constituents (cells and extracellular matrix). AutoFluorescence Spectroscopy (AFS) and Diffuse Reflectance Spectroscopy (DRS) are complementary techniques that can be combined in a multimodal spectroscopy system. They are sensitive to such modifications but provide intensity spectra from which diagnostic information cannot be extracted in a straightforward way. Furthermore, some spectroscopic signals (AF) are inherently weak, and spectral curve shape differences related to the various tissue states are usually subtle with noticeable spectral overlapping and intensity variations [9].

Another difficulty to deal with is the high dimensionality of the raw spectral data. Typically for nanometric spectral resolution, one intensity spectrum corresponds to a vector of several hundreds of components. This problem is further amplified in the case of multiple excitation AFS, multimodal spectroscopy and spatially resolved spectroscopy for which one tissue point can be characterized by several vectors of hundreds of components.

Therefore, powerful and robust spectral data processing algorithms are much needed to identify the most relevant features to characterize the tissue histopathology. Multivariate statistical techniques such as Principal Component Analysis (PCA) [10, 11], Linear Discriminant Analysis (LDA) [12] and Artificial Neural Networks (ANNs) [13] have been previously used in efficient algorithms for cancerous tissue classification based on optical spectroscopy data. Other powerful classifiers are Support Vector Machines (SVMs) [14], which have now emerged as an efficient approach to the classification of spectroscopic data for tissue diagnosis. For instance, Lin et al. [15] used an SVM to differentiate *in vivo* autofluorescence spectra of NasoPharyngeal Carcinoma (NPC) from normal tissue with a sensitivity of 95% and a specificity of 99%, higher than that of PCA-LDA. Palmer et al. [16] used a linear SVM for classifying autofluorescence and diffuse reflectance spectra of breast normal and cancerous tissues *in vitro*. Majumder et al. [17] also applied linear and non-linear SVMs to fluorescence spectra for distinguishing between malignant and normal tissues in the oral cavity. The best sensitivity and specificity values obtained by the SVM-RFE (Recursive Feature Elimination) method on the data sets investigated in the framework of a leave-one-out cross-validation procedure are respectively 93% and 97%. When tested on the spectral data of the uninvolved oral cavity sites from the patients it yielded a specificity of 85%. Nayak et al. [18] developed algorithms based on PCA and ANN for discriminating normal, premalignant and malignant human oral tissues. They achieved sensibility $Se = 100\%$ and specificity $Sp = 92.9\%$ using PCA and $Se = 100\%$ and $Sp = 96.5\%$ with ANN.

Most of the studies to date on the application of optical spectroscopy to skin cancer diagnosis focused firstly, on a local analysis i.e., the analysis involves only some specific wavelength bands of the spectrum using *a priori* knowledge of curve-shape features that allow the distinction between the histological classes of interest, and secondly, on bi-class classification. Multi-class SVM implementations for spectroscopic diagnosis of cancerous tissue are still very limited.

In a first published study [19], our team showed that combining autofluorescence and diffuse reflectance in a bimodal approach increases diagnosis accuracy of pairwise discrimination between four histological classes of mouse skin carcinogenesis when compared to each modality (autofluorescence or diffuse reflectance) used alone. Such a result highlights how informative the different types of light interaction with skin are, i.e., autofluorescence and diffuse reflectance provide the physician with complementary types of information when it comes to cancer diagnosis. Indeed, cancer progression implies functional as well as morphological modifications of skin that autofluorescence and diffuse reflectance can respectively detect. Therefore, fusing the two types of optical information gets a wider insight into the skin modi-

fications. In a second published study [20], our team showed how spatial resolution (the use of several distances between excitation and collection optical fibers) increases diagnosis accuracy as well as the fact that SVM is the most appropriate classification algorithm to our problem at hand (still in a pairwise approach with *a priori* knowledge on the data) acquired through a bimodal approach. In order to get closer to medical interest, we propose in the current study to process data obtained through a spatially-resolved bimodal approach first, with no *a priori* knowledge of the data set (no extraction of spectral features based on spectra visualization) and second, to base our SVM-classification on a multi-class approach (instead of a decomposition scheme). Discrete Cosine Transform (DCT) and Mutual Information (MI) are proposed for spectral feature extraction and selection to produce a data description of lower dimensionality. Such process is more general than the one used previously [19, 20], while allowing an efficient representation of the raw data. The performance of a multi-class SVM, the M-SVM^2, is evaluated for the discrimination between the 4 histological classes of interest.

2. Optical spectroscopy set up

A general description of the instrumentation set up can be found elsewhere [20]. Here some other optical characteristics are provided in order to complete such previous description of the set up and get more insights into optics. Light source is a short-arc Xenon lamp with parabolic mirror providing a 25.4 *mm* - beam diameter (Eurosep, France). Its illumination bandwidth is extended in the UV spectral range and an anti-caloric filter is used in order to have main emission in the 300-800 nm spectral range. Two linearly variable band-pass filters are used to tune the excitation wavelength (autofluorescence) or the desired bandwidth (diffuse reflectance). Finally light is focused into the excitation optical fiber core (200 *μm*-diameter). Spatial resolution is achieved through a multiple optical fibers bundle. The bundle's distal tip was put in gentle contact with mouse skin. The probe contains 37 optical fibers (numerical aperture is 0.22, SEDI, France) arranged in concentric circles within the 2 *mm*-diameter bundle. Three collecting optical fibers were chosen at D1 = 271 *μm*, D2 = 536 *μm* and D3 = 834 *μm*. At the entrance of the multichannel spectrograph (iHR 320, Horiba Jobin Yvon, France) an adaptation bundle is set in order to match each collecting optical fiber of the bundle onto defined rows of photosites on the CCD (Symphony 2048x512 Cryogenic back illuminated UV-sensitive CCD detector; back illumination allows higher sensitivity in the UV spectral range). This allowed us to simultaneously measure up to 6 intensity spectra from 6 collecting optical fibers. Light intensity collected by each optical fiber is diffracted on CCD photosites through diffraction gratings (150 lines/mm blazed at 600 *nm*): sufficient spacing between photosites corresponding to two adjacent collecting optical fibers prevents intensity crosslinks. Fiber tracks are electronically binned on the CCD to detect lower light levels. The overall optical resolution of the acquisition system is 5 *nm*.

3. Data processing

In order to use AFS and DRS as non invasive tools to diagnose skin pre-cancerous and cancerous tissues, dedicated data analysis algorithms need to be developed including raw spectra pre-processing, identification and selection of discriminant spectral features, and classification.

The raw data set consists of 6048 optical spectra measured *in vivo* on $N_s = 252$ spots of mouse skin either sham-irradiated (control group) or UV-irradiated weekly for up to 20 weeks. Briefly (see [19] for details), it consists of 7 AF spectra (7 excitation wavelengths that were chosen because they are long enough to be harmless i.e., they do not carry enough energy to induce mutation in skin cells' DNA: 360, 368, 390, 400, 410,420 and 430 *nm*) and 1 DR spectrum (UV-visible spectral range) at 3 different distances between collecting and exciting optical fibers (Collection to Excitation Fiber Separation CEFS). Four reference classes were
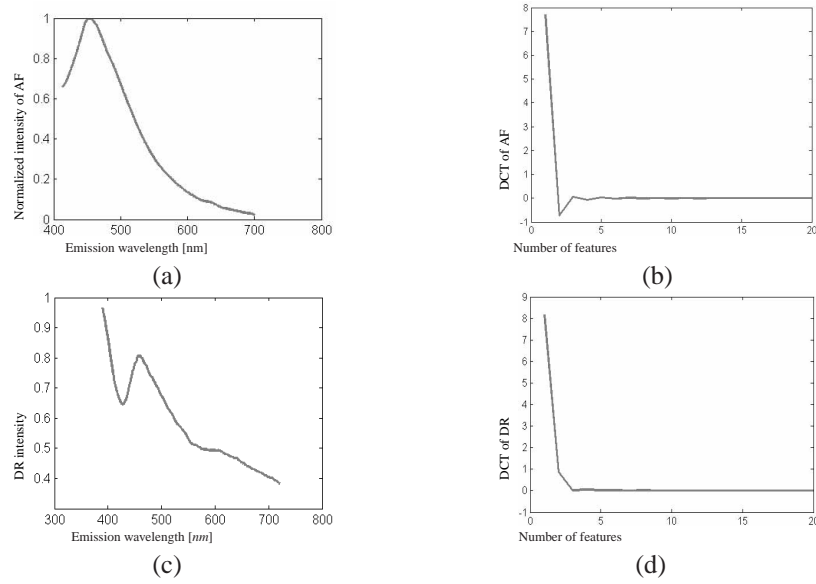
Fig. 1. Examples of emission autofluorescence spectrum for an excitation wavelength of 360 *nm* (a) and diffuse reflectance spectrum (c), both acquired at interfiber distance D1. (b) and (d) show the 20 first DCT coefficients calculated for (a) and (c) respectively.

defined based on the histopathological analysis of the skin samples: Healthy (H), Compensatory Hyperplasia (CH), Atypical Hyperplasia (AH), and Dysplasia (D). After *in vivo* spectral data acquisition, the 252 skin spots were biopsied and classified by an anatomopathologist physician into the 4 histological classes: 84 were considered as H, 47 were classified into CH class, 64 into AH class, and 57 into D class [19].

After acquisition of AF and DR spectra, a method of normalization was applied to remove the absolute intensity information from the spectra that might be affected by many unavoidable experimental factors. In the case of autofluorescence, the sample associated with each site was normalized to its maximum intensity peak from that site. Diffuse Reflectance spectra were obtained by dividing elastic scattering raw spectra by standard spectra acquired at the shortest CEFS (271 μ*m*) on a Lambertian surface (WS-1 Diffuse Reflection Standard, Ocean Optics), uniformly reflective by wavelength across the entire range 250-1500 *nm* (>98%).

3.1. Spectral feature extraction using DCT

The strong capability of DCT to compress energy as well as the availability of a fast implementation of the transform make it a good candidate for pattern recognition applications [21], image compression [22] as well as in feature extraction [23]. The DCT has been used in many practical applications involving 1D data such as speech processing [24]. For k in $\{1,\ldots,N\}$, the *kth* DCT coefficient $x_c(k)$ of a spectrum $(x(n))_{1\leqslant n\leqslant N}$ is defined as [25]:

$$x_c(k) = w(k) \sum_{n=1}^{N} x(n) \cos \left[\frac{\pi(2n-1)(k-1)}{2N} \right] \tag{1}$$

$$\text{where } w(k) = \begin{cases} \frac{1}{\sqrt{N}} & if \quad k = 1 \\ \sqrt{\frac{2}{N}} & 2 \leqslant k \leqslant N \quad otherwise \end{cases}.$$

Looking at examples of AF spectrum (Fig. 1(a)) and DR spectrum (Fig. 1(c)) for which the 20 first DCT coefficients are given respectively in Figs. 1(b) and 1(d), it can be noticed that most of the visually significant information is concentrated in just a few first DCT coefficients. Therefore, the low-frequency DCT coefficients are selected as features. The choice of an appropriate number of DCT coefficients is further studied in Section 3 (Results).

3.2. Feature selection using MI

Let X and Y be two discrete random variables taking their values in $\mathscr{X}$ and $\mathscr{Y}$ respectively. Their mutual information can be defined in terms of their probability density functions (pdfs) $p_1(x)$, $p_2(y)$, and joint pdf $p(x,y)$ as [28]:

$$MI(X,Y) = \sum_{x \in \mathscr{X}} \sum_{y \in \mathscr{Y}} p(x,y) \log \frac{p(x,y)}{p_1(x)p_2(y)}. \tag{2}$$

Based on entropy, the mutual information between X and Y can also be expressed using the conditional probability $p(x|y)$. The entropy H of X is a measure of its randomness and is defined as $H(X) = -\sum_{x \in \mathscr{X}} p_1(x) \log p_1(x)$.

The conditional entropy of Y given X is given by :

$$H(X|Y) = -\sum_{y \in \mathscr{Y}} p_2(y) \sum_{x \in \mathscr{X}} p(x|y) \log p(x|y). \tag{3}$$

The mutual information between X and Y can be computed from the entropy terms defined above by

$$MI(X,Y) = H(X) - H(X|Y). \tag{4}$$

MI is considered to be a good quantitative indicator of the dependence of two random variables [26], well adapted in the case of non linear models. Formally, the mutual information between the nth feature $x(n)$, $n \varepsilon \{1,...,N_d\}$, and the class label $Y \in \{1,...,Q\}$ (Q is the number of classes), $MI(x(n),Y)$ represents the amount of information gained about the class if the feature $x(n)$ is used. A high value of MI indicates that the feature is potentially relevant for the classification (next step) [27] and thus should be selected.

To select a subset of features from an initial data set we have:

1. Maximized the pertinence by calculating the MI between the different pairs feature/Y:
   $argmax\, V, \qquad V = \frac{1}{|N_d|^2} \sum_{i=1}^{N_d} MI(Y,x(i))$

2. Minimized the redundancy by calculating the MI between pairs of features:
   $argmin\, W, \qquad W = \frac{1}{|N_d|^2} \sum_{j=1}^{N_d} \sum_{i=1}^{N_d} MI(x(i),x(j))$

3. Combined between the minimization of the redundancy and the maximization of the pertinence by maximizing: $argmax\,(V - W)$

In practice, the empirical values for the features are provided by the N_s measurement spots which have been assigned to the four histological classes ($y \in \{1,...,4\}$) by the anatomopathologist physician. The detail of the algorithm can be found in [28].
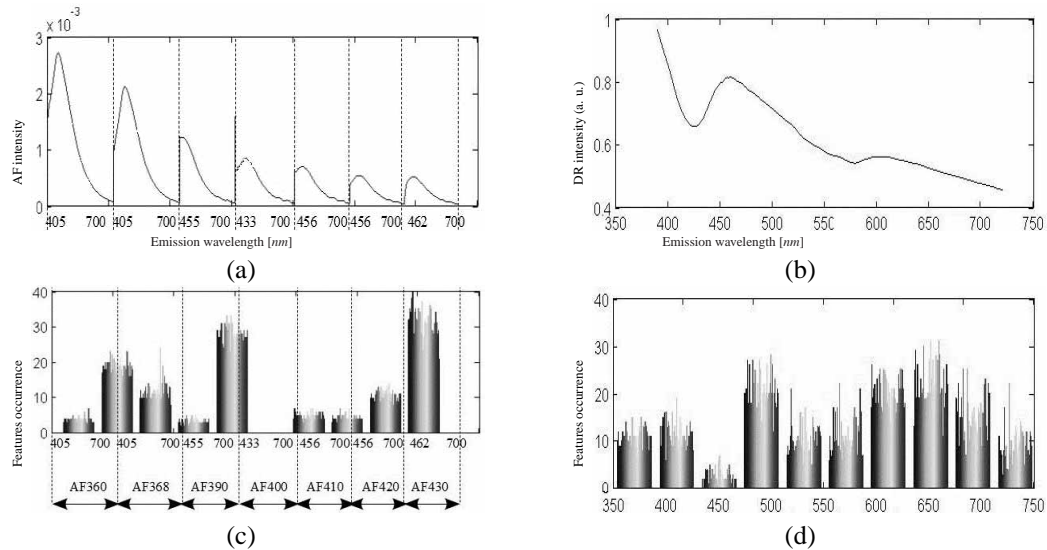
Fig. 2. Examples of 7 AF spectra concatenated end-to-end (a), and an DR spectrum (b). Occurrence histograms of the discriminant features selected with MI for the AF spectra (c) and the DR spectrum (d).

The selection of the discriminant features from AF spectra using MI is applied to a single vector consisting of 7 AF spectra concatenated end-to-end (see Fig. 2(a)). Figures 2(c) and 2(d) show histograms of the selected wavelengths for the seven AF excitations wavelengths and for DR respectively. The discriminant features for AFs spectra were found for AF430 and AF390. AF430 gives the most discriminant intensities around the wavelength ≈ 500 nm and AF390 gives discriminant features around the wavelength=600 nm. Concerning DR spectra, the discriminant features were found around 500 nm and 650 nm (see Fig. 2(b)).

Various studies mentioned that in AF mono-excitation at 410 nm on human oral cancer, haemoglobin absorption peaks ($\approx$420, 545 and 575 nm) related to progressive hyperplastic vascular activity and autofluorescence emissions related to flavins, collagen, NADH and porphyrins ($\approx$633 and 672 nm), can be used as biomarkers to differentiate between various precancerous and cancerous tissue states. More specifically, the presence of increased porphyrin intensity peaks is also associated with hyperplastic modifications implying haemoglobin transformations, or changes in the cellular environment [29]. In our study, pertinent features selected with MI are found using AF mono-excitations at 390 nm and at 430 nm. The difference between our data and those used by Inaguma et al. [29] is the kind of lesions: we are interested in precancer stages, whereas they work on benign and malignant lesions.

3.3. Support Vector Machine based classification

In order to classify automatically the 252 tissue samples in the 4 predefined classes (H, CH, AH, D) based on the spectral features previously extracted, a multi-category supervised classification was performed. The labels were provided by the "gold standard" results of histological analysis. The accuracy of the classification is characterized by means of the recognition rate τ as a function of the number of features.

A great many methods have been used for multi-class classification such as ANNs [30], k-

Nearest Neighbors [31], and SVMs [14]. The latter have rapidly gained in popularity due to their theoretical merits, computational simplicity, and excellent performance in real-world applications [32]. The first studies dealing with the use of SVMs for multi-category classification report results obtained with decomposition methods [33, 34] involving Vapnik's bi-class machine [35]. Multi-class support vector machines (M-SVMs) were only introduced three years later [36]. We have chosen to use one of them, the M-SVM^2 [37, 38]. Its main advantage rests in the fact that an algorithm is available to set automatically the value of its soft margin coefficient. Typically, the clouds of points associated with the different categories are not linearly separable in the description space, so that the use of a nonlinear SVM appears appropriate. We selected a Gaussian kernel:

$$\kappa\left(x,x'\right) = \exp\left(-\mu\left\|x-x'\right\|^2\right) \tag{5}$$

where, x and x' are two data points, and μ is the bandwidth coefficient. The experimental protocol is systematically a 3-fold cross-validation.

4. Results and discussion

4.1. Effect of the number of features on classification performance

Extracting the small set of discriminant features from the raw spectral data is the first key step in our work. If the features are carefully chosen it is expected that they will retrieve the relevant information from the input data in order to perform the desired task. This allows to prevent the curse of dimensionality, and to avoid classification errors resulting from the domination of irrelevant features [40].

4.1.1. DCT-based extraction method

The number of DCT coefficients was chosen so as to maximize the recognition rate. Experimental results are summarized in Fig. 3. For AF spectra alone, a slight increase in recognition rate can be observed when the number of coefficients increases (from 75% for the 20 first coefficients up to 78% for the first 50 coefficients). Regarding DR spectra alone, the recognition rate increases from 40% up to 64% with N_{DR}=147 coefficients; for a number of coefficients higher than N_{DR}, no significant change in recognition accuracy was observed.

For bimodality configuration, a slight decrease in recognition accuracy is observed as we go to higher numbers of coefficients: from 75% for 1 coefficient down to 74% for 10 and over. In other words, more DCT coefficients does not necessarily mean better recognition rate because high frequency components are related to unstable activity features such as noise. According to Figs. 3(a) and 3(b), the best performance of our system is obtained when N=56 for AF alone and when N=147 for DR alone. Therefore, the recognition rate was calculated with $N_{AF} = 56$ coefficients for each AF spectrum leading to 56x7=392 features for the 7 AF spectra corresponding to the 7 excitation wavelengths. Finally, with a total of $N_{AF} = 392$x3 features for the combination of 3 CEFS D1, D2, and D3, we obtained a maximum rate of $\simeq$78%, which is close to that found with D1 alone. This result can be explained by the fact that the number of features is far higher than the number of examples (252). This implies that the curse of dimensionality is a problem that impacts our classifier. To overcome this problem, we worked with $N_{AF} = 2$ per AF excitation (i.e., 2x7=14 features in total). For DR data spectra, we kept working with the number of features that gave the best rate $N_{DR} = 147$ (see Fig. 3(b)).
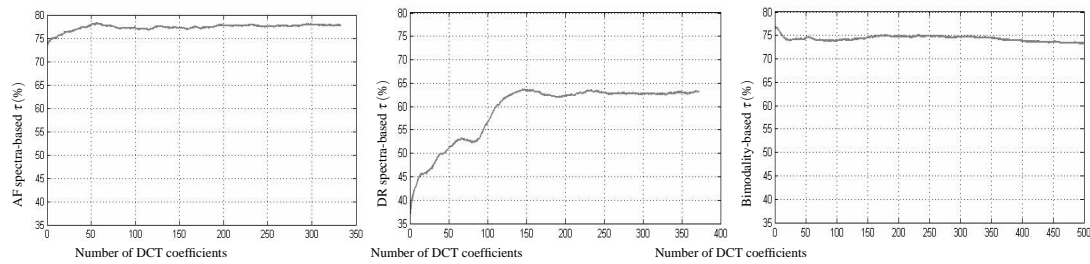
Fig. 3. Recognition rate τ (%) calculated as a function of the number of DCT coefficients retained for (left) AF spectra alone acquired at D1, (center) DR spectra acquired at D1, (right) AF and DR spectra all acquired at D1 (bimodality).
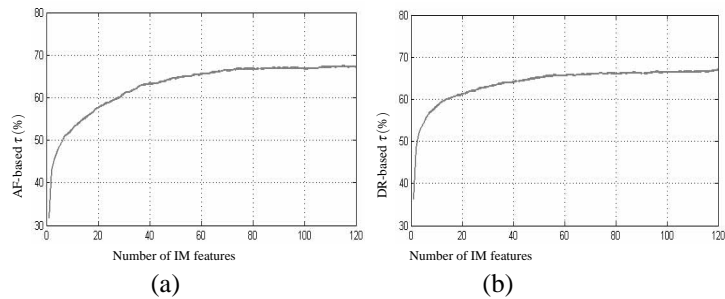

Fig. 4. Recognition rate τ (%) as a function of the number of features selected with MI for (a) AF spectra and (b) for DR spectra acquired at D1.
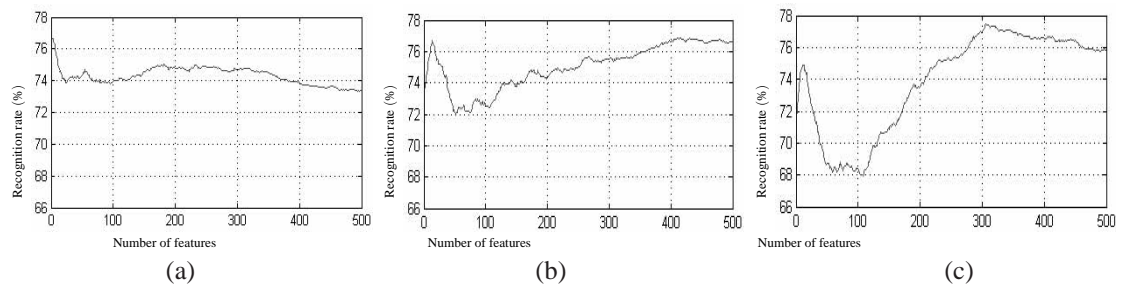

Fig. 5. Recognition rate as a function of the number of DCT coefficients for the bimodal configuration (AF+DR) for the 3 CEFS (a) D1 (b) D2 (c) D3.

4.1.2. MI-based selection method

As in the previous case, the number of features was selected so as to maximize the recognition rate. Experimental results are summarized in Fig. 4.

The results obtained with MI for each monomodal method are similar. Both curves show increasing values in recognition accuracy as we go to higher numbers of features ($\geqslant$ 100 features).

4.1.3. Hybrid method of extraction/selection

Overall, the results obtained (Figs. 3 and 4) highlight the fact that on one hand, DCT-based extraction/selection gives better results for AF spectra than for DR spectra for a smaller number of features, and on the other hand, MI-based extraction/selection outperforms DCT-based one for DR spectra.

The MI-based feature selection method performs best when applied to DR spectra. The performance of DCT-based algorithm is even more valuable when considering the very low number of discriminant features used for classification. By choosing the best performing method for each modality, we propose a hybrid approach combining DCT applied to AF spectra and MI to DR ones.

4.2. Influence of the different CEFS (spatial resolution) and their combinations on the classification performance

Varying the separation distance between source and detection fibers (CEFS) allows for tissue depth selectivity in the spectroscopic measurements. Larger separations are more likely to detect photons that traveled deeper inside the tissue and were multiply scattered to a greater radial distance compared with photons that only underwent minimal scattering events and stayed in the superficial layers [41].

Considering several CEFS does not improve the classification results based on AF spectra (see Fig. 5). On the contrary, it generates an erratic behavior.

The results obtained from the different sets of characteristics for each mono- and bimodal excitations with DCT method, MI method, and hybrid method respectively are given in bargraph Figs. 6(a), 6(b) and 6(c). Each bar represents the recognition rate obtained with the multi-class SVM (M-SVM^2) for each of the 7 possible combinations of CEFS tested: D1 alone, D2 alone, D3 alone, and D1D2, D1D3, D2D3, D1D2D3 with the aforementioned sets of features of respective cardinalities N_{AF} and N_{RD}. τ values for D1 alone are globally better than those for D2 alone or D3 alone. The diagnostic accuracy obtained with D1 is the best for each monomodal excitation AF (71.4% with DCT method) or DR (68.3% with DCT method), and for bimodality (74.6% with DCT method).

Spatial resolution always implies an increase in classification performance. For each modality, an increase in discrimination efficiency between the 4 histological classes for both modalities was obtained using an extended set of characteristics from multiple distances combinations D1D3, D1D2, and D1D2D3. The combination D2D3 gives the lowest rate for each modality compared to D1D3, D1D2, and D1D2D3. An increase in discrimination performance was obtained using an extended set of characteristics under D1D2D3 combination and bimodal configuration compared to each unimodal excitation (AF or DR).

Overall, the DCT-based extraction/selection applied to AF spectra for the autofluorescence gives better results than when applied to DR ones, while MI-based selection gives close results for each modality. The best result of τ=81.7% was obtained for bimodal configuration, with D1D3 and D1D2 combinations using DCT applied to AF spectra and MI to DR spectra.

These findings highlight the fact that the diagnosis performance of our system depends on two factors: first, the choice of the best CEFS combination for each modality, i.e., the effect
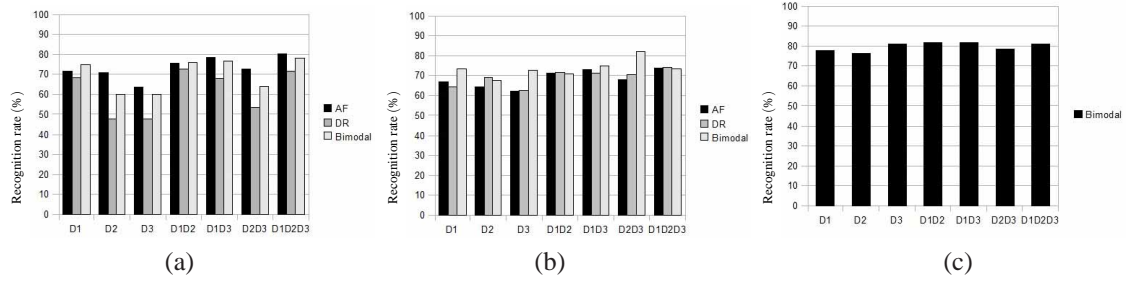
Fig. 6. Recognition rate of the M-SVM² as a function of the 7 combinations of $(D_i)_{1 \leqslant i \leqslant 3}$ distances and the 3 modalities: AF alone (black bar), DR alone (grey bar) and AF+DR together (light grey bar), using (left) DCT method, (center) MI method and (right) hybrid method.
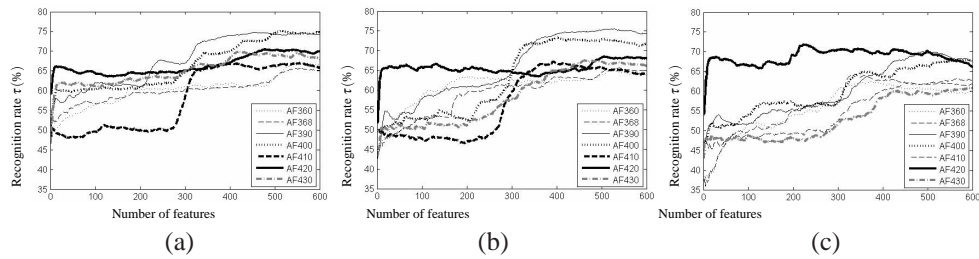

Fig. 7. Recognition rate as a function of the number of DCT coefficients calculated for each of the 7 AF excitation wavelengths (AF360-AF430) and for the 3 CEFS D1 (a), D2 (b), D3 (c).
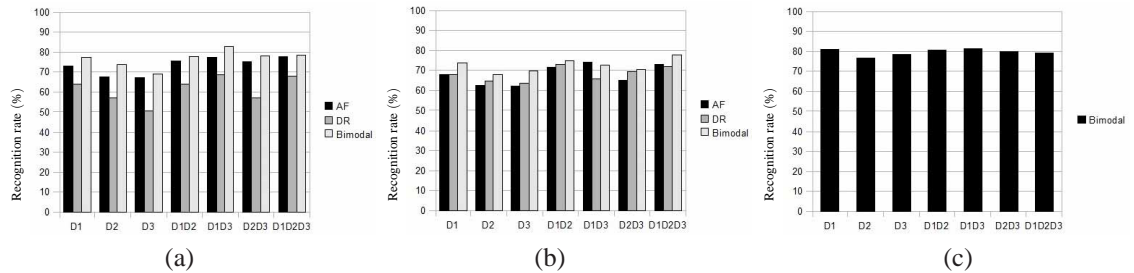
Fig. 8. Recognition rate of the OVA as a function of the 7 combinations of $(D_i)_{1 \leqslant i \leqslant 3}$ distances and the 3 modalities: AF alone (black bar), DR alone (grey bar) and AF+DR together (light grey bar), using (left) DCT method, (center) MI method and (right) hybrid method

of CEFS on the discrimination between the different classes, and second, the selection of an optimal number of coefficients for DCT-based extraction.

4.3. Excitation wavelength effect on the classifier performance

Since DCT-based extraction/selection method gave good results with the 7 AF excitations used together, i.e., preserving useful information and allowing a global analysis of the spectrum, the

recognition rate for each excitation wavelength using DCT method was calculated. Figures 7(a), 7(b) and 7(c) represent the variation of recognition rates for the 7 AF excitation wavelengths and for the 3 CEFS D1, D2 and D3 respectively. All excitation wavelengths except that of 420 *nm* gave similar results, i.e., the higher the number of coefficients, the higher the recognition rate. Distinctively, the 420 *nm* excitation wavelength has stable performance regardless of CEFS. These results confirm those obtained by Van Staveren [42] who selected this specific excitation wavelength to distinguish between different grades of oral cancerous tissues. For this study, the choice of 420 *nm* excitation wavelength was also guided by previous investigations with excitation wavelengths in the range of 400-450 *nm* [43, 44]. At this wavelength, several endogenous tissue fluorophores can be excited, viz. porphyrins, lipo-pigments and flavins, which have a fluorescence emission in the range of 450-650 *nm* [45].

According to Fig. 7, the best rates with 420 *nm* excitation (AF420) are around 65%, hence the importance of multi-excitation that allows to achieve a rate that exceeds 74%.

4.4. *Performance assessment*

Because of the specificity of our experimental data set (classes, measurement conditions), it is highly difficult to compare the present work with works described in the literature. In this subsection, we assess performance with respect to feature extraction/selection and model selection:

1/ Analysis of variance (ANOVA) is applied to explore the effects of categorical factors on one or more quantitative variable(s) by analyzing the mean and variance across different levels of the factors. The important parameter for ANOVA result is the p-value, which is the probability of accepting the hypothesis that the factor has great incidence. It can be used to compute the least significant difference between any of the means in terms of probability of error. The first factor for us is the modality used (AF, DR, Bimodal), the second factor is the method used for data processing (DCT, MI, Hybrid). The respective p-values were 0.001 and 0.02, which is less than the standard threshold value 0.05, so the proposed approach allows a significant improvement in the diagnosis accuracy.

For a deep study on the efficiency of the method used, Table 1 shows the p-value corresponding to different combinations of methods. The bimodal method reaches better performance than unimodality methods (bold figures in the table are less than the threshold of ANOVA), particularly when the extraction of the discriminant features is based on the hybrid approach (last row of the table). The hybrid approach outperforms the MI applied to AF (p-value=6.08 E-05) and the DCT applied to DR (p-value=5.99E-05). The key step of our work is to extract/select the appropriate features. ANOVA test has shown the efficient of the proposed hybrid approach.

The effect of the choice of the classifier on the performance will be characterized in the following.

Table 1. P-value for different combinations

	DCT (AF)	DCT (DR)	DCT (Bimodal)	MI (AF)	MI (DR)	MI (Bimodal)	Hybrid
DCT (AF)	1.00000	0.09980	0.02687	0.13316	0.39251	0.91850	**0.013**
DCT (DR)	0.09980	1.00000	0.14653	0.81089	0.67985	0.05475	**0.00005**
DCT (Bimodal)	**0.02687**	0.14653	1.00000	0.11651	0.12853	**0.02016**	**0.00105**
MI (AF)	0.13316	0.81089	0.11651	1.00000	0.79488	0.07504	**0.00006**
MI (DR)	0.39251	0.67985	0.12853	0.07504	1.00000	0.32763	**0.00975**
MI (Bimodal)	0.91850	0.05475	**0.02016**	0.07504	0.32763	1.00000	**0.00635**
Hybrid	**0.013**	**0.00005**	**0.00105**	**0.00006**	**0.00975**	**0.00635**	1.00000

2/ In a previous work [20], we have compared the performance of different standard classi-

fiers (kNN, LDA, SVM), and the best results were obtained with an SVM. In order to compare the performance of the chosen classifier (M-SVM^2) with another "classical" method, we have recalculated the recognition rate with the One-Versus-All decomposition method (OVA). Figure 8 shows the results obtained using OVA. According to the sample proportion test, the gain appears too low to be statistically significant with high confidence. It is utterly true that the classification accuracies of the M-SVM^2 and the OVA decomposition scheme are similar. In fact, it springs from our experimental results that once the features have been adequately selected, a high accuracy can be obtained with different methods involving multi-class or bi-class SVMs. In that context, we found it more appropriate to choose the M-SVM^2 given its main feature: it is the sole M-SVM developed so far for which an automatic model selection procedure of low computational complexity is available. In comparison, to reach the same accuracy with the OVA approach, we had to optimize globally the soft margin coefficients of the four bi-class SVMs involved. To do so, we applied a four-dimensional grid search, which proved very tedious. Such a grid search would have been intractable had the number of classes been slightly higher. In short, computational complexity is indeed the main argument in favour of the choice of the M-SVM^2.

4.5. Confusion matrices

Table 4.5 shows the classification results in the form of the confusion matrices that compare the pathologist's diagnosis (rows) with the prediction of the M-SVM^2 (columns). The best performance of the latter was in classifying H and CH tissues (100% accuracy), and errors were spread among the remaining classes D and AH.

For both DCT-AF-based and MI-DR-based methods, the highest confusion is found between AH and D. Using AF-DCT approach, 10% of AH are classified as D and 31% of D are classified as AH. DR-MI leads to similar confusion between AH and D, with extra confusion found between H and CH. Both AH and D are malignant classes; H and CH are benign; so, the confusion between AH and D and between H and CH does not alter the diagnosis, whereas the confusion between malignant and benign classes has an effect on the diagnosis accuracy. 6% of AH and 10% of D are classified as H, which corresponds to the incorrect predictions. Finally, the hybrid method proposed here overcomes these confusion problems, with only a small confusion between AH and D (1.5%).

It is also interesting to evaluate the performance of the method in a pairwise discrimination approach considering merging the classes H and CH on one hand (as the benign class), and merging the classes AH and D on the other hand (as the malignant class). These choices were motivated by the results of initial experiments involving different kernels. Table 4.5 shows confusion matrices corresponding to bi-class problem of spectra acquired at D1 and of spectra acquired at D1D2D3 together using a linear kernel. Indeed on a histological point of view, such two classes can be distinguished by the absence (respectively the presence) of atypical features within skin cells (keratinocytes). The method performs better for this pairwise discrimination (recognition rate > 94% both for D1 alone and for D1D2D3) than for the multi-class classification (H, CH, AH, D) (recognition rate around 80%). When considering D1 alone, sensitivity (*Se*) and specificity (*Sp*) of the pairwise discrimination respectively are 95% and 93%. When considering spatial resolution (D1D2D3), $Se = 93\%$ and $Sp = 98\%$. From a clinical point of view, such a result highlights that optical spectroscopy detects atypical features in skin cells which is of utmost medical value. Optical spectroscopy also detects other types of histological features that allow discrimination of two sub-classes within each of the two classes (H and CH for the benign class; AH and D for the malignant class) but with a lower recognition rate.

In order to fully test spatially resolved-bimodal spectroscopy ability to be used in a clinical environment as a decision-making support tool, we also tested our method classification per-

formance on three classes: benign class (H and CH together), atypical class (AH alone) and dysplastic class (D alone). This approach is closer to clinical interests as clinical management will be specific to each of those three histological classes. For the combination of the three distances (D1D2D3), the confusion matrix (Table 4.5) shows that 60.9% and 77.8% of AH, respectively D, samples are well classified. Such results can be compared with the ones previously published by our team [20] when discriminating between AH and D classes: $Se = 56\%$. Therefore, our new data processing method performs better than the previous one (*a priori* knowledge of spectra) even when considering more classes. Again, the benign class is very well recognized with 97.7% of correctly classified samples among the three classes confirming the results previously mentioned with two classes.

These final results confirm that the bimodal approach is of particular interest in improving the diagnosis performance of fibered optical spectroscopy tools applied to skin pre-cancerous states discrimination.

Table 2.   Confusion matrices of the M-SVM^2 calculated from DCT-based extraction/selection applied to AF spectra (top), from MI-based selection applied to DR spectrum (middle) and from hybrid method (both AF-DCT and DR-MI) (bottom) of spectra acquired at D1, D2 and D3. The 3 matrices on the left correspond to test performance given for D1 alone and the 3 matrices on the right provide test performance given for D1D2D3 together.

		DCT-SVM						DCT-SVM		
		H	CH	AH	D			H	CH	AH	D
S	H	83.7%	0.0%	10.2%	6.1%	S	H	88.4%	1.0%	5.3%	5.3%
G	CH	0.0%	100.0%	0.0%	0.0%	G	CH	0.0%	100.0%	0.0%	0.0%
H	AH	3.4%	0.0%	56.9%	39.7%	H	AH	0.0%	0.0%	68.1%	31.9%
	D	0.0%	0.0%	42.9%	57.1%		D	0.0%	2.3%	27.9%	69.8%

		MI-SVM						MI-SVM		
		H	CH	AH	D			H	CH	AH	D
S	H	71.5%	5.9%	16.7%	5.9%	S	H	83.2%	5.6%	6.7%	4.5%
G	CH	9.5%	78.6%	4.8%	7.1%	G	CH	8.9%	84.4%	0.0%	6.7%
H	AH	10.2%	8.2%	53.1%	28.5%	H	AH	8.8%	3.5%	70.2%	17.5%
	D	3.4%	6.8%	32.2%	57.6%		D	1.6%	3.3%	29.5%	65.6%

		Hybrid						Hybrid		
		H	CH	AH	D			H	CH	AH	D
S	H	89.1%	2.2%	5.4%	3.3%	S	H	93.1%	2.3%	2.3%	2.3%
G	CH	0.0%	97.8%	0.0%	2.2%	G	CH	0.0%	95.7%	0.0%	4.3%
H	AH	3.3%	0.0%	68.4%	28.3%	H	AH	3.1%	0.0%	64.6%	32.3%
	D	0.0%	1.8%	32.7%	65.5%		D	1.9%	0.0%	37.7%	60.4%

Test performance with D1   Test performance with D1D2D3

HGS: Histology "Gold Standard"

5.   Conclusion and future work

To our knowledge, the current study is the first one addressing the problem of bimodal spectroscopic multi-class classification of four pre-cancerous stages. Based on epidermis and der-

Table 3. Confusion matrices of the M-SVM[2] calculated from hybrid method (both AF-DCT and DR-MI) of spectra acquired at D1 (on the left) and of spectra acquired at D1D2D3 together (on the right), using a linear kernel.

		Hybrid	
		H and CH	AH and D
HGS	H and CH	93.0%	7.0%
	AH and D	4.9%	95.1%

		Hybrid	
		H and CH	AH and D
HGS	H and CH	97.7%	2.3%
	AH and D	7.3%	92.7%

Test performance with D1   Test performance with D1D2D3
HGS: Histology "Gold Standard"

Table 4. Confusion matrices of the M-SVM[2] calculated from hybrid method (both AF-DCT and DR-MI) of spectra acquired at D1 (on the left) and of spectra acquired at D1D2D3 together (on the right), using a linear kernel.

		Hybrid		
		H and CH	D	AH
	H and CH	97.7%	2.3%	0.0%
HGS	D	11.1%	77.8%	11.1%
	AH	21.7%	39.1%	39.1%

		Hybrid		
		H and CH	D	AH
	H and CH	97.7%	2.3%	0.0%
HGS	D	11.1%	77.8%	11.1%
	AH	8.7%	30.4%	60.9%

Test performance with D1   Test performance with D1D2D3
HGS: Histology "Gold Standard"

mis morphological characteristics, a classification of UV-irradiated mouse skin pre-cancerous stages was performed. In this paper, we have presented a new extraction and selection method to automatically classify skin pre-cancerous tissues using multiple autofluorescence excitation and diffuse reflectance spectroscopies acquired at 3 different distances between collecting and exciting optical fibers.

For feature extraction, two approaches were evaluated separately for each modality: one based on the discrete cosine transform and one based on mutual information. In order to classify automatically the examples in the 4 predefined classes, an M-SVM[2] was applied.

The diagnostic accuracy obtained with D1 is the best for each monomodal excitation AF (74.5% with DCT method) or DR (63.6% with DCT method), and for bimodality (76.7% with DCT method).

Spatial resolution always implies an increase in classification performance. The results obtained demonstrate that the best method applied to AF spectra was DCT (80.2% with D1D2D3), and the best one for DR spectra was MI (74.21% with D1D2D3). Consequently, a hybrid method has been developed, where DCT was used to produce a vector of descriptors from the autofluorescence spectroscopy and MI was used to perform feature selection from diffuse reflectance spectra.

The best result of 81.7% was obtained for a bimodal configuration, with D1D2 and D1D3 combinations using DCT applied to AF spectra and MI to DR spectra (hybrid method).

This study shows that different pre-cancer stages can be distinguished *in vivo* using a multi-class SVM (or a decomposition scheme involving bi-class SVMs) applied on features derived from autofluorescence and diffuse reflectance. These are encouraging results for clinical application since such visible wavelengths are harmless. Combining several excitation wavelengths improves diagnostic sensitivity while combining modalities (DR and multi-excitation AF) im-

proves diagnostic specificity especially when discriminating the three types of hyperplasia. Indeed, combining information from a number of points located at variable distances from the excitation optical fiber will probe the tissue at several depths resulting in a potentially better discrimination between the different histological classes.

In the future, we intend to use another strategy for data fusion based on the principle of sensor fusion (decision-level fusion). Obtaining best performance with a pre-clinical system such as ours (application on the mouse skin) allows to promote performance of the dedicated clinical system. Our previous and present works provide the foundation for the primary focus of future work, which is the optimization of the experimental protocol by minimizing the exposure of skin tissue to light radiation and by reducing the time of the measure step before moving to clinics.

Acknowledgment