

## Diagnóstico de lesiones en la piel a partir de espectros de reflexión difusa empleando algoritmos computacionales: un estudio preliminar

E. E. Orozco<sup>a, b</sup>, J. A. Delgado<sup>a, c</sup>, S. Vázquez<sup>a</sup>, J. Castro<sup>a</sup>, A. E. Villanueva<sup>a</sup> y F. Gutiérrez.

a) Instituto Nacional de Astrofísica Óptica y Electrónica, C.P 72000 Puebla, México. b) Facultad Experimental de Ciencia y Tecnología, Depto. de Física, Universidad de Carabobo, C.P.2005 Valencia-Venezuela; eorozco@uc.edu.ve†, c) Depto. de Física, Centro de Aplicaciones Tecnológicas y Desarrollo Nuclear, CEADEN-CITMA, La Habana, Cuba, d) Centro de Estudios y Prevención del Cáncer (CEPREC), Oaxaca, México.

†autor para la correspondencia

Recibido el 15/04/2009. Aprobado en versión final el 18/02/2010.

**Sumario.** La determinación del espectro de reflexión difusa de la piel humana en el rango espectral de 400 nm-1000 nm empleando espectrómetros de fibras ópticas es una técnica no invasiva ampliamente usada para estudiar los parámetros ópticos de este órgano, brinda información sobre las propiedades de absorción y de esparcimiento de la luz que pueden ser empleadas para estudiar la morfología y fisiología del tejido y poder detectar y diagnosticar enfermedades de la piel en su etapa inicial. En este trabajo se propone un algoritmo computacional para la selección de los atributos más relevantes de los espectros de reflexión difusa en la piel humana obtenidos con un sistema experimental que consta básicamente de un espectrómetro, una fuente de luz blanca y una sonda de fibra óptica bifurcada que permite enviar y coleccionar la luz. Para clasificar la señal espectral se ha diseñado una interfaz gráfica en Matlab2006© que hace uso de las máquinas de soporte vectorial y se ha utilizado un algoritmo de selección de atributos que permite alcanzar una sensibilidad y especificidad superior al 80% y una exactitud en la clasificación del 85%.

**Abstract.** The determination of diffuse reflection spectrum on human skin in the spectral range from 400nm-1000nm using an optical fiber spectrometers is a non-invasive technique widely used to study the optical parameters of this tissue, provides information about the absorption and scattering properties of light that can be employed to study the morphology and physiology of the tissue and to detect and diagnose skin diseases in early stages. In this paper a computational algorithm for the selection of the most important attributes of diffuse reflection spectra of human skin obtained with an experimental system that basically consists of a spectrometer, a white light source and bifurcated fiber optic probe that allows send and collect light. To classify the spectral signal was designed a Matlab2006© graphical interface which use support vector machines and algorithm for selecting attributes that allows to achieve a sensitivity and specificity exceeding 80% and 85% of accuracy in the classification.

**Palabras clave.** Tissue Damage 87.85.em, Pattern Recognition 42.30.Sy, Cancer 87.19.xj

### 1 Introducción

El diagnóstico de lesiones pigmentadas en la piel se hace generalmente mediante la apreciación visual empleando el método ABCD<sup>1</sup> (asimetría, borde, color, dimensión),

principalmente del color y el borde, la biopsia aún se mantiene como la regla de oro para obtener un diagnóstico fiable. Por esta razón en los últimos años se han desarrollado investigaciones empleando técnicas ópticas no invasivas que permiten diagnosticar lesiones y extraer

parámetros ópticos en tejidos biológicos<sup>2-4</sup>.

La Espectroscopia de Reflexión Difusa (ERD) es una técnica óptica no invasiva que ha sido ampliamente usada para la caracterización de los tejidos biológicos<sup>5-8</sup>. Con la ERD se puede obtener información sobre las propiedades de absorción y de esparcimiento del tejido. El principio fundamental para diagnosticar enfermedades con técnicas espectroscópicas consiste en construir algoritmos robustos que permitan extraer las características más importantes de la señal espectral y correlacionarlas con su respectiva patología<sup>9</sup>.

Las Máquinas de Soporte Vectorial (MSV), son una herramienta computacional empleada para discriminar patrones o conjuntos de identidades que comparten alguna característica que las diferencia de otras. El objetivo de este trabajo es implementar un sistema de clasificación empleando MSV para diferenciar el tejido biológico sano del tejido lesionado a partir del espectro de reflexión difusa en la piel, para ello se propone preprocesar la señal espectral y seleccionar los atributos más relevantes con un algoritmo de búsqueda como el Best-First y finalmente colocar todos los procesos computacionales en una interfaz gráfica que permita al usuario leer los datos obtenidos mediante ERD y clasificarlos para diagnosticar lesiones en piel.

## 2 Materiales y método

La colección de los espectros de reflexión difusa en la piel humana fue obtenida de 95 pacientes voluntarios que asistieron a una campaña de prevención del cáncer de piel promovida por el Centro de Estudios y Prevención de Cáncer (CEPREC) en la ciudad de Juchitán de Zaragoza en el estado de Oaxaca, México.

**Sistema Experimental.** El arreglo experimental empleado se muestra en la figura 1 y consiste de un espectrómetro USB4000 equipado con un detector CCD Toshiba de 3648 píxeles, una fuente de luz HL2000 optimizada para el VIS-NIR (360-2000nm), una sonda de fibra óptica bifurcada (R600\7\VIS\125F), un patrón de reflexión (Teflón) y un computador con el software SpectraSuite que calcula automáticamente el porcentaje de luz reflejada mediante la siguiente expresión matemática<sup>10</sup>.

$$\%R(\lambda) = \frac{S(\lambda) - D(\lambda)}{R_{mr}(\lambda) - D(\lambda)} \times 100\% \quad (1)$$

donde  $S(\lambda)$  es la señal del medio analizado (Piel) para cada longitud de onda ( $\lambda$ ),  $D(\lambda)$  es la señal de oscuridad y  $R_{mr}(\lambda)$  es la muestra de referencia.

Es importante mencionar que el espectro de reflexión difusa se ha medido cinco veces en el sitio de la lesión y en una zona adyacente a la misma.

## 3 Máquinas de soporte vectorial

Los fundamentos de la MSV han sido desarrollados por

Vapnik<sup>11</sup> y han ganado mucha popularidad debido a sus características y rendimiento, éstas permiten enfrentar problemas de clasificación en dominios complejos y pueden ser usadas para extraer información relevante a partir de conjuntos de datos y construir algoritmos de clasificación eficientes y rápidos para datos masivos.

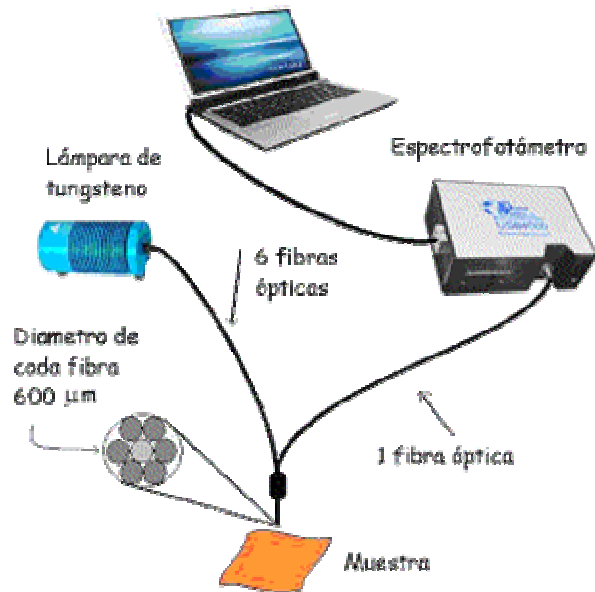
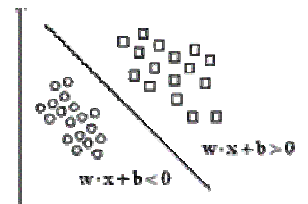


Figura 1. Diagrama esquemático del arreglo experimental

Figura 2. Hiperplano de separación



Las MSV, aplicadas a problemas de clasificación, mapean los datos a un espacio de características alto-dimensional, donde se puede encontrar más fácilmente un hiperplano de separación. Este mapeo puede ser llevado a cabo aplicando una función tipo Kernel que transforma implícitamente el espacio de entrada en un espacio de características de mayor dimensión. El hiperplano de separación es calculado maximizando la distancia de los patrones más cercanos.

Considerando un conjunto de datos de entrenamiento  $\{x_i, y_i\}$  con  $i = 1, \dots, m$ ,  $y_i \in \{-1, 1\}$  y  $x_i \in \mathbb{R}^d$ . Existe un hiperplano, como se muestra en la figura 2, que separa las etiquetas positivas y negativas tales que:

$$w \cdot x_i + b \geq 1 - \xi_i \quad \text{para } y_i = 1 \quad (2)$$

$$w \cdot x_i + b \leq -1 - \xi_i \quad \text{para } y_i = -1 \quad (3)$$

$\xi_i \geq 0, \forall i$ , donde  $w$  es la normal al hiperplano y  $\xi_i$  son las variables introducidas por los errores de clasificación en calidad de violaciones del hiperplano de forma tal que  $\sum \xi_i$  sea la cota de error de clasificación.

## 4 Pre-procesado de la señal espectral y selección de atributos

Una forma general de llevar a cabo el proceso de clasificación consiste en la separación de elementos de un conjunto  $\{x\}$  en diferentes subconjuntos

$x_i = \{1, 2, \dots, m\}$  denominados clases, con base en las características medidas de los elementos de  $x$ . Determinando las propiedades del subconjunto en los cuales puede ser clasificado el conjunto original (modelo), los elementos de éste son comparados con cada modelo para indicar a cuál clase pertenecen.

Una vez obtenidos los cinco espectros de reflexión difusa ( $E_1, \dots, E_5$ ), se procede a eliminar la columna correspondiente a la longitud de onda en cada espectro mediante un programa implementado en Matlab®, posteriormente se calcula el promedio  $\overline{\%R}$  de las cinco medidas realizadas y finalmente se normaliza el promedio obtenido  $\overline{\%R}_N$ , el siguiente paso es seleccionar del total de 3161 elementos (atributos) los más relevantes ( $A_R$ ) de forma que se pueda reducir la cantidad de atributos para emprender el proceso de clasificación empleando MSV. En la figura 3 se describe en forma de esquema este proceso.

El proceso de selección de los atributos más relevantes de la columna normalizada con la información de la reflexión se llevó a cabo empleando el software de código abierto WEKA, para el cual se deben preparar previamente los datos con las características exigida por el programa, este proceso fue realizado mediante un programa implementado en Matlab®. El algoritmo de búsqueda empleado es un algoritmo de búsqueda incremental que utiliza información de búsquedas previas a problemas similares de forma más rápida que realizando cada búsqueda partiendo de cero<sup>12</sup>. En este caso se ha empleado Best-First, el cual busca en el espacio de los subconjuntos de atributos utilizando la estrategia greedy Hill-climbing con backtracking. La dirección de la búsqueda realizada por Best-First fue hacia adelante partiendo del conjunto vacío de atributos.

## 5 Resultados

Los espectros de reflexión difusa de un tejido lesionado y una zona circundante sin lesión, se muestran en la figura 4. En ésta se puede apreciar cómo la señal espectral es menor en el tejido lesionado, además en el espectro del tejido sano se observa el típico patrón "W" relacionado con la presencia de la sangre en la dermis en la región entre 500 nm y 600 nm.

En el proceso de aprendizaje y prueba del sistema de clasificación con MSV cada espectro fue tratado como un vector de 14 atributos, después de haber realizado el respectivo proceso de selección de atributos para obtener los más relevantes.

Una vez generada la función modelo con el conjunto de entrenamiento, procedemos a realizar la evaluación

mediante el clasificador (MSV). Para tener una certera estimación de la calidad del subconjunto de características seleccionadas, se usa la técnica de validación cruzada con  $K$  particiones y determinamos el error promedio que hemos obtenido al realizar las  $K$  pruebas, así obtenemos una aproximación del error del clasificador. La exactitud del clasificador empleando  $k=10$  fue de 84.128.

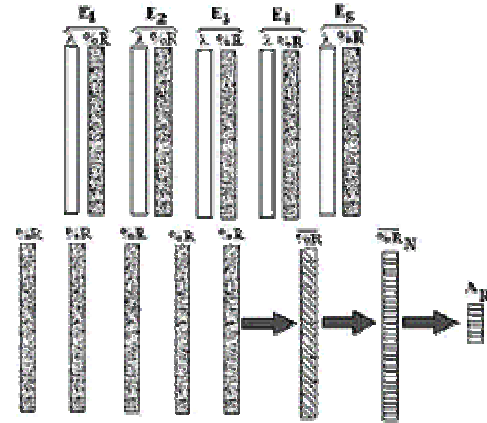


Figura 3. Esquema general del Pre-procesado de los datos.

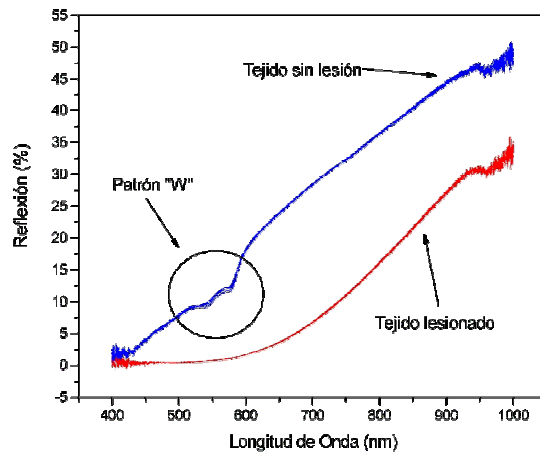


Figura 4. Espectros de reflexión difusa para un tejido sin lesión y uno lesionado.

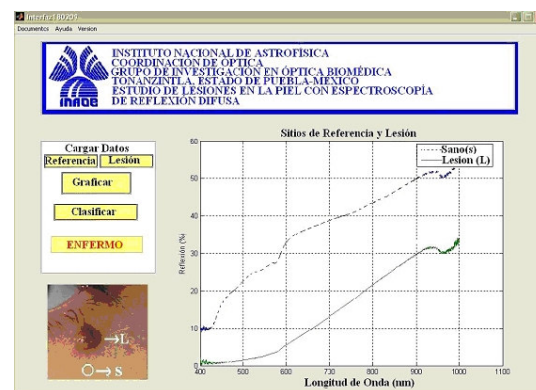


Figura 5. Imagen de la interfaz gráfica desarrollada en Matlab.

El análisis de la validez de una prueba diagnóstica puede obtenerse calculando los valores de la sensibilidad y la especificidad<sup>13</sup>. La sensibilidad es la probabilidad de clasificar correctamente a un individuo enfermo y por lo tanto indica la capacidad para detectar la enfermedad y se puede determinar mediante la siguiente expresión:

$$\text{Sensibilidad} = \frac{VP}{VP + FN} \quad (4)$$

donde VP son los verdaderos positivos y FN los falsos negativos, por su parte la especificidad es la probabilidad de clasificar correctamente a un individuo sano, y se calcula con la siguiente expresión:

$$\text{Especificidad} = \frac{VN}{VN + FP} \quad (5)$$

en este caso VN son los verdaderos negativos y FP los falsos positivos.

Finalmente los resultados obtenidos para clasificar 20 espectros: 10 con lesiones y 10 sin lesiones son los siguientes: Sensibilidad: 81.81%; Especificidad: 88.88%; Exactitud: 85%.

En la figura 5 se muestra la interfaz gráfica que fue diseñada en Matlab para procesar y clasificar los espectros de reflexión difusa. Como se puede apreciar en la figura, la interfaz cuenta con dos botones que permiten cargar los datos de la señal espectral: uno para el sitio de referencia (sano) y otro para la lesión, un tercer botón que permite mostrar en pantalla las gráficas de los datos cargados, así como también la fotografía del área de estudio, y por último, el botón más importante que corresponde a la clasificación del espectro que se ha cargado con la etiqueta de lesión.

## 6 Conclusiones

El estudio presentado demuestra que la técnica de espectroscopia de reflexión difusa es una técnica óptica no invasiva que acompañada de un sistema de clasificación eficiente permite diferenciar el tejido sano del tejido lesionado. El resultado de 85% de exactitud en la clasificación demuestra que las Máquinas de Soporte Vectorial pueden ser empleadas como una herramienta para discriminar lesiones en la piel. Finalmente la Interfaz gráfica desarrollada permite a los usuarios no familiarizados con las herramientas computacionales empleadas, poder clasificar lesiones a partir de los espectros de reflexión difusa.

## Referencias

1. T. B. Fitzpatrick, R. A. Johnson, K. Wolf and D. Suurmond, "Atlas en color y sinopsis de dermatología clínica", McGraw Hill Interamericana, Madrid (2001).
2. Y. N. Mirabal, S. K. Chang, E. N. Atkinson, A. Malpica, M. Follen, and R. R. Richards-Kortum, "Reflectance spectroscopy for *in vivo* detection of cervical precancer" J. Biomed. Opt. 7, 587-594 (2002).
3. F. Koenig, R. Larne, H. Enquist, F. J. McGovern, K. T. Schomacker, N. Kollias, and T. F. Deutsch, "Spectroscopic measurement of diffuse reflectance for enhanced detection of bladder carcinoma," Urology 51, 342-345 (1998).
4. M. C. Skala, G. M. Palmer, K. M. Vrotsos, A. Gendron-Fitzpatrick, and N. Ramanujam, "Comparison of a physical model and principal component analysis for the diagnosis of epithelial neoplasias *in vivo* using diffuse reflectance spectroscopy" Opt. Express 15, 7863-7875 (2007).
5. V. P. Wallace, D. C. Crawford, P. S. Mortimer, R. J. Ott and J. C. Bamber, "Spectrophotometric assessment of pigmented skin lesions: methods and feature selection for evaluation of diagnostic performance", Phys. Med. Biol. 45, 735-751(2000).
6. J. J. Scarisbrick, C. D. O. Pickard, A. C. Lee, G. M. Briggs, K. Johnson, S. G. Bown, M. Novelli, M. R. S. Keshtgar, I. J. Bigio, R. Yu, "Elastic scattering spectroscopy in the diagnosis of pigmented lesions: comparison with clinical and histopathological diagnosis", Proc. SPIE 5141, 147-156 (2003).
7. I. J. Bigio and J. R. Mourant, "Ultraviolet and visible spectroscopies for tissue diagnostics: fluorescence spectroscopy and elastic-scattering spectroscopy", Phys. Med. Biol. 42, 803-813 (1997).
8. M. G. Muller, T. A. Valdez, I. Georgakoudi, V. Backman, C. Fuentes, S. Kabani, N. Laver, Z. Wang, C. W. Boone, R. R. Dasary, S. M. Shapshay, S. M. Feld, "Spectroscopic detection and evaluation of morphologic and biochemical changes in early human oral carcinoma", Cancer 97, 1681-1692 (2003).
9. W. Lin, X Yuan, P Yuen, WI Wei, J Sham. PC Shi, "Classification of *in vivo* autofluorescence spectra using support vector machines". J Biomed. Opt.9, 180-186 (2004).
10. Oceans Optics, SpectraSuite Installation and Operation Manual, Document Number 000-20000-300-02-0607.
11. V. Vapnik, Controlling the Generalization Ability of Learning Processes in The Nature of Statistical Learning Theory, pp. 93-99 (Springer-Verlag, 2000).
12. Sven Koenig, Maxim Likhachev, Yaxin Liu, David Furci, "Incremental Heuristic Search in Artificial Intelligence", Artificial Intelligence Magazine 25, (2), 99-112, (2004).
13. D. G. Altman., J.M. Bland., "Statistics Notes: Diagnostic tests 2: predictive values". British Medical Journal, 9, 309, (1994).