Proceedings of the International School and Conference on Photonics, PHOTONICA09

Optical Biopsy Method for Breast Cancer Diagnosis Based on Artificial Neural Network Classification of Fluorescence Landscape Data

T. Dramićanin[a,*], I. Zeković[a], B. Dimitrijević[a], S. Ribar[b] and M.D. Dramićanin[a]

[a]Institute of Nuclear Sciences "Vinča", University of Belgrade, 11001 Belgrade, Serbia

[b]Faculty of Mechanical Engineering, University of Belgrade, Kraljice Marije 16, 11120 Belgrade, Serbia

Supervised self-organizing map, a type of artificial neural network, is applied for classification of human breast tissue samples utilizing data obtained from fluorescence landscape measurements. Female breast tissue samples were taken soon after the surgical resection, identified and stored at −80 °C until fluorescence measurements. From fluorescence landscapes obtained in UV–VIS region spectral features showing statistically significant differences between malignant and normal samples are identified and further quantified to serve as a training input to neural network. Additional set of samples was used as a test group input to trained network in order to evaluate performance of proposed optical biopsy method. Classification sensitivity of 83.9% and specificity of 88.9% are found.

PACS numbers: 87.64.kv, 84.35.+i, 87.19.xj, 33.50.–j

1. Introduction

The breast cancer is one of the most common malignant tumors among women in the world [1] and if not diagnosed at proper time it delivers high mortality rates. On the other hand, if observed in early stages breast cancer is one of the most treatable forms of cancer. The important task of oncology is the development of methods for the early detection of tumors and tumor pre-stages, because a successful therapy essentially depends on the point in time at which the disease is detected, making possible to improve patient quality of life and survival rates.

Tissue diagnosis using optical spectroscopy has been considered as an alternative technique for the conventional diagnostic methods because of its advantages, such as minimal invasiveness, less time consumption and reproducibility. For more than two decades, various optical spectroscopic techniques including fluorescence spectroscopy have been widely explored as diagnostic tools in the discrimination of normal from abnormal tissues in various organ sites such as breast [2, 3], colon [4], oral [5], and skin [6]. The endogenous fluorophores, such as nicotinamide adenine dinucleotide (NADH), flavin adenine dinucleotide (FAD), collagen, elastin, amino acids, vitamins, lipids and porphyrins, have a significant variation in the concentration in different tissue types. These differences, together with alternations in the local environment within the tissue, are the basis for the discrimination between tumor and normal tissue by fluorescence spectroscopy. To observe majority of fluorescence changes a more sophisticated method of fluorescence diagnosis is developed, called fluorescence landscape spectroscopy (also known as excitation–emission matrix spectroscopy), utilizing multiple-color illumination, with the full fluorescence spectrum recorded for each excitation wavelength. The different excitation wavelengths might be expected to variously excite different fluorophores, resulting in more complex emission patterns with more information relevant to biochemical changes than for single-color excitation, and with presumed greater likelihood of distinguishing malignancy from normal conditions. However, observed data are subtly related in ways that are often difficult to express in the form of diagnostic rules and must be processed for tissue classification purposes.

Artificial neural networks are very useful for handling complex decision tasks such as those involved in medical diagnosis. The networks can capture such relationships between the input findings to generate robust outputs. In addition, networks are always consistent, for they are not prone to human fatigue or bias. Among the various existing neural network architectures and learning algorithms, Kohonen's self-organising map (SOM) [7, 8] is one of most popular neural network models. SOMs converts high-dimensional, non-linear statistical relationships into simple geometric relationships in an n-dimensional array. This reduced representation seeks to best preserve the in-

* corresponding author; e-mail: tatjana@vinca.rs

put data's original topology and density. Although SOMs are often described as neural networks, they can be also described as vector prototyping. A SOM's aim is to find the optimal set of lower dimensional prototype vectors to properly group the high dimensional pattern space (clustering property), while preserving the probability density of the original manifold (topology preservation). For these reasons they are particularly good candidate for fluorescence data conversion into classification rules.

2. Methods and results

Method for breast cancer diagnostic, based on measurements of tissue fluorescence landscapes and data classification by supervised self-organizing maps (SSOM), is developed in several steps. First, fluorescence landscapes are measured on two sets of breast tissue specimens, malignant and normal. Selection of spectral features is made regarding statistical significance of observed differences in fluorescence response. Second, SSOM is constructed and trained with data obtained by quantification of selected spectral features. Then fluorescence measurements are performed on the test group samples and chosen spectral components are quantified to create test data sets. Tissue classifications are then obtained as SSOM outputs after test data inputs. Finally, classification results are compared with histopathology data to calculate classification sensitivity and specificity of the method.

2.1. Fluorescence landscape measurements and data

The breast tissue specimens were obtained from the Institute of Oncology and Radiology of Serbia. The samples were taken soon after the surgical resection, identified and stored at $-80\,^{\circ}\mathrm{C}$ until luminescence characterization. Their sizes varied from $0.2 \times 0.5 \times 0.5$ cm^3 to $0.3 \times 1.0 \times 1.5$ cm^3. According to the histopathological exam, all malignant breast tissue samples included in the present study were infiltrating ductal carcinoma. Tissue specimens were collected after the signed Informed Consent was obtained from patients. The Consent was acquired according to the International Ethical Guidelines for Biomedical Research involving Human Subjects (CIOMS), Geneva 1993 and the Guidelines for Good Clinical Practice (CPMP/ICH/135/95), September 1997.

Fluorescence landscapes were measured at room temperature using Perkin Elmer Fluorescence Spectrophotometer LS45 in two excitation-emission ranges to avoid excitation-emission overlapping. First range covered excitation from 335 to 400 nm and emission from 430 to 625 nm (EEM1), and the second had excitation from 400 to 470 nm and emission from 500 to 640 nm (EEM2). The spectra were collected at 150 nm/min scan rate and were automatically normalized to excitation power by the instrument. Fluorescence landscapes obtained as a difference of averaged measurements on normal and malignant sample sets in both spectral ranges are given in Fig. 1 in a form of contour diagrams, EEM1 on the left and EEM2 on the right. It can be clearly seen that normal and malignant breast tissue differently fluoresce in five spectral regions, two in the EEM1 and three in the EEM2, marked with dashed lines in Fig. 1. Volumes below intensity surface are calculated in each region for all samples using method previously reported [3], and further denoted as $V_{\mathrm{I}-1}$, $V_{\mathrm{II}-1}$, $V_{\mathrm{I}-2}$, $V_{\mathrm{II}-2}$ and $V_{\mathrm{III}-2}$, where roman numbers stand for spectral region marked in Fig. 1. 1 and 2 refer to EEM1 and EEM2, respectively.

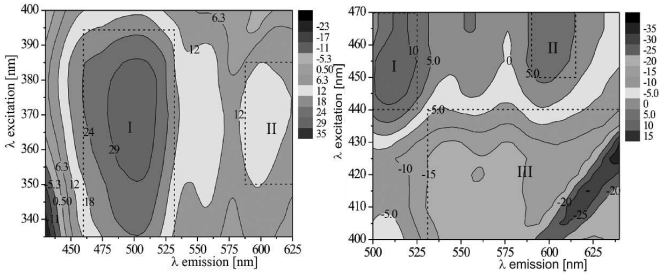

Fig. 1. Difference of averaged fluorescence landscapes of normal and malignant tissue sample groups in two spectral ranges.

Existence of statistically significant differences between spectral volumes of normal and malignant tissue is evaluated through hypothesis testing using the *two-tailed t-test* [9]. The results of hypothesis testing together with mean values and standard deviations are given in Table I. Decision on statistical significance is made in a traditional way [9] on the basis of probability value for null hypothesis, p: > 0.05 Not Significant (NS), 0.01 to 0.05 Significant (S), 0.01 to 0.001 Very Significant (VS), and < 0.001 Extremely Significant (ES).

TABLE I

Results of the statistical hypothesis testing using *two-tailed t-test* (mean value and σ are given in arbitrary units).

	Tissue type	Mean value	σ	p	Decision
$V_{\mathrm{I}-1}$	malignant normal	212700 274428	97880 85416	0.035	S
$V_{\mathrm{II}-1}$	malignant normal	33273 45360	10274 12596	0.002	VS
$V_{\mathrm{I}-2}$	malignant normal	55580 63717	10689 9161	0.012	S
$V_{\mathrm{II}-2}$	malignant normal	20673 27660	7090 4671	< 0.001	ES
$V_{\mathrm{III}-2}$	malignant normal	219508 190385	38294 55399	0.054	NS

Taking into account results of statistical analysis we choosed $V_{\mathrm{II}-1}, V_{\mathrm{I}-2}$, and $V_{\mathrm{II}-2}$ as inputs for SSOM.

2.2. SSOM architecture and training

In order to use SOM in a supervised way (SSOM) the network design comprised a 2-dimensional Kohonen map

(5×5; hexagonal connections between map nodes) and a communication layer, Fig. 2. The later one poses two types of nodes: observation nodes (X) and class coding nodes (C). During the training phase the whole communication layer is used as input layer, while during exploitation phase observation nodes are used for inputs and class coding nodes are used as outputs.

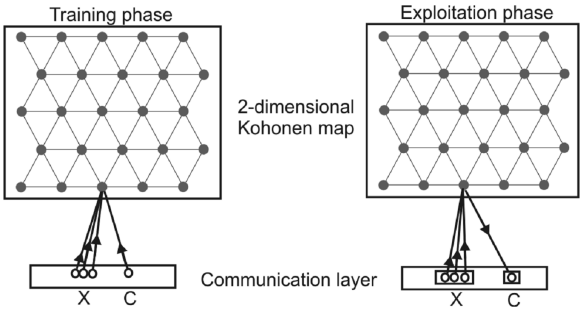

Fig. 2. Schematic of SSOM architecture (communication layer is connected to all nodes in Kohonen map — for the simplicity connections are drawn just to one node).

Training data set contains 2000 vectors with $V_{\mathrm{II}-1}, V_{\mathrm{I}-2}$ and $V_{\mathrm{II}-2}$ values generated as Gaussian distribution with mean values and standard deviations taken from Table II for normal and malignant tissue groups: $X_{in}(i) \in R^n$, $i = 1, 2, \ldots, 2000$, $n = 3$.

2.3. Exploitation phase and results

To test success rate of proposed optical biopsy method we introduced to trained SSOM data obtained from fluorescence landscape measurements on test group of 67 biopsies for which histopathology found 31 malignant and 36 normal samples. SSOM provided diagnostics through class coding nodes and the results are presented in Table II in comparison to histopathology findings.

TABLE II

Accuracy of SSOM based optical biopsy method.

S		Histology	
S	Tissue type	Malignant	Normal
O	malignant	26	4
M	normal	5	32

From these data a sensitivity of 83.9% and specificity of 88.9% is calculated for presented SSOM based optical biopsy method. Five malignant biopsies and four normal biopsies were misclassified, yielding a malignant predictive value of 86.7% and a normal tissue predictive value of 86.5%.

3. Conclusion

In this work the statistical and the neural network approach have been implemented for breast cancer diagnosis from tissue fluorescence landscapes. It is shown that differences in fluorescence of malignant and normal breast tissue are statistically significant in three spectral regions and that fluorescence data from these regions are adequate inputs for the SSOM to provide relatively high diagnostic sensitivity and specificity.

Acknowledgments

This work was supported by the Serbian Ministry of Science and Technological Development (project No. 143010).

References

[1] `http://www.komen.org/bci` (accessed 2009).

[2] R.R. Alfano, G.C. Tang, A. Pradhan, W. Lam, D.S.J. Choy, E. Opher, *IEEE J. Quantum Electron.* **23**, 1806 (1987).

[3] T. Dramićanin, M.D. Dramićanin, V. Jokanović, D. Nikolić-Vukosavljević, B. Dimitrijević, *Photochem. Photobiol.* **81**, 1554 (2005).

[4] K.T. Schomacker, J.K. Frisoli, C.C. Compton, T.J. Flotte, J.M. Richter, N.S. Nishioka, T.F. Deutsch, *Lasers Surg. Med.* **12**, 63 (1992).

[5] S.D. Kamath, K.K. Mahato, *J. Biomed. Opt.* **12**, 14028 (2007).

[6] H.J.C.M. Sterenborg, M. Motamedi, R.F. Wagner, M. Duvic, S. Thomsen, S.L. Jacques, *Lasers Med. Sci.* **9**, 191 (1994).

[7] T. Kohonen, *Self-Organizing Maps*, 3rd ed., Springer, Berlin 2001.

[8] T. Kohonen, *Computer* **21**, 11 (1988).

[9] H. Motulsky, *Intuitive Biostatistics*, Oxford University Press, New York 1995.