Backpropagation Neural Network for Analysis and Classification of Fluorescence Spectroscopy of Squamous Cell Carcinoma in Animal Model

João Marcelo Nogueira
São Carlos Institute of Physics
University of São Paulo
São Carlos, Brazil
jmarcelo@ifsc.usp.br

Marlon Rodrigues Garcia
São Carlos Institute of Physics
University of São Paulo
São Carlos, Brazil
marlongarcia@usp.br

Michelle Barreto Requena
São Carlos Institute of Physics
University of São Paulo
São Carlos, Brazil
requenamichelle@gmail.com

Lilian Tan Moriyama
São Carlos Institute of Physics
University of São Paulo
São Carlos, Brazil
lili@ifsc.usp.br

Sebastião Pratavieira
São Carlos Institute of Physics
University of São Paulo
São Carlos, Brazil
prata@ifsc.usp.br

Daniel Varela Magalhães
São Carlos Institute of Physics
University of São Paulo
São Carlos, Brazil
daniel@sc.usp.br

***Abstract*—The present study aims to evaluate the performance of a backpropagation neural network (BPNN) using the principal component analysis (PCA) of fluorescence spectra for discrimination between normal skin and skin tumor on mice. The fluorescence spectra were acquired from nude mice with induced squamous cell carcinoma (SCC). The artificial neural network (ANN) used in this study is a classical multiplayer feed-forward type with a back-propagation algorithm. The classification results show this technique as promising for healthy and unhealthy tissue classification. During the validation, the network classified 100% of the training set spectra and 90% of the test set.**

***Index Terms*—Fluorescence; Spectroscopy; Artificial neural network; Squamous cell carcinoma; Animal model; Skin cancer**

I. Introduction

According to the World Health Organization (WHO), non-melanoma skin cancer (NMSC) is one of the most common cancers, and the basal cell carcinoma (BCC) is the most frequent NMSC. In the recent years, optical spectroscopy methods have had a significant impact in the field of biomedical diagnostics, being used in the study of many tissue conditions. Fluorescence spectroscopy has been the most widely explored technique, mainly because fluorescence is highly sensitive to the biochemical composition of tissues. In the area of oncology, this technique has been used as a fast, non-invasive and non-destructive option in the diagnosis of tissue injuries, such as skin cancer [1] and neoplasms [2] in addition to improve cancer treatment during photodynamic therapy (PDT), since fluorescence can be used to identify and monitor photosensitizers distribution [3], [4]. Lesions are initially classified using a dermatoscope and ABCD (asymmetry, border irregularity, color unevenness, diameter) scoring criteria [5]. When early detected, many cancers can be treated during their initial stage, offering a better prognosis for the patient.

Biological tissues are complex and comprise many molecules that interact differently with light. Depending on the type of interaction, the molecule can be called chromophores (chemical groups that absorb light), scatterers (structures that change the propagation direction of the incident photon, but conserve its energy) and fluorophores (chemical groups that can convert the light absorbed in fluorescence) [4]. The fluorescence spectrum detected *in vivo* is a superposition of fluorescence spectra of endogenous fluorophores existing in the tissue, distorted by partial re-absorption by other tissue molecules (mainly blood and melanin) [13]. The endogenous fluorophores detected are some structural proteins (such as collagen, elastin, and keratin), protein cross-links, and co-enzymes (nicotinamide adenine dinucleotide, or NADH, and flavins).

It has been shown that tumors can be detected due to changes in their fluorescence properties in comparison with normal tissue. For example, BCC lesions normally reveal fluorescence intensities lower than surrounding normal skin tissues, while SCC usually exhibits higher fluorescence intensities than surrounding skin. In the other hand the malignant melanoma (MM) lesions have no significant spectral shape changes but always have extremely weak fluorescence [1].

Although we know that there are several differences between the fluorescence emission of tumors and healthy tissues, the data analysis can be very complex in order to identify the best spectral features that can be able to distinguish between

The authors acknowledge the support provided by Brazilian Funding Agencies: Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Finance Code 001; CNPq (465360/2014-9 and 306919/2019-2) and São Paulo Research Foundation (FAPESP) grants: 2013/07276-1 (CEPOF); 2014/50857-8 (INCT)

(978-0-7381-2418-6/21/$31.00 ©2021 IEEE)

tumor and normal tissue.

Most of the studies published in the literature that applied artificial intelligence for skin cancer are normally developed with more elaborated optical techniques. Interesting studies have been published using artificial neural networks (ANN) in Raman spectroscopy to diagnose skin disorders [6]–[8]. Also important studies with ANN have been applied together with synchronous fluorescence spectroscopy (SFS) [9]. Although interesting techniques have been applied in conjunction with ANN, there is a gap in the use of ANN with simple optical techniques like fluorescence spectroscopy for skin diseases and disorders.

Backpropagation neural networks (BPNNs) are interesting architectures, already been applied for tumor classification [10], [11], and could also been considered for skin conditions. The BPNNs are feed-forward networks trained by the error propagation back through the networks. The ANN are more efficient methods for pattern recognition and calibration than conventional statistical methods because of their advantages with self-organization, self-study, self-adoption, and unusual fault tolerance. The present study aims to evaluate a classification of spectra, using a BPNN topology for the recognition of healthy and tumor tissues in animal model, only analyzing the intrinsic light-induced fluorescence emission.

II. Material and Methods

The *in vivo* measurements of fluorescence spectroscopy were performed using a spectrometer (USB2000, Ocean Optics Inc., USA) with laser excitation at 408 nm. [15] The experimental procedures were approved by the Ethics Committee for Animal Use from São Carlos Institute of Physics at the University of São Paulo. The xenographic tumor induction was performed using a human squamous cell carcinoma cell line (A-431, ATCC CRL-1555™, Manassas, Virginia, USA), considered a fast-growing model of non-melanoma skin cancer, with nodular and non-cystic tumor formation [12]. The inoculation was done only once on the right flank of the animal using intradermal injection of 50 μL of cell volume resuspended in PBS, with a concentration of 10^6 cells. The measurements were performed when the tumors reached about 20-25 cubic millimeters. For each animal, we collected five random spectra in the tumor and five in the surrounding healthy skin. The dataset was constitute from the endogenous fluorescence spectra of 35 female balb c nude mice with 20 g average weight. It is taken 36 spectra of normal tissue and 31 spectra of induced tumor. Fig. 1 shows the typical spectrum of healthy tissue and SCC tumor.

A. Pre-processing

The files generated by the spectrometer consisted of 2048 points with count values of the intensity of the collected fluorescence, distributed in the range of 300 to 1100 nm. The tissue fluorescence emission was restricted to the range of 400 to 800 nm, resulting in 1156 points of intensity. During a classification, two classes were assigned, one for healthy tissue and one for the SCC tumor. Fluctuations in the power of the excitation laser, differences in the positioning of the collection fiber in the tissue and the consequent coupling of the light during a measurement or even alteration in the skin, are some factors that can promote variations in the fluorescence spectra.

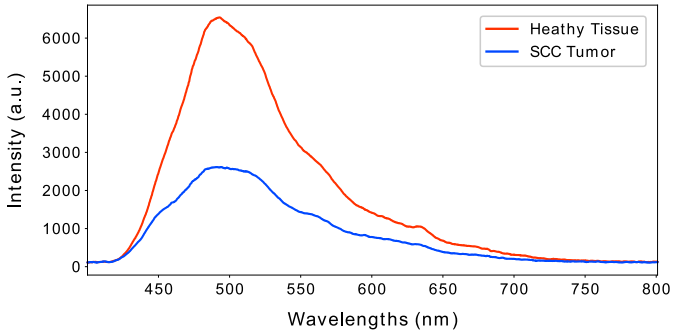

Fig. 1. Typical endogenous fluorescence spectra of healthy skin and SCC tumor.

The spectra were normalized by the maximum-minimum method, to maintain the same order of amplitude within the data sets. A Savitzky-Golay smoothing filter from the SciPy library (savgol filter function) with a 3rd order polynomial and a window with 7 convolution coefficients, allowed smoothing of spectrum, eliminating noise and preserving the spectrum peaks. Then, the first derivative in the spectrum was applied by the Savitzky-Golay method, allowing to correct the differences from the baseline in the spectra and the main sources of variation between the values. Principal component analysis (PCA) was used to compress the spectral data. The application of the PCA allows the analysis of how the input variables can be correlated to a smaller number of components, in order to preserve most of the changes in the input data. The library used to PCA application was the Scikit-learn (function skPCA). In training part, twenty five main components were chosen for RNA training and approximately 90.43 % of the information, was collected for application at the input of the BPNN. Figure 2 shows the different classes for the spectra of healthy and tumor tissue.

B. Backpropagation Neural Network

The BPNN is a multilayered feed-forward network trained by propagation of error back trought the network, referring to a wide family of networks (RNA), whose architecture consists of different interconnected layered systems. Its learning algorithm consists of the Deepest-Descent technique [17]. The network can reduce the error in highly complex non-linear functions. In BPNN, the weights are initialized randomly at the training beginning. There will be a desired output, for which the training is done, the supervisory learning is used here. During the forward pass of signal, according to the initial weights and activation function used, the network gives an output. This output is compared with desired output. If both are not equal, an error occurs. During reverse pass, the error is back-propagated and weights of hidden and output layer are adjusted. The whole process then continues until error is

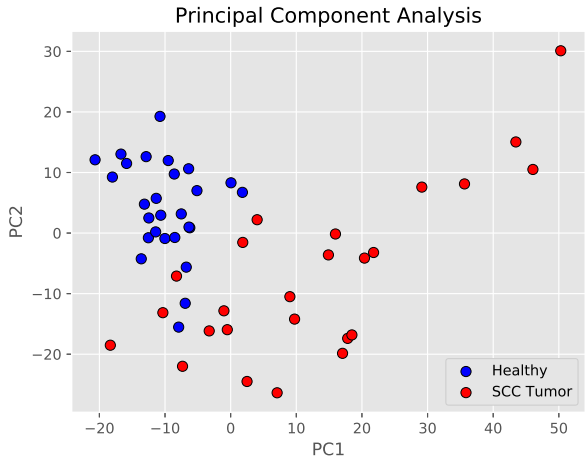

Fig. 2. Principal component analyses of healthy and tumor skin.

zero. The network is trained with known values. After training, network can perform decision making.

In this work the BPNN works with the input layer, the hidden layer and the output layer. The output layer returns values between 0 and 1 indicating the probability for the classification of each label. A training process was performed using a BPNN algorithym written in Python using the open-source libraries NumPy and Scikit-learn. In the training of the BPNN, the spectra were separated into two sets of data at random, including spectra of healthy and tumor tissues, whereas the parameters on the BPNN were changed to attempt optimization of prediction.

III. Results

The evaluation in neural network model was obtained by changing the number of neurons in hidden layer, the learning rate, the momentum factor, and learning epochs. In the first trial, the spectral data were separated, 52 spectra for the training set and 15 spectra for the test set. Fig. 3 shows the resulting accuracy after 800 epochs. If the number of hidden layer neuron is too small, neural network can not find relationships between input and output variables. However, if the number of hidden layer neuron is too large, much more than the sufficient number of layers it causes lost of accuracy on validation. Adding a momentum term can help to avoid a small local minimum in the error surface and speed up the convergence. But if momentum is too large the BPNN will lead to small oscillations in irregular areas of error surface.

Learning epochs also influence the accuracy of BPNN. If the number of learning epoch is too small, the desired convergence will not be reached. The learning rate is an important parameter influencing the stability and training speed of the BPNN. If learning rate is too small, the slow convergence and local minima may happen. However, if learning rate is too large, the convergence speed increases but the performance of neural network can not work well because the network may skip the global minimum. In the hidden layer, 25 neurons were included, the learning rate set to $\eta = 0.12$ and the momentum factor set to $\gamma = 0.5$.

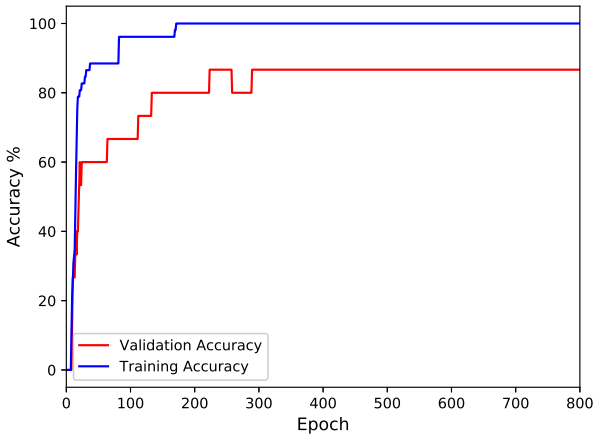

Fig. 3. The training and validation after 800 epochs.

The evaluated performance of the classifier using the BPNN considering the sensitivity specificity, and accuracy for healthy tissue and SCC tumor model use the three terms defined as follow [18]:

a) Sensitivity (true positive fraction): the result indicates positively(disease).

$$Sen. = \frac{TP}{TP + FN} \tag{1}$$

b) Specificity (true negative fraction): the result indicates negatively(non-disease).

$$Spec. = \frac{TN}{TP + FP} \tag{2}$$

c) Accuracy: the probability that the diagnostic test is performed correctly.

$$Acc. = \frac{TP + TN}{TP + TN + FP + FN} \tag{3}$$

Where, TP (True Positives)≡ Correctly classified positive cases, TN (True Negative) ≡ Correctly classified negative cases, FP (False Positives) ≡ Incorrectly classified positive cases, and FN (False Negative) ≡ Incorrectly classified negative cases.

The Sensitivity , specificity and accuracy level is shown on the follow table.

TABLE I
SENSITIVITY , SPECIFICITY AND ACCURACY BETWEEN HEALTHY AND TUMOR TISSUE

Dataset	Sensitivity(%)	Specificity(%)	Accuracy(%)
Training Dataset	100	100	100
Test Dataset	85.50	92.86	90.00

Although the BPNN reached a good prediction in the sensitivity between the healthy and malignant SCC tumor. Other recent studies [1], [19] are also presenting the specificity between different kinds of tumors. Thus, to optimize the results of BPNN prediction other types of tumors should be added to the data set.

IV. Conclusions

In this present study it was demonstrated the implementation of a BPNN for the differentiation of healthy skin and SCC tumor in mice, from fluorescence spectroscopy data sets with 85.5% sensitivity, 92.86% specificity and 90% accuracy. Although other simpler machine learning techniques like K-Nearest Neighbors (KNN) can also been applied, the BPNN was first approached given it's good fault tolerance. Pattern recognition using ANN shows the potential of this classifier as an important tool to improve the diagnosis of skin cancer. The combination of backpropagation algorithm and pre-processing techniques such as smoothing and derivative filters and also the analysis of the main components brings great benefits to the training of the BPNN, acting directly in reducing the number of epochs as well as the number of hidden layers.

References

[1] BORISOVA, E. et al. Using spectroscopy to diagnose skin cancer. SPIE Newsroom, 2014.

[2] RAMANUJAM, N. Fluorescence spectroscopy in vivo. In: MEYERS, R.A. (Ed.). Encyclopedia of analytical chemistry. Chichester: John Wiley. p.20-56, 2000a.

[3] REQUENA M.B., RUSSIGNOLI P.E., VOLLET-FILHO J.D., et al. Use of dermograph for improvement of PpIX precursor's delivery in photodynamic therapy: Experimental and clinical pilot studies. Photodiagnosis Photodyn Ther. 2020;29:101599. doi:10.1016/j.pdpdt.2019.101599.

[4] KURACHI, C.; VOLLET-FILHO, J. D.; BAGNATO, V. S. Detecção óptica no diagnóstico. In: BAGNATO, V.S. (Org.). Novas técnicas ópticas para as áreas da saúde. 1 ed. São Paulo: Livraria da Física, v., p. 81-95, 2008.

[5] David Sbrissa, Sebastião Pratavieira, Ana Gabriela Salvio, Cristina Kurachi, Vanderlei Salvador Bagnato, Luciano Da Fontoura Costa, Gonzalo Travieso, "Asymmetry and irregularity border as discrimination factor between melanocytic lesions," Proc. SPIE 9531, Biophotonics South America, 953122 (19 June 2015); https://doi.org/10.1117/12.2186180

[6] SIGURDSSON, Sigurdur et al. Detection of skin cancer by classification of Raman spectra. IEEE transactions on biomedical engineering, v. 51, n. 10, p. 1784-1793, 2004.

[7] GNIADECKA, Monika et al. Melanoma diagnosis by Raman spectroscopy and neural networks: structure alterations in proteins and lipids in intact cancer tissue. Journal of investigative dermatology, v. 122, n. 2, p. 443-449, 2004.

[8] LUI, Harvey et al. Real-time Raman spectroscopy for in vivo skin cancer diagnosis. Cancer research, v. 72, n. 10, p. 2491-2500, 2012.

[9] LENHARDT, L. et al. Artificial neural networks for processing fluorescence spectroscopy data in skin cancer diagnostics. Physica Scripta, v. 2013, n. T157, p. 014057, 2013.

[10] DANIEL, D. Arul Pon; THANGAVEL, K. Breathomics for gastric cancer classification using back-propagation neural network. Journal of medical signals and sensors, v. 6, n. 3, p. 172, 2016.

[11] ISMAEL, Mustafa R.; ABDEL-QADER, Ikhlas. Brain tumor classification via statistical features and back-propagation neural network. In: 2018 IEEE international conference on electro/information technology (EIT). IEEE, 2018. p. 0252-0257.

[12] HO, D. T. et al. Growth inhibition of an established A431 xenograft tumor by a full-length anti-EGFR antibody following gene delivery by AAV. Cancer Gene Therapy, v. 16, n. 2, p.184-194, 2009.

[13] BACHMANN L., ZEZELL D.,COSTA RIBEIRO A., GOMES L., ITO A., Fluorescence spectroscopy of biological tissues—a review, Appl. Spectr. Rev. 41, pp. 575–590, 2006.

[14] VALDES P. A. et al., "Quantitative fluorescence using 5-aminolevulinic acid-induced protoporphyrin IX biomarker as a surgical adjunct in low-grade glioma surgery," J. Neurosurg., 123 (v.3), pp.771–780, 2015. http://dx.doi.org/10.3171/2014.12.JNS14391 JONSAC 0022-3085

[15] Carolina de Paula Campos, Camila de Paula D'Almeida, Marcelo Saito Nogueira, Lilian Tan Moriyama, Sebastião Pratavieira, Cristina Kurachi "Fluorescence spectroscopy in the visible range for the assessment of UVB radiation effects in hairless mice skin" Photodiagnosis and Photodynamic Therapy Volume 20, December 2017, Pages 21-27

[16] E. Borisova, P. Pavlova, E. Pavlova, P. Troyanova, L. Avramov, Optical biopsy of human skin - A tool for cutaneous tumours' diagnosis, Int. J. Bioautomation. 16 (1 53–72. 2012.

[17] HAYKIN S. Neural Networks and Learning Machines, ser. Neuralnetworks and learning machines.Prentice Hall, no. v. 10. 2009.

[18] POLAT K, AKDEMIR B. GNES S. Computer aided diagnosis of ECG data on the least square support vector machine, Digital Signal Processing 18, pp. 25-32. 2008.

[19] DORJ, U. O., LEE, K. K., CHOI, J. Y., and LEE, M. The skin cancer classification using deep convolutional neural network. Multimedia Tools and Applications, 77(8): 9909-9924, 2018.