Automatic Pattern Recognition Techniques Applied to Medical Hyperspectral Imaging

S. Ortega[1], G.M. Callicó[1] and R. Camacho[2]
[1]Institute for Applied Microelectronics, University of Las Palmas de Gran Canaria, Spain
[2]University Hospital Doctor Negrín of Las Palmas de Gran Canaria, Spain

***Abstract*— Hyperspectral imaging is an emerging technology for medical diagnosis. In this research work, a multidisciplinary team compounds by pathologists and engineers present a proof of concept of employing hyperspectral imaging analysis in order to detect human brain tumour tissue inside pathological slides. The samples were acquired from four different patient diagnosed with brain cancer, specifically with high-grade gliomas. The images where then processed in order to remove the effect caused by the acquisition system. Later, and based on the diagnostic provided by pathologist, a spectral dataset containing only labelled spectra from normal and tumour tissue was created. The data were then processed using three different supervised learning algorithms: Support Vector Machines, Artificial Neural Networks and Random Forests. The capabilities of discriminating between normal and tumour tissue have been evaluated in three different scenarios, where the inter-patient variability of data was or not taken into account. The results achieved in this research study are promising, showing that it is possible to distinguish between normal and tumour tissue exclusively attending to the spectral signature of tissue.**

Keywords - Brain cancer detection, Hyperspectral imaging, Data mining, Supervised Learning

I. INTRODUCTION

Hyperspectral Imaging (HSI) is a technology that combines both spectroscopy and digital imaging, measuring hundreds of narrow bands from the electromagnetic spectrum. Each material has its own interaction with radiation, which can be measured either by using the reflectance or the absorbance values. The response to different wavelengths for a single material is called spectral signature, which allows the discrimination between different types of materials. Although HSI has been widely used in Remote Sensing, it is an emerging technology for clinical diagnosis. Some studies have proven that interaction between electromagnetic radiation and tissue carries useful information for diagnosis proposals [1]. A variety of studies shows that HSI is a helpful tool in the diagnosis of several cancer diseases. Some studies about prostate [2], ovaries [3], breast [4], and tongue [5] cancer detection using HSI have been recently published.

In this study three different supervised learning algorithms are employed in order to automatically discriminate between normal and tumour tissues in pathological slides. The selected algorithms are Support Vector Machines (SVM), Artificial Neural Networks (ANN) and Random Forests (RF) [6][7].

II. MATERIALS AND METHODS

A. Biological Samples

The dataset employed in this study have been previously acquired in another research work [8]. The biological samples used in this research work consist in biopsies from human brain tissue resected during surgery. These biopsies have followed a histological processing, whereby tissue specimens are prepared for sectioning, staining and diagnosis. In this study, 4 different patients were analysed, and 13 diagnosed pathology slides were available. These pathological slides were provided by the Anatomy Pathology department of Doctor Negrín Hospital, at Las Palmas of Gran Canaria.

The hyperspectral capture system consists on a hyperspectral camera coupled with a microscope. This system works in the VNIR spectral range (from 400 nm to 1000 nm) with a spectral resolution of 3nm

B. Experimental description

In order to validate supervised classification algorithms for discriminating between healthy and tumour tissue, three different case studies (CSs) have been proposed. This approaches differs in which patients are included as subject of study:

1) *CS1:* Each patient is processed individually. The inter-patient variability is not taken into account.

2) *CS2:* All data from all patients ir merged in a single dataset, and the processing is performed over all available samples.

3) *CS3:* Data from a new patient is classifier using a model that has been trained using the samples from the other patients.

C. Processing framework

The proposed processing framework is based on a typical supervised classification scheme. Although it has been proven that combining both the spatial and spectral features of the hyperspectral images can improve the accuracy in the predictions, in this research work only the spectral characteristics of the data are taken into account. This way, the inputs of the classifier are the measured spectral signature from healthy and tumour pixels.

Figure 1: Processing framework

This processing framework is based on three major stages:

1) Preprocessing: Aims to compensate the undesirable effects caused by the capture system and the enviromental conditions in the acquired hyprecubes.

2) Supervised classification: Three different algorithms have been employed: SVM, ANN and RF.

3) Model Evaluation: The performance of each classifier have been assesed using *10-fold cross-validation* (CS1 and CS2) and *hold-out* (CS3). Overall accuracy, sensitivity and specificity have been selected as metrics for measuring the model performance.

III. EXPERIMENTAL RESULTS

This section presents the results achieved when applying the previously-described supervised classification framework. These results consist in the performance estimation of each classifier in each Case Study.

For the SVM classifier, two different set ups have been tested: a linear kernel classifier (C1) and a Gaussian (Radial Basis Function, RBF) kernel (C2). Regarding the topology of the neural network (C3), several neural networks architectures varying the number of hidden layers, the number of neurons inside this layers and the activation function selected for each layer have been tested. The experimental results shown that the best performance is obtained using a multilayer neural network with a single hidden layer composed by 16 neurons, using a logistic activation function for this layer. A hyperbolic tangent sigmoid activation function has been selected for the output layer. Finally, an ensemble of 50 different classification trees composes the Random Forest (C4) configuration.

The classification results obtained in this research work are shown on Tables I, II and III.

TABLE I: CLASSIFICATION RESULTS – CS1

Classifier	**Patient number**	**Overall Accuracy**	**Sensitivity**	**Specificity**
C1	1	99.04%	99.29%	98.76%
	2	98.48%	98.75%	98.20%
	3	99.67%	99.87%	99.52%
C2	1	97.34%	97.56%	97.10%
	2	97.18%	97.47%	96.89%
	3	98.78%	99.81%	98.06%
C3	1	99.17%	99.13%	99.20%
	2	99.95%	99.96%	99.94%
	3	99.82%	99.90%	99.76%
C4	1	98.77%	98.67%	98.88%
	2	99.66%	99.73%	99.58%
	3	99.36%	99.77%	99.07%

TABLE II: CLASSIFICATION RESULTS – CS2

Classifier	**Overall Accuracy**	**Sensitivity**	**Specificity**
C1	94.46%	95.15%	93.66%
C2	92.78%	94.55%	90.83%
C3	98.20%	98.72%	97.61%
C4	97.91%	98.06%	97.75%

TABLE III: CLASSIFICATION RESULTS – CS3

Classifier	**Patient number**	**Overall Accuracy**	**Sensitivity**	**Specificity**
C1	1	81.85%	86.94%	77.57%
	2	64.64%	75.96%	59.66%
	3	68.92%	59.33%	83.87%
	4	10.74%	10.74%	-
C2	1	73.86%	79.19%	69.65%
	2	61.19%	68.46%	57.51%
	3	58.44%	50.58%	83.52%
	4	94.50%	94.50%	-
C3	1	48.15%	50.03%	46.19%
	2	47.26%	47.80%	46.99%
	3	33.02%	15.23%	42.72%
	4	99.20%	99.20%	-
C4	1	47.75%	49.60%	46.13%
	2	38.50%	39.53%	37.51%
	3	41.29%	39.27%	47.62%
	4	88.04%	88.40%	-

IV. CONCLUSIONS

This research work presents a proof of concept of using hyperspectral images for detecting brain tumour tissue in pathological slides. Promising results are obtained in CS1 and CS2, showing a good discrimination between healthy and tumour tissue with high specificity and sensitivity values. In CS3 results are not as accurate. This can be caused by the low number of patients (only three) that are used to train the classifier. In future studies, the number of patient should be increased. This kind of tools can help pathologist to analyse the slides without spending a long time for the examination of each sample

REFERENCES

[1] Lu, Guolan, and Baowei Fei. "Medical hyperspectral imaging: a review." Journal of biomedical optics 19.1 (2014): 010901-010901

[2] Akbari, Hamed, et al. "Hyperspectral imaging and quantitative analysis for prostate cancer detection." Journal of Biomedical Optics 17.7 (2012): 0760051-07600510.

[3] Akbari, Hamed, et al. "Cancer detection using infrared hyperspectral imaging."Cancer science 102.4 (2011): 852-857.

[4] Panasyuk, Svetlana V., et al. "Medical hyperspectral imaging to facilitate residual tumor identification during surgery." Cancer biology & therapy 6.3 (2007): 439-446.

[5] Liu, Zhi, Hongjun Wang, and Qingli Li. "Tongue tumor detection in medical hyperspectral images." Sensors 12.1 (2011): 162-174.

[6] Kecman, Vojislav. Learning and soft computing: support vector machines, neural networks, and fuzzy logic models. MIT press, 2001.

[7] Haykin, Simon S., et al. Neural networks and learning machines. Vol. 3. Upper Saddle River: Pearson Education, 2009.

[8] Ortega, S., Callico, G. M., Plaza, M. L., Camacho, R., Fabelo, H., & Sarmiento, R. Hyperspectral database of pathological in-vitro human brain samples to detect carcinogenic tissues. In 2016 IEEE 13th International Symposium on Biomedical Imaging (ISBI) (pp. 369-372). IEEE.