Evaluation of Band Selection Techniques in the Classification of Hyperspectral Images of Brain Tumors

Beatriz Martínez Vega, Gustavo Marrero Callicó, Himar Fabelo Gómez and Samuel Ortega Sarmiento
Institute for Applied Microelectronics (IUMA), University of Las Palmas Gran Canaria (ULPGC), Spain
{bmartinez, gustavo, hfabelo, sortega}@iuma.ulpgc.es

***Abstract*—Hyperspectral imaging (HSI) is an emerging technology in the medical area. HSI is a non-contact, non-ionizing, and label-free imaging modality that can assist neurosurgeons during brain tumor resection without using any contrast agent. This paper describes different methodologies to identify the most representative bands in HS images used for brain tumor detection. This selection process was carried out through different optimization algorithms, specifically *Genetic Algorithm* (GA) and *Particle Swarm Optimization* (PSO). For the evaluation of the selected bands, the supervised *Support Vector Machine* (SVM) classifier was employed.**

Keywords- Hyperspectral imaging; Genetic Algorithm; Particle Swarm Optimization; Support Vector Machines; Feature selection; Machine learning.

I. INTRODUCTION

Hyperspectral imaging (HSI) is a technology that combines conventional imaging and spectroscopy to obtain simultaneously the spatial and the spectral information from an object [1]. Hyperspectral (HS) images provide abundant information that covers hundreds of spectral bands for each pixel of the image [2]. Each pixel contains an almost continuous spectrum (radiance, reflectance or absorbance), acting as a fingerprint (the *so-called* spectral signature) that can be used to characterize the chemical composition of that particular pixel [3].

HSI is characterized by "the curse of dimensionality" [4], having high dimensionality due to the rich amount of data that they contain. However, this fact causes that the computing time required to process the data is extremely high and also, in some cases, the processed data contain redundant information. Thus, it is necessary to employ processing algorithms that reduce the dimensionality of the HS data without losing the relevant information. This dimensional reduction process consists in the transformation of the data, characterized by their high dimensionality, into a significant representation of such data in a reduced dimension. There are two types of methods that allow this transformation: feature selection [4] and feature extraction [5]. In this research work, after an exhaustive study where dimensional reduction was applied to HSI [6], [7], it was decided to use the feature selection methodology. The optimization algorithms employed to find the most relevant bands were Genetic Algorithm (GA), Particle Swarm Optimization (PSO).

II. METHODOLOGY

The HS database used in this research work belongs to the European HELICoiD project [8], [9]. This database is composed by HS images of human brain tissue captured during neurosurgical procedures with a customized intraoperative HS acquisition system. In the next sections, the different proposed approaches are explained.

A. Basic Processing Framework (BPF)

The first proposed processing framework has the goal to evaluate the results obtained with the band selection algorithm (GA and PSO) when employing the entire labeled dataset for the training of the SVM classifier. The labeled database was divided into training and test data performing a leave-one-patient-out cross-validation. Figure 1 shows the detailed procedure. First, the training data was employed in the optimization algorithm, where the initial bands to be used for the classification are randomly selected. After this band selection procedure, a classification model is generated and evaluated with the test dataset, obtaining a classification result that is evaluated using the $OA_{Penalized}$ FoM and $FoM_{Penalized}$ metrics. The procedures of the metrics are shown in Equation 1, Equation 2 and Equation 3 respectively. The value of this metric is stored and then, the procedure is iteratively executed using other bands selected by the optimization algorithm. The algorithm is executed until performs all possible combinations, returning the best metric, or when after a high number of iterations, the metric remains constant. Once the algorithm finishes the execution, it returns the best metric and the identification of the optimal bands to obtain this result.

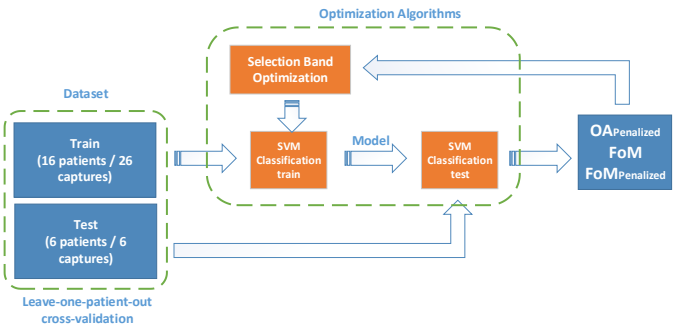


Figure 1. Flow diagram of the BPF for the identification of the most representative bands for the GA and PSO algorithms.

$$OA_{Penalized} = 1 - \frac{OA}{1+\frac{\lambda}{\lambda_{max}}} \quad (1)$$

$$FoM = \frac{1}{2} \cdot \left(\sum_{\substack{i,j \\ i<j}}^{n} \frac{ACC_i + ACC_j}{|ACC_i - ACC_j| + 1} \right) \cdot \binom{n}{2}^{-1} \quad (2)$$

$$FoM_{Penalized} = 1 - \frac{FoM}{1 + \frac{\lambda}{\lambda_{max}}} \quad (3)$$

B. *Optimized Processing Framework (OPF)*

After performing some experiments with the BPF, it was observed that the execution time was really high. Taking into account that the number of generations performed was repeated per each test image, it was necessary to find some techniques that allow reducing the execution time. In order to solve these problems, it was decided to use only 4,000 pixels (1,000 pixels per class) for training the SVM classifier, thus, balancing all classes and dramatically reducing the size of the training database (from ~200,000 to 4,000 pixels). This procedure was done through a methodology based on the *K-means* clustering algorithm [10]. Figure 2 shows the steps followed to obtain the optimized database, in order to eliminate redundant information and also to balance the classes of the training dataset. This modification was applied in the training dataset and followed the same structure as that observed in Figure 1.

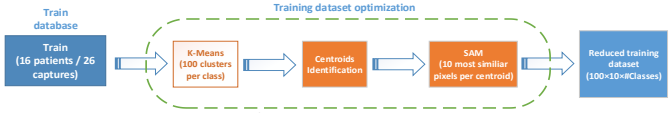


Figure 2. Flow diagram of the training dataset optimization algorithm.

C. *Evaluation methodology for the BPFand OPF*

Once executed the optimization algorithms in the BPF and OPF methods, the optimal selected bands for each test image were identified. Then, the evaluation metrics were computed. The procedure is the one shown in Figure 3. The evaluation metrics used were ***overall accuracy***, ***sensitivity***, ***specificity*** and ***Matthews Correlation Coefficient (MCC)*** for the quantitative evaluation and ***the classification map*** obtained after processing the entire HS cube of the test image for the qualitative evaluation.

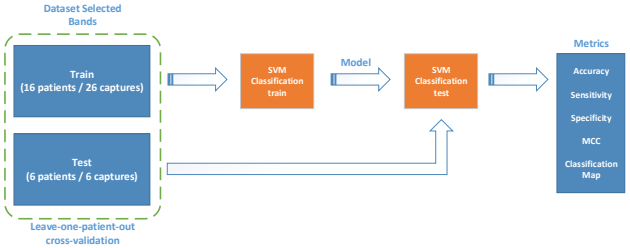


Figure 3. Flow diagram of the evaluation methodology.

III. RESULTS

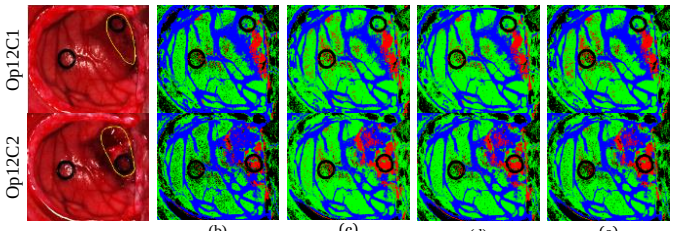
After obtaining the bands selected for each image with the different methodologies, the quality of the bands was evaluated. The optimization algorithms try to find the best results of the $OA_{Penalized}$, FoM and $FoM_{Penalized}$ metrics, performing different combinations at the time of selecting the bands. These experiments were divided into the following structure: 1) Evaluation of the optimal bands for each test image obtained with the GA and PSO algorithms using the BPF (entire dataset) and the OPF (reduced dataset) with the $OA_{Penalized}$, FoM and $FoM_{Penalized}$ metrics; 2) Evaluation of the optimal bands for the complete test image database using the OPF.

The last study was carried out with the best case obtained in the previous step. The best cases were: OPF-GA-FoM, OPF-GA-$FoM_{Penalized}$ and OPF-PSO-FoM. The metric $OA_{Penalized}$ gave poor results, then they were discarded for the rest of the research, and focused on the other metrics.

Table 1 shows the average and standard deviation of the evaluation metrics. It is observed that the results of the **OA** are similar, giving the best case to the OPF-GA-$FoM_{Penalized}$ with 77.9% and the worst case to the OPF-PSO-$FoM_{Penalized}$ with 75.9%. In terms of **sensitivity**, the results of the first two cases remain constant, the only more pronounced difference is that in the tumor class, the OPF-GA-FoM achieves 52.9% and the OPF-GA-$FoM_{Penalized}$ reaches 52.7%. The last case has worse results in the normal and tumor tissue classes, with 77.8% and 47.1%, respectively. Nevertheless, the **specificity** remains constant for all cases. Finally, the **MCC** metric shows that the class with the worst result is the tumor class, with the second case having the best result, with 44.4% and the third case having the worst result with 38.4%. Regarding the **classification maps** shown in Figure 4, it can be observed that the classification maps of the OPF-GA-$FoM_{Penalized}$ (Figure 4.d) present less false positives of the tumor class (tumor area surrounded by yellow line in the RGB representation). Thus, taking into account all the results, the best case obtained was using the GA with the $FoM_{Penalized}$ (48 bands) with the coincident bands in at least one image.

Table 1. Average and standard deviation results of the OPF best cases.

Technique (#bands)	OA AVG (STD) %	Sensitivity AVG (STD)%				Specificity AVG (STD)%				MCC AVG (STD)%			
		NT	TT	HT	BG	NT	TT	HT	BG	NT	TT	HT	BG
GA-FoM (27)	77.3 (16.5)	84.4 (15.2)	52.9 (31.1)	83.4 (21.7)	89.8 (21.1)	87.4 (10.4)	93.4 (8.2)	96.9 (5.0)	85.7 (20.7)	69.1 (19.1)	41.5 (30.9)	82.9 (16.6)	71.3 (22.1)
GA-FoM$_P$ (48)	77.9 (17.0)	85.1 (17.6)	52.7 (29.8)	83.5 (20.9)	92.5 (14.2)	87.3 (12.2)	94.6 (8.3)	96.7 (5.1)	85.3 (18.0)	69.0 (21.0)	44.4 (30.9)	82.4 (16.7)	72.7 (20.2)
PSO-FoM$_P$ (62)	75.9 (17.3)	77.8 (22.2)	47.1 (30.6)	84.5 (20.1)	93.1 (13.0)	87.0 (13.3)	92.9 (11.5)	95.9 (5.7)	84.6 (17.2)	62.5 (19.5)	38.4 (27.7)	82.1 (17.4)	72.2 (19.9)



(a) Synthetic RGB (b) Reference (128) (c) GA-FoM (27) (d) GA-$FoM_{Penalized}$ (48) (e) PSO-$FoM_{Penalized}$ (62)

Figure 4. OPF best cases classifications maps. (a) Synthetic RGB images with a yellow line delineating the area tumor. (b) OPF Reference. (c) OPF-GA-FoM. (d) OPF-GA-$FoM_{Penalized}$. (e) OPF-PSO-$FoM_{Penalized}$.

IV. CONCLUSIONS

This work employed HS human brain cancer images obtained intraoperatively for the development a processing framework able to obtain the most representative spectral bands that allows an accurate classification of the tumor. The obtained results demonstrate that using only 48 bands the classifier is able to improve the classification results. Figure 5 shows the spectral signatures of the healthy tissue (green color), the tumor tissue (red color), the hypervascularized tissue (blue color) and the final bands that were selected (black spots). It can be seen that the selected bands are grouped into small regions throughout the spectral signature. The most important regions are: 440-466 nm,

498-509 nm, 556-575 nm, 593-615 nm, 636-666 nm, 698-731 nm and 884-902 nm.

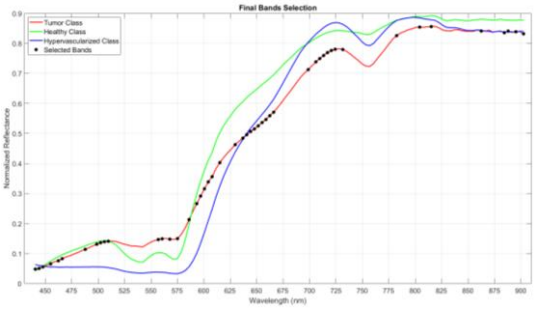


Figure 5. Normalized average spectral signature and final selected bands.

REFERENCES

[1] A. Gowen, C. Odonnell, P. Cullen, G. Downey, and J. Frias, "Hyperspectral imaging-an emerging process analytical tool for food quality and safety control," *Trends Food Sci. Technol.*, vol. 18, no. 12, pp. 590–598, 2007.

[2] J. Wang, X. Wang, K. Zhang, K. Madani, and C. Sabourin, "Morphological band selection for hyperspectral imagery," *IEEE Geosci. Remote Sens. Lett.*, vol. 15, no. 8, pp. 1259–1263, 2018.

[3] P. Ghamisi, J. Plaza, Y. Chen, J. Li, and A. J. Plaza, "Advanced Spectral Classifiers for Hyperspectral Images: A review," *IEEE Geosci. Remote Sens. Mag.*, vol. 5, no. 1, pp. 8–32, 2017.

[4] H. Wan, J. Sun, R. Yang, L. Su, and X. Zhao, "Representative band selection for hyperspectral image classification," *J. Vis. Commun. Image Represent.*, vol. 48, pp. 396–403, 2017.

[5] D. Ravi, H. Fabelo, G. M. Callico, and G. Yang, "Manifold Embedding and Semantic Segmentation for Intraoperative Guidance with Hyperspectral Brain Imaging," *IEEE Trans. Med. Imaging*, 2017.

[6] P. Yu, M. Huang, M. Zhang, and B. Yang, "Optimal Wavelength Selection for Hyperspectral Imaging Evaluation on Vegetable Soybean Moisture Content during Drying," *Appl. Sci.*, vol. 9, no. 2, p. 331, 2019.

[7] B. Ganapathysubramanian, K. Nagasubramanian, S. Jones, A. Singh, S. Sarkar, and A. K. Singh, "Hyperspectral band selection using genetic algorithm and support vector machines for early identification of charcoal rot disease in soybean," *Plant Methods*, vol. 14, no. 1, pp. 1–13, 2018.

[8] H. Fabelo *et al.*, "An Intraoperative Visualization System Using Hyperspectral Imaging to Aid in Brain Tumor Delineation," *Sensors*, vol. 18, no. 2, 2018.

[9] H. Fabelo *et al.*, "Spatio-spectral classification of hyperspectral images for brain cancer detection during surgical operations," *PLoS One*, vol. 13, no. 3, pp. 1–27, Mar. 2018.

[10] T. Kanungo, D. M. Mount, N. S. Netanyahu, Christine D. Piatko, R. Silverman, and A. Y. Wu, "An Efficient k-Means Clustering Algorithm: Analysis and Implementation Tapas," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 7, pp. 881–892, 2002.