Multiclass Brain Tumor Classification Using Hyperspectral Imaging and Supervised Machine Learning

Luisa Ruiz*, Alberto Martín*, Gemma Urbanos*, Marta Villanueva*, Jaime Sancho*, Gonzalo Rosa*,
Manuel Villa*, Miguel Chavarrías*, Ángel Pérez†, Eduardo Juarez*, Alfonso Lagares†, César Sanz*

*Research Center on Software Technologies and Multimedia Systems. Universidad Politécnica de Madrid (UPM)
†Department of Neurosurgery, Hospital Universitario 12 de Octubre; Universidad Complutense de Madrid;
Instituto de Investigación Sanitaria Hospital 12 de Octubre (imas12), Madrid, España
{luisa.rruiz, gemma.urbanos, marta.villanueva.torres, jaime.sancho, miguel.chavarrias, eduardo.juarez, cesar.sanz}@upm.es
{a.martinp, gonzalo.rosa.olmeda, manuel.villa.romero}@alumnos.upm.es {apnunez, alfonsolagares}@salud.madrid.org

Abstract—Hyperspectral Imaging (HSI) can be used as a non invasive medical diagnostic method when used in combination with Machine Learning (ML) algorithms. The significant captured data in HSI can be useful for classifying different types of brain tissues, since they gather reflectance values from different band widths below and beyond the visual spectrum. This allows ML algorithms like Support Vector Machines (SVM) and Random Forest (RF) to classify brain tissues such as tumors. Predicted results can be used to create visualizations and support neurosurgeons before injuring any tissue. This way neurosurgeons can be more precise, reducing any possible damages on healthy tissues. In this work, a proposal for the classification of in-vivo brain hyperspectral images using SVM and RF classifiers is presented. A total of four hyperspectral images from four different patients with glioblastoma grade IV (GBM) brain tumor have been selected to train models and, therefore, classify them. Five different classes have been defined during experiments: healthy tissue, tumor, venous blood vessel, arterial blood vessel and dura mater. Results obtained suggest that SVM usually performs better than RF, generally achieving up to 97% of mean accuracy (ACC). However, RF performance had better results than SVM when classifying images used during training, obtaining almost 100% of mean ACC for all 5 classes described. This study shows the robustness of SVM and the potential of RF for real-time brain cancer detection.

Index Terms—Hyperspectral, Machine Learning, SVM, Random Forest, Medical, classification

I. INTRODUCTION

In recent years, technological advances and their applications in the medical sector have led to improvements in the diagnosis and treatment of many diseases. Nonetheless, cancer continues to be one of the major causes of morbidity and mortality worldwide [1]. In this sense, approximately 2% of cancer deaths are caused by brain tumors. Glioma is the most common cerebral tumor in adults, particularly glioblastoma multiforme, which is an aggressive high-grade brain tumor [2]. Surgery remains an inevitable phase for treating many brain tumors, albeit all risks involved and its high complexity. Nowadays, new existing technologies have been allowing more complete resections with less morbidity. However, resection is not always possible and depends on tumor characteristics. In case of glioblastomas, their high capacity to infiltrate healthy brain tissue makes it difficult to distinguish between malignant tumor borders and healthy tissue [3]. Hyperspectral imaging (HSI) can be used to find a solution to this problem, providing a tool to assist neurosurgeons accurately and help them distinguish between brain tumor and healthy tissue during surgery. HSI is an appropriate technique for medical applications since it requires no contact with the patient, it is non-ionizing as well as non-invasive [4]. This technology collects spectral information with more detail than the human eye, processing information below and beyond the visible spectrum. The goal is to obtain spectral signatures for each pixel using three-dimensional datasets with spatial and spectral information, well known as hyperspectral cubes. Extracting spectral signatures for different types of in-vivo brain tissues helps to construct ground truth maps allowing Machine Learning (ML) algorithms to classify new brain tissues [5]. Ground truth maps, generated with neurosurgeon support, contain reference information from the original image taken during surgery to indicate, with a high level of accuracy, different type of tissue for every pixel. This information can be used to train models, classify images and validate prediction results, which can then be used to create visualizations in the form of classification maps, providing an accurate tool for the neurosurgeon to excise precisely the brain tumor.

In this work, a multi-class brain tumor classification method based on HS imaging and supervised ML techniques is presented. The paper is organized as follows: Section II gives background knowledge while Section III methods and materials used are outlined. Then, Section IV describes the test bench implemented to verify the proposal whilst obtained results and discussions are presented in Section V. Finally, Section VI concludes the paper with the conclusions and the undergoing work.

II. Background

A. Hyperspectral Imaging

Snapshot hyperspectral imaging is a modality of HSI characterized by capturing hyperspectral images in a single time instant without moving the camera. These cameras include a multi-band sensor matrix allowing them real-time HSI cubes acquisition, providing a useful tool in numerous fields [6]. Several studies have also considered the use of HSI for medical applications to analyze and diagnose different kinds of diseases, in particular, cancer detection. Successful studies about breast cancer borders detection have also been presented lately by Kho et al. using HSI to help resections during surgery [7]. Others kind of cancer, melanoma [8], and tongue cancers [9], have been detected using this technology in other studies. In case of brain cancer, other works have been developed supporting the effectiveness of using HSI for tumor detection. HELICoiD project [5], has been developed by Fabelo et al., in which a HSI line scan system based, able to capture 826 spectral bands, for detecting brain tumors during surgery is being refined. Results obtained in the HELICoiD project present great achievements on using HSI and ML techniques for real-time brain cancer detection.

B. Machine Learning

ML is a sub-field of Artificial Intelligence (AI) which is getting lot of attention in recent years. ML algorithms aspire to mimic human brains by learning from observation coming from huge volumes of data. This way, computing models can be trained to analyze and predict new data. Trained models can be used for classification or regression purposes, depending on the problem to be solved. For brain tumor detection, classification ML models have been normally used in order to identify a finite number of brain tissues. When combining data gathered using HSI with ML techniques, predictive models can be trained to accurately discriminate between different brain tissues, using their spectral features to make predictions [5].

Various ML techniques have been used nowadays, including supervised learning, unsupervised learning, semi-supervised learning or reinforcement learning. But for the purpose of this study, supervised learning was the option chosen. The main difference of supervised learning is that every observation of the data is labeled. With classified data, models can be trained since every input has its correspondent output. When they analyze to generalize from the training data in a logical way to unseen cases, algorithms correctly determine class labels for unseen observations. Within this ML technique multiple algorithms can be used to train models. SVM and RF have been used in this work and are described in more detail below:

1) Support Vector Machines: This algorithm maps input vectors into a high dimensional feature space, where a maximal margin hyperplane is build [10]. Different kernel functions can be specified for the decision function depending on the problem. They could be used to implement a multi-class classification on a dataset, providing in advance a subset of labeled data needed for model training.

2) Random Forest: This algorithm is based on an ensemble of decision trees grown with bootstrap samples of the training data [11]. For each tree of the forest, random subsets of features are used at each decision split. When predicting new data, labels are assigned by the forest depending on the most predicted label from all trees of the ensemble.

C. Cross-Validation

Cross-validation prevents overfitting and helps to determine which model will generalize best to a different dataset. Two main types of cross-validation methods can be found, exhaustive cross-validation and non-exhaustive cross-validation:

1) Exhaustive cross-validation: Methods which learn and test every possible way to split the dataset into a training and a validation set; leave-p-out cross-validation and leave-one-out cross-validation are examples of exhaustive cross-validation.

2) Non-exhaustive cross-validation: Methods which will not compute every possible way of separating the dataset. K-fold cross-validation and Holdout method are examples of non-exhaustive cross-validation.

III. Methods and Materials

This section describes which methods have been used to acquire and process data before training ML models. A description about model selection have been included as well as the metrics used to evaluate the performance of the classifiers.

A. Acquisition System

The system used to acquire data is composed by a hyperspectral camera and a light source. The camera is based on a snapshot mosaic model [12], with a scan capacity of 170 cubes/second. Its spatial resolution is 217x409 pixels and the spectral information is distributed within 25 bands along the spectral range between 655 nm and 975 nm. This model is very convenient for medical applications due to its small size and weight, 26x26x26 mm and 32 g. Illumination in the scene is provided by a halogen light source [13] with two fiber optic light guides with 180 cm length and 1.2 cm diameter.

B. Data pre-processing

Before using HSI gathered data, a pre-processing have been applied for every image taken in order to homogenize the spectral signatures of each pixel. The pre-processed stage consists of the following 4 phases:

1) Hyperspectral cube generation: Snapshot raw images consisting of a repetitive mosaic of blocks of 25 (5x5) wavelengths are converted into cubes of 25 bands.

2) Calibration: White and dark reference images are taken with the hyperspectral acquisition system before a surgery begins in the same lighting conditions than those of in-vivo brain resections. White reference images are obtained with a reference reflectance target [14] while dark reference images are obtained covering the camera lens. A calibrated image per band, Ic, is obtained applying Equation 1:

$$I_c = \frac{I - D}{W - D} \qquad (1)$$

where I is a single band of a captured hyperspectral cube, D is the corresponding dark reference for that band and W is the white reference obtained for the same wavelength.

3) Spectral Correction: The wavelength response curves of the IMEC snapshot sensor used, have high sensitivity which produces overlap (cross talks) with neighbor wavelength curves. Due to these non-negligible secondary lobes of wavelength filters, a spectral correction process is required [15]. A spectral-corrected image has been obtained multiplying the signal with a correction matrix, as shown in Equation 2:

$$I_{sc} = I_c \times SCM \qquad (2)$$

where I_c is a single band of a calibrated hyperspectral cube and SCM is the correction matrix (25x25), in which each row is a set of virtual band spectral correction coefficients. The correction matrix have been provided by the manufacturer.

4) Normalization: Illumination changes might lead to differences in brightness levels between similar tissues. Thus, data normalization is necessary to reduce this effect. Normalization coefficients are computed following Equation 3. These coefficients are the root mean square value (RMS) of the spectral signatures. As can be seen in Equation 4, they are employed to transform the spectral corrected cube into a normalized one. Note that the shape of the spectral signatures is preserved in this transformation.

$$c\,[i,j] = \sqrt{\frac{\sum_{k=1}^{B}\left(I_{sc}\,[i,j,k]\right)^2}{B}} \qquad (3)$$

$$I_{Norm}\,[i,j,k] = \frac{I_{sc}\,[i,j,k]}{c\,[i,j]} \qquad (4)$$

where I_{sc} is the spectral corrected cube with dimensions $R \times C \times B$ ($Rows \times Columns \times Bands$).

C. Classification algorithms

The following supervised learning algorithms have been used to train classification models:

1) Support Vector Machines: A python scikit-learn library [16] has been used to train SVM models with a linear kernel.

2) Random Forest: Python scikit-learn implementation [16] has been used to train classification RF models.

D. Model Selection and Metrics

SVM and RF models have been selected using double 5-fold cross-validation and triple 5-fold cross validation, depending on the experiment. Models have been trained with 80% of the available ground-truth data and assessed with the remaining 20%. Best trained models with both supervised ML algorithms have been selected evaluating the overall accuracy (OACC) of the classifiers during the 5-fold cross-validation for every class, using Equation 5:

$$OACC = \frac{\text{Number of correct predictions}}{\text{Total number of predictions}} \qquad (5)$$

To evaluate classifier performances, three metrics have been obtained from confusion matrices: accuracy (ACC), sensitivity

TABLE I
DETAILED NUMBER OF PIXELS FOR EVERY CLASS AND PATIENT IMAGE USED IN THE DATASET OF THE STUDY.

Class	ID029	ID030	ID034	ID035
Healthy-Tissue	3752	2587	1186	3864
Tumor GBM	64	2737	1464	1389
Venous-Blood	98	487	181	837
Arterial-Blood	11	381	176	58
Dura-Mater	1599	1366	780	586
Total	**5524**	**7558**	**3787**	**6734**

(SEN) and specificity (SPE). ACC allows to determine how well models globally classify data. SEN indicates the ability of classifiers to predict correctly a class of interest, while SPE shows models capacity to properly differentiate all classes from the class of interest. These metrics have been calculated for each class following Equations 6, 7 and 8:

$$ACC = \frac{TP + TN}{TP + TN + FP + FN} \qquad (6)$$

$$SEN = \frac{TP}{TP + FN} \qquad (7)$$

$$SPE = \frac{TN}{TN + FP} \qquad (8)$$

where TP means true positives, TN true negatives, FP false positives and FN false negatives.

IV. TEST BENCH

This sections presents the database used for the verification process as well as the conducted experiments.

A. Database

Hyperspectral images have been obtained during real surgery procedures inside the operating theater, at the University Hospital 12 de Octubre of Madrid (Spain), using the acquisition system described above. All images in Table I have been captured from in-vivo brain surfaces after craniotomy and durotomy. After acquiring the images and confirming tumor diagnostic with pathological analysis, neurosurgeons label different tissues of brain images to reference five different classes: healthy tissue, tumor, venous blood vessel, arterial blood vessel and dura mater. More details about the labelling tool used can be found in [5]. The aim is to create a ground-truth map for every hyperspectral image, allowing to train the supervised learning algorithms described above. For our analysis, images corresponding to four patients with glioblastoma grade IV tumor (GBM) have been used: ID0029C02, ID0030C02, ID0034C02 and ID0035C02. These patients have been identified with a unique reference number using the following format: ID patient number - C number of capture. For simplicity, in the rest of the document these patients will be identified as ID029, ID030, ID034 and ID035. The four captures have been taken with the same exposure time, 70 ms. Table I contains the number of pixels labelled for each class and patient.

B. Experiments

1) Experiment A: SVM and RF models have been trained for both ML algorithms using 80% and 20% of patients ground-truth as training and test data, respectively. The best SVM and RF models are determined using a triple cross-validation process to evaluate the best 80% of every patient. In addition, the entirety of each patient image has been predicted with the best models.

2) Experiment B: Four models have been trained for both algorithms with 3 patient images using 100% of their data. Each of these models have been selected after a double cross validation to make sure every pixel has been used as training and test data. Additionally, the remaining image not used during training has been used to make predictions.

V. RESULTS AND DISCUSSION

This section describes the obtained results from experiments A and B. RF results come from models trained with 100 trees. The number of trees has been selected as a trade-off between training time and overall accuracy.

Figure 1 shows the different pre-processed spectral signatures in the dataset. Different classes have been plotted with different colors, where the shaded area represents the standard deviation of the data. These signatures correspond to the average of ground-truth labelled pixels over the four selected patients. Each type of tissue has a characteristic signature which ML models leverage on to differentiate them out of the rest. However, there are discrepancies between the classes healthy tissue, tumor GBM, arterial blood, venous blood and dura mater within different patients. These differences can be represented numerically with the correlation matrix of tumor GBM class in Table II.

Patient ID029 does not have a good correlation with other patients. Although the pathology report confirms a GBM tumor, the region of tumor captured on brain surface has a different grade caused by its infiltrative nature. Patient ID030 also has a good correlation with patient ID034 and ID035, over 67% and 96% respectively. Best results have been obtained with patient ID034, which has a correlation over 60% with patient ID035. Classification results may be affected by these differences, especially in experiment B, where data of the patient to be classified have not been included in the model.

TABLE II
CORRELATION MATRIX OF TUMOR GBM PRE-PROCESSED SPECTRAL SIGNATURE BETWEEN PATIENTS.

	ID029	ID030	ID034	ID035
ID029	1	-0.358	0.286	-0.302
ID030	-0.358	1	0.668	0.96
ID034	0.286	0.668	1	0.609
ID035	-0.302	0.960	0.609	1

A. Experiment A

After selecting models with the best OAAC from the triple cross validation, mean accuracy for all four patients and all 5
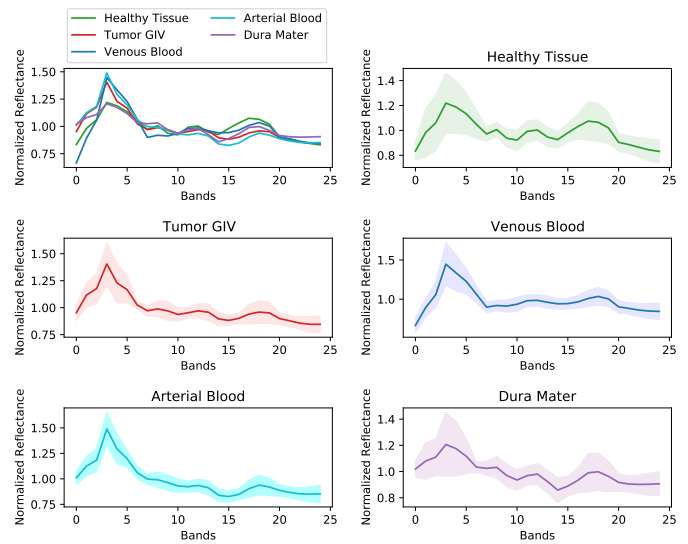

Fig. 1. Mean and standard desviation of the pre-processed spectral signatures of each class in the dataset.

TABLE III
OVERALL ACCURACY OF SVM AND RF AFTER CLASSIFYING THE FOUR PATIENT IMAGES FOR EXPERIMENT A.

	ID029	ID030	ID034	ID035
SVM	0,9756	0,9147	0,9480	0,9837
RF	0,9928	0,9880	0,9939	0,9994

classes can be seen in the top image of Figure 2. Both SVM and RF models have performed above 95% for each class. RF classification method presents better results than SVM in terms of performance. Effectively, as can be seen in Table III where the overall accuracy is shown, RF has achieved better results than SVM for all patients of experiment A.

The detailed performance of the RF model is shown in the middle and bottom images in Figure 2, where SEN and SPE metrics for tumor GBM class represent values over 95% in most cases, except for the SEN value of patient ID029. As can be seen in Table II patient ID029 does not correlate well with other patients, which may explain why the SEN value is lower than the rest of patients.

Additionally, classification maps for patient ID030 are presented in Figure 3 to ease the visualization of prediction results. The top left image is the actual grey scale image of the patient, while the top right is the Ground Truth for that grey scale image. In the bottom side, two classification maps acquired with SVM and RF models are presented. Small differences can be seen between both maps, but RF classifies better Arterial-Blood than SVM when comparing the right side of the Ground Truth with the classification maps.

B. Experiment B

In experiment B models with the best OACC, for both algorithms, have been selected after a double cross validation, providing a total of eight models, four per algorithm. Top
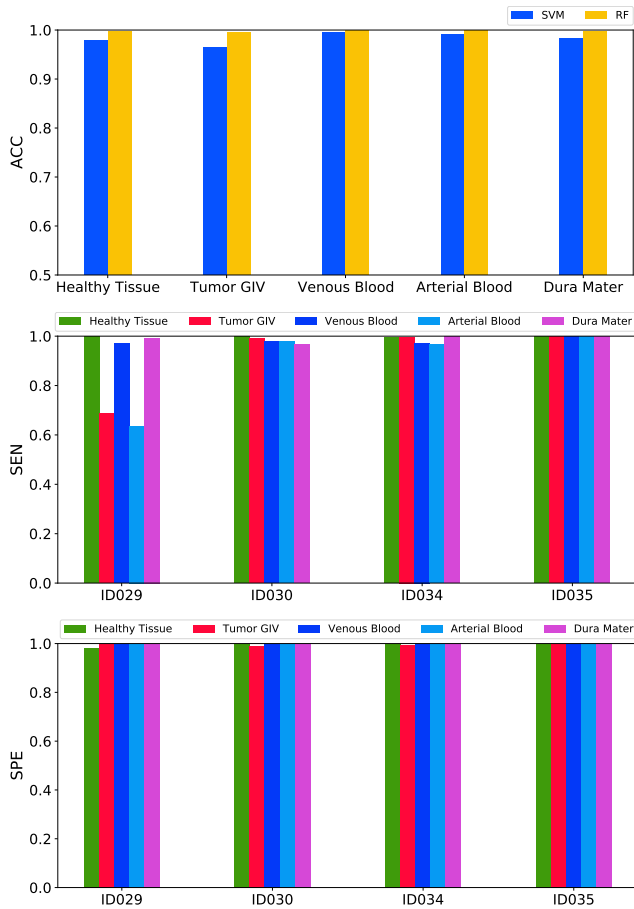
Fig. 2. Experiment A: Mean accuracy of all 5 classes for the four patients after classifying completely each patient image with the best SVM and RF model. Sensitivity and specificity of the tumor GBM class are also displayed for every patient with RF model.
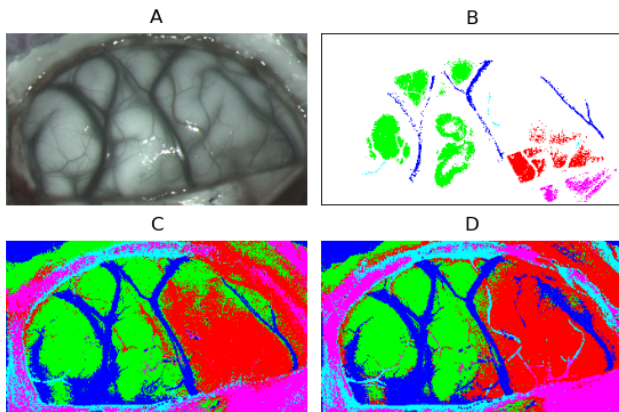

Fig. 3. Experiment A: result patient ID035. (A) Image, (B) Ground Truth map, (C) Classification Map of SVM, (D) Classification map of Random Forest. Green: healthy tissue, Red: Tumor GBM, Blue: Venous Blood Vessels, Cian: Arterial Blood Vessels, Pink: Dura Mater.

image in Figure 4 represents the mean ACC of every class including the four patients for the best SVM and RF models. As can be seen, SVM performs better than RF, where all classes range between 75% and up to 97%. On the other hand,

	ID029	**ID030**	**ID034**	**ID035**
SVM	0,8208	0,7935	0,7927	0,1298
RF	0,2341	0,5583	0,2656	0,0876

RF models have performed worse than SVM in all cases, with values ranging from 55% to 79%. This can also be seen in Table IV where the overall ACC is shown for SVM and RF models after classifying patient images whose pixels are not considered for training.

Since SVM outperforms RF in experiment B, the rest of the plots in Figure 4 only display SVM results. Middle image in Figure 4 shows the SEN of all four SVM models after predicting the patient image not used during training. Classification models for patient images ID030 and ID034 have globally performed better. As can be seen in Table IV, the SVM model trained with data from all patient images but patient ID029 has the best OACC score. However, as can be see in the middle plot of Figure 4, the SEN metric of the tumor class is dramatically low. Looking at Table II, patients ID030 and ID034 correlate well with other patients, explaining why the SEN value is better than the rest of patients.

As can be seen in the bottom plot of Figure 4, the SVM classifier predicts for patient ID035 mostly all data as arterial blood. This fact can be confirmed with the obtained confusion matrix of the patient. Looking at the confusion matrix after classifying patient ID035, most data from healthy tissue, tumor GBM tissue, dura mater and venous blood classes have been classified as arterial blood. When comparing the average spectral signatures per class of the training samples to those obtained from the ground-truth of patient ID035, a high correlation exist between the arterial blood class from the model and all classes from the patient. This explain why the classifier predicted mostly all patient image as arterial blood.

Visualizations of classification maps for patient ID030 are shown in Figure 5. The top left image is the actual grey scale image of the patient, while on the top right the ground truth for that grey scale image is displayed. Both images at the bottom show the classification maps acquired with the SVM and RF models respectively. SVM classified and delimited healthy tissue from the venous blood better than RF.

VI. CONCLUSION AND FUTURE WORK

The use of HSI and ML algorithms provides a non-invasive and effective method for cancer detection. In this paper, HSI and ML (SVM and RF algorithms) have been used to classify in-vivo brain tissues in four patients with glioblastoma grade IV tumor. Both algorithms have shown satisfactory results when classifying brain cancer from only 25 spectral bands captured by a snapshot camera. Results demonstrate that SVM behaves better than RF when predicting new data not used during training, reaching mean ACC values up to 97% for the different brain tissues. Contrarily, RF performance
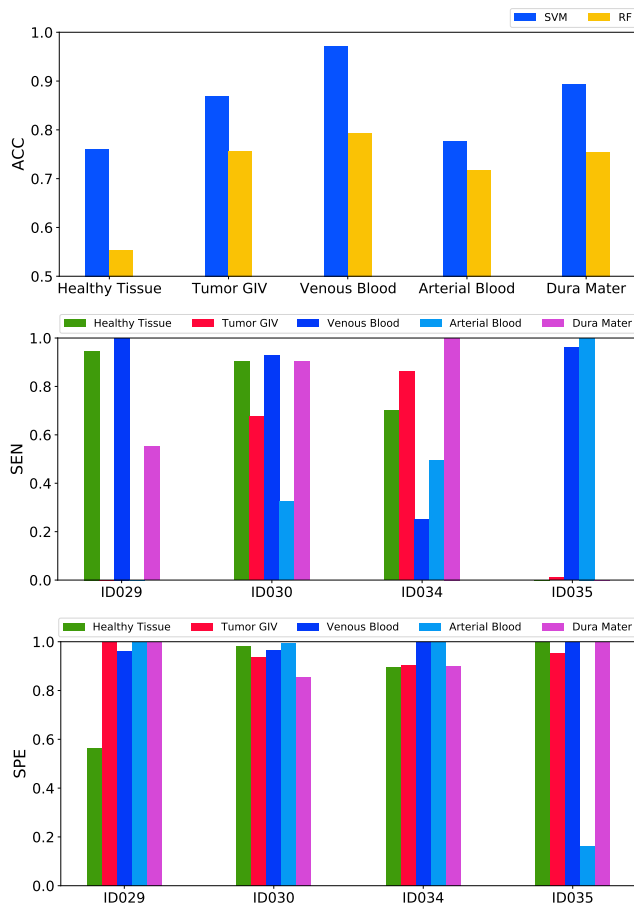
Fig. 4. Experiment B: Mean accuracy of all 5 classes for the four patients after classifying completely each patient image with the best SVM and RF model. Sensibility and specificity of each patient and class with SVM model.
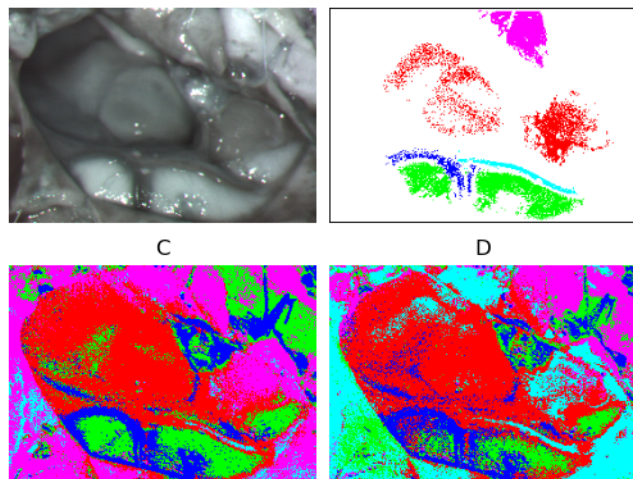

Fig. 5. Experiment B: result patient ID030. (A) Image, (B) Ground Truth map, (C) Classification Map of SVM, (D) Classification map of Random Forest. Green: healthy tissue, Red: tumor GBM, Blue: Venous Blood Vessels, Cian: Arterial Blood Vessels, Pink: Dura Mater.

is better than SVM when prediction is carried out on the same patient data employed for training. Results show that

mean values are above 99% for the different five classes. This study may suggest the use of HSI and ML for real-time brain cancer detection during surgery. In the future, the authors will continue working on improving ML algorithms for real-time classification while exploring Incremental Learning.

REFERENCES

[1] Jacques Ferlay et al. "Cancer incidence and mortality worldwide: Sources, methods and major patterns in GLOBOCAN 2012". In: *International Journal of Cancer* 136.5 (2015), E359–E386.

[2] Andrew H. Kaye. *Essential neurosurgery*. Malden, MA: Blackwell, 2006.

[3] Dr. Pedro Pérez Segura. *Tumores cerebrales*. Available at https://seom.org/info-sobre-el-cancer/tumores-cerebrales.

[4] Guolan Lu and Baowei Fei. "Medical hyperspectral imaging: A review". In: *Journal of biomedical optics* 19 (2014).

[5] Himar Fabelo et al. "Spatio-spectral classification of hyperspectral images for brain cancer detection during surgical operations". In: *PLOS ONE* 13 (2018).

[6] G Wong. "Snapshot hyperspectral imaging and practical applications". In: *Journal of Physics: Conference Series* 178 (2009).

[7] Esther Kho et al. "Hyperspectral imaging for resection margin assessment during cancer surgery". In: *Clinical cancer research* 25.12 (2019), pp. 3572–3580.

[8] Mihaela Antonina Calin et al. "Hyperspectral Imaging in the Medical Field: Present and Future". In: *Applied Spectroscopy Reviews* 49.6 (2014), pp. 435–447.

[9] Liu Zhi et al. "Classification of hyperspectral medical tongue images for tongue diagnosis". In: *Computerized Medical Imaging and Graphics* 31.8 (2007), pp. 672–678.

[10] Olivier Chapelle and Vladimir Vapnik. "Model Selection for Support Vector Machines". In: *MIT Press* (2000), pp. 230–236.

[11] L. Breiman. "Random Forests". In: *Machine Learning* 45 (2001), pp. 5–32.

[12] Ximea Corp. *MQ022HG-IM-SM5X5-NIR - XIMEA*. Available at https://www.ximea.com/en/products/hyperspectral-cameras-based-on-usb3-xispec/mq022hg-im-sm5x5-nir (2020/04/29).

[13] Fiber-lite. *Model Mi-LED: LED Fiber Optic Illuminator*. Available at https://www.setra.com/hubfs/mi-led-data-sheet.pdf (2020/04/29).

[14] Sphere Optics. *Zenith LiteTM Targets*. Available at http://sphereoptics.de/en/wp-content/uploads/sites/3/2014/03/SphereOptics-Ultralight-Targets-Zenith-Lite.pdf (2020/04/29).

[15] Ximea Corp. *xiSpec: USB 3.0 camera series*. Available at http://http://www.ximea.com/downloads/usb3/manuals/xiq_technical_manual.pdf (2020/04/29).

[16] *Python scikit-learn*. Available at https://scikit-learn.org/.