



Tumor Tissue Classification in Hyperspectral Histopathology Images Through Individual and Ensemble of Machine Learning Algorithms

Daia Natalia Martínez-Arboleyda¹(✉), Inés Alejandro Cruz-Guerrero¹,
and Daniel Ulises Campos-Delgado^{1,2}

¹ Facultad de Ciencias, Universidad Autónoma de San Luis Potosí, 78295 San Luis Potosí, San Luis Potosí, Mexico

a282863@alumnos.uaslp.mx

² Instituto de Investigación en Comunicación Óptica, Universidad Autónoma de San Luis Potosí, 78290 San Luis Potosí, Mexico

Abstract. Hyperspectral imaging (HSI) is a versatile modality that can provide a noninvasive tool for cancer detection, image-guided surgery, tissue classification, among other applications. In this work, we demonstrate the integration of hyperspectral imaging, spectral unmixing, and machine learning models to accurately classify tumor and non-tumor tissue from histopathology samples. The studied database contains 494 images from 13 different patients diagnosed with glioblastoma. Our approach is based on identifying characteristic spectral signatures for each hyperspectral image by spectral unmixing, and using them as an input feature vector for machine learning models: support vector machine, random forest, and a voting ensemble. The resulting average accuracy in our evaluation of four folds was 89.4% that improves the reference value of 85.5%, which was the best performance in the state-of-the-art. In this way, our proposed methodology shows promising results to assist in pathological analysis and provide support to healthcare professionals.

Keywords: Hyperspectral imaging · Machine Learning · Binary Classification · Glioblastoma · Spectral Unmixing

1 Introduction

In histopathology, digital imaging is a powerful tool for achieving a systematic evaluation of the studied samples [2]. Nonetheless, different factors affect the evaluations, such as sample preparation and pathologist expertise. In this sense, machine learning has recently been suggested in medical applications to avoid bias in diagnosis [4], and in particular for digital histopathology samples [5]. Unlike traditional images, which only work with three color channels (Red, Green, and Blue - RGB), hyperspectral (HS) images provide much more detailed spectral information about the object and its composition. Hyperspectral imaging (HSI) is an emerging imaging modality that provides a noninvasive tool in medical applications: cancer detection, image-guided surgery, tissue

classifications, etc. HSI is a hybrid modality that combines imaging and spectroscopy, by collecting spectral information at each pixel of a two-dimensional detector array, generates a three-dimensional dataset of spatial and spectral information, known as hypercube. In this way, each pixel in the HS image has a unique spectral signature, which refers to the pattern or distribution of light intensity across different wavelengths. Through analyzing these spectral signatures, these images allow the detection and identification of various pathological conditions [1]. HSI can be incorporated to improve diagnosis, as the intensity of the light transmitted through the sample could be analyzed beyond the visible spectrum. Some examples of this trend include the application of deep learning for the diagnosis of breast cancer cells [6], the use of machine learning for the automatic recognition of colon and esophagogastric cancer [7], and the implementation of an ensemble of deep neural networks to identify cancer samples in brain tissue [13].

In this work, we focus on utilizing spectral unmixing and machine learning models (support vector machine with polynomial and radial basis function kernels, random forest, and a voting ensemble) to classify HS images of histopathology samples of brain tissue. A supervised classification approach was employed, where a pathologist previously labeled the samples in the HSI database as non-tumor and tumor. The database was organized into four folds to address the variance in the dataset during the learning stage due to the imbalance in the classes. The input feature vectors for the machine learning models were obtained by the concatenation of the estimated endmembers from the spectral unmixing stage. The results obtained were compared with those produced in [13], where the same database was considered, along with an ensemble of deep neural networks for classification. Our results show an improvement in the average accuracy with respect to the state-of-the-art method (89.4% vs. 85.5%).

2 Materials and Methods

The proposed methodology in this work consists of six steps: (i) HSI database, (ii) pre-processing, (iii) spectral unmixing, (iv) features extraction, (v) training and testing, and (vi) results, as illustrated in Fig. 1. In this section, we describe each step highlighting their collective contribution to the overall research objective.

2.1 HSI Database

This research work used a database consisting of human biopsies obtained during brain tumor resection procedures. The study protocol and consent procedures were approved by the “Comité Ético de Investigación Clínica - Comité de Ética en la Investigación” (CEIC/CEI) of “Hospital Universitario de Gran Canaria Doctor Negrin” [8]. The HS camera was configured with a magnification.

20× and captured images with a size 1004 × 800 pixels. Further details about the dataset can be found in [8]. After the resection, the samples were diagnosed as glioblastoma (GB) based on the World Health Organization (WHO) classification of nervous system tumors by a pathologist. Regions of interest (ROI) were selected for subsequent analysis, using the mentioned instrumentation [8], resulting on 494 HS images from 13 different patients (P1, P2, ..., P13) with GB. An example of tumor and non-tumor tissue

samples from the database are illustrated in Fig. 1.(i). In fact, not all patients presented simultaneously tumor and non-tumor tissue samples. In Fig. 2, the distribution of tumor and non-tumor HS images is presented. In this way, patients P9 to P13 only contain tumor tissue samples.

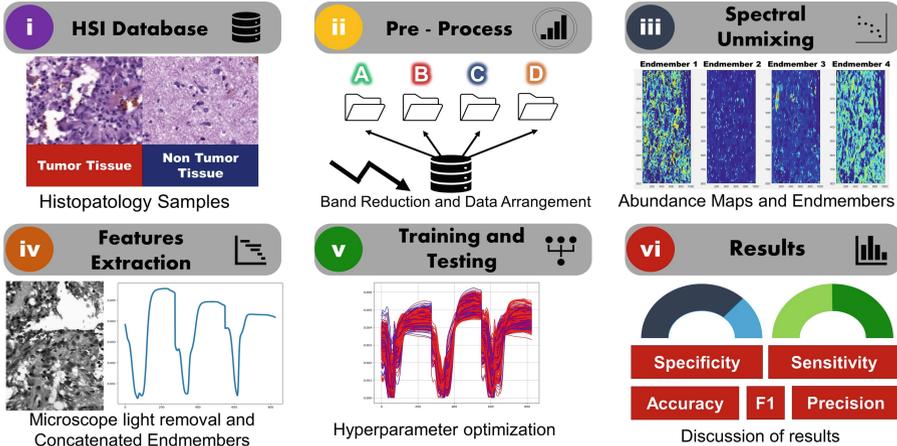


Fig. 1. Diagram of the proposed methodology to classify hyperspectral histopathology images.

2.2 Pre-processing Stage

A pre-processing of the database was performed to adjust the spectral responses of the HS images, and reduce noise and redundant information. The HS images covered a spectral range of 400 to 1000 nm with a resolution of 2.8 nm. Thus, the original HS images had a spatial size of 1004×800 pixels with 826 spectral channels. In [8], a reduction of the spectral bands was proposed to remove redundant information by their high correlation, which led to a decrease from 826 to 275 channels.

As mentioned earlier, not all patients in the database presented both types of classes (non-tumor and tumor), nor they have the same number of HS images for each class (as explained in detail in [8]). This imbalance in class representation, along with the limited number of patients (just 13 different), presented a challenge for the classification scheme. To address these issues and mitigate the potential bias and overfitting during the learning stage, the database was organized into four distinct folds (A, B, C, and D) for cross-validation, with subsets for training, validation, and test. Folds, A, B, and C included nine patients for training, one for validation, and three for testing. Meanwhile, Fold D contained eight patients for training, one for validation, and four for testing. These folds are illustrated in Fig. 3, with different colors (red, green, and blue) indicating each subset. All the folds were randomized to ensure that each sample was presented once in the testing subset, and that patients with both types of labeled data (non-tumor and tumor) were selected for the validation subset. With this four folds approach, the database is presented in four different ways to the ML algorithms for cross-validation.

This perspective follows previous works in [8] and [9], which served as guides for this study.

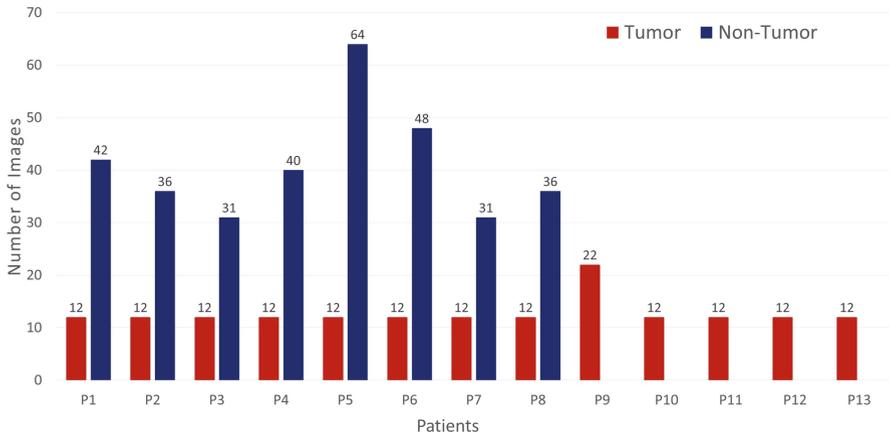


Fig. 2. Distributions of tumor and non-tumor HS images acquired from 13 patients (P1, P2,...,P13).

2.3 Spectral Unmixing

In the field of HS imaging, spectral unmixing (SU) plays a crucial role in analyzing and extracting information about the fundamental materials present in a scene or image. By identifying specific spectral signatures known as endmembers, which represent the fundamental materials, the unmixing technique uses them as reference points. This process enables the computation of concentrations or abundances of each component at pixel level. The resulting abundance maps, like the ones depicted in Fig. 1. (iii), provide valuable insights into the spatial distribution of materials within the image. This technique is a valuable tool to understand the composition and characteristics of the scene or image captured through HSI [10]. Hence, the identification of the fundamental signatures in the HS image by spectral unmixing allows the implementation of machine learning models to perform classification tasks. In addition, the computational cost of the classification task can be reduced through the application of spectral unmixing.

In this study, the SU stage was carried out by using the EBEAE algorithm [10]. Other unmixing algorithms such as sparse non-negative matrix factorization, non-smooth non-negative matrix factorization, or their variants could also be applied [11], but EBEAE was chosen due to its robustness to different types and levels of noise and its low complexity. EBEAE has the ability to estimate the abundances of individual endmembers by minimizing the approximation error and maximizing their entropy. In addition, EBEAE normalizes the cost function to mitigate the dependence on the size of the database. The EBEAE hyperparameters were determined according to the recommendations provided in [10], and they were set to, $\rho = 1$ (*similarity weight*), $\mu = 0$ (*entropy weight*), error threshold $\epsilon = 1 \cdot 10^{-6}$, and a maximum number of 10 iterations (*maxiter*). The linear unmixing process assumed four characteristic components, that is, $N = 4$, as in [13].

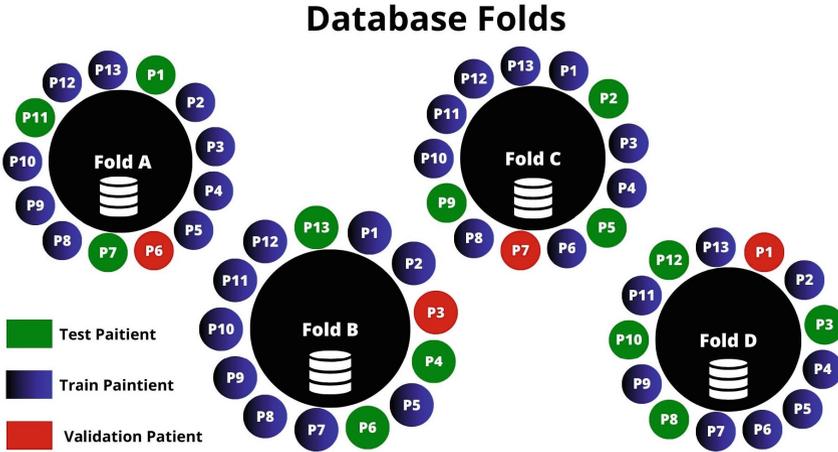


Fig. 3. Database folds for the 13 patients (P1, P2,...,P13) with subsets for training, validation, and testing.

2.4 Features Extraction

In the histopathology samples, there are areas without brain tissue where the microscope light can transmit directly. These areas without tissue do not contain information and need to be eliminated to avoid bias in the classification process. To address this issue, the first step is to identify the endmember with a flat spectral signature that corresponds to no-tissue regions, where the microscope light is transmitted directly. Next, this endmember must be eliminated since it does not provide characteristic information of the sample. This phenomenon can be easily identified due to the sum-to-one restriction applied by EBEAE [10], which results in a spectral signature with uniform information across most spectral bands. In Fig. 4.(i), the green-colored endmember, denoted as endmember 3, corresponds to the microscope light (roughly flat spectral signature). Another challenge is that in EBEAE, the order of the resulting endmembers could vary for each HS image. Therefore, a mathematical method needs to be devised to identify this flat endmember. For this goal, the Euclidean distance between each resulting endmember and a flat spectral signature is computed by recalling that there are 275 spectral bands:

$$E_i = \left\| p_i - \frac{1}{275} \mathbf{1} \right\| \quad \forall i \in \{1, 2, 3, 4\} \quad (1)$$

where p_i represents the i -th estimated endmembers by EBEAE, $\mathbf{1}$ is a vector with just unitary entries, and $\|\cdot\|$ denotes the Euclidean norm. Consequently, for each HS image, we eliminate the endmember with the smallest value E_i .

Next, the remaining three endmembers were organized in a specific order to define an overall feature vector for each HS image. To achieve this, we selected a HS image as a reference in the database and utilized the order of endmembers in this image as a guide to arrange the remaining ones. In this stage, we used again the Euclidean distance as a similarity index among endmembers. At the end of this stage, the arranged and

concatenated endmembers (as visualized in Fig. 4.(ii)) were used as a single feature vector per HS image. As a final step, the feature vectors were randomly mixed in each set of the different folds to ensure that the processes were not biased towards any specific order.

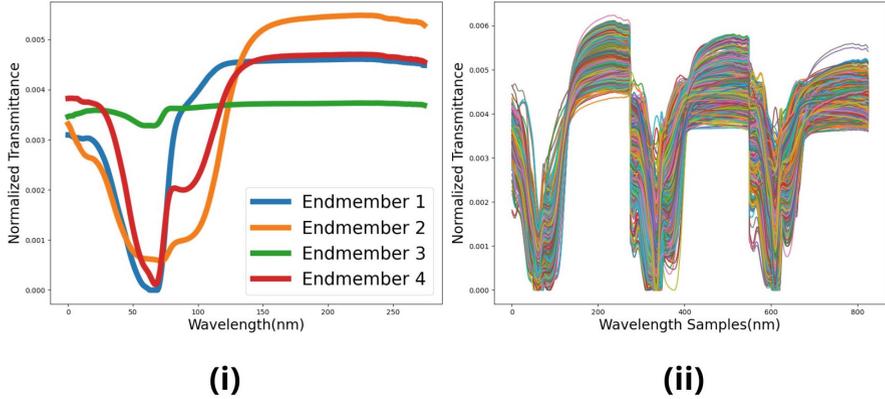


Fig. 4. Features extraction step: (i) Example of the estimated endmembers; and (ii) features vectors by concatenating the endmembers with the same order.

2.5 Training and Testing

In this study, a supervised approach is employed for the classification task, since each HS in the database is already labeled as tumor or non-tumor (binary problem) by the pathologist [8]. Machine learning models have varying strengths and weaknesses, and their suitability depends on factors, such as the source of the data, the complexity of the problem, and the specific requirements of the task [12]. For instance, certain classifiers may excel in handling linearly separable data, while others may be more effective for non-linear relationships or high-dimensional data. Now, by looking at the endmembers and input feature vectors in Fig. 4, there is some similarity in the input data for classification and the vectors are high-dimensional, so a complex nonlinear problem could be expected. As a result, we focus on two well-known strategies to deal with this complex scenario: Support Vector Machine (SVM) and Random Forest (RF). In the following, we present a short description of these classification models:

- SVM is a powerful, versatile, and popular machine learning model. SVMs are particularly well suited for the classification of complex small or medium-sized datasets [12]. For nonlinear classification problems, the selection of the kernel can help to adjust the decision boundaries, for example, polynomial and radial basis functions. Each of these kernels has hyperparameters that can be adjusted to fit the training dataset.
- RF is an ensemble of decision trees, generally trained with the bagging method and based on the wisdom of the crowds paradigm [12]. Therefore, the RF classifier is also

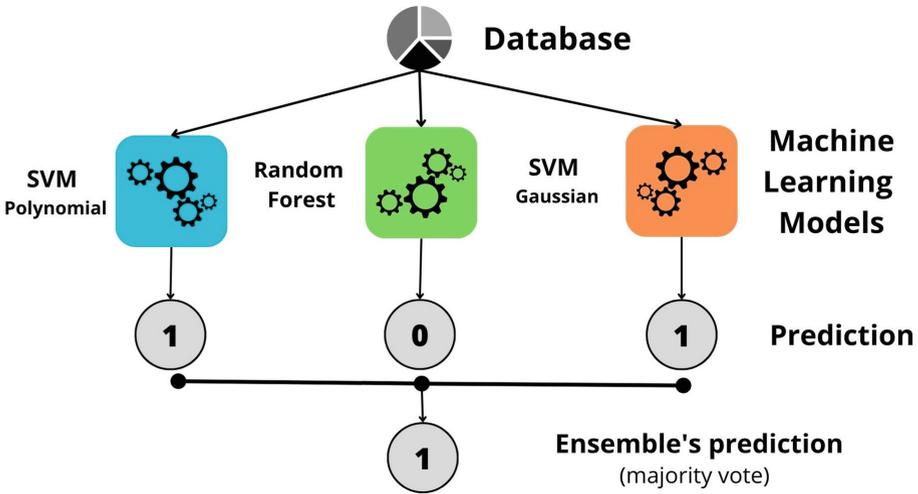


Fig. 5. Structure of the voting ensemble.

a popular machine learning model that can capture complex relationships between features and targets.

Three classification models were initially considered: SVM with a polynomial kernel (SVM-Polynomial), SVM with a Gaussian radial basis function kernel (SVM-Gaussian), and RF. Furthermore, to utilize the specific properties of each classifier and enhance the decision's robustness, an ensemble of them was also explored, resulting in the consideration of a voting methodology. In this methodology, the prediction is based on the class that receives most votes [12]. Typically, the voting ensemble outperforms the best classifier within the ensemble in terms of accuracy. The structure of the voting ensemble is depicted in Fig. 5. To have the best hyperparameters, we used a Bayesian Optimizer in this tuning process, and the final values are listed in Table 1. Two principal hyperparameters are associated with SVM-Polynomial: the degree of the polynomial kernel function (*PolynomialDegree*), and the regularization weight C . The degree of the kernel establishes a compromise between model complexity and computational time, and the regularization weight C shows an inverse relationship with the margin size of the decision regions and must be strictly positive. In the case of SVM-Gaussian, there are two hyperparameters: scaling coefficient of the radial basis function (*Gamma*), and once more, the regularization weight C . In this case, *Gamma* is related to the influence radius of the support vectors. Finally, for RF, there are three hyperparameters: the number of estimators (*Estimators*), the minimum samples split (*MinSamplesSplit*), and maximum depth (*MaxDepth*). The number of estimators corresponds to the number of decision trees in the forest, while *MinSamplesSplit* denotes the minimum number of samples required to split an internal node. Lastly, *MaxDepth* indicates the maximum depth of each decision tree [12].

3 Results

In this section, we discuss the results obtained from the supervised classification process of the HS images in the database. Every fold proposed in Sect. 2.2 was trained with different hyperparameters to get the best results in each one. The pre-processing stages were carried out in MATLAB. Meanwhile, the machine learning classification was performed and evaluated in Python using Scikit-learn [14] to take advantage of the libraries and tools for this goal.

Table 1. Optimized hyperparameters of the machine learning models.

Classifier	Partitions				
	Parameter	Fold A	Fold B	Fold C	Fold D
SVM-Polynomial	<i>PolynomialDegree</i>	3	2	3	3
	<i>C</i>	1	0.126	120	0.21
SVM-Gaussian	<i>Gamma</i>	110	101	800	500
	<i>C</i>	8000	9	8	5
Random Forest	<i>Estimators</i>	100	140	88	100
	<i>MinSamplesSplit</i>	0.015	2	0.001	1
	<i>MaxDepth</i>	10	18	3	20

Table 2. Accuracy scores of the machine learning models in the test subset.

Classifier	Folds				Average Accuracy
	A	B	C	D	
SVM-Polynomial	95.22	79.27	94.53	95.22	91.06
SVM-Gaussian	93.85	79.67	94.46	92.83	90.20
Random Forest	90.00	77.25	88.92	85.66	85.46
Voting Ensemble	96.51	79.67	94.31	93.09	90.89
Average	93.89	78.96	93.05	91.70	89.40
Reference in [13]	96	79	77	90	85.5

Table 2 presents the accuracy score (%) obtained in each fold with every classifier in the test subset. Hence, Fold A reached the best performance with the Voting Ensemble at 96.51%, and the average of all machine learning models was 89.40%. However, the best machine learning model by averaging four folds was SVM-Polynomial, followed by the Voting Ensemble. Meanwhile, Fold B had the worst performance out of the four folds. To validate the results, we compared our work with [13], which used the same database and applied a similar partition methodology, but the classification model is based on an ensemble of deep neural networks. In the following, the metrics to evaluate the

performance of the machine learning models in the test subset were: accuracy, precision, F1-score, sensitivity, and specificity.

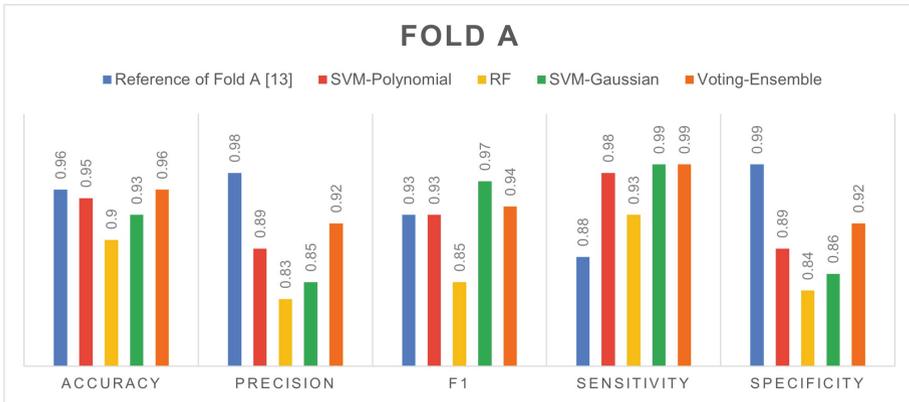


Fig. 6. Performance metrics for Fold A with the proposed machine learning models compared to [13] in the test subset.

In Fig. 6, Fold A shows good results and high similarity with the reference scores in [13]. Among the proposed classifiers, RF exhibited the poorest performance in all metrics. The reference in [13] showed a clear improvement in precision and specificity, but in accuracy, all the proposed machine learning models were close to each other. Meanwhile, in F1-score, SVM-Gaussian achieved top-performance, and in sensitivity, all the proposed models improved the reference. Hence, the resulting high sensitivity indicates the ability to accurately identify true positives.

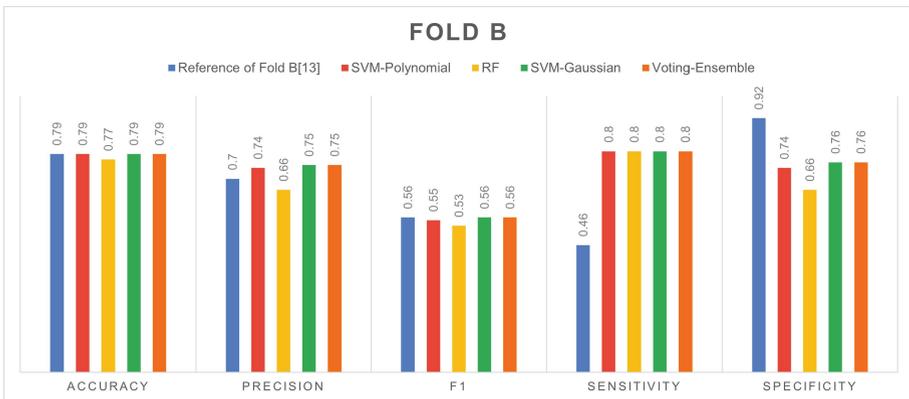


Fig. 7. Performance metrics for Fold B with the proposed machine learning models compared to [13] in the test subset.

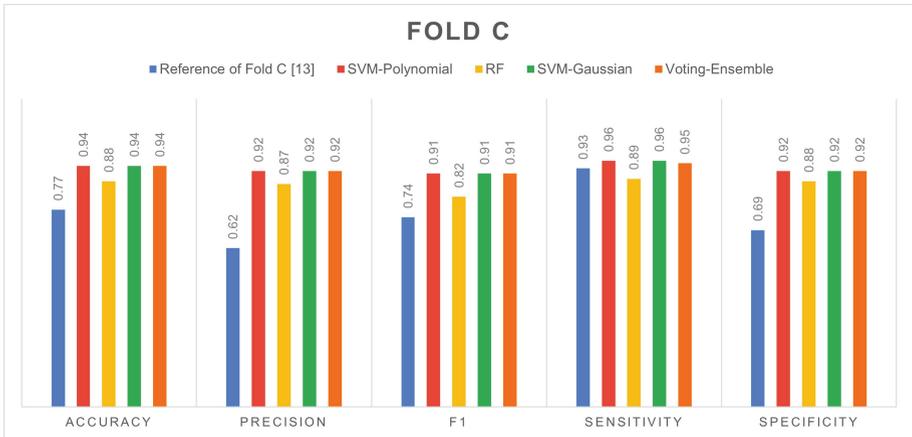


Fig. 8. Performance metrics for Fold C with the proposed machine learning models compared to [13] in the test subset.

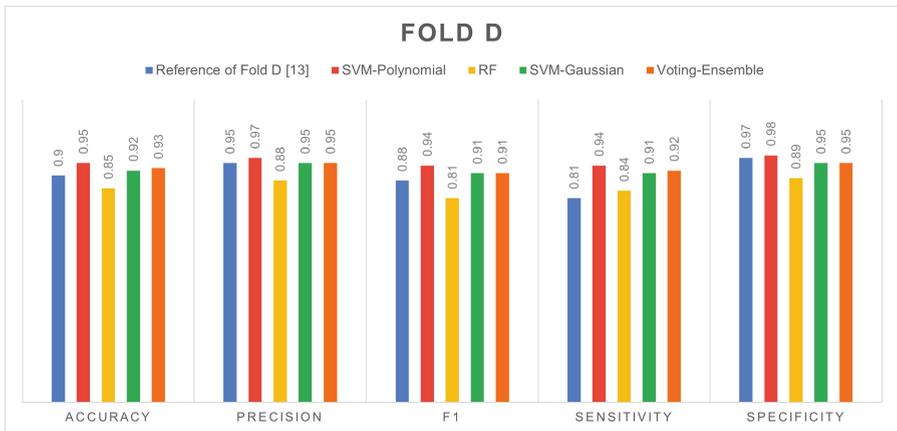


Fig. 9. Performance metrics for Fold D with the proposed machine learning models compared to [13] in the test subset.

As mentioned earlier, the lowest performance among the four partitions was observed in Fold B, where the results are illustrated in Fig. 7. All four machine learning models and the reference in [13] achieved roughly the same performance in accuracy, precision, and F1-score. But the proposed models improved sensitivity, and the reference in [13] reached the best specificity. In overall, the sensitivity rates were relatively higher than accuracy, precision, F1-score, and specificity suggesting that the proposed models were more successful in identifying true positives. This observation could be attributed to the limited number of samples and the imbalanced nature of the database.

Finally, Folds C and D demonstrated the superior performance of our machine learning models compared to the reference in [13]. Figure 8 shows the results for Fold C,

where all four proposed machine learning models outperformed the reference. For Fold D, the performance metrics are illustrated in Fig. 9, where SVM-Polynomial achieved the highest metrics among all, and improved both the reference and other classifiers in every metric. It can be inferred that the more robust results observed in Fold D might be attributed to the difference in the number of samples of each class in every fold, which can be visualized in Figs. 2 and 3.

4 Conclusions

In this work, the effectiveness of machine learning models in conjunction with SU techniques has been demonstrated for the classification of HS histopathological images for tumor tissue diagnosis. The performance metrics obtained from our proposals have surpassed in most cases the reference performance in [13]. In our study, we obtained an average accuracy of 89.40%, which improves the previous result of 85.5% in [13]. Moreover, the proposed machine learning models are based on shallow classifiers (SVM and RF) compared to the scheme in [13]. On average, the best machine learning models were SVM-Polynomial and the Voting Ensemble (see Table 2). These results open up exciting possibilities for clinical applications, where our methodology could be utilized to assist in pathological analysis and support healthcare professionals to improve decision-making processes and patient diagnosis.

Nevertheless, a limitation of this proposal is that specific hyperparameters and independent training stages were required for each fold. As future work, we will explore common hyperparameters in the four folds that could maintain or exceed our performance metrics. Besides that we will investigate other pathologies that can be analyzed using HSI and this classification methodology.

References

1. Lu, G., Fei, B.: Medical hyperspectral imaging: a review. 10.1117/1. JBO.19.1.010901
2. Madabhushi, A.: Digital pathology image analysis: opportunities and challenges. *Imaging Med.* **1**(1), 7 (2009)
3. Halicek, M., Fabelo, H., Ortega, S., Callico, G.M., Fei, B.: In-vivo and ex-vivo tissue analysis through hyperspectral imaging techniques: revealing the invisible features of cancer. *Cancers* **11**(6), 756 (2020). <https://doi.org/10.3390/cancers11060756>
4. Batta, M.: Machine learning algorithms-a review. *Int. J. Sci. Res. (IJSR)* **9**, 381–386 (2020)
5. Magoulas, G.D., Prentza, A.: Machine learning in medical applications. In: Paliouras, G., Karkaletsis, V., Spyropoulos, C.D. (eds.) ACAI 1999. LNCS, vol. 2049, pp. 300–307. Springer, Heidelberg (2001). https://doi.org/10.1007/3-540-44673-7_19
6. Sortega, S., et al.: Hyperspectral imaging and deep learning for the detection of breast cancer cells in digitized histological images. *Proc. SPIE* **11320** (2020). <https://doi.org/10.1117/12.2548609>
7. Collins, T., et al.: Automatic recognition of colon and esophagogastric cancer with machine learning and hyperspectral imaging. *Diagnostics* **11**(10), 1810 (2021). <https://doi.org/10.3390/diagnostics11101810>
8. Ortega, S., et al.: Hyperspectral imaging for the detection of glioblastoma tumor cells in H&E slides using convolutional neural networks. *Sensors* **20**(7), 1911 (2020). <https://doi.org/10.3390/s20071911>

9. Ortega, S., et al.: Hyperspectral superpixel-wise glioblastoma tumor detection in histological samples. *Appl. Sci.* **10**(13), 4448 (2020). <https://doi.org/10.3390/app10134448>
10. Campos-Delgado, D.U., et al.: Extended blind end-member and abundance extraction for biomedical imaging applications. *IEEE Access* **7**, 178539–178552 (2019)
11. Gan, J., Liu, T., Li, L., Zhang, J.: Non-negative matrix factorization: a survey. *Comput. J.* **64**(7), 1080–1092 (2021)
12. Geron, A.: *Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow*. O'Reilly Media, Inc., Sebastopol (2022)
13. Cruz-Guerrero, I.A., et al.: Hybrid brain tumor classification scheme of histopathology hyperspectral images using linear unmixing and deep learning. <https://doi.org/10.2139/ssrn.4292742>
14. Raschka, S., Liu, Y.H., Mirjalili, V., Dzhulgakov, D.: *Machine Learning with PyTorch and Scikit-Learn: Develop Machine Learning and Deep Learning Models with Python*. Packt Publishing Ltd., Birmingham (2022)