Received: 19 August 2023 | Revised: 9 February 2024 | Accepted: 18 March 2024

Healthcare Technology Letters

DOI: 10.1049/htl2.12084

The Institution of Engineering and Technology WILEY

LETTER

Hybrid brain tumor classification of histopathology hyperspectral images by linear unmixing and an ensemble of deep neural networks

Inés A. Cruz-Guerrero[1,2] | **Daniel Ulises Campos-Delgado**[1] | **Aldo R. Mejía-Rodríguez**[1] | **Raquel Leon**[3] | **Samuel Ortega**[3] | **Himar Fabelo**[3] | **Rafael Camacho**[4] | **Maria de la Luz Plaza**[4] | **Gustavo Callico**[3]

[1]Facultad de Ciencias, Universidad Autonoma de San Luis Potosí (UASLP), San Luis Potosi, Mexico

[2]Department of Biostatistics and Informatics, Colorado School of Public Health, University of Colorado Anschutz Medical Campus, Colorado, USA

[3]Institute for Applied Microelectronics (IUMA), University of Las Palmas de Gran Canaria, Las Palmas de Gran Canaria, Spain

[4]Department of Pathological Anatomy, University Hospital Doctor Negrin of Gran Canaria, Las Palmas de Gran Canaria, Spain

Correspondence
Daniel Ulises Campos-Delgado, Facultad de Ciencias, Universidad Autonoma de San Luis Potosí (UASLP), Av. Chapultepec No. 1570, Col. Privadas del Pedregal, 78295, S.L.P., Mexico.
Email: ducd@fciencias.uaslp.mx

Funding information
Consejo Nacional de Humanidades, Ciencias y Tecnologías, Grant/Award Numbers: 254637, 865747; Ministerio de Universidades & European Union (Maria Zambrano 2021-2023), Grant/Award Number: ING-ARQ/6431200-(SI-1821); Spanish Government & European Union (TALENT-HExPERIA), Grant/Award Number: PID2020-116417RB-C42; European Union, Grant/Award Number: MCIN/AEI/10.13039/501100011033; Agencia Canaria de Investigacion, Innovación y Sociedad de la Informacion del Gobierno de Canarias, Grant/Award Number: ACIISI; Asociacion Universitaria Iberoamericana de Postgrado

Abstract

Hyperspectral imaging has demonstrated its potential to provide correlated spatial and spectral information of a sample by a non-contact and non-invasive technology. In the medical field, especially in histopathology, HSI has been applied for the classification and identification of diseased tissue and for the characterization of its morphological properties. In this work, we propose a hybrid scheme to classify non-tumor and tumor histological brain samples by hyperspectral imaging. The proposed approach is based on the identification of characteristic components in a hyperspectral image by linear unmixing, as a features engineering step, and the subsequent classification by a deep learning approach. For this last step, an ensemble of deep neural networks is evaluated by a cross-validation scheme on an augmented dataset and a transfer learning scheme. The proposed method can classify histological brain samples with an average accuracy of 88%, and reduced variability, computational cost, and inference times, which presents an advantage over methods in the state-of-the-art. Hence, the work demonstrates the potential of hybrid classification methodologies to achieve robust and reliable results by combining linear unmixing for features extraction and deep learning for classification.

1 | INTRODUCTION

Traditional histopathology evaluation has been an important tool in cancer diagnosis for more than a century. With this tool, the evaluation is carried out by manual analysis of the morphological and texture characteristics present in the sample, as well as the reaction to different types of reagents [1, 2]. However, traditional histological evaluations can be very subjective. A major advance is CAD, where digital data from samples is used to speed up diagnosis [2, 3]. An example of CAD is the use of RGB images to improve pathological studies, since histopathology images are characterized by repetitive patterns on various scales that are suitable for automated recognition [2, 3].

Recently, CAD has evolved considerably through digital image analysis, and new tools have emerged that allow the acquisition of multimodal information, such as HSI, which provides co-registered spatial and spectral information of samples [4, 5].

Abbreviations: CAD, Computer aided diagnosis; CNN, Convolutional neural network; DL, Deep learning; EBEAE, Extended blind endmember and abundance extraction; H&E, Hematoxylin and eosin; HS, Hyperspectral; HSI, Hyperspectral imaging; GB, Glioblastoma; NN, Neural network; RGB, Red, green and blue; SU, Spectral unmixing; SVM, Support vector machine; VNIR, Visible and near-infrared.

HSI captures information along the electromagnetic spectrum mainly from the visible to near-infrared regions, that is, HSI provides from tens to thousands of two-dimensional images, each one acquired at a certain wavelength [6–8]. In HSI, the information contained at each spatial point or pixel is known as spectral signature, and it characterizes the components of the sample [5, 6]. Although HSI has demonstrated its potential benefits in histopathology, this technology faces several key challenges. One is the high dimensionality of the data, which increases the difficulty in the feature extraction step and makes algorithms computationally expensive. To address this issue, classification and identification methods based on dimensional reduction have been proposed to improve performance, while reducing the complexity of the method [9, 10]. However, another important challenge is the presence of mixtures of components in the same pixel, a situation that complicates the analysis and classification task [11]. Due to this phenomenon, SU techniques have been used in the literature to identify the components that make up mixed pixels [12]. Thus, SU allows unmixing the elements in the sample, so obtaining the spectral response of pure materials (endmembers) and calculating their relative proportions (abundances) from the HS data [12, 13]. This processing step could generate classification maps from the resulting abundances. In the state of the art, there are examples of SU in medical image processing for classification and detection tasks. In [14], two approaches for the classification of craniotomy HS images were presented for the detection of type IV GB, while [15] uses SU to study H&E for histopathological analysis of cellular tissues. Finally, [16] describes a method for the analysis of Raman microspectroscopy data by SU.

Lately, DL methods have been extensively used in computer vision tasks due to the rapid technological advancements in digital hardware [17]. Therefore, multiple solutions that incorporate DL have been adopted in medical image analysis, which have been reviewed in [18, 19]. In [20], a UNet was suggested for nuclei segmentation in histopathology images, and in [21], an ensemble of deep multi-scale networks for the classification of breast cancer images. In fact, machine learning and DL have also been implemented in conjunction with HSI to extract features and classify tissue samples through (i) convolutional neural networks (CNN) for breast cancer cells, glioblastoma tumor cells and colorectal resections [22–24], (ii) super-pixel representations and support vector machines for glioblastoma tumor detection [25], and (iii) multi-layer perceptrons for esophageal adenocarcinoma specimens [26]. These examples demonstrate the scope and opportunity area of machine learning and DL to effectively analyze histopathology HS samples. However, this research line is relatively incipient [18], since the applications of HSI in the biomedical field are still experimental. One disadvantage of DL is that these algorithms could be computationally expensive and require complex hyperparameter tuning [17]. However, one promising strategy is transfer learning to alleviate the training stage and decrease intra and inter-patient variability [27].

In an attempt to overcome potential problems in DL architectures, some efforts have focused on using SU and DL to perform HSI classification [28, 29]. These methodologies propose the decomposition of HSI into abundance maps and endmembers, which serve as a features engineering step for the DL architectures. These proposals have two advantages. The first one is that they reduce the dimensionality of the input data to the DL architectures [28, 29]. The second advantage is that by this dimensional reduction, the proposed DL classifiers can have simple architectures, facilitating their implementation in high performance computing platforms, and, hence, reducing their execution time. Even, in some cases, this joint framework could potentially improve the classification results by reducing the chances of overfitting [28, 29].

Considering the advantages of combining SU and DL for classification, this paper proposes a methodology that applies both techniques to histopathology HSI. The proposed hybrid scheme incorporates the information obtained from the SU stage into the classification process from a new perspective. Hence, unlike most proposals in the literature that rely on the abundance maps for classification [28, 29], our proposal focuses on the characteristic spectral information of each HS image that is obtained through SU, that is, just the endmembers are used as input features. In this way, the most representative spectral signatures per pixel are used to characterize the overall HS image. This approach also provides robustness to the classification results compared to the methods in the state-of-the-art when there is only a label at an image-level, that is, there is no pixel-level ground-truth available. In our proposal, a preprocessing chain is applied first to the histopathological brain HS images, then SU is used to obtain endmembers and abundance maps for features engineering. The endmembers dataset is used to train an ensemble of deep NNs, which performs a binary classification of the HS images to distinguish between tumor and non-tumor tissue. The ensemble of NNs is trained using a transfer learning methodology [27]. Our overall approach has three main benefits: (i) decrease the input data needed in the classification stage, which in turn reduces the computational cost of training; (ii) reduce the presence of overfitting, and (iii) decrease the variability of the classification results by the ensemble of deep NNs. Our proposal is compared with two state-of-the-art approaches based on machine learning and DL to illustrate its advantages in terms of accuracy and computational cost [23, 25].

2 | HYBRID CLASSIFICATION SCHEME

Figure 1 illustrates the proposed hybrid classification approach, which consists of seven steps (A) to (G). These steps are presented in detail next.

2.1 | Histopathology hyperspectral database

The database for this work consisted of a set of HS images acquired from histological samples of human brain tissue, as described in [23, 25]. HS images were captured using an Olympus BX-53 microscope (Olympus, Tokyo, Japan) equipped with

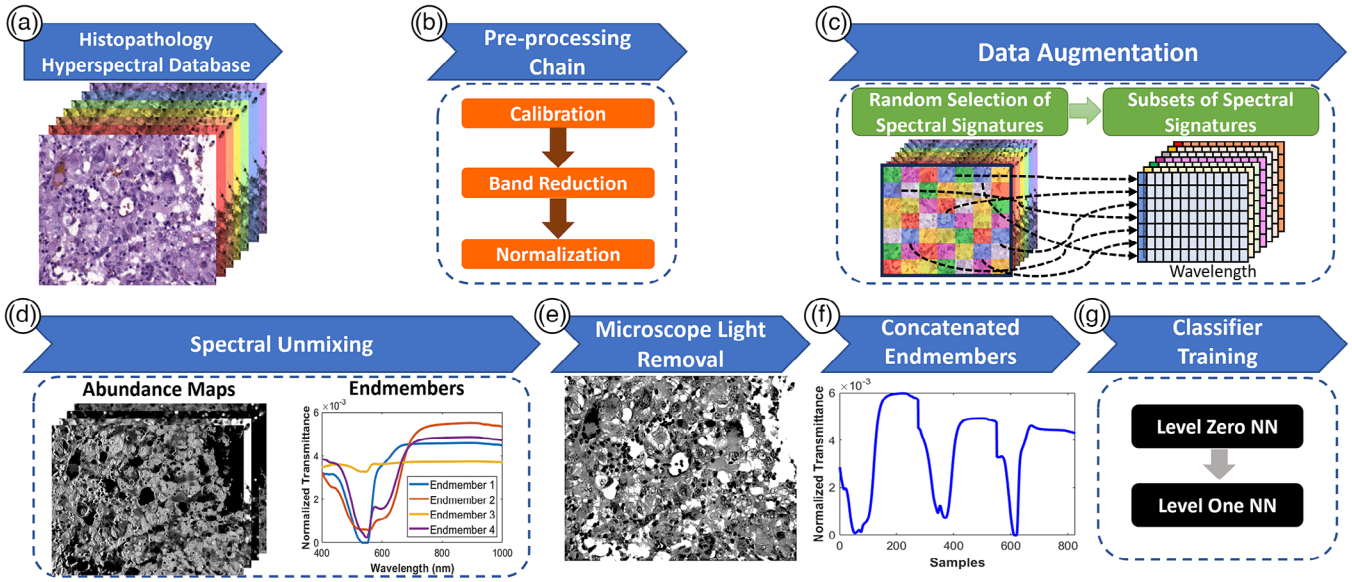


FIGURE 1 Overall diagram of the proposed hybrid scheme for processing and classification of HS images.

a scanning HS camera of type Hyperspec VNIR A-Series from HeadWall Photonics (Fitchburg, MA, USA). The HS camera was set at a 20× magnification, producing a HS image of size 375 μ m × 299 μ m. The acquisition system is presented in more detail in [30]. The spectral range of the HS images covers 400 to 1000 nm with a spectral resolution of 2.8 nm; so the resulting HS images have 1004 × 800 pixels with 826 spectral channels. Figure 2a shows an example of HS images for the two classes studied in this work: *non-tumor* and *tumor*.

The database was obtained from histological samples of 13 patients (P1,…,P13) with GB tumors (grade IV). Different regions of interest were selected in the samples, based on pathologist annotations, from which a maximum of 10 HS images per region (as described in [23]) were acquired, producing 494 HS images. Out of these HS images, 328 corresponded to the non-tumor class, while 166 were classified as tumor. The distribution of tissue on the slides per patient is shown in Figure 2b. As can be seen, the number of images labeled as non-tumor is different in each patient. Only HS images from the tumor class were available for some patients, that is, patients P9 to P13. This distribution of the data established three main problems: (i) the limited number of patients and samples; (ii) not all samples contain both classes (non-tumor and tumor), with only eight patients having both types; and (iii) the data set is not balanced between both classes, as more non-tumor images are available. As a consequence, we chose a cross-validation procedure to evaluate our hybrid classification scheme to avoid bias and overfitting during the learning stage. For cross-validation, four folds were selected by subsets of independent patients for training, validation and test (see Figure 2c). A random assignment of the samples in each fold was followed to guarantee that each sample was presented once in the testing subset and by selecting patients with both types of labeled data (non-tumor and tumor) in the validation subset. Therefore, at least, three patients with just tumor information were assigned to the training subset. A similar cross-validation distribution was used in previous works with the same database [23, 25].

2.2 | Pre-processing chain

The acquired HS images required a preprocessing chain, which is based on two main steps. The first consisted of transforming the images from radiance to normalized transmittance using a white reference image captured in a blank area of the pathological slide. Subsequently, neighboring spectral bands were averaged to remove redundant information, due to the high correlation between adjacent spectral bands, reducing the number of spectral bands from 826 to 275. At the end of the preprocessing chain, a normalization to sum one was applied to the spectral signatures per pixel of the HS images. This normalization step standardize all the spectral signatures and prepare them for subsequent stages.

Prior to the normalization stage, the HS data was visualized to assess the complexity of the classification problem. This task was carried out by selecting different regions corresponding to cell nuclei and their external structures in images of both classes and different patients. An example of this visualization is shown in Figure 3. After obtaining data from the different regions (selected with red, blue, green, and black circles in Figures 3a and 3b), the spectral signatures of the selected areas were analyzed by their mean and standard deviation for both classes, see Figure 3c. In this figure, we can observe the great similarity between the spectral signatures of non-tumor and tumor classes. This similarity can be measured by the correlation coefficient between the mean spectral response of each class. Therefore, a high correlation can be observed, very close to a perfect

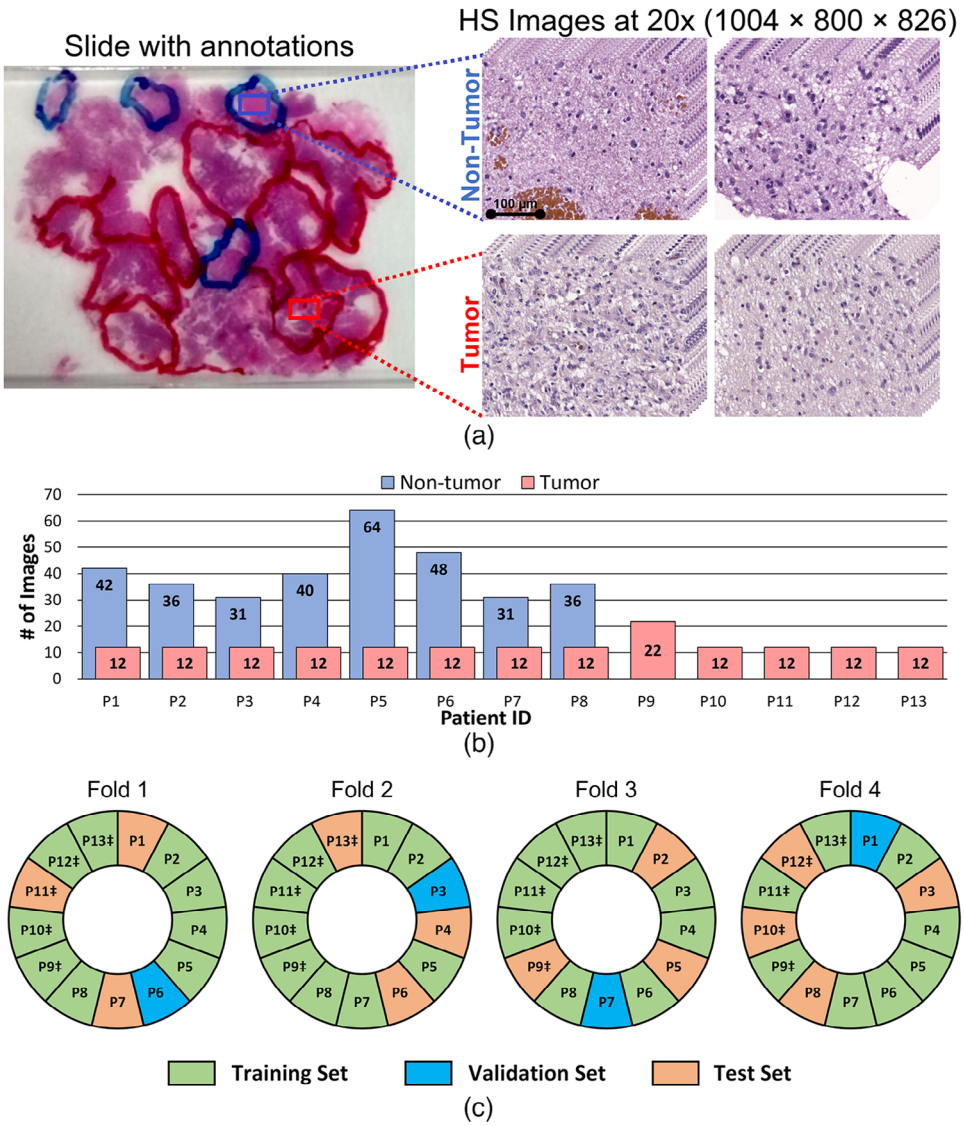


FIGURE 2 Database overview: (a) Example of a pathological sample with annotations of the regions with tumor (red line) and non-tumor (blue line) tissue macroscopically made by the clinical experts on the slide after diagnosis, and examples of non-tumor and tumor HS images used in this study for classification (imaged at 20×); (b) number of tumor and non-tumor HS images acquired from each sample per patient; and (c) distribution of the data for each patient in the proposed partitions (‡ patients with only tumor).

alignment (correlation > 0.98), due to the H&E staining process. In other words, because the same reagents were applied to both classes, the spectral response tends to be very similar in both cases. This observation implied that differentiating between both two types of tissue was not feasible by simply analyzing the spectral signatures, so a more involved classification scheme was required.

2.3 | Data augmentation

In general, the size of the database used for training a classifier has a direct impact on its results. Typically, increasing the size of the database can enhance performance, up to a certain point beyond which further augmentation does not yield significant improvements [31, 32]. Thus, a larger database provides more diverse examples, which can help the classifiers to learn robust features and generalize better to unseen data [31, 32]. However, the quality of the data is equally important as the quantity [32]. A large database with poor quality data might not produce better results than a smaller one with high quality [32]. Therefore, the database has to be representative and accurately labeled to avoid biases and errors in the classification task [32].

In this work, we have 494 HS images from the studied database; these data are used to train, validate, and test the classifier. As shown in Figure 2, each fold has approximately 320 HS images for training, 50 for validation, and 124 for testing. However, this reduced database poses a limitation for the learning process of the hybrid classifier. Therefore, to overcome this

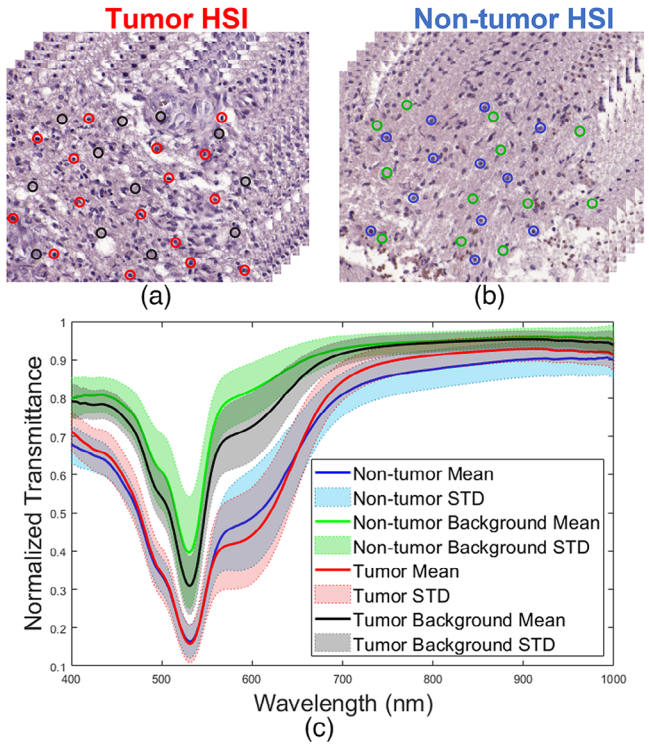


FIGURE 3 Preview example of the spectral information in the HS images: (a) and (b) HS image with tumor and non-tumor tissue, respectively (the examples of selected regions are labeled with a circular marker to analyze their spectral information); and (c) mean average spectral signatures and standard deviations: tumor cells (red), non-tumor cells (blue), tumor background tissue (black), and non-tumor background tissue (green).

condition, data augmentation was performed to ensure sufficient data for proper learning. For this goal, the pixels of each HS image were divided into ten random subsets, as shown in Figure 1c, and subsequently SU was applied to the spectral signatures of these pixels. This division of HS images into subsets increases the input features for classification. Each subset possesses a distribution of information similar to that of the original HS image. As a result, the data augmentation step generated 10 times more features for each HS image, which translates to $10 \times 494 = 4940$ input features in total, representing the information of the entire database. In addition, up to 20 random subsets were also evaluated for each HS image without improvements in the initial classification tests. These subsets were now used to compute the SU stage, to obtain the distinctive features for training, validating, and testing the classifier. Thus, the proposed data augmentation step enables more consistent training with reduced variation when classifying high-resolution images.

2.4 | Spectral unmixing

SU is employed as a preliminary step to the classifier to obtain characteristic input features. Thus, the abundances and characteristic endmembers were estimated from each subset of pixels created in the data augmentation stage. In a general sense, endmembers are representative of each component in the HS image, while the abundance maps indicate the proportion of such components at each pixel. The SU stage was carried out by the EBEAE algorithm described in [13], using the N-FINDR scheme [33] as the endmembers initialization method. Other linear unmixing algorithms such as sparse non-negative matrix factorization, non-smooth non-negative matrix factorization or their variants could be also applied [34], but EBEAE was chosen due to its robustness to different types and levels of noise, low complexity and its excellent performance when evaluating various types of multi- and hyperspectral biomedical data. The hyper-parameters of EBEAE were set to $\rho = 1$ (*similarity weight*), $\mu = 0$ (*entropy weight*), error threshold $\epsilon = 1 \times 10^{-6}$, and a maximum number of iterations (*maxiter*) of 10. The linear unmixing process assumed four characteristic components, that is, $N = 4$. These hyper-parameters were tuned manually following the guidelines in [13] and [14].

2.5 | Microscope light removal

There is an intrinsic problem with the histological samples, since there are areas without tissue, so there is direct transmission of light from the microscope. To rule out this problem, after the SU stage, the endmember with a flat spectral signature corresponding to these regions is eliminated. This phenomenon can be visualized in Figure 1d, specifically in the information contained in endmember 3, which due to the sum-to-one restriction applied by EBEAE, results in a spectral signature with uniform information in practically all bands. As a result, recalling that there are 275 bands in the HS images, this endmember can be identified by the following optimization procedure:

$$\mathbf{p}^* = arg\,min_{\forall \tilde{\mathbf{p}} \in \mathcal{P}} \left\| \tilde{\mathbf{p}} - \frac{1}{275}\mathbf{1} \right\|, \tag{1}$$

where $\mathcal{P} = \{\mathbf{p}_1, \ldots, \mathbf{p}_N\}$ represents the set of estimated endmembers by EBEAE, $\mathbf{1}$ is a vector with just unitary entries, and $\|.\|$ denotes the Euclidean norm. In this way, endmember $\mathbf{p}^*$ corresponds to the spectral signature of the regions that do not have cellular tissue information in the HS image, so this endmember is discarded as an input feature during classification.

2.6 | Concatenated endmembers

After the endmembers were estimated in each subset, they have to be ordered for a fair comparison, since EBEAE does not guarantee a consistent ordering. For this purpose, a subset is selected randomly as a reference for each fold, and its order of endmembers is used as a guideline for the remaining ones, where the Euclidean distance acts as a similarity index. The ordered endmembers were concatenated into an augmented vector containing the spectral information, as shown

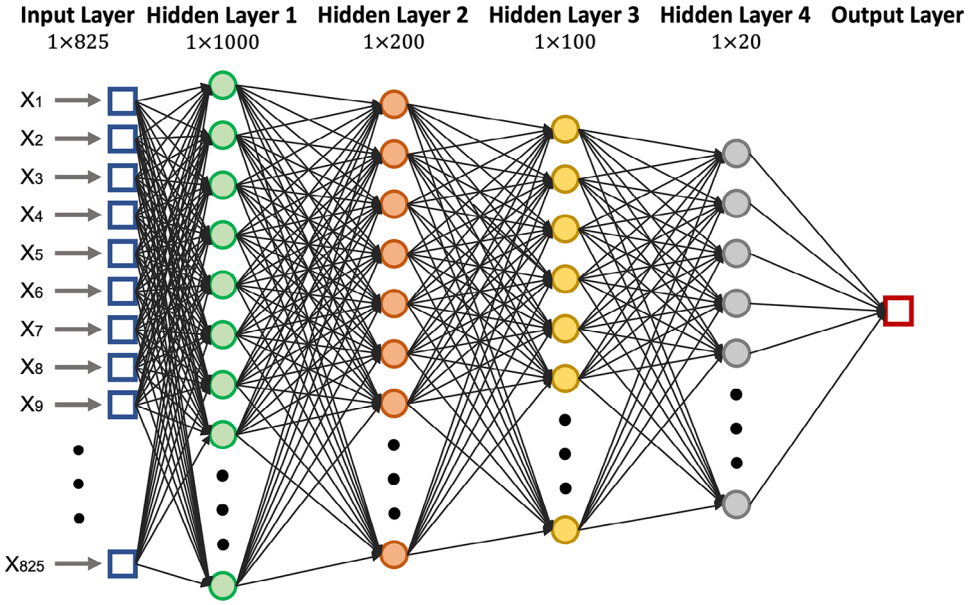


FIGURE 4 Proposed architecture of the three level-zero NNs.

in Figure 1f. As a result, there is only one vector of dimension $3 \times 275 = 825$ to represent each subset. This severe features extraction step is based on the assumption that the concatenated spectral components in tumorous and non-tumorous tissue samples provides distinctive information for classification. Another important factor for choosing just endmembers and discarding the abundance maps is that only one global label is available for each HS image. Therefore, the spatial properties of each class are unknown, which makes their use challenging, and just using endmembers provided remarkable error metrics with low complexity, as it is shown in Section 3. Nonetheless, in our initial experiments, we also trained classifiers with abundance maps yielded significantly lower results compared to endmembers.

2.7 | Classifier training

The classification stage considered a deep NNs ensemble to decrease the variability of a single NN architecture, that is, to add robustness to this process, while seeking to improve the overall accuracy of the proposed methodology by a transfer learning approach [27, 35, 36]. The ensemble was built by joining three identical level-zero NNs. This number of NNs was experimentally selected to keep the classification performance stable, since increasing in more than three level-zero NNs did not present a significant difference. The proposed architecture is presented in Figure 4, and included an input layer of 825 features (concatenated endmembers), and four hidden ones with $\{1000, 200, 100, 20\}$ neurons, respectively, to build a hierarchical learning structure in the network. In our methodology, the optimization of the number of layers and neurons per layer was also conducted experimentally, while keeping the network simple as possible and ensuring maximum performance. Each internal layer had an activation function *LeakyReLu* (with slope for negative values $\alpha = 0.1$), L_2 regularization of 0.0002, and a dropout stage of 50%. These two last elements were included to avoid overfitting during the learning process. The output layer had only a sigmoid activation function (binary classification problem), and the binary crossentropy cost function was used to adjust the weights [37]. To reduce the variability in the results, the uniform variance scaling method was used as an initialization method for all layers [38], both in the level-zero NNs and in the ensemble. The training of both architectures used a minibatch of 100 elements with a learning rate of 0.001, and the number of epochs was set to 500 and 100 epochs for the zero and one level NNs, respectively. As previously described, the input data to the level-zero NNs considered the training sets in each fold in Figure 2c by using only the concatenated endmembers of the subsets related to each HS image with dimension 825. After training the NNs, they are saved and coupled to the deep ensemble or level-one network.

The deep NNs ensemble or level-one network uses the three pre-trained level-zero networks by following a transfer learning methodology (i.e. during the training of the level-one network, the weights of the level-zero NNs were frozen) [27]. Figure 5 presents the general architecture of the deep ensemble, where the outputs of the level-zero NNs were discarded (see Figure 4) to couple with the level-one network. So, in the deep NNs ensemble, the outputs of the three level-zero networks generated a vector of 60 input elements. This vector is processed by a NN formed by three hidden layers of $\{300, 200, 32\}$ neurons with LeakyReLu activation function (with slope for negative values $\alpha = 0.1$ in each layer), L_2 regularization of 0.0002, and a dropout stage of 50%. Once more, the binary crossentropy cost function was used to tune the weights. Similar to the level-zero networks, the level-one NNs were optimized to achieve

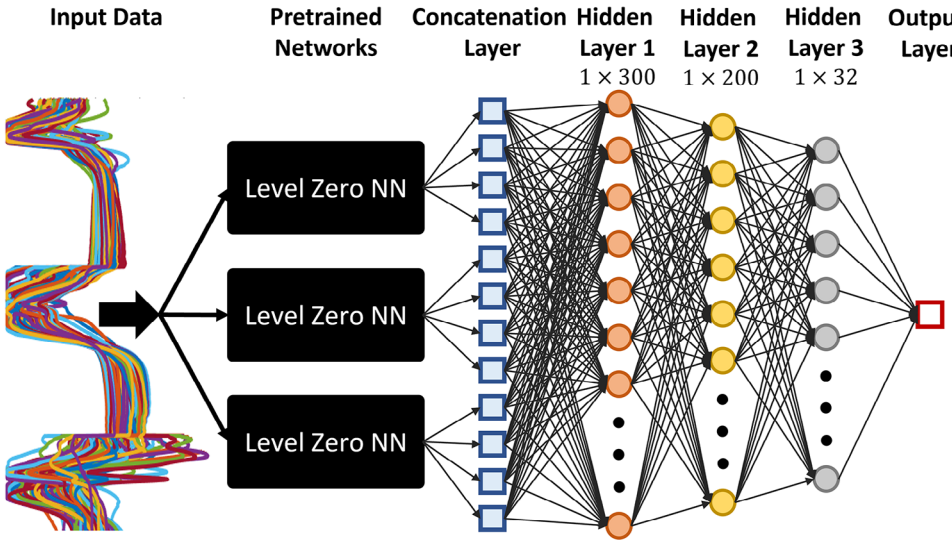


FIGURE 5 Proposed architecture of the NNs ensemble, or level-one NN.

maximum performance through a simple architecture. An initial hidden layer of 300 neurons was chosen to enhance the representability and interconnection of the layers with the data from the three level-zero NNs. The aim of this selection was to ensure optimal information flow and effective integration across the network. In addition, a sigmoid activation function was chosen in the output layer, which produced a bounded output between zero and one, which is rounded to generate the classification result. Finally, the overall deep NNs ensemble returned a binary value, zero for the non-tumor classes and one for tumor classes, which was verified by the labeling performed by the clinical experts.

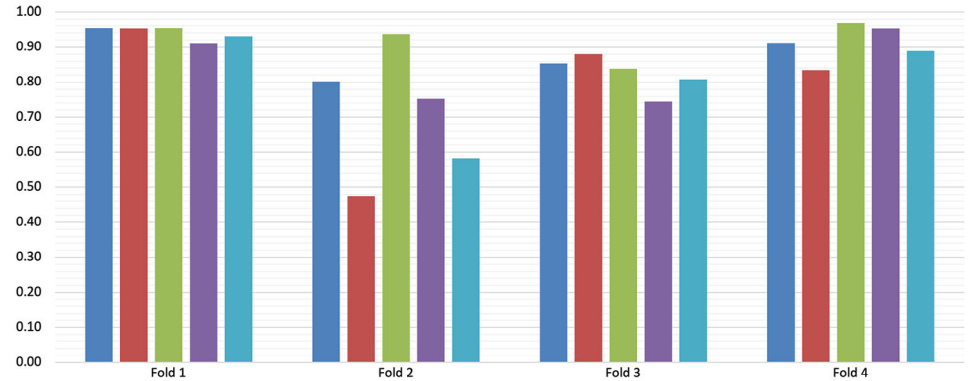
3 | RESULTS

In this section, we describe the results obtained using the proposed hybrid classification methodology. In this evaluation, an independent classifier is trained for each proposed fold in Figure 2c). The pre-processing and SU stages were carried out in MATLAB®, and the resulting data per patient was saved for later use. Meanwhile, the deep NNs ensemble was trained and evaluated in Python using TensorFlow modules [39]. Both tasks were performed on a computer with a 4.2 GHz Intel Core i7 quad-core processor and 16 GB of RAM. The classification results were evaluated by considering the labeling of the clinical experts as ground-truth, and the following metrics: accuracy, sensitivity, specificity, precision, and F1-score [40, 41]. These metrics are standard for evaluating performance in classification tasks with respect to true-positive and true-negative values. The learning process for the level-zero and level-one NNs was carried out by all the available training data in each fold. The accuracy, sensitivity, specificity, precision, and F1-score of the training phase were on average above 94% ± 3%, with minimal variability across different folds. On the other hand, the validation results for each fold were reduced on average to 77% ± 30% across all metrics, translating to a 17% decrease compared to the training stage, with an increase in variability among the folds. However, the validation results were linked to a single patient in all folds (see Figure 2c), so the decline in performance is likely attributed to the lack of generality in the validation data. This same behavior was observed in the comparison methods from the state-of-the-art [23, 25].

3.1 | Test results

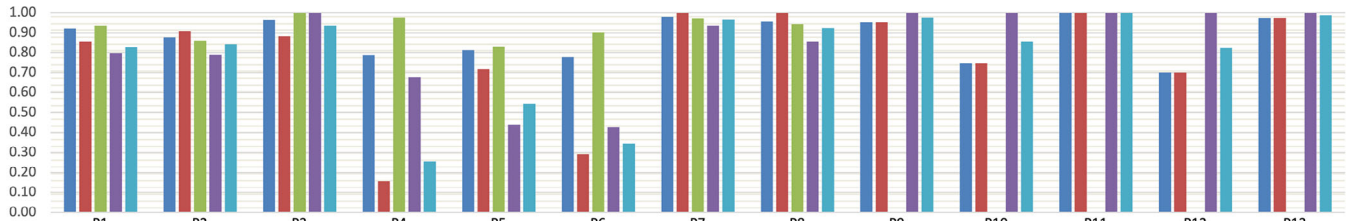
In this section, we present the classification results for the test set for each fold in Figure 2c. These results are shown in Figure 6. In the first instance, the resulting accuracy, sensitivity, specificity, precision and F1-score were 88% ± 7%, 78% ± 17%, 80% ± 16%, 84% ± 11%, and 92% ± 6%, respectively. In this way, the performance metrics showed improvement compared to the validation results with lower variability. As shown in Figure 6, the best performance metrics were reached in Fold 1, followed by Fold 4. Meanwhile, the lowest values for sensitivity, precision and F1-score were obtained in Fold 2.

Finally, Figure 7 presents the classification results for each patient evaluated with this methodology. Figure 7 shows that only four patients have an accuracy lower than 80% (P4, P6, P10 and P12), two were between 80% and 90% (P2 and P5), while the rest obtained a performance greater than 90%; and a similar performance trend is observed in the other metrics. When analyzing cases with the lower performance, patients P4, P6, P10, and P12 were classified with the models trained with the data from Folds 2 and 4. However, these models showed good performance in the rest of the patients (P3, P8 and P13), being Fold 2 the one with the lowest performance in all metrics. It should be clarified that the specificity can only be calculated for patients with both types of classes (patients P1 to P8), so patients P9 to P13 presented a sensitivity of 100%. The overall high performance in the performance metrics suggests that our proposal



	Fold 1	Fold 2	Fold 3	Fold 4
Acc.	0.95	0.80	0.85	0.91
Sen.	0.95	0.48	0.88	0.83
Spe.	0.95	0.94	0.84	0.97
Prec.	0.91	0.75	0.75	0.95
F1	0.93	0.58	0.81	0.89

FIGURE 6 Experimental results of the test set in each fold by using the hybrid classification scheme.



	P1	P2	P3	P4	P5	P6	P7	P8	P9	P10	P11	P12	P13
Acc	0.92	0.88	0.97	0.79	0.81	0.78	0.98	0.96	0.95	0.75	1.00	0.70	0.98
Sen	0.86	0.91	0.88	0.16	0.72	0.29	1.00	1.00	0.95	0.75	1.00	0.70	0.98
Spe	0.94	0.86	1.00	0.98	0.83	0.90	0.97	0.94	-	-	-	-	-
Prec	0.80	0.79	1.00	0.68	0.44	0.43	0.94	0.86	1.00	1.00	1.00	1.00	1.00
F1	0.83	0.84	0.94	0.26	0.55	0.35	0.97	0.92	0.98	0.86	1.00	0.82	0.99
	Fold 1	Fold 3	Fold 4	Fold 2	Fold 3	Fold 2	Fold 1	Fold 4	Fold 3	Fold 4	Fold 1	Fold 4	Fold 2

FIGURE 7 Performance metrics per patient in the test set for the NNs ensemble. P9 to P13 have only tumor class samples, so it is not possible to obtain the specificity metric.

could serve as a potential CAD to detect histopathology HS images with tumor presence.

3.2 | Comparison with state-of-the-art methods

3.2.1 | Performance metrics

Table 1 shows the results of the hybrid classification scheme compared to two methodologies published in the state-of-the-art with the same database and machine learning approaches [23, 25]. The first one is a SVM classification, which was trained by a superpixels scheme [25]; where specifically the results at the image level were used. Meanwhile, the second comparison method is a patch-based CNN [23]. Both algorithms were trained with the same data except for patient P6, since an annotation error was detected during the CNN implementation. These annotation errors were corrected both in the SVM work with superpixels [25] and in this work. Therefore, to provide a fair comparison, patient P6 was not included in the calculation of the total means and standard deviations. Hence, the mean accuracy obtained with the hybrid classification scheme was 3% higher than the CNN-based approach, and 6% higher than the SVM with superpixels, including a reduction in the standard deviation. This behavior is repeated in the specificity metric, where the proposed approach obtained 82% $\pm$ 33% against 78% $\pm$ 34% and 77% $\pm$ 16% of the SVM-based, and CNN-based approaches, respectively. On the other hand, when examining the sensitivity results, the SVM-based approach reached the best performance with 91% $\pm$ 22%, followed by the CNN-based approach with 88% $\pm$ 13%, and finally the proposed methodology with 83% $\pm$ 24%. The sensitivity performance for the SVM-based approach is consistent, since it was optimized with respect to this metric [25]. However, the proposed methodology in this paper maintains a balance between sensitivity and specificity at an overall level. In this way, our results indicate that the classification of HS images can be achieved by their spectral characteristics without observable overfitting. Additionally, the proposal presented a clear improvement with respect to CNN- and SVM-based approaches [23, 25], both of which use spatial and spectral information.

TABLE 1 Comparison of results between the hybrid classification scheme, and previous works using a superpixel SVM-based approach [25], and a CNN-based methodology [23], highlighting in bold the best results.

	Hybrid classification scheme			SVM-based approach [25]			CNN-based approach [23]		
Patient	Accuracy	Sensitivity	Specificity	Accuracy	Sensitivity	Specificity	Accuracy	Sensitivity	Specificity
P1	0.92	0.86	0.94	**1.00**	**1.00**	**1.00**	0.93	0.91	0.96
P2	0.88	0.91	0.86	**0.94**	0.83	**1.00**	0.89	**0.99**	0.83
P3	0.97	0.88	**1.00**	**1.00**	**1.00**	**1.00**	0.85	0.91	0.80
P4	**0.79**	0.16	**0.98**	**0.79**	**1.00**	0.73	0.57	0.57	0.58
P5	**0.81**	0.72	**0.83**	0.68	**1.00**	0.63	0.69	0.81	0.64
P6*	0.78	0.29	0.90	**0.90**	**0.50**	**1.00**	–	–	–
P7	**0.98**	**1.00**	**0.97**	0.35	**1.00**	0.10	0.66	0.71	0.63
P8	**0.96**	**1.00**	0.94	**0.96**	0.83	**1.00**	**0.96**	0.96	0.96
P9	0.95	0.95	NA	0.95	0.95	NA	**0.99**	**0.99**	NA
P10	0.75	0.75	NA	0.25	0.25	NA	**0.89**	**0.89**	NA
P11	**1.00**	**1.00**	NA	**1.00**	**1.00**	NA	0.92	0.92	NA
P12	0.70	0.70	NA	**1.00**	**1.00**	NA	0.92	0.92	NA
P13	0.98	0.98	NA	**1.00**	**1.00**	NA	0.99	0.99	NA
Avg.	**0.89**	0.83	**0.82**	0.83	**0.91**	0.78	0.86	0.88	0.77
Std.	**0.10**	0.24	0.33	0.27	0.22	0.34	0.14	**0.13**	**0.16**

NA = metrics not applicable, this is because both classes were not present in these patients.
*Data not available for the CNN-based approach, these data are excluded from the calculation of the mean (Avg.) and standard deviation (Std.).

TABLE 2 Comparison of training time, inference time, and storage size between the hybrid classification scheme, and previous works using a superpixel SVM-based approach [25], and a CNN-based methodology [23].

Method	Training time (min)	Inference time per image (s)	Storage size
CNN-based approach [23]	NA	72	768 GB
SVM-based approach [25]	≈240	1440	0.870 GB
Hybrid classification scheme	AD proposal 7	18	0.023 GB

NA = data not available for the CNN-based approach.

3.2.2 | Data size and computational time

In addition to these results, the hybrid classification scheme significantly reduced the data size and computational time for the learning stage, and inference time for evaluation. One side of this trend can be observed by analyzing the input data for each methodology, which is presented in Table 2. Hence, the CNN-based approach used 49,565 image patches of 87×87 pixels with 275 bands, which requires an approximated size in memory of 768 GB. On the other hand, the data for the SVM-based approach used 426,260 superpixels with 275 bands, and this information spanned approximately 0.87 GB of memory space. However, our hybrid classification scheme significantly reduced the required information, that is, only 4,940 vectors with 875 elements each, which translates to 22.9 MB; this is a ×38 reduction than the SVM-based scheme in [25]. Therefore, the proposed approach involved a reduction in memory and computational overhead to perform data classification. This property is also reflected in the training time of each methodology in each fold (see Table 2), since the SVM-based approach required 4 hours to perform the training, while the proposal of zero-level NNs needed approximately 7 min, respectively. Note that in the case of the CNN-based approach [23], it was necessary a more powerful processing equipment, since the computational time required more than a day to complete the training.

3.2.3 | Inference times

Finally, when analyzing the inference times presented in Table 2, the proposed hybrid classification scheme achieved the lowest value compared to the CNN-based, taking 25% of the inference time of this DL scheme. On the other hand, in the case of the SVM-based method, the difference is much more significant, requiring only 1.25% of the time. These improved results are due to the lower complexity of the proposed hybrid classification scheme. Hence, the optimization of the deep NNs ensemble had 3,229,925 total trainable parameters. Of these parameters, 1,048,320 corresponded to the ones of each level-zero NN, while the level-one NN had only 84,965. Additionally, when analyzing the HS image evaluation, the main complexity in the Big $\mathcal{O}$ analysis lies in the SU stage, which has a value of $\mathcal{O}\left(maxiter \cdot K \cdot (L + N^4)\right)$, where L represents the

number of spectral bands and K the number of pixels to be analyzed within the HS image. Hence, the SU stage could be considered the bottleneck in the hybrid classification scheme. Similarly, the main limitation in the SVM-based approach is the bottleneck caused by the superpixel generation. Although the superpixel algorithm is highly parallelizable, the current implementation in [25] is iterative, leading to high inference times. In addition, in the original paper where this algorithm was used for HS histological data, the authors mention that the number of superpixels per image was overestimated, which further increased the computational time required for the generation of superpixels. Nevertheless, either the proposed hybrid scheme or the comparison methods are feasible for parallelization, which would reduce training and inference times considerably.

4 | CONCLUSIONS

In this study, we demonstrated the potential of a hybrid classification scheme that combines SU and an ensemble of deep NNs for the identification of histopathology samples affected by GB tumor. The proposed methodology relies on SU as a features engineering step, and an ensemble of deep NNs for binary classification. Our results showed improvements in the proposed method against two approaches in the state-of-the-art with the same database in terms of performance metrics, data size, and computational and inference times [23, 25]. The proposal reached accuracy and specificity with an average of 89% and 82%, respectively, with a reduced variability. Meanwhile, sensitivity was the only metric (average of 83%) in which the comparison methods were superior to the proposed one. Furthermore, our hybrid scheme significantly reduced the computational cost to perform the classification step, resulting in shorter learning stages and inference times without compromising performance. These processing results open up the possibility that the proposed approach could be used in clinical applications to identify regions in histological brain samples for further analysis by a pathologist.

Despite the improvements achieved by the proposed method, our work presents some limitations. The most significant is the reduced amount of data available for training and evaluating the model. Although this issue is partially mitigated by the data augmentation strategy, there is a risk of overfitting. The second limitation lies in the macroscopic labeling, which restricts the effective identification of the characteristic components of each class at the cellular level. Lastly, there is awareness of spectral alterations caused by physical effects at different magnifications used for analysis in histopathological samples, especially at 50x and 100x levels. However, the means to attenuate these alterations without losing relevant information remain limited. In future work, we plan to increase the database with patients who jointly present both classes (tumor and non-tumor) in the histological samples to improve the classification metrics and further ensure the generalization of the proposed solution in a broader dataset. In addition, we will investigate the influence of magnification and its adverse effects on the classification process.

We also suggest exploring new directions in line with the proposed hybrid classification scheme. This goal involves considering classifier alternatives, such as attention-based methods and even quantum learning techniques. Furthermore, we propose exploring various deep learning strategies applied in remote sensing that could be feasible for implementation in HS histopathology imaging. Similarly, the possibility of expanding the proposed hybrid scheme to perform pixel-level classification in an HS image will be evaluated. This expansion is considered once pixel-level labeling has been generated to achieve a more precise identification of the characteristic components of each class, increasing accuracy at the expense of a rise in computational cost.

AUTHOR CONTRIBUTIONS

Ines Alejandro Cruz—Guerrero: Conceptualization; investigation; methodology; software; visualization; writing—original draft. **Daniel Ulises Campos—Delgado**: Conceptualization; formal analysis; funding acquisition; investigation; methodology; supervision; writing—original draft. **A. R. Mejia—Rodriguez**: Conceptualization; validation; writing—review and editing. **Raquel Leon**: Conceptualization; data curation; writing—review and editing. **Samuel Ortega**: Data curation; validation; writing—review and editing. **Himar Fabelo**: Data curation; validation; writing—review and editing. **Rafael Camacho**: Validation; writing—review and editing. **Maria de la Luz Plaza**: Resources; validation; writing—review and editing. **Gustavo Callico**: Funding acquisition; project administration; validation; writing—review and editing.

ACKNOWLEDGEMENTS

This work was supported by the CONAHCYT through a Basic Science Grant No. 254637 and by the Spanish Government and European Union as part of support program in the context of TALENT-HExPERIA project, under contract PID2020-116417RB-C42. The work of Inés A. Cruz-Guerrero was supported by CONAHCYT through a doctoral grant (#865747) and by CUMex through a grant from the CUMex-AUIP Mobility Scholarship Program. Additionally, the work of Raquel Leon was supported with a pre-doctoral grant given by the ACIISI of the "Gobierno de Canarias", and, Himar Fabelo was beneficiary of the FJC2020-043474-I funded by MCIN/AEI/10.13039/501100011033 and by the European Union "NextGenerationEU/PRTR".

CONFLICT OF INTEREST STATEMENT

The authors declare no conflicts of interest.

DATA AVAILABILITY STATEMENT

The data that supported this study will be made openly available in The Cancer Imaging Archive (TCIA) at https://doi.org/10.7937/z1k6-vd17. For further details and updates, we encourage interested parties to visit the website: https://en.iuma.ulpgc.es/gustavo.

ORCID

Daniel Ulises Campos-Delgado https://orcid.org/0000-0002-1555-0131

REFERENCES

1. Mobadersany, P., Yousefi, S., Amgad, M., Gutman, D.A., Barnholtz-Sloan, J.S., Velázquez Vega, J.E., Brat, D.J., Cooper, L.A.: Predicting cancer outcomes from histology and genomics using convolutional networks. Proc. Natl. Acad. Sci. U.S.A. 115(13), E2970–E2979 (2018)
2. Kurc, T., Bakas, S., Ren, X., Bagari, A., Momeni, A., Huang, Y., Zhang, L., Kumar, A., Thibault, M., Qi, Q., Wang, Q., Kori, A., Gevaert, O., Zhang, Y., Shen, D., Khened, M., Ding, X., Krishnamurthi, G., Kalpathy-Cramer, J., Davis, J., Zhao, T., Gupta, R., Saltz, J., Farahani, K.: Segmentation and classification in digital pathology for Glioma research: Challenges and deep learning Approaches. Front. Neurosci. 14, 1–15 (2020)
3. Jimenez-del Toro, O., Otálora, S., Andersson, M., Eurén, K., Hedlund, M., Rousson, M., Müller, H., Atzorim, M.: Analysis of Histopathology Images. Biomed Texture Anal. 281–314 (2017)
4. 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 (2019)
5. Johansen, T.H., Møllersen, K., Ortega, S., Fabelo, H., Garcia, A., Callico, G.M., Godtliebsen, F.: Recent advances in hyperspectral imaging for melanoma detection. Wiley Interdiscip. Rev. Comput. Stat. 12(1), e1465 (2020)
6. ElMasry, G., Sun, D.W.: Principles of Hyperspectral Imaging Technology. Hyperspectral Imaging Food Qual. Anal. Control 3–43 (2010)
7. Li, S., Song, W., Fang, L., Chen, Y., Ghamisi, P., Benediktsson, J.A.: Deep learning for hyperspectral image classification: An overview. IEEE Trans. Geosci. Remote Sens. 57(9), 6690–6709 (2019)
8. Martinez, B., Leon, R., Fabelo, H., Ortega, S., Piñeiro, J.F., Szolna, A., Hernandez, M., Espino, C., O'shanahan, A.J., Carrera, D., Bisshopp, S., Sosa, C., Marquez, M., Camacho, R., de la Luz Plaza, M., Morera, J., Callico, G.M.: Most relevant spectral bands identification for brain cancer detection using hyperspectral imaging. Sensors 19(24), 5481 (2019)
9. Bayramoglu, N., Kaakinen, M., Eklund, L., Heikkila, J.: Towards virtual h&e staining of hyperspectral lung histology images using conditional generative adversarial networks. In: Proceedings of the IEEE International Conference on Computer Vision Workshops, pp. 64–71. IEEE, Piscataway (2017)
10. Panda, A., Pachori, R.B., Sinnappah-Kang, N.D.: Classification of chronic myeloid leukemia neutrophils by hyperspectral imaging using euclidean and mahalanobis distances. Biomed. Signal Process. Control 70, 103025 (2021)
11. Keshava, N., Mustard, J.F.: Spectral unmixing. IEEE Signal Process Mag. 19(1), 44–57 (2002)
12. Bioucas-Dias, J.M., Plaza, A., Dobigeon, N., Parente, M., Du, Q., Gader, P., Chanussot, J.: Hyperspectral unmixing overview: Geometrical, statistical, and sparse regression-based approaches. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 5(2), 354–379 (2012)
13. Campos-Delgado, D.U., Gutierrez-Navarro, O., Rico-Jimenez, J.J., Duran-Sierra, E., Fabelo, H., Ortega, S., Callico, G., Jo, J.A.: Extended blind end-member and abundance extraction for biomedical imaging applications. IEEE Access 7, 178539–178552 (2019)
14. Cruz-Guerrero, I.A., Leon, R., Campos-Delgado, D.U., Ortega, S., Fabelo, H., Callico, G.M.: Classification of hyperspectral in vivo brain tissue based on linear unmixing. Appl. Sci. 10(16), 5686 (2020)
15. Sikkandar, M.Y., Jayasankar, T., Kavitha, K., Prakash, N., Sudharsan, N.M., Hemalakshmi, G.: Three factor nonnegative matrix factorization based he stain unmixing in histopathological images. J. Ambient Intell. Hum. Comput. 12(6), 6505–6513 (2021)
16. Lobanova, E., Lobanov, S.: Efficient quantitative hyperspectral image unmixing method for large-scale raman micro-spectroscopy data analysis. Anal. Chim. Acta 1050, 32–43 (2019)
17. Zhang, Y., Ni, Q.: Recent advances in quantum machine learning. Quantum Eng. 2(1), e34 (2020)
18. Signoroni, A., Savardi, M., Baronio, A., Benini, S.: Deep learning meets hyperspectral image analysis: A multidisciplinary review. J. Imaging 5(5), 52 (2019)
19. Litjens, G., Kooi, T., Bejnordi, B.E., Setio, A.A.A., Ciompi, F., Ghafoorian, M., Van Der Laak, J.A., Van Ginneken, B., Sánchez, C.I.: A survey on deep learning in medical image analysis. Med. Image Anal. 42, 60–88 (2017)
20. Huang, X., Chen, J., Chen, M., Wan, Y., Chen, L.: Fre-net: Full-region enhanced network for nuclei segmentation in histopathology images. Biocybern. Biomed. Eng. 43(1), 386–401 (2023)
21. Karthik, R., Menaka, R., Siddharth, M.: Classification of breast cancer from histopathology images using an ensemble of deep multiscale networks. Biocybern. Biomed. Eng. 42(3), 963–976 (2022)
22. Ortega, S., Halicek, M., Fabelo, H., Guerra, R., Lopez, C., Lejeune, M., Godtliebsen, F., Callico, G.M., Fei, B.: Hyperspectral imaging and deep learning for the detection of breast cancer cells in digitized histological images. In: Medical Imaging 2020: Digital Pathology, vol. 11320, pp. 113200V. International Society for Optics and Photonics, Bellingham, Washington DC (2020)
23. Ortega, S., Halicek, M., Fabelo, H., Camacho, R., Plaza, M.d.l.L., Godtliebsen, F., Callicó, G.M., Fei, B.: Hyperspectral imaging for the detection of glioblastoma tumor cells in h&e slides using convolutional neural networks. Sensors 20(7), 1911 (2020)
24. Collins, T., Bencteux, V., Benedicenti, S., Moretti, V., Mita, M.T., Barbieri, V., Rubichi, F., Altamura, A., Giaracuni, G., Marescaux, J., et al.: Automatic optical biopsy for colorectal cancer using hyperspectral imaging and artificial neural networks. Surg. Endoscopy 36(11), 8549–8559 (2022)
25. Ortega, S., Fabelo, H., Halicek, M., Camacho, R., de la Luz Plaza, M., Callicó, G.M., Fei, B.: Hyperspectral superpixel-wise glioblastoma tumor detection in histological samples. Appl. Sci. 10(13), 4448 (2020)
26. Maktabi, M., Wichmann, Y., Köhler, H., Ahle, H., Lorenz, D., Bange, M., Braun, S., Gockel, I., Chalopin, C., Thieme, R.: Tumor cell identification and classification in esophageal adenocarcinoma specimens by hyperspectral imaging. Sci. Rep. 12(1), 1–14 (2022)
27. Kora, P., Ooi, C.P., Faust, O., Raghavendra, U., Gudigar, A., Chan, W.Y., Meenakshi, K., Swaraja, K., Plawiak, P., Rajendra Acharya, U.: Transfer learning techniques for medical image analysis: A review. Biocybern. Biomed. Eng. 42(1), 79–107 (2022)
28. Fang, B., Bai, Y., Li, Y.: Combining spectral unmixing and 3d/2d dense networks with early-exiting strategy for hyperspectral image classification. Remote Sens. 12(5), 779 (2020)
29. Guo, A.J., Zhu, F.: Improving deep hyperspectral image classification performance with spectral unmixing. Signal Process. 183, 107949 (2021)
30. Ortega, S., Fabelo, H., Camacho, R., de la Luz Plaza, M., Callicó, G.M., Sarmiento, R.: Detecting brain tumor in pathological slides using hyperspectral imaging. Biomed. Opt. Express 9(2), 818 (2018)
31. Vabalas, A., Gowen, E., Poliakoff, E., Casson, A.J.: Machine learning algorithm validation with a limited sample size. PloS one 14(11), e0224365 (2019)
32. Barbedo, J.G.A.: Impact of dataset size and variety on the effectiveness of deep learning and transfer learning for plant disease classification. Comput. Electron. Agric. 153, 46–53 (2018)
33. Winter, M.E.: N-findr: An algorithm for fast autonomous spectral end-member determination in hyperspectral data. In: Imaging Spectrometry V, vol. 3753, pp. 266–275. International Society for Optics and Photonics, Bellingham, Washington DC (1999)
34. Gan, J., Liu, T., Li, L., Zhang, J.: Non-negative Matrix Factorization: A Survey. Comp. J. 64(7), 1080–1092 (2021)
35. Witten, I.H., Frank, E., Hall, M.A., Pal, C.J.: Chapter 12 - Ensemble learning. In: Witten, I.H., Frank, E., Hall, M.A., Pal, C.J. (Eds.) Data Mining (4th ed.), pp. 479–501. Morgan Kaufmann, Burlington, MA (2017)
36. Hastie, T., Tibshirani, R., Friedman, J.: Model Inference and Averaging. Springer, New York (2009)
37. Ho, Y., Wookey, S.: The real-world-weight cross-entropy loss function: Modeling the costs of mislabeling. IEEE Access 8, 4806–4813 (2019)

38. He, K., Zhang, X., Ren, S., Sun, J.: Delving deep into rectifiers: Surpassing human-level performance on imagenet classification. arXiv, p. 1502.01852 (2015)
39. Géron, A.: Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems. O'Reilly Media, Inc., Sebastopol, CA (2019)
40. Boughorbel, S., Jarray, F., El Anbari, M.: Optimal classifier for imbalanced data using Matthews Correlation Coefficient metric. PLOS ONE 12(6), e0177678 (2017)
41. Chicco, D., Jurman, G.: The advantages of the matthews correlation coefficient (mcc) over f1 score and accuracy in binary classification evaluation. BMC Genom. 21(1), 1–13 (2020)

How to cite this article: Cruz-Guerrero, I.A., Campos-Delgado, D.U., Mejía-Rodríguez, A.R., Leon, R., Ortega, S., Fabelo, H., Camacho, R., Plaza, M.D.L.L., Callico, G.: Hybrid brain tumor classification of histopathology hyperspectral images by linear unmixing and an ensemble of deep neural networks. Healthc. Technol. Lett. 11, 240–251 (2024). https://doi.org/10.1049/htl2.12084