



Contents lists available at ScienceDirect

Computer Methods and Programs in Biomedicine Update

journal homepage: www.sciencedirect.com/journal/computer-methods-and-programs-in-biomedicine-update



Unifying heterogeneous hyperspectral databases for *in vivo* human brain cancer classification: Towards robust algorithm development

Alberto Martín-Pérez^{a,1,*} , Beatriz Martínez-Vega^{b,1,*} , Manuel Villa^a, Raquel Leon^b ,
Alejandro Martínez de Terneró^a , Himar Fabelo^{b,c,d}, Samuel Ortega^{b,e,f}, Eduardo Quevedo^b ,
Gustavo M. Callico^{b,*}, Eduardo Juárez^{a,*}, César Sanz^a

^a Research Center on Software Technologies and Multimedia Systems (CITSEM), Universidad Politécnica de Madrid, Madrid, 28031, Spain

^b Research Institute for Applied Microelectronics (IUMA), University of Las Palmas de Gran Canaria, Las Palmas de Gran Canaria, 35017, Spain

^c Fundación Canaria Instituto de Investigación Sanitaria de Canarias (FIISC), Las Palmas de Gran Canaria, 35019, Spain

^d Research Unit, Hospital Universitario de Gran Canaria Dr. Negrín, Las Palmas de Gran Canaria, 35019, Spain

^e Nofima, Norwegian Institute of Food Fisheries and Aquaculture Research, Tromsø, NO-9091, Norway

^f Department of Mathematics and Statistics, UiT The Arctic University of Norway, Tromsø, NO-9091, Norway

ARTICLE INFO

Keywords:

Brain tumor
Hyperspectral imaging
Machine learning
Deep learning
Neurosurgery

ABSTRACT

Background and objective: Cancer is one of the leading causes of death worldwide, and early and accurate detection is crucial to improve patient outcomes. Differentiating between healthy and diseased brain tissue during surgery is particularly challenging. Hyperspectral imaging, combined with machine and deep learning algorithms, has shown promise for detecting brain cancer *in vivo*. The present study is distinguished by an analysis and comparison of the performance of various algorithms, with the objective of evaluating their efficacy in unifying hyperspectral databases obtained from different cameras. These databases include data collected from various hospitals using different hyperspectral instruments, which vary in spectral ranges, spatial and spectral resolution, as well as illumination conditions. The primary aim is to assess the performance of models that respond to the limited availability of *in vivo* human brain hyperspectral data. The classification of healthy tissue, tumors and blood vessels is achieved through the utilisation of different algorithms in two databases: *HELICoID* and *SLIMBRAIN*.

Methods: This study evaluated conventional and deep learning methods (*KNN*, *RF*, *SVM*, *1D-DNN*, *2D-CNN*, *Fast 3D-CNN*, and a *DRNN*), and advanced classification frameworks (*LIBRA* and *HELICoID*) using cross-validation on 16 and 26 patients from each database, respectively.

Results: For individual datasets, *LIBRA* achieved the highest sensitivity for tumor classification, with values of 38 %, 72 %, and 80 % on the *SLIMBRAIN*, *HELICoID* (20 bands), and *HELICoID* (128 bands) datasets, respectively. The *HELICoID* framework yielded the best *F1 Scores* for tumor tissue, with values of 11 %, 45 %, and 53 % for the same datasets. For the *Unified dataset*, *LIBRA* obtained the best results identifying the tumor, with a 40 % of sensitivity and a 30 % of *F1 Score*.

1. Introduction

In 2022, the World Health Organisation (WHO) globally registered around 321,731 new cases of brain and other Central Nervous System (CNS) tumors with 248,500 deaths [1]. CNS cancers have increased on a ten-year average annual percentage basis (2008–2017) by 0.7 % and 0.5 % in ages 0–14 and 15–19 respectively [2]. Physicians usually perform

an aggressive treatment including surgery, radiotherapy and chemotherapy, which overall prognosis is unfavorable since enhancing survival rates continues to be a major challenge [3]. Therefore, innovative approaches to improve the diagnosis of tumor tissue, to provide more information for surgical guidance and to improve the outcomes of the surgery are still needed.

A problem that arises during neurosurgical interventions is the so-

* Corresponding authors.

E-mail addresses: a.martinp@upm.es (A. Martín-Pérez), bmartinez@iuma.ulpgc.es (B. Martínez-Vega), gustavo@iuma.ulpgc.es (G.M. Callico), eduardo.juarez@upm.es (E. Juárez).

¹ These authors contributed equally to this work

<https://doi.org/10.1016/j.cmpbup.2025.100183>

Available online 5 February 2025

2666-9900/© 2025 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

called brain shift phenomena, which describes the deformation of the brain tissue structure as the brain moves after craniotomy [4]. Tools like neuronavigators, Intraoperative Magnetic Resonance (IMR), Intraoperative Ultrasound (IOUS), or drugs like 5-aminolevulinic acid (5-ALA) are commonly used to address the tumor brain location. However, each of these approaches present inherent limitations [5–8]. To provide a quicker alternative to address the brain shift problem, IOUS is frequently employed, however, it provides artifacts and low resolution images [9,10]. Another effective alternative is 5-ALA, but it is invasive and only suitable for high grade tumors in human adults [11,12]. Therefore, faster and non-invasive techniques are needed to address the tumor brain areas during surgical workflows.

Hyperspectral (HS) Imaging (HSI) is a technology which combines conventional imaging and spectroscopy. This combination provides simultaneous spectral and spatial information of an object [13]. HSI offers a wealth of information covering a large number of contiguous spectral bands, from the ultraviolet to the infrared spectral ranges. Thus, the reflectance measured in a pixel acts as a fingerprint, the so-called spectral signature, which allows characterizing the chemical composition of the material in that pixel [14]. HSI is a technique well suited for medical applications for its non-invasive, non-ionising and label-free nature. This technology has been evaluated in diverse applications, such as histopathological tissue analysis [15–17], head and neck cancer [18,19], colon and esophagogastric cancer [20–22], and brain cancer [23–29], among others.

In recent years, advances in HS sensors have enabled data acquisition through different techniques, including Scanning-Based (SB) and Wide-Field (WF) methods [30]. SB techniques involve acquiring the spectrum for each pixel (pushbroom cameras). Although these techniques provide extremely high spectral resolution, the acquisition time is high, which can take up to 1 min [31]. Besides, since the spatial dimensions are obtained by moving the Field-OfView (FOV) of the camera, spatial aliasing can occur as object movements can take place during the scanning. In contrast, WF methods capture the entire spatial scene in a single exposure with 2D sensors, which can be achieved by stepping through wavelengths filters to complete the HS cube (wavelength scan) or by simultaneously acquiring the spatial and spectral information (snapshot) [32]. These techniques do enable real-time solutions by providing a live sequence of HS images. However, the spatial and spectral resolution is much lower. Since it is important to have both good resolution and low acquisition time during brain tumor interventions, there is no specific HS sensor to use. Sancho et al. [33] used HS snapshot cameras and Fabelo et al. employed pushbroom cameras [34].

HSI, combined with Machine Learning (ML) algorithms has been used to differentiate pathological from healthy tissue in *in vivo* human brain [28], offering another intraoperative tool to address the tumor brain area problem. For example, several studies have differentiated Glioblastoma (GB) from healthy tissue using HSI and ML under different scenarios, either at a macroscopic scale [31,33] or by coupling HS cameras to intraoperative microscopes [27,35]. Furthermore, other researchers have used HS data to train Deep Learning (DL) models and classify tumor [24,29,36,37]. These studies acquire HS images from the surface of the brain, using reflected light to reveal diagnostic information related to tissue pathology.

The aim of this work is to compare and analyse the performance of different classification algorithms performed on two independent and unified databases: traditional ML, DL approaches and complex frameworks. The rest of the study is structured as follow: Section 2 describes the acquisition systems, the proposed general processing framework, and the classification algorithms to be evaluated. Section 3 presents the obtained results and the discussions. Finally, in Section 4 the conclusions and the next steps to be followed in the future are explained.

2. Materials and methods

A comprehensive overview of two independent brain HS databases,

namely *HELICoiD* and *SLIMBRAIN* is presented in Section 2.1, providing a summary of the number of labelled pixels for each database in Section 2.2.4. The data management performed to pre-process the HS cubes of both databases, as well as the data downsampling method used are detailed in Section 2.2. In Section 2.2.2 the data partitioning and optimization of ML are explained. Then, in Section 2.3 the various ML classification algorithms are described. Finally, the evaluation metrics used to assess the performance of the classification models are presented in Section 2.4.

2.1. Data acquisition

2.1.1. *HELICoiD* database

The system employed for the acquisition of the *HELICoiD* database was developed during a European research project under the same name [31]. In the course of this project, 36 *in vivo* brain HS images from 22 different patients were obtained. The description of the acquisition system used is described in further detail in [38], but the data used for this work was taken from a pushbroom camera gathering information within the visual near infrared (VNIR), specifically within the 400-1000 nm. The study protocol and consent procedures were approved by the *Comité Ético de Investigación Clínica-Comité de Ética en la Investigación* (CEIC/CEI) of the University Hospital of Gran Canaria Doctor Negrin (CEIC code: 130069) and the *National Research Ethics Service (NRES) Committee South Central Oxford C for the University Hospital of Southampton*. Written informed consent was obtained from all the subjects.

Ground Truth (GT) annotations were performed by combining pathological assessment with neurosurgical criteria and the use of the Spectral Angle Mapper (SAM) algorithm. The labelled pixels were obtained using a customised labelling tool developed to this end [38]. Three different classes were defined: Healthy Tissue (HT), Tumor Tissue (TT) and Blood Vessels (BV).

The dataset employed in this study was collected during two acquisition data campaigns carried out at the University Hospital of Gran Canaria Doctor Negrin (Spain) [38]. Composed of 26 HS images obtained from 16 adult patients, this dataset includes six HS images from four different patients diagnosed with GB. The remaining patients were affected by other tumor types or underwent surgery due to other diseases. From these latter patients, only samples of HT and BV were registered and employed in this study. Fig. 1 shows three HS images captured from subjects affected by GB within this dataset.

2.1.2. *SLIMBRAIN* database

The *SLIMBRAIN* database is a multimodal image database created to identify *in vivo* human brain tumors [39]. The main types of images available include HS, RGB and depth information for more than 193 patients. The guidelines of the Declaration of Helsinki were followed and the acquisition of HS images was approved by the *Research Ethics Committee* of Hospital Universitario 12 de Octubre, Madrid, Spain (protocol code 19/158, 28 May 2019).

Although the acquisition system used to acquire the database is described in [39], the primary camera utilized was a near-infrared (NIR) HS snapshot camera, capable of gathering 25 spectral bands within the near-infrared (NIR) range, specifically, within the 660–952 nm. Furthermore, the procedure to label the HS data is the same as followed for the *HELICoiD* database, described in Section 2.1.1. For this work, 26 HS images from 26 adult patients were used, all diagnosed by GB. However, only 13 images have pathological pixels labelled since they have the tumor on the surface. Hence, the other 13 patients are used to only obtain HT and BV samples. Fig. 1 shows an example of 3 HS images of this dataset with their corresponding GT maps overlaid.

2.2. Data management

Fig. 2 illustrates the general processing chain followed in this study. The pre-processing framework differs based on the HS database

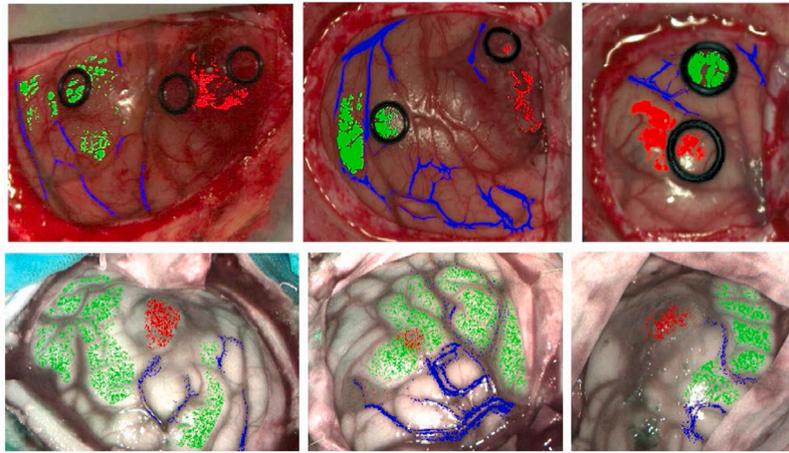


Fig. 1. Synthetic RGB images with overlapped ground truths of three patients from the *HELICoID* (top row) and *SLIMBRAIN* (bottom row) databases affected by GB. Green, red, and blue pixel colors indicate healthy tissue, tumor tissue, and blood vessels, respectively.

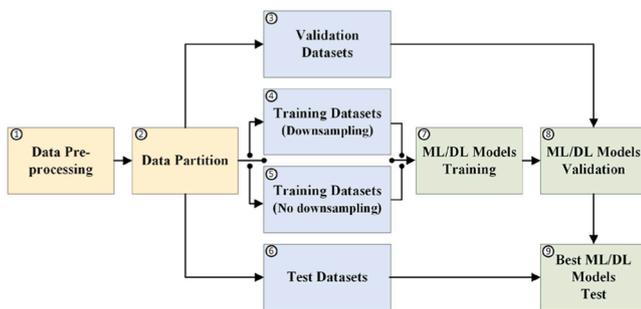


Fig. 2. Generic processing chain block diagram for classification using ML and DL models in each databases. The yellow boxes [1–2] indicate the pre-processing stage and the partitioning method closest to the clinical case. Blue boxes [3–6] represent the datasets obtained for train, validation and testing. Green boxes [7–9] illustrate the ML and DL algorithms used for the classification.

employed and will be detailed in the following section.

2.2.1. Pre-processing chain

This section explains in detail the pre-processing block used in each database (Fig. 2 (1)), which is divided into three general steps: calibration, filtering, and removal of extreme bands. First, after HS data was captured, the raw radiance (I_{raw}) from the HS cameras was calibrated using Eq. 1 to obtain the reflectance R . Obtaining R requires white (I_{white}) and dark (I_{dark}) reference images taken under the same conditions as I_{raw} . The I_{white} of both databases was acquired with a Spectralon tile (SphereOptics GmbH, Herrsching, Germany) and the I_{dark} was acquired by keeping the camera shutter closed. Both I_{white} and I_{dark} are required in the calibration process to minimize the influence of the light source on the sample and to remove the electrical noise of the HS sensor, respectively.

$$R = \frac{I_{raw} - I_{dark}}{I_{white} - I_{dark}} \quad (1)$$

Then, after calibrating the images, the data from the snapshot camera (*SLIMBRAIN* database) require a spectral correction to be applied using Eq. 2. This process must be done to deal with crosstalk in the sensor due to fluctuations in the incident light angles. To mitigate this issue, the manufacturer provides a spectral correction matrix (SCM) in Eq. 2 for adjusting the response of the snapshot sensor. Eq. 2 ends up producing the spectrally corrected reflectance of the snapshot HS images, R_{sc} .

$$R_{sc} = R_{snapshot} \times SCM \quad (2)$$

Then, R and R_{sc} are spectrally filtered to reduce noise. A Gaussian smoothing filter and a HySime filter, introduced by Bioucas-Dias et al. [39,40], were used for the *HELICoID* and the *SLIMBRAIN* HS images, respectively. Specifically, the first stage of the HySime algorithm was applied, in which the noise is estimated. In this process, each band of the HS image is assumed to be a linear combination of the rest of the bands, and thus the noise is the error in the linear combination. The estimated noise is then subtracted from the original HS image.

The extreme bands from the *HELICoID* data are then removed, as they are noisy due to the low performance of the sensor in that region. Thus, the spectral range available in that database is 440–902 nm. Hence, to evaluate classification algorithms under similar conditions with the two databases, the latest bands between 900 and 950 nm were removed from the *SLIMBRAIN* database data. Then, for the *HELICoID* database only, after the removal of the extreme bands, leaving a total of 645 spectral bands, the spectral wavelengths are decimated from 645 to 128 bands. In [41], it is shown how the decimation of bands is performed, where the average classification performance does not decrease until 128 bands are reached. Furthermore, an interpolation is also performed to have exactly the same wavelengths as in the *SLIMBRAIN* database. Finally, for the three datasets (*H20*, *H128* and *SLIMBRAIN*), each spectral signature is independently adjusted to the range between 0 and 1 by applying a min-max normalization.

2.2.2. Data partition

The datasets are divided into training, validation, and test sets using K-fold cross-validation at patient level to minimize bias and enhance model robustness [31,34,36,42] (Fig. 2. (2) to (6)). The *H20* and *H128* datasets, each with 16 patients, include four TT class patients. These datasets are partitioned into 4-folds, ensuring patients with all classes (HS images) appear in all sets, while patients with only HT and BV are included only in training sets (see the supplementary material document, Section 1). Patient identifiers use the format: patient number and capture number (e.g., 1C2). Detailed partitioning is in Table S1 of the supplementary material.

The *SLIMBRAIN* dataset includes 26 patients, 13 with TT class. To avoid bias and reduce computation time, an 8-folds partition is employed. Remaining patients are used only in training folds. The *SLIMBRAIN* dataset training captures use a similar nomenclature, with exposure time indicated (e.g., 1C2). Detailed partitioning is in Table S2 of the supplementary material. All tumor-labeled HS images are included in at least one set.

The *Unified dataset* combines the *H20* and *SLIMBRAIN* datasets (42

patients, 17 with TT class) and requires at least 10-folds for partitioning. Detailed partitioning is in Table S3 of the supplementary material.

2.2.3. Training data downsampling

Supervised classifiers rely on data quality and quantity for effective training, but unbalanced data or redundancy can potentially slow down training and degrade classification performance. In order to mitigate these limitations, the downsampling methodology proposed by Martínez et al. [41] is employed (Fig. 2 (4)). This methodology is based on an unsupervised K-Means clustering applied jointly with the SAM metric. In particular, the pixels of the training datasets are divided into three groups corresponding to the three classes (HT, TT and BV). K-Means clustering is applied independently to each group of labelled pixels to obtain 100 different clusters per class (300 clusters in total). This cluster number is used in order to match the number of clusters presented in [41] that serves as a basis for our study.

The training data reduction is executed twice: once for each independent datasets and once for the *Unified dataset*. Employing SAM, centroids identify the most representative pixels of each class. For the independent databases, the 34 most similar pixels per centroid are selected, totaling 3,400 pixels per class, based on the minimum TT class pixels (3,405). However, for the *Unified dataset*, the 115 most similar pixels per centroid are selected, totaling 11,500 pixels per class, based on the minimum TT class pixels (11,552). Train, validation, and test partitions are detailed in Section 2.2.2. Classification models are also trained with all pixels for performance comparison.

2.2.4. Summary datasets

This study analyzes four datasets from the two databases: (i) *H128* with 128 bands (440–902 nm) from the *HELICoID* database, (ii) *H20* with 20 bands (660–900 nm) from the *HELICoID* database for comparison with *SLIMBRAIN*, (iii) *SLIMBRAIN* with 20 bands identical to *H20*, and (iv) *Unified dataset* combining *H20* and *SLIMBRAIN*. Fig. 3 shows the mean spectral signatures for the TT class after the preprocessing described in Sub-sub Section 2.2.1 for the first three datasets.

The spectral signatures of the *SLIMBRAIN* dataset exhibit high standard deviation within pixels of the same class. Hahn et al. [43] found that the Fabry-Pérot sensor's spectral correction matrix is insufficient for accurate spectrum reconstruction, leading to significant measurement errors. They suggested creating an individual matrix postcharacterization, which requires a dedicated optical system not available during the creation of the *SLIMBRAIN* database. However, the mean spectral signatures of the *H128* and *H20* datasets are smoother than those of the *SLIMBRAIN* dataset. Reflectance differences between *H128* and *H20* result from the normalization step performed after

selecting the number of bands and spectral range. Further information about the amount of pixels available from each database as well as the spectral signatures for the HT and BV tissues are presented in Table S4 and Fig. S2.

2.3. Model optimization and classification algorithms

The next step in the generic processing chain is the ML and DL models evaluation (Fig. 2 (7)–(9)). This section explains the classification models used in the experiments. First, Fig. 4 shows the process that the supervised classifiers perform in order to obtain the optimized hyperparameters.

Models were evaluated using a validation set during hyperparameter optimization, regardless of dataset and downsampling (Section 2.2.3). Although grid and random search are popular for hyperparameter tuning, they focus more on exploitation than exploration. Hyperparameter optimization for ML models is computationally intensive, but Bayesian Optimization (BO) improves effectiveness [44]. Using a surrogate function reduces computational costs [45]. Therefore, Sequential Model-Based Optimization (SMBO), a BO formalization, was employed to optimize hyperparameters. SMBO maximized the weighted average area under the ROC curve (ROC-AUC) for all class pairwise combinations. Optimized models were evaluated on the validation and test sets using metrics from Section 2.4. The evaluated classifiers included K-Nearest Neighbors (KNN), Random Forest (RF), Support Vector Machines (SVMs), 1D deep neural network (1D-DNN), 2D convolutional

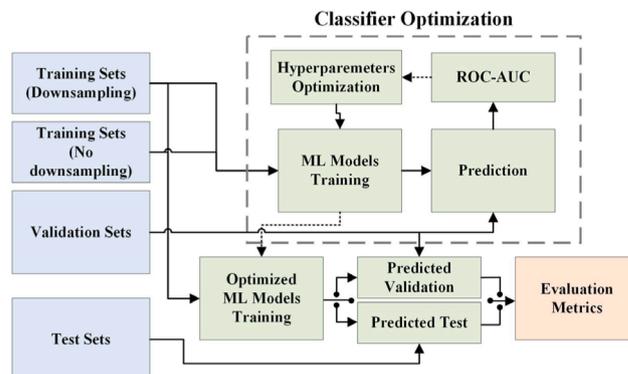


Fig. 4. Block diagram of the process used to train, optimize, and evaluate ML classification models, with the aim of finetuning and improving hyperparameters.

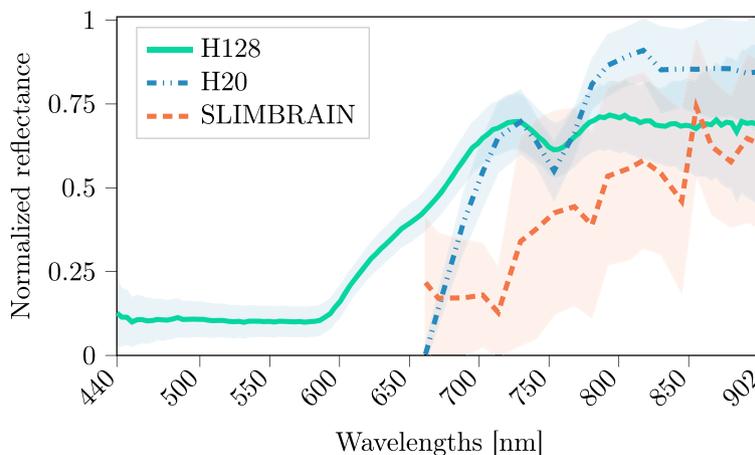


Fig. 3. Mean and standard deviation of the spectral signatures from tumor tissue including all patients for each of the 3 datasets employed in this study. The graph displays the mean with the central curves and the standard deviation with the shaded region.

neural network (2D-CNN), 3D convolutional neural network (3D-CNN), and a recurrent neural network (RNN), commonly used in medical HS imaging [31,34,36,42,48,50].

2.3.1. Supervised learning algorithms

This section describes the three supervised ML algorithms used in this work. First, SVM discriminates samples by predicting the optimal hyperplane in an n -dimensional space. Common kernels used in the state-of-the-art include linear and Radial Basis Function (RBF) [24,25,34]. Second, RF uses the bootstrapping resampling technique to construct a decision tree ensemble, with final predictions based on the majority label [46]. RF has been used to classify *in vivo* human brain images in recent studies [24,25]. Finally, KNN classifies samples by proximity to neighboring data points [47] using metrics like Euclidean and Cosine distances, previously used for the classification of brain tumors using HS data [31]. The hyperparameters of these algorithms are optimized as detailed in Table S5 of the supplementary material.

2.3.2. Deep learning algorithms

In this study, four DL models are employed. First, the *1D-DNN* and *2D-CNN* models described in [36] are used, which operate on spectral information alone or utilises both spectral and spatial information, respectively. Furthermore, the *Fast 3D-CNN* model outlined in [48] is used as well, which allows for the learning of combined features from the spatial and spectral dimensions [49]. Finally, a recent deep RNN (*DRNN*) model is employed, as outlined by Cebrian et al. [50]. These RNN models are founded on the interrelationships between the spectral extent of the HSI, as articulated in [51]. For the *2D-CNN*, *Fast 3D-CNN*, and *DRNN* models, data are structured as in Fabelo et al. [36], with 11×11 pixel patches centred on each labelled ground truth pixel, allowing overlap, and with a batch size of 12. It is noteworthy that all models employ the Adam optimiser, with 40 epochs for the *1D-DNN*, 50 for the *2D-CNN*, *Fast 3D-CNN*, and *DRNN* as done in previous works [36]. The learning rate (lr) was optimised in accordance with Table S5 in the supplementary material. Python code with the implementation for the DL models used is also available in the supplementary material.

2.3.3. Classification frameworks

This study analyzes two classification frameworks, *HELICoiD* and *LIBRA*, designed to enhance ML and DL classification results by combining various algorithms. The *HELICoiD* framework combines spatial and spectral features of HS images using supervised and unsupervised ML methods. It involves four main steps: dimensional reduction, spatial-spectral supervised classification, unsupervised clustering segmentation, and majority voting. PCA-guided KNN filtering is used for spatio-spectral homogenization, and the filtered classification map is combined with an unsupervised segmentation map through majority voting to produce the final thematic map. The *HELICoiD* framework utilizes a supervised linear SVM with optimized hyperparameters for each training group, as described in further detail in [31].

On the other hand, *LIBRA* is a hierarchical classification framework aiding neurosurgeons in brain tumor surgeries, as described by Villa et al. [52]. It employs multiple SVMs and ensemble methodologies to improve classification. *LIBRA* filters relevant samples into groups and uses genetic algorithms to implement Breiman's bagging process [53]. This filtering is accomplished by generating groups of samples ($N = 20$) belonging to each class. Since our problem has three classes, a total of 60 groups are generated. Then, genotypes are iteratively selected, using $M = 10$ generations with $G = 10$ genotypes each, to find the one that maximizes the *F1 Score* of the tumor class, thus, facilitating efficient classification with reduced complexity.

2.4. Performance metrics

The evaluation of the algorithms includes both qualitative and quantitative analysis. Four classification metrics, accuracy (*ACC*),

sensitivity (*SEN*), *F1 Score*, and Matthews Correlation Coefficient (*MCC*) are used to compare algorithms trained on different datasets. These metrics require true positives (*TP*), false positives (*FP*), true negatives (*TN*), and false negatives (*FN*) from the confusion matrix, as shown in Eqs. 3–6. Additionally, specificity (*SPE*), and *ROC-AUC* metrics are provided in the supplementary material (Equations S1 and S4).

$$ACC = \frac{TP + TN}{TP + TN + FP + FN} \quad (3)$$

$$SEN \text{ or } (TPR) = \frac{TP}{TP + FN} \quad (4)$$

$$F1 \text{ Score} = \frac{2 \times TP}{2 \times TP + FP + FN} \quad (5)$$

$$MCC = \frac{TP \times TN - FP \times FN}{\sqrt{(TP + FP) \times (TP + FN) \times (TN + FP) \times (TN + FN)}} \quad (6)$$

Qualitative analyses include classification maps for ML and DL algorithms, probability maps for *LIBRA*, and density maps for *HELICoiD* framework. Classification maps use colors to indicate the class with the highest probability per pixel, while probability maps display a mix of colors based on class probabilities, and density maps show color distribution based on pixel count within K-means clusters. All maps use green, red, and blue to represent HT, TT, and BV classes, respectively. Quantitative metrics for complex frameworks are evaluated before any smoothing or filtering procedures are applied to the maps.

3. Results

Bar plots with 95 % confidence intervals illustrate the mean results from K-fold cross-validation (Section 2.2.2). The validation metrics *ACC*, *SEN*, and *F1 Score* for the TT class are used to validate the main focus of the study. Results for the additional metrics and classes are in the supplementary material (Figs. S3-S7), while Tables S6-S10 show the optimized hyperparameters using the validation set across datasets. Sections 3.2 and 3.3 present quantitative and qualitative results for the best models. Further test metrics (Fig. S8 and Fig. S9) and thematic maps (Fig. S10 to Fig. S20) are also in the supplementary material. Finally, Section 3.4 shows results from unifying the *SLIMBRAIN* and *H20* datasets.

3.1. Quantitative validation results

The validation results obtained for the TT class are illustrated in Fig. 5. It shows four bar plots representing the mean results for the *ACC*, *SEN*, *F1 Score*, and *MCC* metrics, respectively. Three colors are used to represent each dataset, orange, blue, and green for the *SLIMBRAIN*, *H20*, and *H128* datasets, respectively. On one hand, light colors with diagonal lines starting from the bottom left to the top right indicate that all pixels of the training set have been employed. On the other hand, darker colors with diagonal lines starting from the top left to the bottom right indicate that the training samples have been downsampled as specified in Section 2.2.3. Additionally, the bars are grouped according to the classifier used, detailed in the X-axis.

Downsampling generally results in an increase in the mean value of all the metrics in most cases. Exceptions along with *SEN* values are minimal, and decreases in the mean value are less than 7 %. This implies that downsampling improves the results in validation and has a negligible impact on the mean outcomes.

Nevertheless, downsampling means that the dispersion of the results of all metrics across partitions increases, regardless of the dataset under inspection. This may be because reducing the number of training pixels causes algorithms not having enough samples to generalize well to new patients, as training pixels get reduced to 3,400 pixels per class. Even so, it is observed that the mean value of any metric increases, suggesting

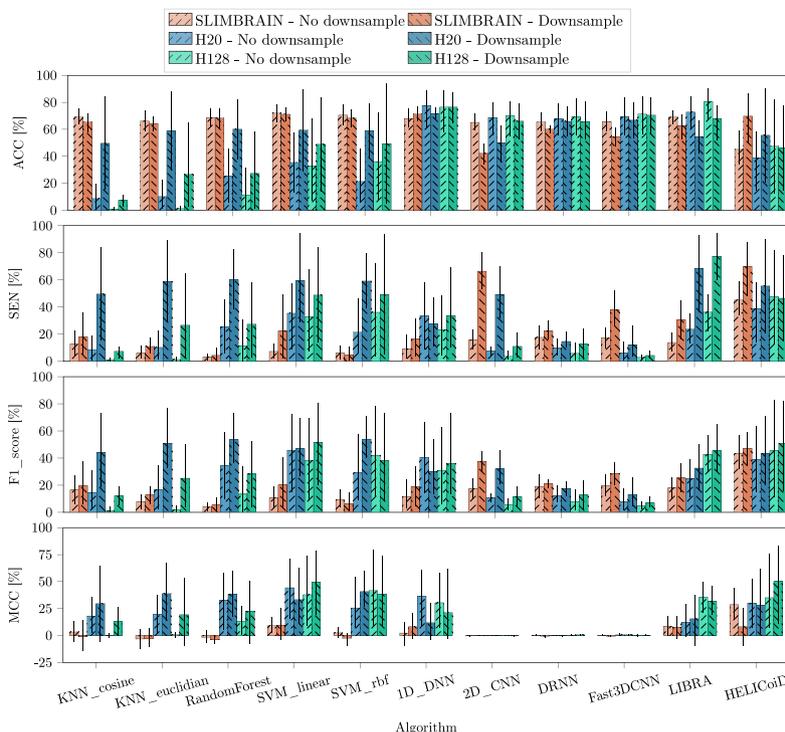


Fig. 5. Bar plots illustrating the mean and confidence interval of 95 % with the results obtained on the validation sets after training with and without downsampling for the *SLIMBRAIN*, *H20*, and *H128* datasets. Plots illustrate the *ACC*, *SEN*, *F1 Score*, and *MCC*, respectively, for the tumor tissue.

that downsampling eliminates noisy samples that have a direct influence on a worse classification

After evaluating the performance of several classifiers in each of the datasets, it is concluded that first, although the dispersion increases after downsampling, there is a noticeable improvement in classifier performance in all datasets. Second, in the context of the traditional algorithms, the *SVM_linear* exhibits the best performance considering the three datasets and the results obtained. The best *SEN* value is reached in all three datasets, being $\approx 20\%$ in *SLIMBRAIN*, $\approx 60\%$ in *H20*, and $\approx 50\%$ in *H128*. Regarding the *F1 Score*, it is observed that with the *H128* dataset the best value reaches $\approx 52\%$. With the *SLIMBRAIN* and *H20* datasets, the classifiers that obtain the best results are *KNN_Euclidean* and *SVM_RBF*, respectively. Moreover, the best results obtained for *MCC* are also obtained with *SVM_linear* for the *SLIMBRAIN* and *H128* datasets, with $\approx 10\%$ and $\approx 50\%$ mean values, respectively. However, for the *H20* dataset the *SMV_RBF* classifier obtained the highest mean value of $\approx 35\%$ after performing downsampling, slightly lower than that obtained with *SVM_linear* when not performing downsampling. It is important to note that the disparity between the classifiers is small, being 7 % for *SLIMBRAIN* and 6 % for *H20* dataset. Third, with respect to the DL classifiers, the best performing classifier is *1D_DNN*, since it has obtained better mean results across the three datasets regardless of the metric under evaluation. It is important to note that the elevated values obtained for *SEN* with the *2D_CNN* algorithm are attributable to the fact that it predicted the majority of the pixels as tumor, as in *F1 Score* the value reach around a 40 %, which takes into account the FP. A notable observation regarding the *MCC* metric for all networks except *1D_DNN* is the proximity of its mean values to 0. Subsequent investigation into the underlying causes of this phenomenon has revealed that these networks, given their incorporation of spatial information during training, appear to encounter difficulties in attaining effective learning and generalization capabilities, hence influencing the results obtained. This low performance could be related not only to the inherent limitations of the network architectures, but also to the limited number of patients available during the training process. Finally, the *HELICoiD* and *LIBRA* frameworks are also evaluated in the test evaluation in Section 3.2.

3.2. Quantitative test results

Based on the validation results, the performance of the *SVM_linear*, the *1D_DNN* algorithm and both classification frameworks generated in the downsampling validation process will be examined using the test datasets. For this case, the *ACC*, *SEN*, *F1 Score*, and *MCC* metrics are provided for all the classes in Fig. 6. These bar plots indicate with colors the different datasets employed, where each group of bars indicate the tissue under analysis. Furthermore, the different patterns in the bars indicate each classifier.

Most algorithms have obtained better results with the *H20* and *H128* datasets than with the *SLIMBRAIN* dataset. This is likely due to the noise introduced by the snapshot sensor. Both *LIBRA* and *HELICoiD* frameworks achieve better average results in the TT class compared to *SVM_linear* or *1D_DNN*. This suggests that when dealing with a dataset like *SLIMBRAIN*, which might have higher noise on its measurements, it is essential to employ more complex algorithms that can better leverage the data. Interestingly, with the *H20* and *H128* datasets, complex algorithms do not show substantial improvements compared to other classifiers.

For the HT and BV classes with the *H128* dataset, the results of the classifiers are generally high. With this dataset, minimum mean values of 80 % are reached for all classes and for all classifiers (with the exception of *LIBRA*). As for the *H20* dataset, these tend to move between average values of 30–50 %, with the exception of *LIBRA* in the BV class. With respect to the *SLIMBRAIN* dataset, it has a little more variation depending on the classifier used. With the *1D_DNN* classifier, between 12 % and 75 % is obtained, with the *SEN* and *F1 Score* metrics obtaining the best values. Regarding the TT class, the classifiers show a higher variability between each other for all datasets, from 10 % to 80 % on the mean values.

Further analysis of the classification frameworks reveals that the *HELICoiD* framework generally has higher average values than the rest of the classifiers for both the *H20* and *H128* datasets in the HT class. Results with the *SLIMBRAIN* dataset reveals that *LIBRA* is the best classifier for identifying healthy pixels. For *ACC* and *SEN* metrics, *LIBRA*

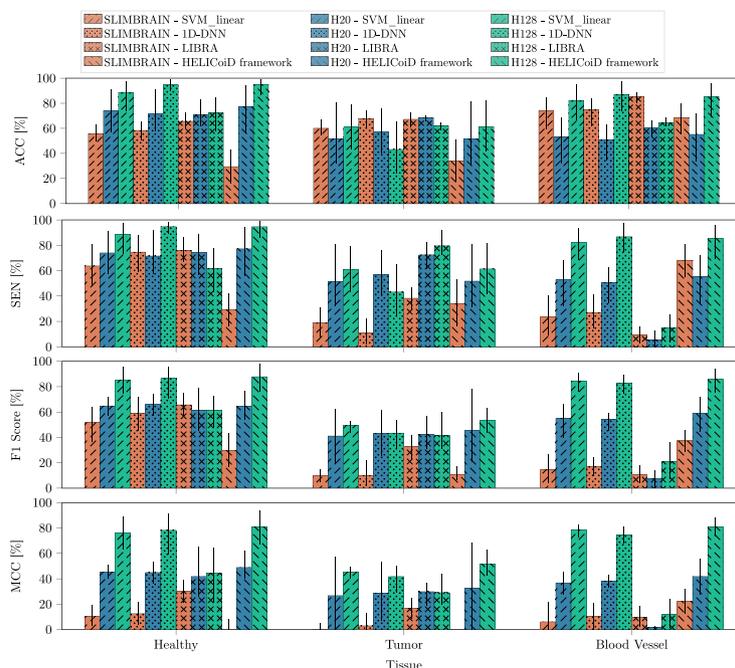


Fig. 6. Bar plots illustrating the mean and confidence interval of 95 % with the results obtained on the test sets after training with downsampling for the *SLIMBRAIN*, *H20*, and *H128* datasets. Plots illustrate the *ACC*, *SEN*, *F1 Score*, and *MCC*, respectively, for the healthy tissue, the tumor tissue, and for the blood vessel tissue.

is the best for all three datasets in TT. For *F1 Score*, the *HELICoiD* framework is slightly better than any other classifier for any dataset.

3.3. Qualitative test results

This section presents the thematic maps corresponding to the models evaluated in Section 3.2 obtained from the analysis of the three datasets. As an example, qualitative results from two patients for each dataset are presented in Fig. 7, including their GT overlaid on the RGB image and the thematic maps of the classifiers. The rest of the thematic maps can be found in the supplementary material (from Fig. S10 to Fig. S19). The thematic maps are composed by classification maps (ML and DL algorithms), probability maps (*LIBRA* framework) and density maps (*HELICoiD* framework). Black rectangles in the images illustrate the areas near the tumor pixels labeled by the neurosurgeons. Thematic maps show HT, TT, and BV tissues in green, red, and blue colors, respectively.

SLIMBRAIN dataset show that none of the algorithms were able to delineate the TT from the HT. This fact is observed when looking inside the black rectangles of either capture F1-67C1 or F5-71C2, where most algorithms classified TT as HT. Furthermore, *SVM_linear* shows more false positives from the TT class on its classification map outside the rectangle at the bottom right compared to *1D_DNN*, *LIBRA* and *HELICoiD* frameworks. Observing capture F5-71C2, *SVM_linear* and *1D_DNN* both encounter problems when classifying BV located beyond the lower-right corner of the black rectangle, classifying most pixels as BV or TT instead of HT tissue. However, it can be seen that *LIBRA* and *HELICoiD* frameworks are able to better delineate these BV pixels, introducing less BV false positives.

Regarding the *H20* dataset, patients F1-12C1 and F3-20C1 are shown respectively. It can be seen at first glance that, with respect to F1-12C1, *SVM_linear* and *1D_DNN* are able to identify the delimitation of the tumor region within the black rectangle, while *LIBRA* visualize a small area and *HELICoiD* framework clearly identifies the HT class and some areas of the TT class. *SVM_linear* and *1D_DNN* identify some small areas of the TT class of the F3-20C1, having a better delimitation *1D_DNN*. However, *LIBRA* and *HELICoiD* frameworks have not been able to identify all the tumor region.

As for the *H128* dataset, it is observed that *SVM_linear* for F1-12C1

identifies the TT class that is within the black rectangle, however, it also shows a large number of false positives. For F3-20C1, it is not able to identify the tumor region. Observing the *1D_DNN* classifier, the tumor region is identified for F1-12C1 but for F3-20C1 only HT class is identified. However, a minimum false positives are provided. *LIBRA* is not able to visualize the TT class while the *HELICoiD* framework shows the TT clearly and with minimum amount of false positives in the tumor region of F1-12C1. None of the algorithms were able to identify TT for the F3-20C1 capture using *H128* dataset.

Generally, *SVM_linear* and *1D_DNN* classifiers have difficulties identifying blood vessels. This behaviour can be qualitatively observed with probability maps using *SLIMBRAIN* presented from Fig. 7 (b)–(j), where the complex frameworks can better delineate the BV presented in Fig. 7 (a) and (f). Furthermore, the *HELICoiD* framework not only segments veins in *SLIMBRAIN* more effectively with the classification metrics, but also performs as the best algorithm in the *H20* dataset. It is worth noting that the *SLIMBRAIN* or the *H20* datasets lack the highest absorption peaks of the oxygenated hemoglobin (HbO₂) at $\lambda = 542$ nm and $\lambda = 572$ nm, which might be useful to identify the different classes with more precision. Comparing the classification maps obtained with the *H20* dataset (see Fig. 7 (b)–(j)) against those obtained with the *H128* dataset (see Fig. 7 (b)–(j)), *SVM_linear*, *1D_DNN*, and the *HELICoiD* framework are able to better delineate the BV when *H128* is used.

3.4. Test results using the unified dataset

The bar plots presented in Fig. 8 illustrate the mean and confidence interval of 95 % for the performance metrics obtained on test sets after training with and without downsampling on the *Unified dataset*. The plots display the *ACC*, *SEN*, *F1 Score*, and *MCC* for HT, TT, and BV tissue across different algorithms: *SVM_linear*, *1D_DNN*, *LIBRA*, and the *HELICoiD* framework, both with and without downsampling.

A notable observation is that downsampling appears to have a varying impact on different algorithms and tissues. For HT, all algorithms exhibit high *ACC*, ranging from approximately 75 % to 85 %, with downsampling having minimal effect on performance. *SEN*, *F1 Score* also remain consistently high across algorithms for HT, with *SEN* values around 80 % and *F1 Scores* between 70 % and 80 %, indicating

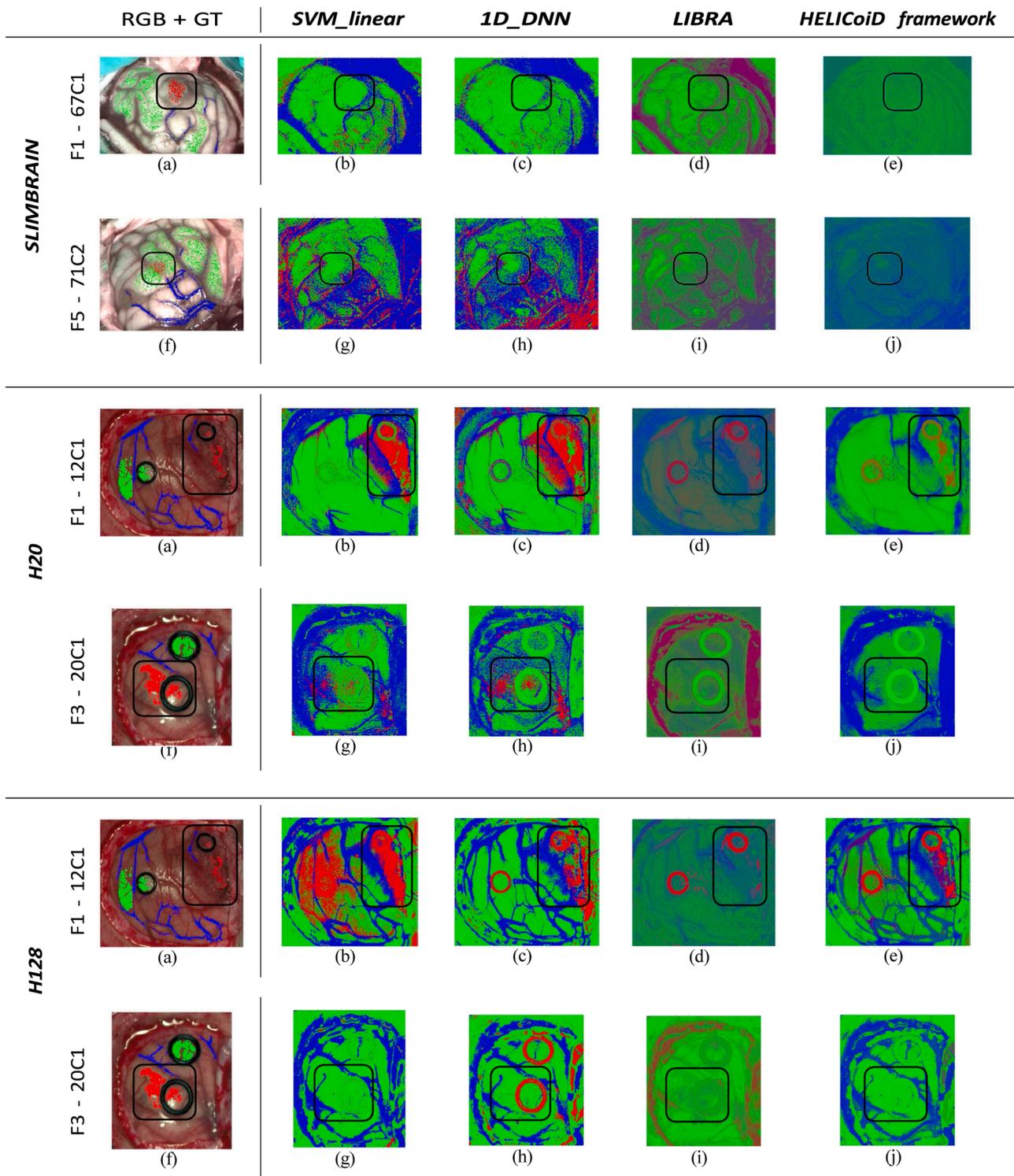


Fig. 7. Thematic maps obtained with the classifiers evaluated on the test sets for each dataset. Black rectangles indicate the location of the tumor.

robust detection capability without substantial loss due to down-sampling. However, *MCC* values remain low, below to 30 %. Conversely, for *TT*, down-sampling significantly affects the performance metrics. Algorithms without down-sampling generally show poor *SEN*, with values below 10 %, *F1 Scores* around 10 %, and *MCC* close or below to 0 %, reflecting challenges in detecting tumors without balancing the dataset. However, algorithms like *LIBRA* demonstrates improved *SEN* (around 45 %) and *F1 Scores* (approximately 35 %) with down-sampling, suggesting its efficacy in handling class imbalance for tumor detection.

Generally, *BV* tissue exhibit reduced performance when down-sampling is done, regardless of the algorithm and metric observed.

Importantly, no thematic maps are presented since they mostly identify *HT* and *BV*, illustrating difficulties classifying *TT*. However, some examples are presented in Fig. S20 available in the supplementary material.

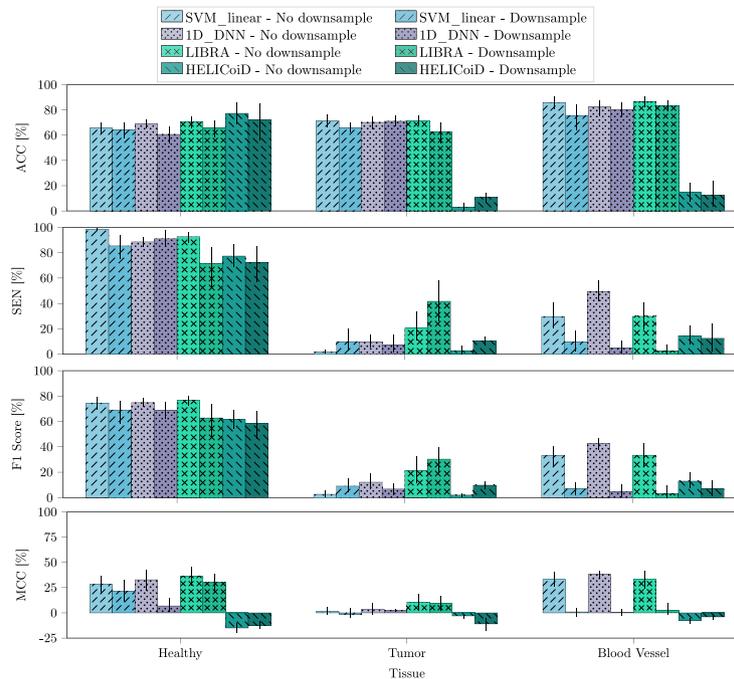


Fig. 8. Bar plots illustrating the mean and confidence interval of 95 % with the results obtained on the test sets after training with and without downsampling when using the *Unified dataset*. Plots illustrate the *ACC*, *SEN*, *F1 Score*, and *MCC*, respectively, for the HT, the TT, and for the BV tissue.

4. Conclusions

This paper presents a study with the performance of classification algorithms for brain tumor detection in two independent *in vivo* brain HS databases, namely *HELICoID* and *SLIMBRAIN*. For fair comparisons between both datasets, classification models were trained and evaluated with 20 spectral bands within the 660–900 nm range, hence matching the spectral information captured with the HS pushbroom camera used to collect *HELICoID* with the data measured with the HS snapshot camera utilized for *SLIMBRAIN*. We also used all 128 bands with spectral information between 450 and 900 nm as reference, since they were initially included in the *HELICoID* HS cubes. This leads to three datasets being employed, *SLIMBRAIN*, *H20* with 20 bands, and *H128* with 128 bands. Quantitative results in the validation sets indicate that, in general, classification models demonstrate improved average performance on the TT class, but with an increased result dispersion, when downsampling the training set to balance the classes in the datasets and accelerate the training process.

However, the performance is limited due to several factors. Primarily, the limited number of samples results in a restriction of the variability of the data. While the balancing of the data indeed improves the results, it also implies a reduction in the data and, consequently, in the variability. Another limitation is the spectral range, with *H128* providing superior results in comparison to *H20*, emphasizing the necessity for the identification of relevant spectral ranges in this particular application.

Moreover, an analysis of the qualitative results with the thematic maps yielded insights into the impact of these factors. Results for the test sets may suggest that the models encounter challenges in extrapolating to unseen patients. Overall, results show large inter-patient variability, specially true when evaluating models on the *SLIMBRAIN* dataset. Nonetheless, complex classification frameworks like *LIBRA* appear to more effectively harness data from *SLIMBRAIN* compared to conventional ML algorithms or DL approaches. This is particularly evident in the identification of blood vessels, despite the fact that crucial absorption peaks of HbO₂ at $\lambda = 542$ nm and $\lambda = 572$ nm are not present in the *SLIMBRAIN* data which could improve identify the different classes.

This extended analysis yields several key conclusions. Firstly, prior study of the material is crucial for appropriate tool selection. The performance of classifiers is influenced by the sensor type and spectral range, with superior results from higher quality HS equipment. Secondly, enlarging the databases is essential for achieving more robust results. It was noted that reducing the number of training pixels while maintaining the variability of spectral signatures increases mean values. However, both the quantity and spectral quality of pixels are vital for training robust models. Moreover, the presence of false positives in the maps is likely due to the spatial resolution of the data. Thus, single color maps observed in the *SLIMBRAIN* database are due to spectrally noisy input images, resulting in grainy segmentation. High spatial resolution ensures each pixel represents a single tissue type, whereas low spatial resolution pixels may contain information from multiple tissues, leading to mixed spectral signatures and misclassifications. The results from the *Unified dataset* indicate that while HT remains robust across various algorithms and sampling strategies, TT shows improvement except in 1D-DNN classifiers. *LIBRA*, in particular, obtains a higher TT class classification performance compared to the rest of the models, which indicates that although it is not yet applicable for clinical applications, it works to manage data variability and data imbalance more effectively. This emphasizes the importance of addressing class imbalance in datasets, especially for detecting less prevalent classes like tumors, where downsampling can enhance detection capabilities. Despite initial weak results, the preliminary study demonstrates that robust models can be developed using different databases and HS cameras, underlining the need to identify the relevant spectral range. In summary, the study underscores the importance of advanced ML/DL models and complex frameworks in enhancing brain tumor detection. Although HS data from *in vivo* human brain tissues is scarce, the work highlights the necessity of additional training data to develop more robust models, while ensuring that increased model complexity does not hinder the development of real-time solutions essential for neurosurgeons. Additionally, it points out that lower quality HS equipment, such as the snapshot camera from the *SLIMBRAIN* database, may result in less accurate classification since wavelengths that may provide more information are not captured (as can be appreciated with the results of the *H128* database).

Therefore, several future investigations are proposed to improve the results: first, increase the number of patients and labelled data in order to have more relevant information to increase the differentiation between tumor and healthy tissue. Then, new DL algorithms shall be developed with innovative approaches in this research area, especially highlighting Graph Convolutional Networks (GCNs) and combinations between GCNs and CNNs [54]. Finally, perform an exhaustive study of different preprocessing techniques (e.g., Standard Normal Variate [55], Contrast Enhancement [56]), since adequate data preprocessing is the fundamental basis to guarantee the quality and efficiency of the classification algorithms [20].

Ethics statement

Data from both databases have been acquired under specific ethic protocols. Written informed consent was obtained from all the subjects.

HELICoID database: The study protocol and consent procedures were approved by the *Comité Ético de Investigación Clínica-Comité de Ética en la Investigación* (CEIC/CEI) of the University Hospital of Gran Canaria Doctor Negrin (CEIC code: 130069) and the *National Research Ethics Service (NRES) Committee South Central Oxford C for the University Hospital of Southampton*.

SLIMBRAIN database: The guidelines of the Declaration of Helsinki were followed and the acquisition of HS images was approved by the *Research Ethics Committee* of Hospital Universitario 12 de Octubre, Madrid, Spain (protocol code 19/158, 28 May 2019).

CRedit authorship contribution statement

Alberto Martín-Pérez: Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Beatriz Martínez-Vega:** Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Methodology, Investigation, Data curation, Conceptualization. **Manuel Villa:** Writing – review & editing, Software, Methodology, Conceptualization. **Raquel Leon:** Writing – review & editing, Methodology, Conceptualization. **Alejandro Martínez de Ternerero:** Writing – review & editing, Software, Methodology, Conceptualization. **Himar Fabelo:** Writing – review & editing, Methodology, Conceptualization. **Samuel Ortega:** Writing – review & editing, Methodology, Conceptualization. **Eduardo Quevedo:** Writing – review & editing, Resources, Conceptualization. **Gustavo M. Callico:** Writing – review & editing, Supervision, Resources, Conceptualization. **Eduardo Juarez:** Writing – review & editing, Supervision, Resources, Conceptualization. **César Sanz:** Writing – review & editing, Supervision, Resources, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This work has been supported by Spanish Government and European Union (FEDER funds) in the context of TALENT-HEXPERIA (HyperSpectral Imaging for Artificial intelligence applications), TALENT-HIPSTER (High Performance Systems and Technologies for E-health and fish faRming) projects, under contracts PID2020-116417RB-C42 and PID2020-116417RB-C41, respectively. OASIS-HARMONIE (Hyperspectral AI-based Registration of Multimodal systems for Brain Cancer Navigation and Environmental monitoring) and OASIS-RETIRO (AI-Based Real-Time Information Sensing and Analysis) projects, under the contracts PID2023-148285OB-C43 and PID2023-148285OB-C44, respectively, all projects are funded by MICIU/AEI/10.13039/

501100011033. The work also has been funded by the European project STRATUM under grant agreement no. 101137416. Additionally, this work was completed while Beatriz Martínez-Vega was beneficiary of a predoctoral grant given by the “Agencia Canaria de Investigación, Innovación y Sociedad de la Información (ACIISI)” of the “Consejería de Economía, Conocimiento y Empleo”, which is part-financed by the European Social Fund (FSE) (POC 2014-2020, Eje 3 Tema Prioritario 74 (85 %)). Also, Manuel Villa was beneficiary of the Programa Propio I+D+i 2021 of Universidad Politécnica de Madrid.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at [doi:10.1016/j.cmpbup.2025.100183](https://doi.org/10.1016/j.cmpbup.2025.100183). It includes raincloud plots with the results obtained during validation for every ML and DL algorithm on the healthy, tumor and blood vessel classes. The first row of those plots show the results using all training pixels, whereas the second row show results when applying downsampling. These plots present the *SEN*, the *F1 Score*, *MCC*, *SPE*, *ACC* and *ROC-AUC* values obtained. Moreover, further raincloud plots with the results obtained during test for the best ML and DL algorithm on validation, after training with downsampling, are presented. These plots show the *SEN*, the *F1 Score*, *MCC*, *SPE*, *ACC* and *ROC-AUC* values obtained. Furthermore, classification maps obtained with the classifiers evaluated on the test set of every fold for the databases are illustrated.

The implementation for the DL models used in the experiments is available in a Python package, which can be downloaded from <https://gitlab.citsem.upm.es/public-projects/hyperspectral/hyperspectral-neural-networks>. This Python package has been written by the authors to read HS data from both databases. However, the data should follow the same structure as in the *SLIMBRAIN* database. The repository also includes Python scripts showing how to train and evaluate the *1D-DNN* and *2D-CNN* models, but the example scripts can easily be used to train and evaluate the other networks (e.g. *Fast 3D-CNN* or *DRNN*). The package also provides helpful tools to generate the thematic maps.

References

- [1] H. Sung, J. Ferlay, R.L. Siegel, M. Laversanne, I. Soerjomataram, A. Jemal, F. Bray, Global cancer statistics 2020: globocan estimates of incidence and mortality worldwide for 36 cancers in 185 countries, *CA Cancer J. Clin.* 71 (3) (2021) 209–249.
- [2] K.D. Miller, Q.T. Ostrom, C. Kruchko, N. Patil, T. Tihan, G. Cioffi, H.E. Fuchs, K. A. Waite, A. Jemal, R.L. Siegel, J.S. Barnholtz-Sloan, Brain and other central nervous system tumor statistics, 2021, *CA Cancer J. Clin.* 71 (5) (2021) 381–406.
- [3] A.C. Tan, D.M. Ashley, G.Y. López, M. Malinzak, H.S. Friedman, M. Khasraw, Management of glioblastoma: state of the art and future directions, *CA Cancer J. Clin.* 70 (4) (2020) 299–312.
- [4] I.J. Gerard, M. Kersten-Oertel, K. Petrecca, D. Sirhan, J.A. Hall, D.L. Collins, Brain shift in neuronavigation of brain tumors: a review, *Med. Image Anal.* 35 (2017) 403–420.
- [5] D.A. Orringer, A. Golby, F. Jolesz, Neuronavigation in the surgical management of brain tumors: current and future trends, *Expert. Rev. Med. Devices* 9 (5) (2012) 491–500.
- [6] T. Noh, M. Mustroph, A.J. Golby, Intraoperative imaging for high-grade glioma surgery, *Neurosurg. Clin.* 32 (1) (2021) 47–54.
- [7] B. Albayrak, A. Samdani, P. Black, Intra-operative magnetic resonance imaging in neurosurgery, *Acta Neurochir. (Wien)* 146 (6) (2004) 543–557.
- [8] R.U. Gandhe, C.P. Bhave, Intraoperative magnetic resonance imaging for neurosurgery—an anaesthesiologist’s challenge, *Indian J. Anaesth.* 62 (6) (2018) 411.
- [9] T. Selbekk, A.S. Jakola, O. Solheim, T.F. Johansen, F. Lindseth, I. Reinertsen, G. Unsgård, Ultrasound imaging in neurosurgery: approaches to minimize surgically induced image artefacts for improved resection control, *Acta Neurochir. (Wien)* 155 (6) (2013) 973–980.
- [10] R. Sastry, W.L. Bi, S. Pieper, S. Frisken, T. Kapur, W. Wells III, A.J. Golby, Applications of ultrasound in the resection of brain tumors, *J. Neuroimag.* 27 (1) (2017) 5–15.
- [11] I.W.H. Chung, S. Eljamel, Risk factors for developing oral 5-aminolevulinic acid-induced side effects in patients undergoing fluorescence guided resection, *Photodiagnosis. Photodyn. Ther.* 10 (4) (2013) 362–367.
- [12] N. Ferraro, E. Barbarite, T.R. Albert, E. Berchmans, A.H. Shah, A. Bregy, M.E. Ivan, T. Brown, R.J. Komotar, The role of 5-aminolevulinic acid in brain tumor surgery: a systematic review, *Neurosurg. Rev.* 39 (4) (2016) 545–555.

- [13] A. Gowen, C. O'Donnell, P. Cullen, G. Downey, J. Frias, Hyperspectral imaging – an emerging process analytical tool for food quality and safety control, *Trends. Food Sci. Technol.* 18 (12) (2007) 590–598.
- [14] J. Wang, X. Wang, K. Zhang, K. Madani, C. Sabourin, Morphological band selection for hyperspectral imagery, *IEEE Geosci. Rem. Sens. Lett.* 15 (8) (2018) 1259–1263.
- [15] L. Ma, X. Zhou, J.V. Little, A.Y. Chen, L.L. Myers, B.D. Sumer, and B. Fei, "Hyperspectral microscopic imaging for the detection of head and neck squamous cell carcinoma on histologic slides," in *Medical Imaging 2021: Digital Pathology* (J. E. Tomaszewski and A. D. Ward, eds.), vol. 11603, p. 116030P, International Society for Optics and Photonics, SPIE, 2021.
- [16] Y. Oiknine, M. Abuleil, E. Brozgol, I.Y. August, I. Barshack, I. Abdulhalim, Y. Garini, A. Stern, Compressive hyperspectral microscopy for cancer detection, *J. Biomed. Opt.* 28 (9) (2023) 096502.
- [17] X. Qin, M. Zhang, C. Zhou, T. Ran, Y. Pan, Y. Deng, X. Xie, Y. Zhang, T. Gong, B. Zhang, L. Zhang, Y. Wang, Q. Li, D. Wang, L. Gao, D. Zou, A deep learning model using hyperspectral image for esofna cytology diagnosis in pancreatic ductal adenocarcinoma, *Cancer Med.* 12 (16) (2023) 17005–17017.
- [18] X. Zhou, L. Ma, H.K. Mubarak, J.V. Little, A.Y. Chen, L.L. Myers, B.D. Sumer, and B. Fei, "Automatic detection of head and neck squamous cell carcinoma on pathologic slides using polarized hyperspectral imaging and deep learning," in *Medical Imaging 2022: Digital and Computational Pathology* (J. E. Tomaszewski, A. D. Ward, and R. M. L. M.D., eds.), vol. 12039, p. 120390G, International Society for Optics and Photonics, SPIE, 2022.
- [19] K. Young, E. Ma, S. Kejrival, T. Nielsen, S.S. Aulakh, A.C. Birkeland, Intraoperative in vivo imaging modalities in head and neck cancer surgical margin delineation: a systematic review, *Cancers* 14 (14) (2022).
- [20] B. Martínez-Vega, M. Tkachenko, M. Matkabi, S. Ortega, H. Fabelo, F. Balea-Fernandez, M. La Salvia, E. Torti, F. Leporati, G.M. Callico, C. Chalopin, Evaluation of preprocessing methods on independent medical hyperspectral databases to improve analysis, *Sensors* 22 (22) (2022).
- [21] M. Maktabi, H. Köhler, M. Ivanova, B. Jansen-Winkel, J. Takoh, S. Niebisch, S. M. Rabe, T. Neumuth, I. Gockel, C. Chalopin, Tissue classification of oncologic esophageal resectates based on hyperspectral data, *Int. J. Comput. Assist. Radiol. Surg.* 14 (2019).
- [22] T. Collins, M. Maktabi, M. Barberio, V. Bencteux, B. Jansen-Winkel, C. Chalopin, J. Marescaux, A. Hostettler, M. Diana, I. Gockel, Automatic recognition of colon and esophagogastric cancer with machine learning and hyperspectral imaging, *Diagnostics* 11 (10) (2021).
- [23] S. Puustinen, H. Vrzáková, J. Hyttinen, T. Rauramaa, P. Fält, M. Hauta-Kasari, R. Bednarik, T. Koivisto, S. Rantala, M. von und zu Fraunberg, J.E. Jääskeläinen, A.-P. Elomaa, Hyperspectral imaging in brain tumor surgery—evidence of machine learning-based performance, *World Neurosurg.* 175 (2023) e614–e635.
- [24] G. Urbanos, A. Martín, G. Vázquez, M. Villanueva, M. Villa, L. Jimenez-Roldan, M. Chavarrías, A. Lagares, E. Juárez, C. Sanz, Supervised machine learning methods and hyperspectral imaging techniques jointly applied for brain cancer classification, *Sensors* 21 (11) (2021) 3827.
- [25] A. Martín-Pérez, M. Villa, G. Vázquez, J. Sancho, G. Rosa, P. Sutradhar, M. Chavarrías, A. Lagares, E. Juárez, C. Sanz, Hyperparameter optimization for brain tumor classification with hyperspectral images, in: 2022 25th Euromicro Conference on Digital System Design (DSD), 2022, pp. 835–842.
- [26] S. Puustinen, H. Vrzáková, J. Hyttinen, T. Rauramaa, P. Fält, M. Hauta-Kasari, R. Bednarik, T. Koivisto, S. Rantala, M. von und zu Fraunberg, et al., Hyperspectral imaging in brain tumor surgery—evidence of machine learning-based performance, *World Neurosurg.* (2023).
- [27] R. Vandebriel, S. Luthman, K. Vunckx, M. Jayapala, W. Charle, L. Solie, S.D. Vleeschouwer, T. Giannantonio, A. Alperovich, and X. Zhang, "Integrating hyperspectral imaging in an existing intraoperative environment for detection of intrinsic brain tumors," in *Advanced Biomedical and Clinical Diagnostic and Surgical Guidance Systems XXI* (C. Boudoux and J. W. Tunnell, eds.), vol. 12368, p. 123680D, International Society for Optics and Photonics, SPIE, 2023.
- [28] N. Kifle, S. Teti, B. Ning, D.A. Donoho, I. Katz, R. Keating, R.J. Cha, Pediatric brain tissue segmentation using a snapshot hyperspectral imaging (shsi) camera and machine learning classifier, *Bioengineering* 10 (10) (2023).
- [29] T. Giannantonio, A. Alperovich, P. Semeraro, M. Atzori, X. Zhang, C. Hauger, A. Freytag, S. Luthman, R. Vandebriel, M. Jayapala, L. Solie, and S. de Vleeschouwer, "Intra-operative brain tumor detection with deep learning-optimized hyperspectral imaging," in *Optical Biopsy XXI: Toward Real-Time Spectroscopic Imaging and Diagnosis* (R. R. Alfano and A. B. Seddon, eds.), vol. 12373, p. 123730F, International Society for Optics and Photonics, SPIE, 2023.
- [30] Y.W. Wang, N.P. Reder, S. Kang, A.K. Glaser, J.T. Liu, Multiplexed optical imaging of tumor-directed nanoparticles: A review of imaging systems and approaches, *Nanotheranostics*. 1 (2017) 369–388.
- [31] H. Fabelo, S. Ortega, R. Lazcano, D. Madroñal, G.M. Callico, E. Juárez, R. Salvador, D. Bulters, H. Bulstrode, A. Szolna, J.F. Piñeiro, C. Sosa, A.J. O'Shanahan, S. Bisshopp, M. Hernández, J. Morera, D. Ravi, B.R. Kiran, A. Vega, A. Báez-Quevedo, G.-Z. Yang, B. Stanculescu, R. Sarmiento, An intraoperative visualization system using hyperspectral imaging to aid in brain tumor delineation, *Sensors* 18 (2) (2018).
- [32] G. Lu, B. Fei, Medical hyperspectral imaging: a review, *J. Biomed. Opt.* 19 (1) (2014) 010901.
- [33] J. Sancho, M. Villa, M. Chavarrías, E. Juárez, A. Lagares, C. Sanz, Slimbrain: Augmented reality real-time acquisition and processing system for hyperspectral classification mapping with depth information for in-vivo surgical procedures, *J. Syst. Arch.* 140 (2023) 102893.
- [34] H. Fabelo, et al., Spatio-spectral classification of hyperspectral images for brain cancer detection during surgical operations, *PLoS. One* 13 (3) (2018) e0193721.
- [35] S. Puustinen, J. Hyttinen, G. Hisuin, H. Vrzáková, A. Huotari, P. Fält, M. Hauta-Kasari, A. Immonen, T. Koivisto, J.E. Jääskeläinen, and A.-P. Elomaa, "Towards clinical hyperspectral imaging (hsi) standards: Initial design for a microneurosurgical hsi database," in *2022 IEEE 35th International Symposium on Computer-Based Medical Systems (CBMS)*, pp. 394–399, 2022.
- [36] H. Fabelo, M. Halicek, S. Ortega, M. Shahedi, A. Szolna, J.F. Piñeiro, C. Sosa, A. J. O'Shanahan, S. Bisshopp, C. Espino, M. Márquez, M. Hernández, D. Carrera, J. Morera, G.M. Callico, R. Sarmiento, B. Fei, Deep learning-based framework for in vivo identification of glioblastoma tumor using hyperspectral images of human brain, *Sensors* 19 (4) (2019).
- [37] F. Manni, F. van der Sommen, H. Fabelo, S. Zinger, C. Shan, E. Edström, A. Elmi-Terander, S. Ortega, G. Marrero Callico, P.H.N. de With, Hyperspectral imaging for glioblastoma surgery: Improving tumor identification using a deep spectral-spatial approach, *Sensors* 20 (23) (2020).
- [38] H. Fabelo, S. Ortega, A. Szolna, D. Bulters, J.F. Piñeiro, S. Kabwama, A. J. O'Shanahan, H. Bulstrode, S. Bisshopp, B.R. Kiran, D. Ravi, R. Lazcano, D. Madroñal, C. Sosa, C. Espino, M. Márquez, M. De La Luz Plaza, R. Camacho, D. Carrera, M. Hernández, G.M. Callico, J. Morera Molina, B. Stanculescu, G.-Z. Yang, R. Salvador, E. Juárez, C. Sanz, R. Sarmiento, In-vivo hyperspectral human brain image database for brain cancer detection, *IEEE Access*. 7 (2019) 39098–39116.
- [39] A. Martín-Pérez, M. Villa, G.R. Olmeda, J. Sancho, G. Vázquez, G. Urbanos, A.M. de Terro, M. Chavarrías, L. Jimenez-Roldan, A. Perez-Núñez, A. Lagares, E. Juárez, and C. Sanz, "Slim brain database: a multimodal image database of in-vivo human brains for tumour detection." preprint on webpage at <https://www.researchsquare.com/article/rs-3629358/v1>, 11 2023.
- [40] J.M. Bioucas-Dias, J.M.P. Nascimento, Hyperspectral subspace identification, *IEEE Trans. Geosci. Rem. Sens.* 46 (8) (2008) 2435–2445.
- [41] B. Martínez, R. Leon, H. Fabelo, S. Ortega, J.F. Piñeiro, A. Szolna, M. Hernandez, C. Espino, A.J. O'Shanahan, D. Carrera, S. Bisshopp, C. Sosa, M. Márquez, R. Camacho, M.d.l.L. Plaza, J. Morera, G.M. Callico, Most relevant spectral bands identification for brain cancer detection using hyperspectral imaging, *Sensors* 19 (24) (2019).
- [42] S. Ortega, M. Halicek, H. Fabelo, R. Camacho, M.d.l.L. Plaza, F. Godtliedsen, G. M. Callico, B. Fei, Hyperspectral imaging for the detection of glioblastoma tumor cells in h&e slides using convolutional neural networks, *Sensors* 20 (7) (2020).
- [43] R. Hahn, F.-E. Hämmerling, T. Haist, D. Fleischle, O. Schwanke, O. Hauler, K. Rebner, M. Brecht, W. Osten, Detailed characterization of a mosaic based hyperspectral snapshot imager, *Opt. Eng.* 59 (12) (2020) 125102–125102.
- [44] I. Dewancker et al., "Bayesian optimization primer," 2016.
- [45] J. Bergstra et al., "Algorithms for hyper-parameter optimization," in *Advances in Neural Information Processing Systems* (J. Shawe-Taylor et al., eds.), vol. 24, Curran Associates, Inc., 2011.
- [46] L. Breiman, Random forests, *Mach. Learn.* 45 (2001) 5–32.
- [47] T. Cover, P. Hart, Nearest neighbor pattern classification, *IEEE Trans. Inf. Theory*. 13 (1) (1967) 21–27.
- [48] H. Ayaz, D. Tormey, I. McLoughlin, M. Ahmad, S. Unnikrishnan, Hyperspectral brain tissue classification using a fast and compact 3d cnn approach, in: 2022 IEEE 5th International Conference on Image Processing Applications and Systems (IPAS) Five, 2022, pp. 1–4.
- [49] D. Eggert, M. Bings, S. Westermann, N. Gessert, A.O.H. Gerstner, N.A. Mueller, J. Bewarder, A. Schlaefer, C. Betz, W. Laffers, In vivo detection of head and neck tumors by hyperspectral imaging combined with deep learning methods, *J. Biophotonics*. 15 (3) (2022) e202100167.
- [50] P.L. Cebrían, A. Martín-Pérez, M. Villa, J. Sancho, G. Rosa, G. Vázquez, P. Sutradhar, A. Martínez de Terro, M. Chavarrías, A. Lagares, E. Juárez, and C. Sanz, "Deep recurrent neural network performing spectral recurrence on hyperspectral images for brain tissue classification," in *Design and Architecture for Signal and Image Processing* (M. Chavarrías and A. Rodríguez, eds.), Cham, pp. 15–27, Springer Nature Switzerland, 2023.
- [51] A. Sherstinsky, Fundamentals of recurrent neural network (rnn) and long short-term memory (lstm) network, *Physica D Nonlinear Phenomena* 404 (2020) 132306.
- [52] M. Villa, A. Martín-Pérez, G. Vázquez, G. Rosa-Olmeda, J. Sancho, P. Sutradhar, M. Chavarrías, L. Jimenez-Roldan, A. Pérez-Núñez, A. Lagares, E. Juárez, C. Sanz (2024). LIBRA: Low spectral resolution brain tumor classifier for medical hyperspectral imaging. <https://www.researchsquare.com/article/rs-4668541/v1>.
- [53] L. Breiman, Bagging predictors, *Mach. Learn.* 24 (2) (1996) 123–140.
- [54] P. Yang, X. Zhang, A dual-branch fusion of a graph convolutional network and a convolutional neural network for hyperspectral image classification, *Sensors* 24 (14) (2024).
- [55] M. Zeaiter and D. Rutledge, "3.04 preprocessing methods," in *Comprehensive Chemometrics* (S. D. Brown, R. Tauler, and B. Walczak, eds.), pp. 121–231, Oxford: Elsevier, 2009.
- [56] D. Vijayalakshmi, M.K. Nath, A systematic approach for enhancement of homogeneous background images using structural information, *Graph. Models*. 130 (2023) 101206.