Hybrid models for classifying histological images: An association of deep features by transfer learning with ensemble classifier

Cléber I. de Oliveira[1] · Marcelo Z. do Nascimento[2] · Guilherme F. Roberto[2] · Thaína A. A. Tosta[3] · Alessandro S. Martins[4] · Leandro A. Neves[1]

Abstract

The use of a convolutional neural network with transfer learning is a strategy that defines high-level features, commonly explored to study patterns in medical images. These features can be analyzed via different methods in order to design hybrid models with more useful and accurate solutions for clinical practice. In this paper, a computational scheme is presented to define hybrid models through deep features by transfer learning, selection by ranking and a robust ensemble classifier with five algorithms. The obtained models were applied to classify histological images from breast, colorectal and liver tissue. The strategy developed here allows knowing important results and conditions to improve models of computer-aided diagnosis, even exploring classic CNN models. The features were defined using layers from the AlexNet and ResNet-50 architectures. The attributes were organized into subsets of the most relevant features and submitted to a k-fold cross-validation process. The best hybrid models were obtained with deep features from the ResNet-50 network, using distinct layers (*activation_48_relu* and *avg_pool*) and a maximum of 35 descriptors. These hybrid models provided 98.00% and 99.32% of accuracy values, with emphasis on histological images of breast cancer, indicating the best solution among those available in the specialized Literature. Also, these models provided more relevant results for classifying UCSB and LG datasets than regularized techniques and CNN architectures, exploring data augmentation or not. The computational scheme with detailed information regarding the main hybrid models is a relevant contribution to the community interested in the study of machine learning techniques for pattern recognition.

Keywords Deep features · Transfer learning · Hybrid models · Histological images · Pattern recognition

✉ Cléber I. de Oliveira
  oliveira.cleber@gmail.com

Extended author information available on the last page of the article

1 Introduction

In the fields of image processing and computer vision, techniques for feature extraction require special treatment for processing natural data in its raw form [1]. Thus, a machine learning-based system requires careful engineering, for instance, to define features with enough representativeness to enable the detection or pattern classification [2]. Techniques based on deep learning (DL) minimize some difficulties encountered in this process by making the feature engineering stage an automated process [3]. A DL-based system involves multiple layers of processing in order to provide data representation with different levels of abstraction, such as those based on convolutional neural networks (CNN) [3].

CNN models allowed significant advances in image processing due to the proposals presented in [4] and [5]. In these approaches, the increasing number of layers reduced the error rates in classification and pattern recognition tasks, with emphasis on computer-aided diagnosis (CAD) for histological images. AlexNet and deep residual network (ResNet) are some examples of CNN strategies widely explored in this system category due to the relevant results achieved to accurately classify different types of cells and tissue structures. These classic architectures have also been tested on important datasets, making them generalizable to histological images, as well as comparable and robust to variations in the staining process, an important challenge in this context [6–10]. The conditions presented here are useful for the improvement of existing CAD systems, especially when the classic approaches are investigated in order to verify their capacities of providing relevant features to reach more optimized and comprehensive solutions for specialists.

Considering this context, the ResNet architecture deserves to be highlighted for image classification tasks, such as investigating histological images, because it minimizes the well-known problem of the vanishing gradient [7, 9, 11, 12]. Even so, classic CNN models can contain millions of trainable parameters, which can make them unfeasible with few samples. This situation is observed in the context of histological images. A commonly explored alternative to overcome this limitation is the use of transfer learning with hybrid models [7, 13–15], via pre-training carried out on the ImageNet dataset [16]. This alternative reached relevant results in several domains [9, 13, 17], especially considering the use of feature maps from specific layers of the CNN architectures [15, 18, 19]. Thus, when the hybrid models are observed, some issues have to be investigated to ensure the success of the solution, mainly involving the definition of layers, selection methods and classification strategy [7, 14, 15, 18, 20–22]. For instance, hybrid models were defined considering that the initial layers provide the identification of local patterns, such as color, edge and shape. On the other hand, deeper layers provide the generalization of global patterns, such as texture and semantics [18].

In order to develop the previously cited models, the computational scheme can be based on feature selection algorithms with a single classifier or an ensemble classification [10, 23, 24]. However, feature selection with an ensemble classifier combines the strengths of these strategies in order to provide more stable and relevant solutions [25–28], with more accurate and fully useful CAD systems. Moreover, the feature selection process plays a critical role in identifying complex patterns in relevant contexts, such as those explored here, with the most optimized and comprehensible solutions [29]. This process can be designed via techniques categorized as filters, wrapper, and embedded, but there is no universal approach to define the best results for all contexts [25–28]. On the other hand, filter algorithm like ReliefF is fully capable of detecting feature dependencies, indicating the best schemes in different experiments [27, 30–32]. The ReliefF algorithm is relatively fast, with an asymptotic time complexity of $\mathcal{O}(n^2 * m)$, where n is the number of instances and m is the number of features,

and the selected features do not depend on an induction algorithm [25]. Also, feature sets of different sizes can be obtained via this algorithm, based on any desired criteria. Consequently, when the most relevant subsets of features are indicated via an ensemble classifier, the hybrid models are more generalizable and robust in order to reach more optimized and comprehensive solutions.

In this paper, a computational scheme is described to provide hybrid models via the association of deep features by transfer learning, selection by ranking, and a robust ensemble classifier. The obtained models were analyzed to classify breast, colorectal and liver tissue images stained with hematoxylin-eosin (H&E). The proposal considered deep features provided by the layers from the AlexNet and ResNet-50 architectures. In the AlexNet architecture, the computational scheme explored all convolutional layers. In the ResNet-50 network, the analyzed layers were those able to provide local and global image patterns, such as *max_pooling2d_1*, *activation_4_relu*, *activation_48_relu*, *activation_49_relu* and *avg_pool*. The deep features were organized into subsets and submitted to a *k*-fold cross-validation process. A systematic analysis was carried out in order to rank and define the most relevant subsets via an ensemble classifier with five algorithms. Thus, this study provides the following contributions:

1. A computational scheme able to provide hybrid models representing the main associations of deep features by transfer learning, ReliefF algorithm and an ensemble classifier with five algorithms;
2. An optimized hybrid model that provided the best performance for distinguishing breast cancer, based on only 35 deep features from the intermediate layer (*activation_48_relu*) of the ResNet-50;
3. Hybrid models based on AlexNet's deep features that outperform CNN architectures by directly classifying the UCSB dataset, with or without data augmentation;
4. Hybrid models via ResNet-50's deep features that showed more relevant results for classifying UCSB and LG datasets than regularized techniques and CNN architectures;
5. Solutions based on a reduced number of features and without overfitting, useful for developing CAD systems focused on H&E images or even as more robust baseline schemes commonly explored in this type of investigation.

In the second section of this paper, relevant works on the classification of H&E images exploring hybrid models are described. The methodology is presented in Section 3 and, in Section 4, the results are presented and discussed. Finally, the conclusion is presented in Section 5.

2 Related work: An overview

The use of hybrid models based on handcrafted features (HC) or deep features by transfer learning has indicated important advances in different contexts. Regarding models based on HC, Watanabe et al. [33] presented an approach via Gist descriptors, principal component analysis and linear discriminant analysis to classify liver histological images. The system was able to provide an accuracy of 93.70%. In the proposal of [34], the authors presented associations of sample entropy and a fuzzy strategy to classify colorectal tissue, and the achieved accuracy was 91.39%. The authors in [35] presented a histological image classifier for the breast and colorectal tissues. The model used percolation attributes and color-normalized images. The accuracy values were 86.20% to distinguish breast tumors and 90.90% to classify colorectal tumors. In another study [36], an approach was proposed in order to detect

malignant tumors in representative histological images of breast cancer. The proposed method reached an accuracy of 86.20% employing the texture descriptors, morphological attributes and intensity.

When strategies exploring CNN architectures are taken into account, the proposal of [6] considered an adversarial stain transfer technique for classifying histological images of colorectal tissue. The authors used the U-Net model for the stain-transfer network, exploring the fully connected layer from the AlexNet architecture. The accuracy metric obtained by the model was 87.50%. The authors of [37] classified breast tissues via a model based on 13 CNN layers with the SVM classifier. In this approach, the accuracy was a value of 83.30%. Considering the tensor decomposition for multiple-instance, the proposal [38] achieved an accuracy value of 84.67% for classifying the same type of tissue.

In addition, Kausar et al. [39] described a classifier based on color normalization, haar wavelet decomposition and a 16-layer CNN. In this proposal, the maximum accuracy value was 91% to distinguish breast tissue samples. In another study [40], a model based on deep learning with a stacked denoising autoencoder was proposed in order to analyze H&E breast cancer images. In this strategy, the accuracy value was 94.41%. The authors of [41] explored the RefineNet and DenseNet architectures, through the deep-reverse active learning technique. The model was applied to classify H&E histological images as representatives of breast cancer with an accuracy of 97.63%. In the proposal of [42], the authors developed a modular cGAN classification framework for colorectal tumor detection. This approach used the U-Net and Inception V3 models, via pre-training on the ImageNet dataset, providing an accuracy value of 94.02%. Saxena et al. [8] described a ResNet-50 model with kernelized weighted to distinguish H&E breast tissue samples. The achieved performance was an accuracy value of 60.30%. Recently, Lee and Wu [43] presented the DIU-Net architecture with a color conversion scheme in the training step. When applied to breast tumors, the model indicated an accuracy value of 94.09%.

Strategies to optimize deep learning models can also be found in the study of H&E images. Deep learning techniques have been used for detecting the preneoplastic and neoplastic lesions in human colorectal histological images [44]. The model provided an accuracy of 93.28%. In another study [45], the authors proposed a classifier using the U-Net and GoogLeNet networks with color normalization. The model was able to classify images of colorectal tissue with an accuracy of 85%. Also, a model based on ResNet, transfer learning and deep-tuning was defined to classify the same type of tumor [9]. The strategy provided an accuracy of 86.67%. Considering this type of image, Dabass et al. [46] and Dabass et al. [47] presented models based on 31-layer CNN and a hybrid CNN with attention learning, respectively. The system described in [46] achieved an accuracy of 96.97%, while the strategy developed by [47] provided an accuracy value of 97.50%.

Hybrid models are also observed with deep learning and classic classification techniques applied to histological images [10, 14, 17, 48]. For instance, Kumar and Sharma [17] developed a strategy via Xception and VGG-16 architectures, exploring different types of classifiers (logistic regression, SVM and decision tree) and artificial data augmentation. The model was applied to classify H&E breast cancer samples. The best accuracy was a value of 82.45%. In another study [10], the authors described a composition of a fractal neural network with an ensemble classification based on experiments with the ResNet-50, ResNet-101, Inception V3 and Xception architectures. Handcrafted features were also used, such as lacunarity, fractal dimension and percolation. The models were applied to investigate breast cancer, colorectal cancer, lymphoma and liver tissue H&E images. The authors concluded that the combination was able to provide accuracy rates from 89.66% to 99.62%. Also, Longo et al. [14] indicated a hybrid model involving handcrafted attributes (lacunarity and fractal dimension) with deep

features from the ResNet-50, Inception-V3 and VGG-19 networks. The authors explored multiple classifiers for H&E images of breast cancer, colorectal tumor, and liver tissue. The achieved accuracy values ranged from 93.10% to 99.25%.

Finally, it is possible to verify research issues to analyze the discriminative power of specific layers of a CNN. Thus, the authors of [7] proposed a study based on the fully connected layers from the AlexNet, VGG-16 and VGG-19 networks. Deep features were combined with HC descriptors. With a k-nearest neighbors classifier, the model achieved an accuracy of 84.20% in distinguishing H&E breast cancer images. Also, Younas et al. [21] described an ensemble framework of deep neural networks in order to distinguish polyps in colorectal images. The authors used the GoogLeNet, Xception and Resnet-50 networks, all pre-trained on the ImageNet dataset, with a combination via ensemble classifier. Younas et al. [21] state that the system was able to surpass the performances reported in the Literature addressed to distinguish the classes of colorectal cancer, adenoma, hyperplasia and adenocarcinoma.

From the previously presented works, it is possible to note the benefits of using hybrid models based on multiple strategies, exploring transfer learning, deep features and ensemble classification. In this context, the hybrid models applied to histological images provided can be highlighted, such as those discussed in [7, 10, 14, 17, 21, 37, 47]. The proposals based on deep learning models lead to significant results for tissue with distinct magnifications. Despite the valuable contributions, multiple associations have been defined without the full limits of classic architectures in order to design hybrid models. Moreover, hybrid models designed with the most relevant deep features from classical models, via selection by ranking, multiple convolutional layers and robust ensemble classification have not been fully explored in several types of H&E images, such as the breast, colorectal and liver tissue analysis. This type of investigation and models provide more optimized and comprehensive solutions for the specialists and CAD systems, especially when the results are relevant in datasets with few samples and without overfitting.

3 Methodology

The hybrid models were obtained through a computational scheme that explores deep features obtained from CNN architectures, pre-trained on the ImageNet dataset [16], selection by ranking and an ensemble classifier to investigate different types of H&E images. This scheme investigated sets of layers from the CNN models and collects the corresponding deep features, a process carried out during the execution of each network. Then, the most relevant deep features were obtained from a systematic analysis, based on selection by ranking (ReliefF algorithm) with the k-fold cross-validation strategy. Finally, an ensemble classification with five algorithms was applied to identify the more relevant associations. The obtained hybrid models were applied for distinguishing the different lesion patterns present in H&E histological images. An overview of this scheme is illustrated in Fig. 1, with details presented in the next subsections.

3.1 Software packages and environment for the experiments

In this work, the approach for processing CNN architecture and extracting deep features was implemented using the deep network design and transfer learning toolboxes, available on the MATLAB R2019a package [49]. The layers explored here follow the nomenclatures defined in these toolboxes. The algorithms employed for selecting and classifying features
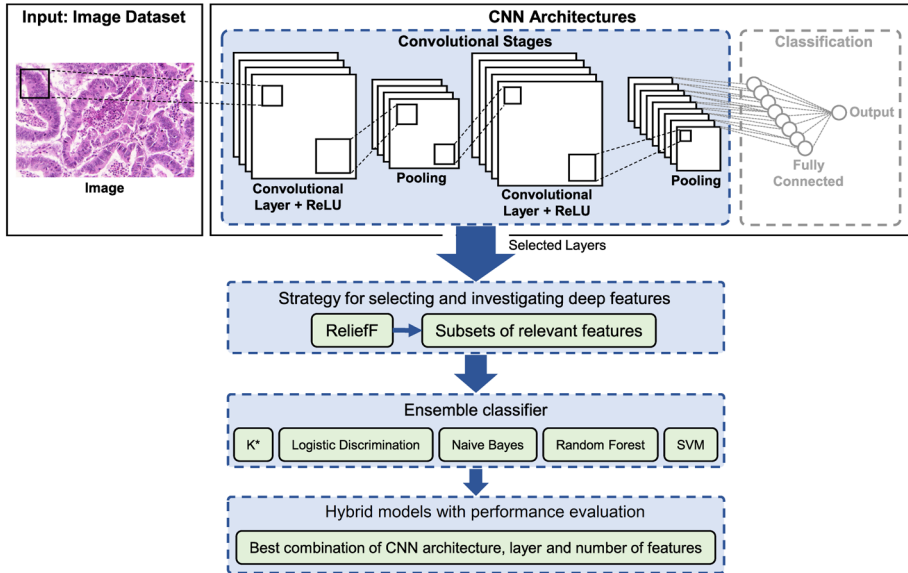
Fig. 1 Illustrative overview of the proposed scheme for classification of H&E histological images

are available on the Weka 3.6.15 package [50]. From the CNN models, the deep features were explored considering the stochastic gradient descent with momentum optimizer using the default parameters: an initial learning rate of 0.0001; a learning rate drop period of 10; a learning rate drop factor of 0.1; an L2 regularization value of 0.0001 and a mini-batch size of 4. The experiments were done by splitting the entire dataset into 80% training and 20% test data. The experiments were performed on an AMD Ryzen 5 3600X 6-Core CPU at 3.79 GHz with 64 GB of RAM and an NVIDIA GeForce GTX 1660 SUPER.

3.2 Image datasets

The proposed approach was evaluated through three different public datasets of H&E histological images: breast cancer; colorectal cancer; and liver tissue. For breast cancer, the images were provided by the Center of Bio-image Informatics, University of California, Santa Barbara (UCSB) [51]. The dataset consists of 58 breast histological images, divided into 32 benign and 26 malignant. The second dataset, colorectal cancer (CR), was provided by [52]. This dataset has 74 benign and 91 malignant samples, totaling 165 images. The third dataset is named liver gender (LG), which was provided by the Atlas of Gene Expression in Mouse Aging Project (AGEMAP) [53]. This dataset consists of liver samples from mice separated as males and females. Thus, these two classes represent the gender of the sample collected, totaling 265 examples: male with 150 images and female with 115 samples. In this work, the quantities of images were adjusted in order to balance the dataset, considering the smallest number of samples available in each group of each dataset. The removed samples were randomly chosen. This procedure prevented a dominant group from affecting the result. Table 1 presents the details related to the datasets explored in this study. All investigated datasets have two classes and images are exclusively stained with H&E. Some examples of these images are presented in Fig. 2.

Table 1 Details related to the histological datasets explored through the obtained hybrid models

Dataset	Image	Samples	Resolution	Classes
UCSB	Breast tumors	52	896×768	Benign *versus* Malignant
CR	Colorectal tumors	148	From 567×430 to 775×522	Benign *versus* Malignant
LG	Liver tissue	230	417×312	Male *versus* Female

3.3 CNN architectures and layer selection: exploring transfer learning

The proposed scheme considered the classic CNN models, such as AlexNet and ResNet-50 architectures, that were pre-trained on the ImageNet dataset [16]. For the training and optimization of these CNN models, a large dataset is necessary. However, for classifying small datasets, such as those explored here, it is difficult to determine the appropriate local minima for the cost function, and the network may suffer from overfitting. To overcome these limitations, the use of pre-trained models has been widely explored in recent studies [54, 55], considering transfer learning. This strategy can provide high-level deep features, even on datasets with few labeled samples [56]. Also, it is important to highlight that the architectures explored here are still widely used in many investigations regarding CAD systems, especially due to the significant results obtained in classifying different cell types and tissue structures in H&E images, as well as to minimize the variations in the staining process of this type of image [6–10]. Thus, the initial AlexNet layers were explored to extract low-level features such as edges and textures, while the later layers were defined to recognize higher-level patterns and structures. Regarding the use of ResNet, the initial layers were investigated to extract low-level features, while the deeper layers were indicated to recognize more complex and higher-level features. These features were used in order to define hybrid models based on different conditions for classifying the H&E images, exploring transfer learning in order to minimize overfitting and the vanishing gradient problem [57].

The AlexNet model consisted of five convolutional layers, three pooling layers, two fully connected layers, and one softmax layer [57]. This architecture used a dropout regularization scheme and rectified linear units (ReLU) to reduce overfitting [58], as well as local response normalization (LRN) to minimize the vanishing gradient problem [57]. On the other hand, the ResNet-50 model consisted of four blocks, each one with convolution layers and residual blocks. The first block had nine convolution layers and three residual blocks. The second block had 12 convolution layers and four residual blocks. The third block had six residual blocks and 18 convolution layers. The fourth block had the same number of convolution layers and residual blocks as the first block [59]. In this architecture, the layers received values resulting from the ReLU activation function and the input values of these functions. Thus, the ResNet-50 architecture used shortcut connection identity containing batch standardization groups to skip layers, allowing to minimize issues involving overfitting and the vanishing gradient problem [59]. An overview of the CNN models is presented in Table 2.

Table 2 An overview of the CNN models explored in this study

Models	Layers	Parameters	Input size
AlexNet	8	6.2×10^7	$224 \times 224 \times 3$
ResNet-50	50	2.6×10^7	$256 \times 256 \times 3$
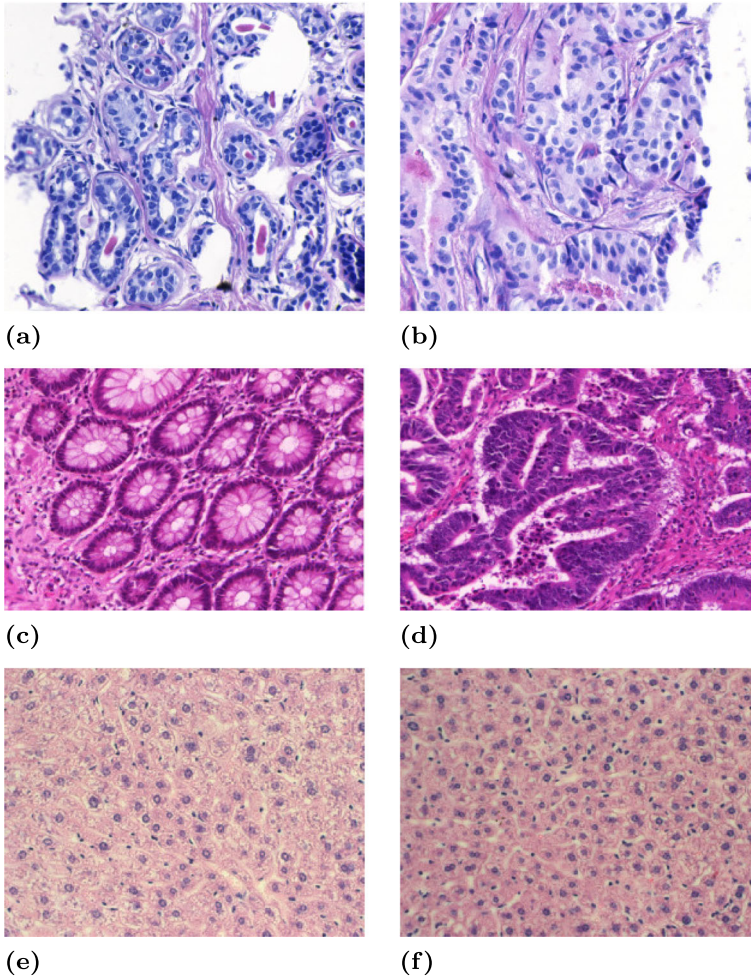
Fig. 2 Examples of H&E histological images: breast UCSB [51], benign (a) and malignant (b); CR [52], benign (c) and malignant (d); LG [53], male (e) and female (f)

According to the CNN models previously described, the deep features were obtained considering the strategy presented by [13]. For the ResNet-50 architecture, the proposed scheme explored two initial layers and the last three layers of the network. The initial layers provided deep features responsible for quantifying images' edges and colors. The deeper layers were used to identify global patterns, such as texture and semantics [18]. The *max_pooling2d_1* layer corresponded to the max pooling (with step size equal to 2×2) from the first convolution layer, which had a kernel size of 7×7 and 64 different filters. The layer, with the most features, was the *activation_4_relu* with the corresponding function $\mathcal{F}(x) + x$ from the first residual block, which was useful to evaluate the accuracy of the model with a set of dense features. The *activation_48_relu* and *activation_49_relu* layers belong to the final segment of the ResNet-50 model, being part of the last residual block and the last activation layer over the network, respectively. Also, from the average pooling layer, which had a core size

of 7×7, applied on *activation_49_relu*, the last layer chosen was *avg_pool* due to the lower number of features.

Regarding the deep features via AlexNet, the investigation was performed with the five convolutional layers of the network, excluding the fully connected and softmax layers. It is important to note that the first four convolution layers were selected based on the ReLU activation function of each layer, removing features with negative values. In addition, the *pool5* layer corresponded to the max pooling of the last convolution layer in the network.

The names of the layers with the total features used to design the hybrid models are shown in Table 3.
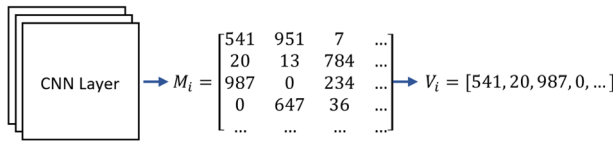
3.4 Strategy for investigating and selecting the most relevant deep features

The layers of a CNN architecture are represented by n-dimensional arrays, here named as $M_i[...]$, in which i defined each one of the five layers of each CNN model under investigation. Each column of an M_i array was sequentially arranged in a vector of deep features $V_i[...]$, where M_i and V_i have the same dimension. It is important to note that the order of the deep features was preserved in relation to the observed in each M_i, making it possible to reconstruct each array through the values and the dimension of V_i. An illustration of this representation is shown in Fig. 3.

After defining the V_i vectors, each set was distributed into S subsets, according to (1). The limited amount of 100 deep features was defined based on the models described in [25, 60]. Thus, each S_m subset was defined by the best-ranked m elements of each V_i under investigation, considering the ReliefF algorithm [61–63]. This algorithm was chosen due to its powerful and widely used feature selection method for machine learning and data mining problems, as employed in [14, 22, 25, 27]. In the proposed approach, this algorithm identified and ranked the most significant features within an original dataset to enhance the predictive capability of the hybrid models. The algorithm was applied to estimate the feature weights with the observed difference between instances that are similar, penalizing those that provide distinct values to neighbors of the same group. In addition, the algorithm rewarded features that indicated different values to neighbors of distinct groups [25, 64]. This process was carried out through a random sampling of instances, and the weights of the features were accumulated. Finally, the features were ranked according to their weights in order to indicate

CNNs	Layer Name	Total of Features
AlexNet	*relu1*	290,400
AlexNet	*relu2*	186,624
AlexNet	*relu3*	64,896
AlexNet	*relu4*	64,896
AlexNet	*pool5*	9,216
ResNet-50	*max_pooling2d_1*	193,600
ResNet-50	*activation_4_relu*	774,400
ResNet-50	*activation_48_relu*	25,088
ResNet-50	*activation_49_relu*	100,352
ResNet-50	*avg_pool*	2,048

Table 3 Information related to the layers and corresponding deep features to define the hybrid models

Fig. 3 Illustration of the organization of deep features obtained from a layer in a formatted feature vector

the most relevant predictors in each convolutional layer.

$$m \in \{5 \leq m \leq 100, \frac{m}{5} \in \mathbb{N}\}, \tag{1}$$

where m indicates the number of deep features in each subset.

The analysis of each S_m subset was performed through the k-fold cross-validation strategy in order to evaluate the generalization capacity of the models. In addition, $k = 5$ was defined in all tests due to the reduced number of available samples in each histological dataset. Finally, a robust ensemble classification was applied to calculate the accuracy rate (2) in each k-fold. The average accuracy rate in each S_m subset was given by (3). Therefore, the best association of V_i and a corresponding subset S_m was defined through the highest average accuracy rate (Acc_Avg) in each evaluated dataset. Consequently, the obtained results correspond to the most relevant deep features, via transfer learning, for pattern recognition in the investigated H&E images. Figure 4 illustrates the described steps for the feature selection and classification processes.

$$Acc_j = \frac{TP + TN}{TP + FP + TN + FN}, \tag{2}$$

in which: j refers to the number of the fold corresponding to the cross-validation iteration; TP, true positive rate, defines an outcome where the model correctly predicts the positive group; TN, true negative rate, indicates an outcome where the model correctly predicts the negative group; FP, false positive rate, represents an outcome where the model incorrectly predicts the positive group; and, FN, false negative rate, defines an outcome where the model incorrectly predicts the negative group.

$$Acc\_Avg = \frac{1}{k} \sum_{j=1}^{k} Acc_j. \tag{3}$$

3.5 Definition of the ensemble classifier

The use of different classifiers is a strategy commonly applied in machine learning-based solutions, offering successful analyses of histological images [65, 66], especially for giving more representativeness for the problem under investigation and minimizing the overfitting. However, the combination presented here has not been used in the specialized Literature focusing on H&E imaging. For this purpose, the ensemble classifier was based on five algorithms of different categories: K* [67], logistic discrimination (LD) [68], naive Bayes (NB) [69], random forest (RF) [70] and SVM [71]. Thus, the decisions were based on the common behaviors of the classifiers, making them more reliable and avoiding overfitting. The classifications were combined through the sum rule, which can be summarized as the sum of prediction probabilities obtained in each classifier [72]. This rule was used due to the good
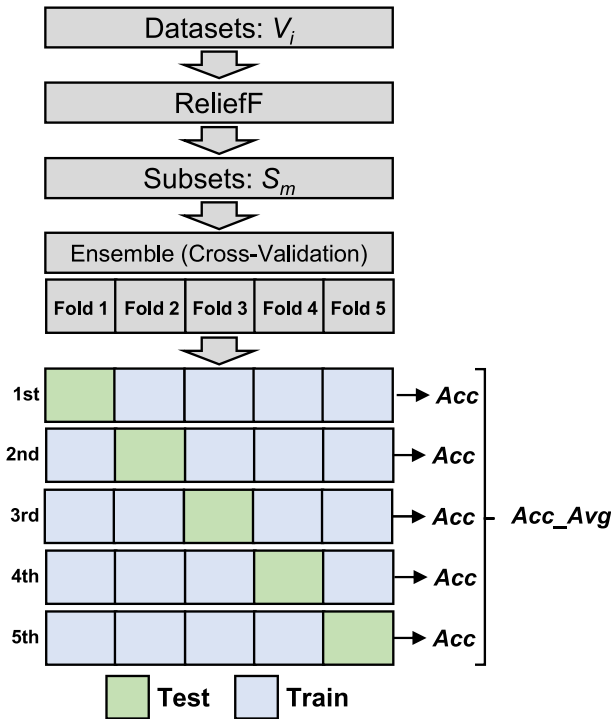
Fig. 4 Illustration of the feature selection and classification processes proposed in this study

results reported by [10]. The decision is given by the ensemble, allowing to define which associations were the most relevant to distinguish the investigated histological image groups.

4 Results and discussion

The proposed scheme was tested in three sets of histological images, as presented in Section 3.2. The evaluated comparisons were: benign *versus* malignant for the UCSB and CR datasets; and male *versus* female for the LG dataset. Considering the investigated layers (Table 3), the selected features by ranking were evaluated via the Mann-Whitney U test in order to measure the significance of each subset in distinguishing the groups investigated here. Each test considered the empirical cumulative distribution function of the descriptors with the corresponding *p-values*[73], analyzing the 100 best-ranked attributes via the ReliefF algorithm. Features with *p-values* of 0.05 or less were considered statistically significant. The main results were observed using the networks: ResNet-50 with *activation_48_relu* (UCSB) and *avg_pool* (CR and LG); AlexNet with *relu2* (UCSB), *relu3* (CR) and *pool5* (LG). The cumulative distribution function of each set is shown in Fig. 5. In these cases, it is noted that more than 80% of the data are statistically separable (*p-values* $\leq$ 0.05), a condition observed in the UCSB dataset. In the other datasets, the statistically separable data represent the highest percentage, with approximately 95% of the features. Thus, Figs. 6 and 7 show the accuracy rates in relation to the number of deep features after applying the proposed ensemble classifier (Sections 3.4 and 3.5).
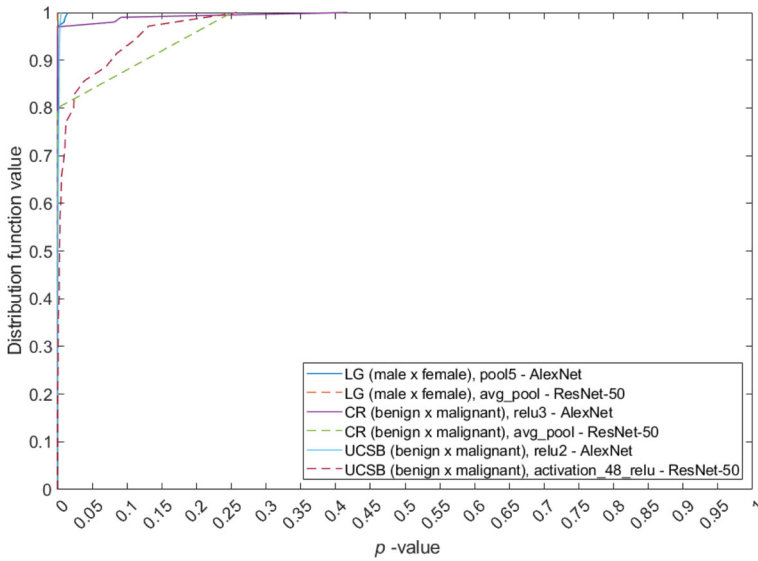
Fig. 5 The empirical cumulative distribution function (y-axis) of the 100 best-ranked features with the corresponding *p-values* (x-axis), considering the main results for classifying the UCSB, CR and LG datasets

From the main results, the proposed scheme identified the highest accuracy rate with the lowest number of features in each scenario, representing the main hybrid models. For instance, the hybrid model using 100 deep features from the deepest layer of the AlexNet (*pool5*) achieved an accuracy rate of 98.70% in the LG dataset. However, the hybrid model exploring the deepest layer of the ResNet-50 network (*avg_pool*) presented an accuracy rate of 99.32% with only 5 deep features. This last association (ResNet-50's deep features via *avg_pool*) was also responsible for providing the most relevant features for the CR dataset, with an accuracy rate of 98.00%. In this case, the hybrid model was defined with 35 deep features. These
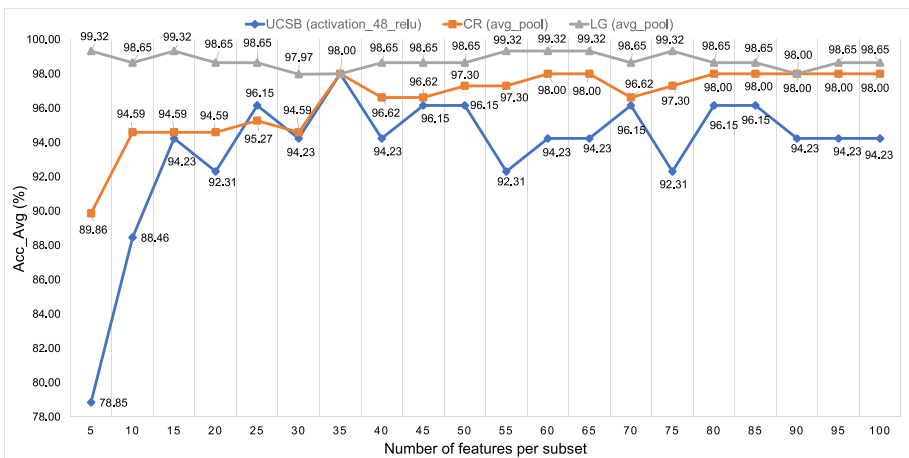
Fig. 6 The main results using ResNet-50's deep features with ReliefF algorithm and ensemble classifier: UCSB with the layer *activation_48_relu*; CR with the layer *avg_pool*; and LG with the layer *avg_pool*
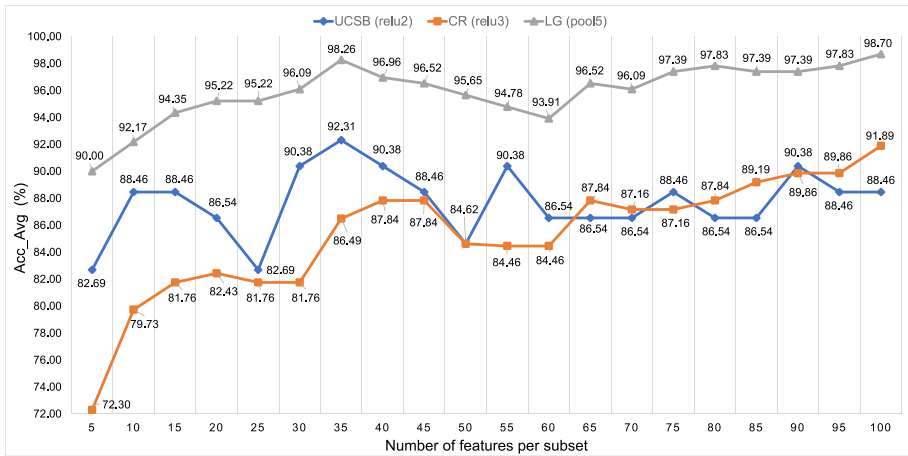
Fig. 7 The main results via AlexNet's deep features with ReliefF algorithm and ensemble classifier: UCSB with the layer *relu2*; CR with the layer *relu3*; and LG with the layer *pool5*

behaviors are in accordance with the investigations available in the Literature, that deeper layers tend to provide higher-level features [20, 74–77]. However, when the UCSB dataset is observed, the best hybrid model was based on only 35 deep features from an intermediate layer (*activation_48_relu*) of the ResNet-50 architecture. This model provided an accuracy rate of 98%. Thus, this study contributes to the Literature by indicating the detailed conditions of this fact for the pattern recognition of breast cancer via H&E images. Moreover, models exploring the *relu2* and *relu3* layers, belonging to the intermediate segment of the AlexNet architecture, were also responsible for providing expressive results in the UCSB and CR datasets, with accuracy values from 91.89% (CR dataset) to 98.70% (LG dataset). These results indicate that the proposed scheme was able to define the main layers and the corresponding features to quantify global and local patterns from different histological images [18].

Considering the conditions previously discussed, Table 4 summarizes the main hybrid models, with the layers that provided the most relevant deep features, the total of attributes used, and the accuracy rates in each histological dataset. It is verified that deep features from the ResNet-50 architecture define the best hybrid models, with a reduced number of descriptors (up to 35 features). This is another contribution since these conditions enabled

Table 4 Summary of the main hybrid models for classifying different histological images, with information regarding the network, layers and the number of deep features

Dataset	Network	Layer Name	Total of Features	Acc_Avg (%)
UCSB	AlexNet	*relu2*	35	92.31
	ResNet-50	*activation_48_relu*	35	98.00
CR	AlexNet	*relu3*	100	91.89
	ResNet-50	*avg_pool*	35	98.00
LG	AlexNet	*pool5*	100	98.70
	ResNet-50	*avg_pool*	5	99.32

Table 5 Accuracy rates (%) provided directly by the AlexNet and ResNet-50 architectures, exploring UCSB, CR and LG datasets without data augmentation

Dataset	AlexNet	ResNet-50
UCSB	53.33 ± 0.05	$\mathbf{60.00 \pm 0.14}$
CR	57.78 ± 0.06	$\mathbf{78.89 \pm 0.03}$
LG	86.96 ± 0.05	$\mathbf{88.41 \pm 0.05}$

the use of CNN in datasets with few samples and without overfitting, fact guaranteed through a robust ensemble classifier composed of five algorithms from different categories.

4.1 Comparisons with techniques for classification and pattern recognition

In this work, some consolidated machine-learning techniques were applied in order to evaluate the main hybrid models via direct comparisons. The results were provided directly by the AlexNet and ResNet-50 networks, as well as via regularized classification techniques: Lasso (least absolute shrinkage and selection operator) and Ridge regression [78–80]. Regularization approaches are widely used to reduce error by fitting a function appropriately to the given training set and avoiding overfitting. The Lasso technique minimizes the objective function by adding a penalty term to the sum of the absolute values of the coefficients. On the other hand, the Ridge strategy minimizes the objective function by adding a penalty term to the sum of squares of the coefficients. The experiments using the regularized techniques were performed in the Scikit-Learn 0.18.1 package [81].

It is important to highlight that experiments with a data augmentation approach were also performed, in order to increase the number of available samples and introduce variability in each set. The transfer learning toolbox for artificial data augmentation (available in MATLAB R2019a package [49]) was used in this process. The strategies used were: artificial random reflections of 50% across the x and y-axes; random rotations of up to 1 degree; and, random horizontal and vertical translations up to 1 pixel. These values were employed to minimize possible degradation of the classification rates due to the background of the image. These strategies allowed doubling the total number of samples available for the training and validation stages. In this type of test, the accuracy rates provided directly by the CNN models were used in the comparisons. The classifications were repeated three times to define the averages and standard deviations in each dataset. The values provided directly by the AlexNet and ResNet-50 architectures without data augmentation are shown in Table 5. These results were obtained from the first epoch of each network in order to avoid overfitting. Also, the performances after applying the data augmentation are shown in Tables 6 and 7 for the AlexNet and ResNet-50 networks, respectively. The accuracy values were defined with the number of training epochs ranging from 1 to 30. The most significant rates are highlighted in bold.

Table 6 Accuracy rates (%) provided directly by AlexNet exploring the UCSB, CR and LG datasets with data augmentation

Dataset	epoch 1	epoch 10	epoch 20	epoch 30
UCSB	60.00 ± 0.08	73.33 ± 0.12	76.67 ± 0.12	$\mathbf{83.33 \pm 0.12}$
CR	68.89 ± 0.15	80.78 ± 0.08	88.89 ± 0.02	$\mathbf{95.55 \pm 0.03}$
LG	73.92 ± 0.09	$\mathbf{98.55 \pm 0.01}$	95.53 ± 0.06	98.55 ± 0.01

Table 7 Accuracy values (%) provided directly by ResNet-50 exploring the UCSB, CR and LG datasets with data augmentation

Dataset	epoch 1	epoch 10	epoch 20	epoch 30
UCSB	53.33 ± 0.12	80.00 ± 0.01	**86.67 ± 0.05**	83.33 ± 0.12
CR	83.33 ± 0.08	95.56 ± 0.02	97.78 ± 0.02	**98.89 ± 0.02**
LG	89.85 ± 0.09	**99.28 ± 0.01**	97.83 ± 0.02	97.10 ± 0.01

In addition, regularized classification techniques were applied to the attributes to establish the main hybrid models, such as: *relu2* (186,624); *relu3* (64,896); *pool5* (9,216); *activation_48_relu* (25,088); 2,048 (*avg_pool*). This type of experiment was useful to indicate the advantages and limits of the main hybrid models. Regarding the solutions with regularized classification approaches, which can define subsets with high-quality features and increase the generalization of each model, the tests were performed through the SVM and logistic discrimination (LD) strategies with the Lasso and Ridge regularizations [68, 78, 82]. The accuracy rates provided by the regularized techniques are shown in Table 8 for the CR, UCSB, and LG datasets. The most relevant combinations were highlighted in bold.

Considering the hybrid models based on deep features from the AlexNet architecture (Table 4), the solutions indicated higher accuracy rates than those achieved via the CNN architectures (AlexNet and ResNet-50, Table 5) classifying directly the datasets without the data augmentation. These conditions illustrate the quality of the proposed scheme and the solutions obtained to improve CAD systems focused on H&E images (UCSB, CR and LG), in scenarios without data augmentation, even via deep features from a classic CNN. On the other hand, when data augmentation (Tables 6 and 7) and comparisons with regularized techniques (Table 8) are considered, the solutions based on deep features from the AlexNet network were more limited, surpassing the convolutional networks with data augmentation only in the UCSB dataset and the regularized techniques only in the LG dataset. These conditions clearly indicate the limits of hybrid models through the AlexNet architecture.

Table 8 Solutions and accuracy rates (%) via regularization strategies, considering the same feature sets used in main hybrid models

Classifier	Dataset	Layer and Network	Accuracy (%)
SVM with Lasso	UCSB	*relu2*, AlexNet	86.50
	CR	*avg_pool*, ResNet-50	97.29
	LG	***avg_pool*, ResNet-50**	**98.64**
SVM with Ridge	UCSB	*activation_48_relu*, ResNet-50	92.30
	CR	*relu3*, AlexNet	87.83
	LG	*avg_pool*, ResNet-50	97.82
LD with Lasso	UCSB	*activation_48_relu*, ResNet-50	75.00
	CR	***avg_pool*, ResNet-50**	**97.97**
	LG	*avg_pool*, ResNet-50	90.31
LD with Ridge	**UCSB**	**activation_48_relu, ResNet-50**	**94.23**
	CR	*avg_pool*, ResNet-50	96.62
	LG	*avg_pool*, ResNet-50	97.39

When the best hybrid models (the highest rates in Table 4 based on deep features of the ResNet-50) are compared with those via available approaches in Table 5, the hybrid solutions provided the best performances in the three datasets. For instance, the classification considering the UCSB dataset indicated the most relevant difference: ResNet-50 applied directly provided an accuracy rate of 60.00% against a rate of 98.00% via hybrid model (35 deep features from the *activation_48_relu* layer of the ResNet-50 with ReliefF algorithm and an ensemble classifier). Regarding the experiments exploring datasets with data augmentation (Tables 6 and 7), the highest difference (11.33%) can also be observed in the UCSB dataset, with an accuracy rate of 86.67% via ResNet-50 classifying directly the H&E images versus 98.00% of the hybrid model. With respect to the LG dataset, the results were 99.32% (hybrid model based on 5 deep features from the *avg_pool* of the ResNet-50 with ensemble classifier) against 99.28% (ResNet-50 classifying directly the H&E images with data augmentation), a difference of 0.04%. For the CR dataset, the best hybrid model (35 deep features from the *avg_pool* with ensemble classifier) provided a lower accuracy rate (0.89% difference) in relation to the achieved performance via ResNet-50 with data augmentation. This condition illustrates an important limit of this hybrid model. Thus, from these experiments, it is possible to define that the best hybrid models are better options to classify UCSB and LG datasets than the ones explored so far, regardless of the combination. This generalization was not observed in the context of colorectal images.

In relation to the best hybrid models against the regularized techniques (Table 8), LD with Ridge indicated an accuracy rate of 94.23% versus 98.00% of the hybrid model (35 deep features from the *activation_48_relu* with ensemble classifier) for the UCSB dataset. In the CR dataset, the LD and Lasso strategy provided an accuracy value of 97.97%, slightly lower in relation to the hybrid model (98.00%), an association of 35 deep features from the *avg_pool* layer. When the LG dataset is observed, the best hybrid model based on 5 deep features of *avg_pool* with ensemble classifier also outperformed the SVM and Lasso strategy, with accuracy values of 99.32% and 98.64%, respectively. From these comparisons, it is noted that the proposed scheme with the best hybrid models is a more robust option in relation to the regularized solutions, indicating the best performances.

In addition, Friedman's test was applied to evaluate the classifications provided by the best hybrid models, considering an overview regarding all datasets (Tables 6, 7 and 8). Friedman's test is a non-parametric statistic approach, able to rank k associations in a way that the main solution acquires rank 1 and the kth solution acquires rank k [83]. Thus, the average ranking is shown in Table 9 by taking into consideration the accuracy rates.

Table 9 Average ranking considering the best associations for UCSB, CR and LG datasets

Ranking	Solution	Score
1st	Hybrid models	1.50
2nd	ResNet-50 network with data augmentation	2.17
3rd	SVM with Lasso	4.00
4th	LD with Ridge	4.33
5th	SVM with Ridge	5.00
6th	AlexNet network with data augmentation	5.33
7th	LD with Lasso	5.67

It can be observed that the hybrid models appear in the first position of the average ranking (Table 9), even in comparison with the achieved results by important techniques. This fact indicates the potential of the hybrid models in the different tested conditions, with the best solutions for the UCSB (a hybrid model based on 35 deep features from the *activation_48_relu* with the ensemble classifier) and LG (a hybrid model based on 5 deep features from the *avg_pool* with the ensemble classifier) datasets. Other comparisons could be carried out to verify whether these results are maintained in more conditions and configurations, or even make adjustments to define the limits of each model. However, the presented experiments were able to provide a relevant overview concerning the main hybrid models when compared to the consolidated approaches commonly explored in the Literature for the classification and pattern recognition processes.

4.2 An illustrative overview of the obtained models in relation to the Literature

Different techniques have been presented in the Literature in order to investigate histological images, such as those for the UCSB, CR and LG datasets. The models were based on multiple combinations, exploring DL techniques, HC approaches, or different ensembles of descriptors and classifiers. An illustrative overview is important to show the quality of this study, with a proposed scheme and corresponding hybrid models not observed in multiple H&E images. This contextualization is shown in Tables 10, 11 and 12 for the UCSB, CR and LG datasets, respectively.

Taking into account this illustrative overview, it is noted that the achieved results are among those best ranked in the specialized Literature, even without exploring complex combinations with handcrafted features, deep-tuning, color normalization, ensemble of CNN models and others, such as described by [10, 14, 41, 47, 84]. Concerning the results presented in Table 10, the hybrid model (35 deep features from the *activation_48_relu* of the ResNet-50) provided

Table 10 Accuracy rates (%) provided by different approaches for breast histology image classification (UCSB)

Method	Approach	Accuracy (%)
Proposed hybrid model (via ResNet-50)	*activation_48_relu* layer, ReliefF and Ensemble (DL), with 35 deep features	98.00
Li et al. [41]	RefineNet and Atrous DenseNet (DL)	97.63
Feng et al. 2018 [40]	Stacked denoising autoencoder (DL)	94.41
Lee JS et al. [43]	DIU-Net (DL)	94.09
Longo et al. [14]	Inception-V3, Fractal Dimension and Lacunarity (DL+HC)	93.10
Kausar et al. [39]	Color normalization, Haar wavelet decomposition and 16-layer CNN (DL)	91.00
Roberto et al. [10]	Fractal neural network with two CNN models, 300 fractal features (fractal dimension, lacunarity and Percolation), DL+HC	89.66
Bouziane et al. [36]	Morphological, intensity and texture features (HC)	86.20
Papastergiou et al. [38]	Spatial decomposition, tensors (DL)	84.67
Araújo et al. [37]	Color normalization, 13-layer CNN and SVM (DL)	83.30
Saxena et al. [8]	ResNet-50 and kernelized weighted (DL)	60.30

Table 11 Accuracy rates (%) defined by different approaches for colorectal histology image classification (CR)

Method	Approach	Accuracy (%)
Roberto et al. [10]	Fractal neural network with two CNN models, 300 fractal features (fractal dimension, lacunarity and Percolation), DL+HC	99.39
Proposed hybrid model (via ResNet-50)	*avg_pool layer*, ReliefF and Ensemble (DL), with 35 deep features	98.00
Dabass et al. [47]	Hydrid CNN with attention learning (DL)	97.50
Dabass et al. [46]	31-layer CNN (DL)	96.97
Tavolara et al. [42]	GAN and U-Net (DL)	94.02
Sena et al. [44]	12-layer CNN (DL)	93.28
Dos Santos et al. [34]	Sample entropy and fuzzy (HC)	91.39
Roberto et al. [35]	Percolation (HC)	90.90
BenTaieb and Hamarneh [6]	U-Net and AlexNet (DL)	87.50
Zhang et al. [9]	ResNet deep-tuning (DL)	86.67
Awan et al. [45]	Color normalization, U-Net and GoogLeNet (DL)	85.00
Bianconi et al. [48]	ResNet, VGG, spatial and spectral features (DL+HC)	94.37

the best performance, surpassing those provided by recent studies, such as RefineNet and Atrous DenseNet [41], DIU-Net [43], Inception-V3 [14] and fractal neural networks [10]. Numerical differences in accuracy rates were up to 37.70% [8]. These facts show the robustness of the proposed method in order to provide a relevant association for classifying breast cancer via H&E images.

For the CR (see Table 11) and LG (see Table 12) datasets, the main hybrid models reached classification rates subtly lower than those provided by some strategies available in the Literature. For instance, the hybrid model via ResNet-50 (35 deep features from the *activation_48_relu* with ensemble classifier) achieved an accuracy rate of 98% to distinguish CR images, against 99.39% from a highly complex system (best model) with two CNN models and 300 fractal features (fractal dimension, lacunarity and percolation) [10]. Despite this, the proposed hybrid model via ResNet-50 outperformed other relevant schemes indicated for

Table 12 Accuracy values (%) achieved in different approaches for gender classification from liver images (LG)

Method	Approach	Accuracy (%)
Nanni et al. [84]	Ensemble of CNNs with texture features (HC) + SVM classifier	100.00
Roberto et al. [10]	Fractal neural network with two CNN models, 300 fractal features (fractal dimension, lacunarity and Percolation), DL+HC	99.62
Proposed hybrid model (via ResNet-50)	*avg_pool layer*, ReliefF and Ensemble (DL), with 5 deep features	99.32
Longo et al. [14]	Inception-V3, Fractal Dimension and Lacunarity (DL+HC)	99.25
Watanabe et al. [33]	GIST descriptors, PCA and LDA (HC)	93.70

CR [6, 9, 34, 35, 42, 44–48] and LG [14, 33] datasets, listed in Tables 11 and 12, respectively. Moreover, the hybrid model considering 5 deep features from the *avg_pool* with ensemble classifier indicated an accuracy rate of 99.32%, against a complex framework based on an ensemble of multiple CNN architectures, texture features (HC) and SVM classifier [84]. The accuracy rate was 100%, with some combinations exploring a single classifier, which can result in higher accuracy rates. On the other hand, it is necessary to evaluate situations in which the classifier is adjusted to the training data, including the bias-variance tradeoff [85]. The hybrid models presented here solve this problem by minimizing the possible overfitting with a robust ensemble classifier. In this case, the numerical difference concerning the accuracy rate was only 0.68% in relation to the results obtained in [84].

Finally, it is important to highlight that most of these proposals lead to an almost ideal model since the mentioned strategies used different types of features and combinations that were capable of quantifying the histological images. Thus, the best solution for distinguishing breast cancer and the valuable information defined in this study contribute to the community interested in the development and improvement of models for classifying patterns in H&E images.

5 Conclusion

In this paper, hybrid models were obtained through a computational scheme exploring deep features by transfer learning, selection by ranking and a robust ensemble classifier with five algorithms. The models were applied to classify histological images stained with H&E from breast, colorectal and liver tissue considering benign *versus* malignant groups (UCSB and CR datasets) and pattern recognition in liver tissue images from mice separated into male and female classes (LG dataset). The best results were obtained through the ResNet-50 architecture in the *activation_48_relu* (UCSB) and *avg_pool* (CR and LG) layers, with a proposed scheme able to define the highest accuracy rate with a reduced number of features in each scenario (up to 35 attributes). The results were accuracy values of 98.00% (UCSB and CR) and 99.32% (LG).

The hybrid model via the *pool5* layer (AlexNet network) achieved an accuracy value of 98.70% in the LG dataset. In the same dataset, the best hybrid model with the deepest layer of the ResNet-50 network (*avg_pool*) achieved 99.32%. This association also provided the most relevant features for the CR dataset, with an accuracy value of 98.00%. The models that explore the deepest CNN layers are the most commonly used in important approaches available in the Literature. However, the tested conditions in this study show that deep features from the *activation_48_relu* layer (ResNet-50) provided a model with the best rate in the UCSB dataset. Thus, these facts show the capacity of the proposed scheme to optimize the transfer learning process and present the relevant hybrid models for classification and pattern recognition in H&E images.

The main results were compared to the obtained performances with consolidated machine-learning techniques, CNN models directly applied to classify the datasets, as well as results via regularized classification techniques (Lasso and Ridge regression). Experiments with a data augmentation approach were also evaluated. In this context, it was demonstrated that the main hybrid models, based on deep features from the AlexNet, indicated higher accuracy rates than those achieved via convolutional architectures (AlexNet and ResNet-50) classifying directly the datasets without data augmentation. With data augmentation, the hybrid models based on deep features from the AlexNet were more limited, with relevant results only in

the UCSB dataset. In relation to the best hybrid models, based on deep features from the ResNet-50, the obtained solutions were better options to classify the UCSB and LG datasets in comparison with the CNN models, exploring data augmentation or not. This generalization was not observed for the CR dataset. In addition, when the comparisons with the regularized techniques were considered, the hybrid model (via AlexNet) provided relevant results only in the LG dataset. On the other hand, the best hybrid models were more robust options, indicating the best performances in the three datasets. This is another important contribution of this study.

In this context, when all comparison conditions are considered (CNN models applied directly to the images, data augmentation or regularized approaches), it is concluded that the hybrid models, based on the deep features of the ResNet-50, are the more relevant solutions for two of the three investigated datasets: UCSB, hybrid model based on 35 deep features from the *activation_48_relu* layer with ReliefF algorithm and ensemble classifier; LG, hybrid model based on 5 deep features from the *avg_pool* layer with ReliefF algorithm and ensemble classifier. The information presented here allows the use of hybrid models via CNN strategies in datasets with a reduced number of samples, without overfitting. Also, these conditions can be used to improve CAD systems focused on H&E images or even as more robust baseline schemes in this type of investigation.

Finally, taking into account an illustrative overview of the obtained models in relation to the Literature, it is observed that the achieved results are among the best ranked, with emphasis on the UCSB context. The proposed scheme provided the best solution among those available in the Literature, based on only 35 deep features from the *activation_48_relu* (intermediate layer), ReliefF algorithm and ensemble classifier. For the CR and LG datasets, the best hybrid models provided subtly lower performances, indicating a possible limit of the proposal.

In future works, it is intended to: 1) use the main associations for pattern recognition in different types of H&E images; 2) explore the main solutions with interpretable CNN models; 3) map each region of the image that provided the most relevant features, investigating the explainable artificial intelligence.

Data availability Not applicable.

Code availability Not applicable.

Declarations

Conflicts of interest / Competing interests The authors have no conflicts of interest to declare that are relevant to the content of this article.

References

1. Yang M, Kpalma K, Ronsin J (2008) A survey of shape feature extraction techniques
2. Goodfellow I, Bengio Y, Courville A (2016) Deep learning. MIT press

3. Gad AF (2018) Convolutional neural networks. In: Practical Computer Vision Applications Using Deep Learning with CNNs, Springer, pp 183–227

4. Hinton GE (2007) Learning multiple layers of representation. Trends Cogn Sci 11(10):428–434

5. Le QV (2013) Building high-level features using large scale unsupervised learning. In: 2013 IEEE international conference on acoustics, speech and signal processing, IEEE, pp 8595–8598

6. BenTaieb A, Hamarneh G (2017) Adversarial stain transfer for histopathology image analysis. IEEE Trans Med Imaging 37(3):792–802

7. Sethy PK, Behera SK (2022) Automatic classification with concatenation of deep and handcrafted features of histological images for breast carcinoma diagnosis. Multimedia Tools and Applications 81(7):9631–9643

8. Saxena S, Shukla S, Gyanchandani M (2020) Breast cancer histopathology image classification using kernelized weighted extreme learning machine. International Journal of Imaging Systems and Technology

9. Zhang R, Zhu J, Yang S, Hosseini MS, Genovese A, Chen L, Rowsell C, Damaskinos S, Varma S, Plataniotis KN (2022) Histokt: Cross knowledge transfer in computational pathology. ICASSP 2022–2022 IEEE International Conference on Acoustics. Speech and Signal Processing (ICASSP), IEEE, pp 1276–1280

10. Roberto GF, Lumini A, Neves LA, do Nascimento MZ, (2021) Fractal neural network: A new ensemble of fractal geometry and convolutional neural networks for the classification of histology images. Expert Syst Appl 166:114103

11. He K, Zhang X, Ren S, Sun J (2016) Deep residual learning for image recognition. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 770–778

12. Kim YJ, Bae JP, Chung JW, Park DK, Kim KG, Kim YJ (2021) New polyp image classification technique using transfer learning of network-in-network structure in endoscopic images. Sci Rep 11(1):3605

13. Tan C, Sun F, Kong T, Zhang W, Yang C, Liu C (2018) A survey on deep transfer learning. In: International conference on artificial neural networks, Springer, pp 270–279

14. Longo LHDC, Martins AS, Do Nascimento MZ, Dos Santos LFS, Roberto GF, Neves LA (2022) Ensembles of fractal descriptors with multiple deep learned features for classification of histological images. 2022 29th International Conference on Systems. Signals and Image Processing (IWSSIP), IEEE, pp 1–4

15. Ghandour C, El-Shafai W, El-Rabaie S (2023) Medical image enhancement algorithms using deep learning-based convolutional neural network. Journal of Optics pp 1–11

16. Deng J, Dong W, Socher R, Li LJ, Li K, Fei-Fei L (2009) Imagenet: A large-scale hierarchical image database. In: 2009 IEEE conference on computer vision and pattern recognition, Ieee, pp 248–255

17. Kumar S, Sharma S (2021) Sub-classification of invasive and non-invasive cancer from magnification independent histopathological images using hybrid neural networks. Evolutionary Intelligence pp 1–13

18. dos Santos FP, Ponti MA (2019) Alignment of local and global features from multiple layers of convolutional neural network for image classification. 2019 32nd SIBGRAPI Conference on Graphics. Patterns and Images (SIBGRAPI), IEEE, pp 241–248

19. Coccia M (2020) Deep learning technology for improving cancer care in society: New directions in cancer imaging driven by artificial intelligence. Technol Soc 60:101198

20. dos Santos FP, Ponti MA (2018) Robust feature spaces from pre-trained deep network layers for skin lesion classification. 2018 31st SIBGRAPI Conference on Graphics. Patterns and Images (SIBGRAPI), IEEE, pp 189–196

21. Younas F, Usman M, Yan WQ (2022) An ensemble framework of deep neural networks for colorectal polyp classification. Multimedia Tools and Applications pp 1–22

22. Tenguam JJ, Longo LHDC, Silva AB, De Faria PR, Do Nascimento MZ, Neves LA (2022) Classification of h&e images exploring ensemble learning with two-stage feature selection. 2022 29th International Conference on Systems. Signals and Image Processing (IWSSIP), IEEE, pp 1–4

23. Abraham B, Nair MS (2020) Computer-aided detection of covid-19 from x-ray images using multi-cnn and bayesnet classifier. Biocybernetics and biomedical engineering 40(4):1436–1445

24. Novitasari DCR, Hendradi R, Caraka RE, Rachmawati Y, Fanani NZ, Syarifudin A, Toharudin T, Chen RC (2020) Detection of covid-19 chest x-ray using support vector machine and convolutional neural network. Commun Math Biol Neurosci 2020:Article–ID

25. Urbanowicz RJ, Meeker M, La Cava W, Olson RS, Moore JH (2018) Relief-based feature selection: Introduction and review. J Biomed Inform 85:189–203

26. Manhrawy II, Qaraad M, El-Kafrawy P (2021) Hybrid feature selection model based on relief-based algorithms and regulizer algorithms for cancer classification. Concurrency and Computation: Practice and Experience 33(17):e6200

27. Ghosh P, Azam S, Jonkman M, Karim A, Shamrat FJM, Ignatious E, Shultana S, Beeravolu AR, De Boer F (2021) Efficient prediction of cardiovascular disease using machine learning algorithms with relief and lasso feature selection techniques. IEEE Access 9:19304–19326

28. Zebari R, Abdulazeez A, Zeebaree D, Zebari D, Saeed J (2020) A comprehensive review of dimensionality reduction techniques for feature selection and feature extraction. Journal of Applied Science and Technology Trends 1(2):56–70

29. Bolón-Canedo V, Sánchez-Marono N, Alonso-Betanzos A, Benítez JM, Herrera F (2014) A review of microarray datasets and applied feature selection methods. Inf Sci 282:111–135

30. Li M, Ma X, Chen C, Yuan Y, Zhang S, Yan Z, Chen C, Chen F, Bai Y, Zhou P et al (2021) Research on the auxiliary classification and diagnosis of lung cancer subtypes based on histopathological images. Ieee Access 9:53687–53707

31. Burçak KC, Uğuz H (2022) A new hybrid breast cancer diagnosis model using deep learning model and relieff. Traitement du Signal 39(2):521–529

32. Silva AB, De Oliveira CI, Pereira DC, Tosta TA, Martins AS, Loyola AM, Cardoso SV, De Faria PR, Neves LA, Do Nascimento MZ (2022) Assessment of the association of deep features with a polynomial algorithm for automated oral epithelial dysplasia grading. In: 2022 35th SIBGRAPI Conference on Graphics, Patterns and Images (SIBGRAPI), IEEE, vol 1, pp 264–269

33. Watanabe K, Kobayashi T, Wada T (2016) Semi-supervised feature transformation for tissue image classification. PLoS ONE 11(12):e0166413

34. Dos Santos LFS, Neves LA, Rozendo GB, Ribeiro MG, do Nascimento MZ, Tosta TAA, (2018) Multidimensional and fuzzy sample entropy (sampenmf) for quantifying h&e histological images of colorectal cancer. Comput Biol Med 103:148–160

35. Roberto GF, Nascimento MZ, Martins AS, Tosta TA, Faria PR, Neves LA (2019) Classification of breast and colorectal tumors based on percolation of color normalized images. Computers & Graphics 84:134–143

36. Bouziane A, Boumali S, Berkane N, Guendouz FS (2020) A hybrid approach for automatic breast cancer detection. In: 2020 International Conference on e-Health and Bioengineering (EHB), IEEE, pp 1–4

37. Araújo T, Aresta G, Castro E, Rouco J, Aguiar P, Eloy C, Polónia A, Campilho A (2017) Classification of breast cancer histology images using convolutional neural networks. PLoS ONE 12(6):e0177544

38. Papastergiou T, Zacharaki EI, Megalooikonomou V (2018) Tensor decomposition for multiple-instance classification of high-order medical data. Complexity 2018

39. Kausar T, Wang M, Idrees M, Lu Y (2019) Hwdcnn: Multi-class recognition in breast histopathology with haar wavelet decomposed image based convolution neural network. Biocybernetics and Biomedical Engineering 39(4):967–982

40. Feng Y, Zhang L, Yi Z (2018) Breast cancer cell nuclei classification in histopathology images using deep neural networks. Int J Comput Assist Radiol Surg 13(2):179–191

41. Li Y, Xie X, Shen L, Liu S (2019) Reverse active learning based atrous densenet for pathological image classification. BMC Bioinformatics 20(1):1–15

42. Tavolara TE, Niazi MKK, Arole V, Chen W, Frankel W, Gurcan MN (2019) A modular cgan classification framework: Application to colorectal tumor detection. Sci Rep 9(1):1–8

43. Lee JS, Wu WK (2022) Breast tumor tissue image classification using diu-net. Sensors 22(24):9838

44. Sena P, Fioresi R, Faglioni F, Losi L, Faglioni G, Roncucci L (2019) Deep learning techniques for detecting preneoplastic and neoplastic lesions in human colorectal histological images. Oncol Lett 18(6):6101–6107

45. Awan R, Al-Maadeed S, Al-Saady R, Bouridane A (2020) Glandular structure-guided classification of microscopic colorectal images using deep learning. Computers & Electrical Engineering 85:106450

46. Dabass M, Vig R, Vashisth S (2018) Five-grade cancer classification of colon histology images via deep learning. In: CRC Press, p 18

47. Dabass M, Vashisth S, Vig R (2022) A convolution neural network with multi-level convolutional and attention learning for classification of cancer grades and tissue structures in colon histopathological images. Comput Biol Med 147:105680

48. Bianconi F, Kather JN, Reyes-Aldasoro CC (2020) Experimental assessment of color deconvolution and color normalization for automated classification of histology images stained with hematoxylin and eosin. Cancers 12(11):3337

49. MATLAB (2019) 9.6.0.1072779 (R2019a). The MathWorks Inc., Natick, Massachusetts

50. Witten IH, Frank E (2002) Data mining: practical machine learning tools and techniques with java implementations. ACM SIGMOD Rec 31(1):76–77

51. Gelasca ED, Byun J, Obara B, Manjunath B (2008) Evaluation and benchmark for biological image segmentation. In: 2008 15th IEEE International Conference on Image Processing, IEEE, pp 1816–1819

52. Sirinukunwattana K, Pluim JP, Chen H, Qi X, Heng PA, Guo YB, Wang LY, Matuszewski BJ, Bruni E, Sanchez U et al (2017) Gland segmentation in colon histology images: The glas challenge contest. Med Image Anal 35:489–502

53. AGEMAP NIoA (2020) The atlas of gene expression in mouse aging project (agemap). https://ome.grc.nia.nih.gov/iicbu2008/agemap/index.html, acesso em: 04/05/2020

54. Rajesh G, Anirudh V, Archana R, Kumar PP, Manoj K (2023) An improved skin cancer classification method using deep convolutional neural networks and transfer learning models. Journal of Engineering Sciences 14(05)

55. Viet-Linh T (2023) Deep convolutional neural network-based transfer learning method for health condition identification of cable in cable-stayed bridge. Journal of Materials and Engineering Structures 10(1):5–18

56. Lu S, Lu Z, Zhang YD (2019) Pathological brain detection based on alexnet and transfer learning. Journal of computational science 30:41–47

57. Krizhevsky A, Sutskever I, Hinton GE (2012) Imagenet classification with deep convolutional neural networks. In: Advances in neural information processing systems, pp 1097–1105

58. Nair V, Hinton GE (2010) Rectified linear units improve restricted boltzmann machines. In: Proceedings of the 27th international conference on machine learning (ICML-10), pp 807–814

59. Al Rahhal MM, Bazi Y, Abdullah T, Mekhalfi ML, AlHichri H, Zuair M (2018) Learning a multi-branch neural network from multiple sources for knowledge adaptation in remote sensing imagery. Remote Sensing 10(12):1890

60. Ribeiro MG, Neves LA, do Nascimento MZ, Roberto GF, Martins AS, Tosta TAA, (2019) Classification of colorectal cancer based on the association of multidimensional and multiresolution features. Expert Syst Appl 120:262–278

61. Kononenko I, Robnik-Sikonja M, Pompe U (1996) Relieff for estimation and discretization of attributes in classification, regression, and ilp problems. Artificial intelligence: methodology, systems, applications pp 31–40

62. Kononenko I, Šimec E, Robnik-Šikonja M (1997) Overcoming the myopia of inductive learning algorithms with relieff. Appl Intell 7(1):39–55

63. Robnik-Šikonja M, Kononenko I (2003) Theoretical and empirical analysis of relieff and rrelieff. Mach Learn 53(1):23–69

64. Cui X, Li Y, Fan J, Wang T (2022) A novel filter feature selection algorithm based on relief. Appl Intell 52(5):5063–5081

65. Sagi O, Rokach L (2018) Ensemble learning: A survey. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery 8(4):e1249

66. Kassani SH, Kassani PH (2019) A comparative study of deep learning architectures on melanoma detection. Tissue Cell 58:76–83

67. Cleary JG, Trigg LE (1995) K*: An instance-based learner using an entropic distance measure. In: Machine Learning Proceedings 1995, Elsevier, pp 108–114

68. Le Cessie S, Van Houwelingen JC (1992) Ridge estimators in logistic regression. J Roy Stat Soc: Ser C (Appl Stat) 41(1):191–201

69. Lewis DD (1998) Naive (bayes) at forty: The independence assumption in information retrieval. In: Machine Learning: ECML-98: 10th European Conference on Machine Learning Chemnitz, Germany, April 21–23, 1998 Proceedings 10, Springer, pp 4–15

70. Breiman L (2001) Random forests. Machine learning 45(1):5–32

71. Alpaydin E (2009) Introduction to machine learning. MIT press

72. Kittler J, Hatef M, Duin RP, Matas J (1998) On combining classifiers. IEEE Trans Pattern Anal Mach Intell 20(3):226–239

73. King AP, Eckersley RJ (2019) Chapter 6 - inferential statistics iii: Nonparametric hypothesis testing. In: Eckersley RJ (ed) King AP. Statistics for Biomedical Engineers and Scientists, Academic Press, pp 119–145

74. Majtner T, Yildirim-Yayilgan S, Hardeberg JY (2016) Combining deep learning and hand-crafted features for skin lesion classification. 2016 Sixth International Conference on Image Processing Theory. Tools and Applications (IPTA), IEEE, pp 1–6

75. Esteva A, Kuprel B, Novoa RA, Ko J, Swetter SM, Blau HM, Thrun S (2017) Dermatologist-level classification of skin cancer with deep neural networks. nature 542(7639):115–118

76. dos Santos FP, Ribeiro LS, Ponti MA (2019) Generalization of feature embeddings transferred from different video anomaly detection domains. J Vis Commun Image Represent 60:407–416

77. Shi Z, Hao H, Zhao M, Feng Y, He L, Wang Y, Suzuki K (2019) A deep cnn based transfer learning method for false positive reduction. Multimedia Tools and Applications 78(1):1017–1033

78. Ng AY (2004) Feature selection, l 1 vs. l 2 regularization, and rotational invariance. In: Proceedings of the twenty-first international conference on Machine learning, p 78

79. Kolter JZ, Ng AY (2009) Regularization and feature selection in least-squares temporal difference learning. In: Proceedings of the 26th annual international conference on machine learning, pp 521–528

80. Schölkopf B, Smola AJ, Bach F, et al (2002) Learning with kernels: support vector machines, regularization, optimization, and beyond. MIT press

81. Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, Blondel M, Prettenhofer P, Weiss R, Dubourg V, et al (2011) Scikit-learn: Machine learning in python. Journal of machine learning research 12(Oct):2825–2830

82. Fan RE, Chang KW, Hsieh CJ, Wang XR, Lin CJ (2008) Liblinear: A library for large linear classification. Journal of machine learning research 9(Aug):1871–1874

83. Demšar J (2006) Statistical comparisons of classifiers over multiple data sets. Journal of Machine learning research 7(Jan):1–30

84. Nanni L, Brahnam S, Ghidoni S, Maguolo G (2019) General purpose (genp) bioimage ensemble of handcrafted and learned features with data augmentation. arXiv preprint arXiv:1904.08084

85. Dong X, Yu Z, Cao W, Shi Y, Ma Q (2020) A survey on ensemble learning. Front Comp Sci 14:241–258

Authors and Affiliations

Cléber I. de Oliveira[1] ⓘ **· Marcelo Z. do Nascimento[2] · Guilherme F. Roberto[2] · Thaína A. A. Tosta[3] · Alessandro S. Martins[4] · Leandro A. Neves[1]**

Marcelo Z. do Nascimento
marcelo.zanchetta@gmail.com

Guilherme F. Roberto
guilhermefroberto@gmail.com

Thaína A. A. Tosta
tosta.thaina@gmail.com

Alessandro S. Martins
alessandro@iftm.edu.br

Leandro A. Neves
leandro.neves@unesp.br

[1] Department of Computer Science and Statistics (DCCE), São Paulo State University (UNESP), Rua Cristóvão Colombo, 2265, 15054-000 São José do Rio Preto, São Paulo, Brazil

[2] Faculty of Computer Science (FACOM) - Federal University of Uberlândia (UFU), Avenida João Neves de Ávila 2121, Bl.B, 38400-902 Uberlândia, Minas Gerais, Brazil

[3] Science and Technology Institute (ICT), Federal University of São Paulo (UNIFESP), Avenida Cesare Mansueto Giulio Lattes, 1201, 12247-014 São José dos Campos, São Paulo, Brazil

[4] Federal Institute of Triângulo Mineiro (IFTM), Rua Belarmino Vilela Junqueira sn, 38305-200 Ituiutaba, Minas Gerais, Brazil