Breast Cancer Histopathological Image Classification using Convolutional Neural Networks

Fabio A. Spanhol, Luiz S. Oliveira
Federal University of Parana
Department of Informatics (DInf)
Curitiba, PR - Brazil
Email: {faspanhol, lesoliveira}@inf.ufpr.br

Caroline Petitjean, and Laurent Heutte
University of Rouen
LITIS Lab
Saint Etienne du Rouvray, France
Email: {caroline.petitjean, laurent.heutte}@univ-rouen.fr

***Abstract*—The performance of most conventional classification systems relies on appropriate data representation and much of the efforts are dedicated to feature engineering, a difficult and time-consuming process that uses prior expert domain knowledge of the data to create useful features. On the other hand, deep learning can extract and organize the discriminative information from the data, not requiring the design of feature extractors by a domain expert. Convolutional Neural Networks (CNNs) are a particular type of deep, feedforward network that have gained attention from research community and industry, achieving empirical successes in tasks such as speech recognition, signal processing, object recognition, natural language processing and transfer learning. In this paper, we conduct some preliminary experiments using the deep learning approach to classify breast cancer histopathological images from BreaKHis, a publicly dataset available at http://web.inf.ufpr.br/vri/breast-cancer-database. We propose a method based on the extraction of image patches for training the CNN and the combination of these patches for final classification. This method aims to allow using the high-resolution histopathological images from BreaKHis as input to existing CNN, avoiding adaptations of the model that can lead to a more complex and computationally costly architecture. The CNN performance is better when compared to previously reported results obtained by other machine learning models trained with hand-crafted textural descriptors. Finally, we also investigate the combination of different CNNs using simple fusion rules, achieving some improvement in recognition rates.**

I. Introduction

NOWADAYS, cancer is a massive public health problem around the world. According to the International Agency for Research on Cancer (IARC), part of the World Health Organization (WHO), there were 8.2 million deaths caused by cancer in 2012 and 27 million of new cases of this disease are expected to occur until 2030 [1]. Among the cancer types, breast cancer (BC) is second most common for women, excluding skin cancer. Besides, the mortality of BC is very high when compared to other types of cancer. Even in face of recent advances in the comprehension of the molecular biology of BC progression and the discovery of new related molecular markers, the histopathological analysis remains the most widely used method for BC diagnosis [2]. Despite significant progress reached by diagnostic imaging technologies, the final BC diagnosis, including grading and staging, continues being done by pathologists applying visual inspection of histological samples under the microscope. Recent advances in image processing and machine learning techniques allow to build Computer-Aided Detection/Diagnosis (CAD/CADx) systems that can assist pathologists to be more productive, objective and consistent in diagnosis. Classification of histopathology images into distinct histopathology patterns, corresponding to the non-cancerous or cancerous condition of the analyzed tissue, is often the primordial goal in image analysis systems for cancer automatic aided diagnosis applications. The main challenge of such systems is dealing with the inherent complexity of histopathological images.

The automatic imaging processing for cancer diagnosis has been explored as a topic of research for more than 40 years [3] but is still challenging due to the complexity of the images to analyze. For example, Kowal *et al.* [4] compare and test different algorithms for nuclei segmentation, where the cases are classified as either benign or malignant on a dataset of 500 images, and report accuracies ranging from 96% to 100%. Filipczuk *et al.* [5] present a BC diagnosis system based on the analysis of cytological images of fine needle biopsies, to discriminate the images as either benign or malignant. Using four different classifiers trained with a 25-dimensional feature vector, they report a performance of 98% on 737 images. Similarly to [4] and [5], George *et al.* [6] propose a diagnosis system for BC based on the nuclei segmentation of cytological images. Using different machine learning models, such as neural networks and support vector machines, they report accuracy rates ranging from 76% to 94% on a dataset of 92 images. Zhang *et al.* [7] propose a cascade approach with rejection option. In the first level of the cascade, authors expect to solve the easy cases while the hard ones are sent to a second level where a more complex pattern classification system is used. They assess the proposed method on a database proposed by the Israel Institute of Technology, which is composed of 361 images and report results of 97% of reliability. In another work [8], the same authors assess an ensemble of one-class-classifiers on the same database achieving a recognition rate of 92%.

Most of these recent works related to BC classification are focused on Whole-Slide Imaging (WSI) [7], [8], [6], [4], [9]. However, the broad adoption of WSI and other forms of digital pathology still facing obstacles such as the high cost of implementing and operating the technology, insuffi-

cient productivity for high-volume clinical routines, intrinsic technology-related concerns, unsolved regulatory issues, as well as "cultural resistance" from the pathologists [10].

Until recently, most of the works on BC histopathology image analysis were carried out on small datasets, which are usually not available to the scientific community. Contributing to mitigate this gap, Spanhol *et al.* [11] introduced a dataset composed of 7,909 breast histopathological images acquired on 82 patients. In the same study, the authors evaluated six different textural descriptors and different classifiers and reported a series of experiments with accuracy rates ranging from 80% to 85%, depending on the image magnification factor. Based on the results presented in [11], it is undeniable that the texture descriptors can offer a good representation to train classifiers. However, some researchers advocate that the main weakness of the current machine learning methods lies exactly on this feature engineering step [12], [13]. To them, machine learning algorithms should be less dependent on feature engineering by being able to extract and organize the discriminative information from the data, in other words, should be capable of learning the representation.

The idea of representation learning is not new but it emerged only recently as a viable alternative due to the appearance and popularization of the Graphic Processing Units (GPUs) which are capable of delivering high computational throughput at relatively low cost, achieved through their massively parallel architecture. Among the different approaches, the Convolutional Neural Network (CNN) introduced by LeCun in [14], has been widely used to achieve state-of-the-art results in different pattern recognition problems [15], [16]. In the case of texture classification it has not been different. Hafemann *et al.* [17] have shown, for images of microscopic and macroscopic texture, that CNN is able to surpass traditional textural descriptors. Besides, the traditional approach to extract appropriate features for classification tasks in pathological images requires considerable efforts and effective expert domain knowledge, frequently leading to highly customized solutions, specific for each problem and hardly applicable in other contexts [18].

In light of this, in this work we evaluate the deep learning approach for the problem of BC histopathological image classification. Besides assessing different CNN architectures, we also investigate different methods to deal with high-resolution texture images without changing the CNN architecture used for low-resolution images. A set of comprehensive experiments on the BreaKHis dataset proposed in [11] shows that the CNN achieves better results than the best results obtained by the other machine learning models trained with textural descriptors. The best performance, though, are obtained by combining different CNNs using simple fusion rules, such as Max, Product, and Sum, leading to an improvement in classification accuracy of 6% when compared to the experiments reported in [11].

The remaining of this paper is organized as follows: Section II briefly introduces the BreaKHis database. Section III covers a short introduction to deep learning using CNN. Section IV describes the architecture of the CNN used in our experiments. Section V reports our experiments and discusses our results. Finally, Section VI concludes the work presenting some insights for further researches.

II. BreaKHis Database

The BreaKHis database [11] contains microscopic biopsy images of benign and malignant breast tumors. Images were collected through a clinical study from January 2014 to December 2014. All patients referred to the P&D Lab, Brazil, during this period of time, with a clinical indication of BC were invited to participate in the study. The institutional review board approved the study and all patients gave written informed consent. All the data were anonymized.

Samples are generated from breast tissue biopsy slides, stained with hematoxylin and eosin (HE). The samples are collected by surgical (open) biopsy (SOB), prepared for histological study and labeled by pathologists of the P&D Lab. The preparation procedure used in this work is the standard paraffin process, which is widely used in clinical routine. The main goal is to preserve the original tissue structure and molecular composition, allowing to observe it in a light microscope. The complete preparation procedure includes steps such as fixation, dehydration, clearing, infiltration, embedding, and trimming [19]. To be mounted on slides, sections of around 3 μm are cut using a microtome. After staining, the sections are covered with a glass coverslip. Then the anatomopathologists identify the tumoral areas in each slide, by visual analysis of tissue sections under a microscope. Final diagnosis of each case is produced by experienced pathologists and confirmed by complementary exams such as immunohistochemistry (IHC) analysis.

An Olympus BX-50 system microscope with a relay lens with magnification of $3.3\times$ coupled to a Samsung digital color camera SCC-131AN is used to obtain digitized images from the breast tissue slides. Images are acquired in 3-channel RGB (Red-Green-Blue) TrueColor (24-bit color depth, 8 bits per color channel) color space using magnifying factors of $40\times$, $100\times$, $200\times$ and $400\times$, corresponding to objective lens $4\times$, $10\times$, $20\times$, and $40\times$.

Figure 1 shows four images — with the four magnification factors (a) $40\times$, (b) $100\times$, (c) $200\times$, and (d) $400\times$ — acquired from a single slide of breast tissue containing a malignant tumor (breast cancer). Highlighted rectangle (manually added for illustrative purposes only) is the area of interest selected by pathologist to be detailed in the next higher magnification.

To date, the database is composed of 7,909 images divided into benign and malignant tumors. Table I summarizes the image distribution.

III. Deep Learning Approach Using CNN

Image classification based on visual content, especially microscopic images from histopathologic sections, is a challenging task, facing issues such as the usually large amount of inter-intraclass variability, the presence of rich geometrical structures due to structural-morphological diversity, and complex textures. Figure 2 shows typical complex textures

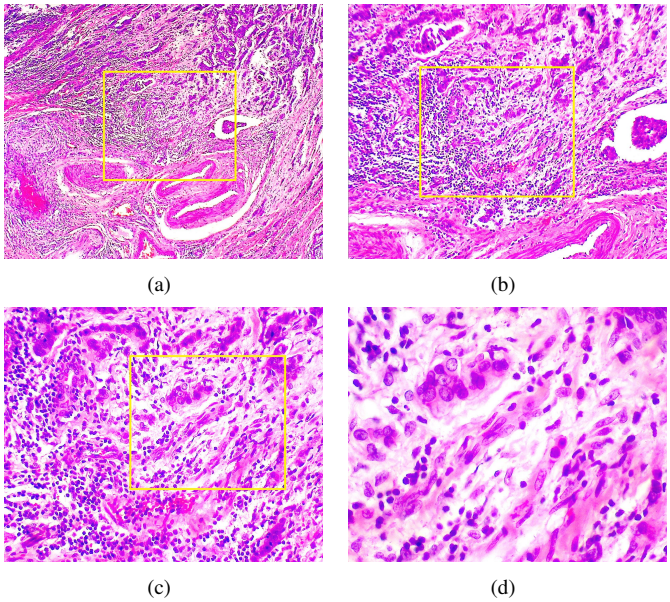


Figure 1. A slide of breast malignant tumor (stained with HE) seen in different magnification factors: (a) $40\times$, (b) $100\times$, (c) $200\times$, and (d) $400\times$. Highlighted rectangle (manually added for illustrative purposes only) is the area of interest selected by pathologist to be detailed in the next higher magnification factor.

Table I
IMAGE DISTRIBUTION BY MAGNIFICATION FACTOR AND CLASS

Magnification	Benign	Malignant	Total
$40\times$	625	1,370	1,995
$100\times$	644	1,437	2,081
$200\times$	623	1,390	2,013
$400\times$	588	1,232	1,820
Total	2,480	5,429	7,909
# Patients	24	58	82

found in histopathological images. Deep learning explores the possibility of learning features directly from input data, avoiding hand-crafted features [12]. The key concept of deep learning is to discover multiple levels of representation aiming that higher-level features represent more abstract semantics of the data [13]. As a particular deep learning technique, Convolutional Neural Networks (CNNs) [13] have achieved success in image classification problems, including medical image analysis [20], [21], [22], [23]. In summary, a CNN consists of multiple trainable stages stacked on top of each other, followed by a supervised classifier and sets of arrays named *feature maps* represent both input and output of each stage [24]. Input can be signals such as image, audio, and video. For example, considering color images, at the input each feature map is a 2D array storing a color channel of the input image. The output consists of a set arrays where each feature map represents a particular feature extracted at locations of the associated input.

A deep net is trained by feeding it input and letting it compute layer-by-layer to generate the final output for comparison with the correct answer. After computing the error at the output, this error flows backward through the net by back-propagation. At each step backward the model parameters are tuned in a direction that tries to reduce the error. This process sweeps over the data improving the model as it goes. Typically, training is an iterative process that involves multiple passes of the input data until the model converges.

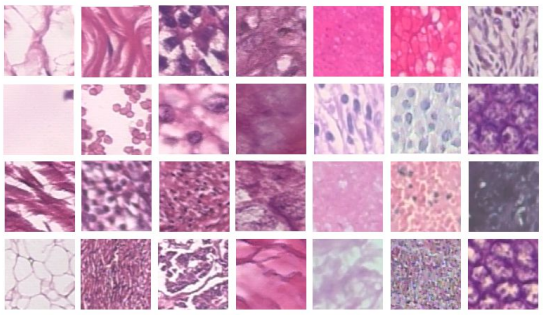


Figure 2. Examples of real textures present in histopathological images (HE staining).

There are three main types of layers used to build CNN architectures: *convolutional layer*, *pooling layer*, and *fully-connected layer*. Normally, a full CNN architecture is obtained by stacking several of these layers. An example of typical CNN architecture with two feature stages is shown in Figure 3.

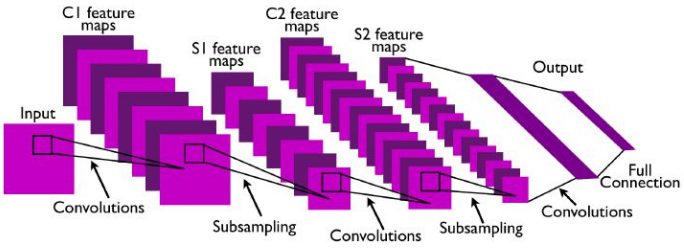


Figure 3. Example of typical CNN architecture with two feature stages. Extracted from [24].

In a CNN, the key computation is the convolution of a feature detector with an input signal. Convolutional layer computes the output of neurons connected to local regions in the input, each one computing a dot product between their weights and the region they are connected to in the input volume. The set of weights which is convolved with the input is called *filter* or *kernel*. Every filter is small spatially (width and height), but extends through the full depth of the input volume. For inputs such as images typical filters are small areas (e. g., 3×3, 5×5, or 8×8) and each neuron is connected only to this area in the previous layer. The weights are shared across neurons, leading the filters to learn frequent patterns that occur in any part of the image. The distance between the applications of filters is called *stride*. Whether stride hyperparameter is smaller than the filter size the convolution is applied in overlapping windows.

Convolution with a collection of filters, like the learned filters (also named feature maps or activation maps) in Fig-

ure 6, improves the representation: at the first layer of a CNN, the features go from individual pixels to simple primitives like horizontal and vertical lines, circles, and patches of color. In contrast to conventional single-channel image processing filters, these CNN filters are computed across all of the input channels. Due to its translation-invariant property, convolutional filters yield a high response wherever a feature is detected.

It is common the insertion of a pooling (subsampling) layer between two successive convolutional layers. The main objective of this practice is to reduce progressively the spatial size of the representation. Thus, reducing the number of parameters and computations required by the network helps in the overfitting control. The pooling layer downsamples the volume spatially, independently in each depth slice of the input volume. Thus, the pool operator resizes the input along width and height, discarding activations. In practice, the *max pooling* function, which applies a window function to the input patch, and computes the maximum in that neighborhood, have been shown better results [25]. However, the pooling units can perform other functions like *L2-norm pooling* or *average pooling*.

In a fully-connected layer, neurons have full connections to all activations in the previous layer and theirs activations can be computed using a matrix multiplication followed by a bias offset. This type of layer is standard in a regular neural network. The last fully-connected layer holds the net output, such as probability distributions over classes [26], [27].

IV. Using an Existing Deep Neural Network Architecture

In order to classify images from BreaKHis dataset, we have evaluated some previously existing deep neural network architectures. We started with LeNet [28], a CNN known to work well on digit classification tasks. However, on the histopathological images assessed, LeNet classification performance were considerably inferior to our previous results reported in [11], achieving about 72% of accuracy.

Therefore, we have chosen a more complex model, specially designed to classify color images. Among a few tested, the model which presented the best performance was a variant based on the AlexNet [26]. The original AlexNet was proposed by Alex Krizhevsky to accurately classify images from CIFAR-10[1], a dataset consisting of 60,000 32×32 color images (50,000 for training, 10,000 for testing) in 10 mutually exclusive classes ('truck', 'plane, 'cat', 'dog, 'bird', etc.), with 6,000 images per class. This architecture is composed of multiple layers of convolution, pooling, Rectified Linear Unit (ReLU) nonlinearities, and local contrast normalization with a linear classifier on top of it all as shown in Figure 4.

A. CNN Architecture

In the end, the CNN architecture that provided the best results in our experiments contains the following layers and parameters:

- Input layer: this layer loads input and produces output used to feed convolutional layers. Some transformations such as mean-subtraction (used in this work and described in Section IV-B) and feature-scaling can be applied. In our case, inputs are images and the parameters are defining the image dimension (32×32 or 64×64 pixels) and the number of channels (3 for RGB).
- Convolutional layers: a convolution layer convolves the input image with a set of learnable filters, each producing one feature map in the output image. There are three convolutional layers in this model. The receptive fields (kernels) are of size 5×5, the zero-padding is set to 2 and the stride is set to 1. The first two convolutional layers learn 32 filters each one and they are initialized from a Gaussian distribution with standard deviation of 0.0001 and 0.01, respectively. The last layer learns 64 filters and it is initialized from a Gaussian distribution with standard deviation of 0.0001.
- Pooling layers: these layers are responsible for downsampling the spatial dimension of the input. There is one pooling-layer after each convolutional layer. All of them are set to use a 3×3 receptive field (spatial extent) with a stride of 2. The first pooling layer uses the most common max operation over the receptive field and the other two perform average pooling.
- ReLU layers: in spite of the ReLU activation function is actually a non-linear element-wise operator, we will treat it, for convenience, explicitly as a layer. There are three ReLU layers in this model. Given an input value x, the ReLU layer computes the neuron's output $f(x)$ as x if $x > 0$ and $(\alpha \times x)$ if $x <= 0$. The parameter α specifies whether to leak the negative part by multiplying it with the slope value (0.01 or so) rather than setting it to 0. The default value of α is 0. So, when this parameter is not set, it is equivalent to the standard ReLU function $f(x) = \max(0, x)$, in other words, the activation is simply thresholded at zero.
- Inner-product layers or fully connected layers: they treat the input as a simple vector and produce an output in the form of a single vector. There are two inner-product layers in this model. The last one, a fully-connected output layer with softmax activation, depends on the number of classes in the classification problem, i.e., 2 output filters for our binary classification problem.

Table II summarizes the parameters of the CNN layers, where CONV+POOL$_{max}$ stands for Convolutional Layer followed by Max-pooling layer, CONV+POOL$_{avg}$, Convolutional Layer followed by Average-pooling layer, and FC by fully-connected layer.

B. Training Strategies

The proposed method aims at dealing with the high-resolution of the images generally used for histopathological BC classification. As pointed out in [17], adapting the existing deep neural network models for larger images can result in more complex architectures, with larger sets of parameters

[1] http://www.cs.toronto.edu/~kriz/cifar.html

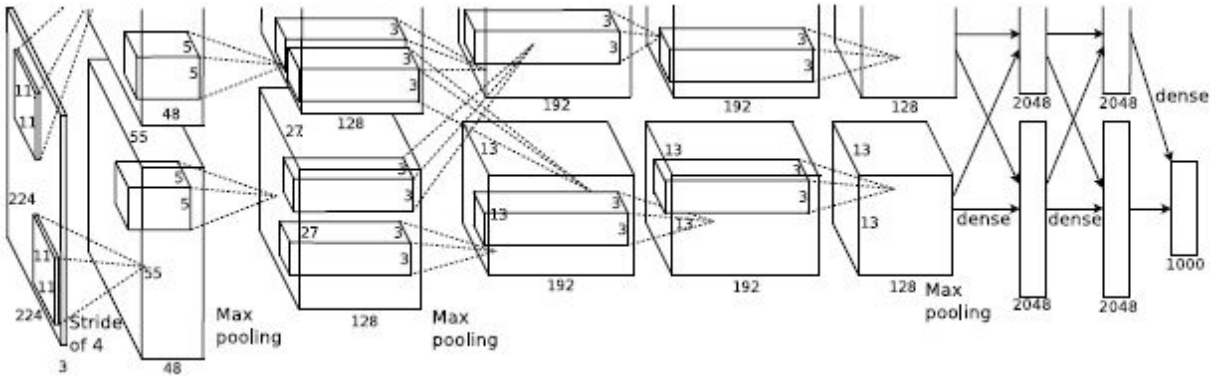


Figure 4. AlexNet CNN architecture. Extracted from [26].

Table II
SUMMARY OF THE CNN LAYERS.

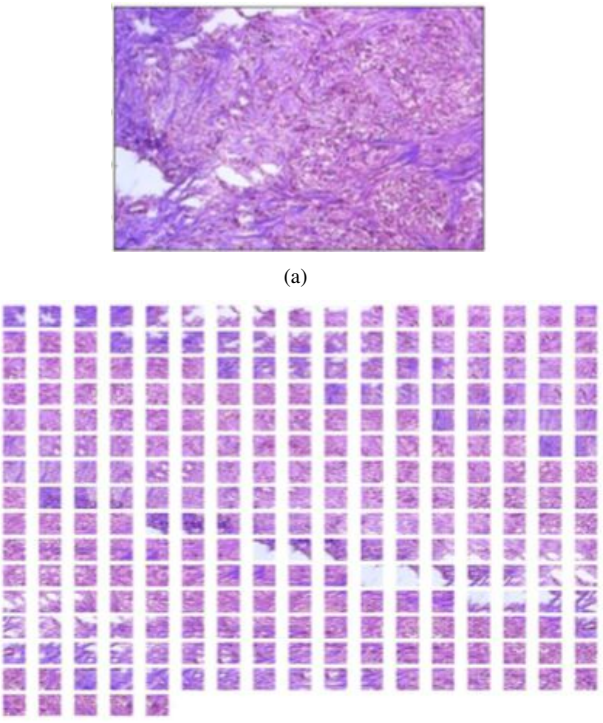
	Layers				
	1	2	3	4	5
Type	CONV+POOL$_{max}$	CONV+POOL$_{avg}$	CONV+POOL$_{avg}$	FC	FC
Channels	32	32	64	64	2
Filter Size	5×5	5×5	5×5	–	–
Convolution Stride	1×1	1×1	1×1	–	–
Pooling Size	3×3	3×3	3×3	–	–
Pooling Stride	2×2	2×2	2×2	–	–
Padding Size	2×2	2×2	2×2	–	–

(more and larger layers), which can substantially increase the complexity of the model. As a consequence, the time that is necessary to fine-tune and train the parameters of the architecture can become very high. To deal with this problem, the proposed method is based on the extraction of random patches for training, and the combination of these patches for recognition.

To learn the parameters of the CNN described in the previous section, only small patches of the images are used for training. The main idea is to extract from the high resolution images patches with sizes that are close to those of the CIFAR dataset. Since we are dealing with textures, the main premise is that these patches can contain enough information for training a model, provided an appropriate set of patches is extracted from each image.

Based on the results reported by Hafemann *et al.* in [17], where the best results were achieved by reducing the dimensionality of the images, in this work the original 700×460 images were reduced to 350×230, resampling using pixel area relation. Afterward, we extracted patches using two different strategies. In the first one, we have used a sliding window with 50% of overlapping while in the second case the patches were extracted randomly with none overlap control between patches. Also based on the results reported in [17], we have assessed two different image patch sizes (32×32 and 64×64). Figure 5 shows the resized image as well as the 32×32 image patches.

In practice, this method brings translation-invariance to



(a)

(b)

Figure 5. (a) Example of breast malignant tumor acquired at 40× magnification and (b) 32×32 patch images.

the model and acts as regularization, preventing the model from overfitting the training set. The sliding window strategy, allowing 50% of overlap between patches of 32×32 and 32×32, results in 260 and 54 patches by image, respectively. On the other hand, considering the random extraction strategy, for both patch sizes, we have fixed an arbitrary number of 1000 patches to be extracted from each input image. Table III summarizes the patch images strategies we have evaluated in our work.

Table III
SUMMARY OF PATCH IMAGE GENERATION STRATEGIES

#	Patch Size	Strategy	Number of Patches
1	32×32	Sliding Window	260
2	64×64	Sliding Window	54
3	32×32	Random	1000
4	64×64	Random	1000

Training protocol used here is the purely supervised type, frequent in practical systems for speech and image recognition. As usual in supervised mode, the Stochastic Gradient Descent (SGD) method [29], with backpropagation to compute gradients and a mini-batch size of 1, was used to update the network's parameters, starting with a learning rate of 10^{-6}, in conjunction with a momentum term of 0.9 and a weight decay of 4^{-5}. The CNN was trained for 80 000 iterations.

The model was trained using the extracted patches as input. However, the adopted architecture assumes a standard pre-processing to demean the input image (for brightness normalization), either subtracting a deterministic *mean* image or subtracting the *mean* pixel value of each channel. Thus, we compute a mean image of the all extracted patches grouping by magnification factor. Finally, we subtract this mean image from each input patch prior to feeding it to CNN.

C. Classification

For the recognition, patch results are combined for the whole image. Since the models are trained on patches of the images, we require a strategy to divide the original test images into patches, run them through the model and combine the results. The optimal result could be achieved by extracting all possible patches from the images, but this is too computationally intensive. Instead, we chose to extract the grid patches of the images, that is, the set of all non-overlapping patches, which in practice demonstrated reasonable balance between classification performance and computational cost.

Running the model, each patch outputs the probability of each possible class given the patch image. To combine the results of all the patches of a given test image, we tested three different fusion rules and the best results were achieved by the Sum rule [30]. In other words, the prediction for a given test image is the class that maximizes the sum of the probabilities on all patches of the image.

V. EXPERIMENTAL RESULTS

Following the experimental protocol proposed in [11], the BreaKHis dataset has been divided into training (70%) and testing (30%) set. To guarantee the classifier generalizes to unseen patients, the dataset was split so that patients used to build the training set are not used for the testing set. The results presented in this work are the average of five trials. This protocol was applied independently to each of the four magnifications available.

When discussing medical images, there are two ways to report the results. In the first one the decision is patient-wise, therefore, the recognition rate is computed at the patient level. Let N_P be the number of cancer images of patient P. For each patient, if N_{rec} cancer images are correctly classified, one can define a patient score as

$$\text{Patient Score} = \frac{N_{rec}}{N_P} \tag{1}$$

and the global patient recognition rate as

$$\text{Patient Recognition Rate} = \frac{\sum \text{Patient Score}}{\text{Total Number of Patients}} \tag{2}$$

In the second case, the recognition rate is computed at the image level (i.e. the patient information is not taken into account), thus providing a means to estimate solely the image classification accuracy of the CNN models. Let N_{all} be the number of cancer images of the test set. If the system classifies correctly N_{rec} cancer images, then the recognition rate at the image level is:

$$\text{Image Recognition Rate} = \frac{N_{rec}}{N_{all}} \tag{3}$$

The CNN models were trained on a NVIDIA® Tesla® K40m GPU [31] using the Caffe framework [32]. These models will be made available in the Caffe format at *http://web.inf.ufpr.br/vri/breast-cancer-database*. Training took about 40 minutes for the sliding window strategy and 3 hours for the random patch strategy, which contains a much bigger training set.

One of the advantages of using deep learning techniques is that they do not require the design of feature extractors by a domain expert, but instead let the model learn them. We can visualize the feature detectors that the model learns on the first convolutional layer, considering the weights on the learned feature maps. Figure 6 displays the 96 feature maps learned on the first convolutional layer of the CNN. We can see that the model learns filters for horizontal and vertical edges, and learns also filters that resemble Gabor filters (edge detectors) [33], [34].

Table IV reports the accuracy of the CNNs at both patient and image levels, as defined in Equations 2 and 3.

To better assess these results, Table V reproduces the best results, at the patient level, reported in [11] for the BreaKHis database. These results were achieved by different classifiers trained with Parameter-Free Threshold Adjacency Statistics

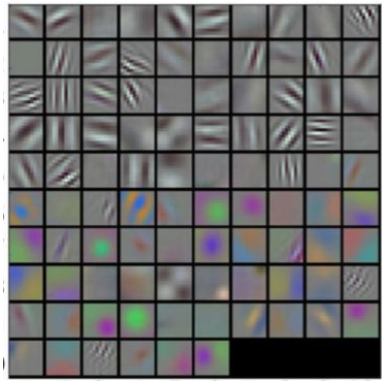


Figure 6. Feature maps learned by the first convolutional layer.

Table IV
MEAN RECOGNITION RATES AND STANDARD DEVIATIONS (PATIENT AND IMAGE LEVELS) OF THE CNN TRAINED WITH THE STRATEGIES PRESENTED IN TABLE III.

Accuracy at	Strategy	Magnification Factors			
		40×	100×	200×	400×
Patient Level	1	80.5 ± 1.6	81.0 ± 3.0	**85.3 ± 3.8**	81.0 ± 1.5
	2	81.0 ± 1.9	82.8 ± 2.8	83.7 ± 2.8	81.1 ± 3.2
	3	81.7 ± 2.9	83.5 ± 5.0	82.9 ± 3.6	81.4 ± 5.1
	4	**88.6 ± 5.6**	**84.5 ± 2.4**	83.3 ± 3.4	**81.7 ± 4.9**
Image Level	1	79.9 ± 2.6	80.8 ± 3.7	**84.0 ± 3.2**	80.7 ± 1.8
	2	80.6 ± 2.1	81.0 ± 3.0	82.7 ± 1.9	**80.8 ± 3.1**
	3	81.8 ± 3.3	82.3 ± 4.9	82.4 ± 2.8	80.3 ± 4.0
	4	**89.6 ± 6.5**	**85.0 ± 4.8**	82.8 ± 2.1	80.2 ± 3.4

(PFTAS) features [35], [36], and using the same protocol as in this study. The performance at image level is not reported in [11].

Table V
BEST RESULTS AT PATIENT LEVEL REPORTED IN [11].

Descriptor	Classifier	Magnification Factors			
		40×	100×	200×	400×
PFTAS	1-NN	80.9 ± 2.0	80.7 ± 2.4	81.5 ± 2.7	79.4 ± 3.9
	QDA	**83.8 ± 4.1**	**82.1 ± 4.9**	84.2 ± 4.1	82.0 ± 5.9
	RF	81.8 ± 2.0	81.3 ± 2.8	83.5 ± 2.3	81.0 ± 3.8
	SVM	81.6 ± 3.0	79.9 ± 5.4	**85.1 ± 3.1**	**82.3 ± 3.8**

From Table IV we may notice that training the CNN with a large number of 64×64 image patches extracted randomly from the image (strategy #4) seems a suitable strategy for low magnification factors such as $40\times$ and $100\times$. In the case of the $40\times$ magnification factor, the CNN was able to achieve an accuracy of about 5% better than the best result reported in Table V. For higher magnification factors, though, training the CNN with a large number of image patches brings no benefit. In those cases, all strategies achieve similar results, which are also comparable to the ones reported in [11].

Since each network was trained with different inputs (i.e., size and number of patches), each classifier builds its own representation, which gives us the perspective of improving such results through the combination of classifiers. As stated before, the CNNs have a final fully-connected layer with softmax activation that allows us to interpret the outputs of the networks as estimation of the posterior probabilities. Therefore, different combination rules may be applied. In this work, we report the results obtained when combining the four patch image generation strategies, using the well-known *Sum*, *Product* and *Max* rules (see [30] for details).

Table VI
COMBINATION OF CNNS USING DIFFERENT FUSION RULES (AT PATIENT AND IMAGE LEVELS)

Accuracy at	Fusion Rule	Magnification Factors			
		40×	100×	200×	400×
Patient Level	Sum	88.4 ± 7.6	88.4 ± 4.8	83.8 ± 2.8	85.3 ±5.6
	Product	89.2 ± 7.4	88.4 ± 4.8	83.8 ± 2.8	85.3 ±5.6
	Max	**90.0 ± 6.7**	**88.4 ± 4.8**	**84.6 ± 4.2**	**86.1** ±6.2
Image Level	Sum	85.4 ± 5.2	83.3 ± 4.3	**83.1 ± 1.9**	**80.8 ± 3.0**
	Product	85.5 ± 5.3	83.4 ± 4.3	83.0 ± 1.8	**80.8 ± 3.0**
	Max	**85.6 ± 4.8**	**83.5 ± 3.9**	82.7 ± 1.7	80.7 ± 2.9

Regarding the performance at image level, Table VI shows that all combination rules produce very similar results and that none of them surpass the individual results reported in Table IV. On the other hand, the combination brings interesting improvements for all magnification factors (except the $200\times$) at patient level. The most noticeable result is for the $100\times$ magnification factor where the improvement is of about 4% and 6% when compared to the best CNN and the best result reported in [11], respectively. In these cases, the *Max* rule outperforms the *Sum* and *Product* rules.

VI. CONCLUSIONS

In this paper, we have presented a set of experiments conducted on the BreaKHis dataset using a deep learning approach to avoid hand-crafted features. We have shown that we could use an existing CNN architecture, in our case AlexNet, that has been designed for classifying color images of objects, and adapt it to the classification of BC histopathological images. We have also proposed several strategies for training the CNN architecture, based on the extraction of patches obtained randomly or by a sliding window mechanism, that allow to deal with the high-resolution of these textured images without changing the CNN architecture designed for low-resolution images. Our experimental results obtained on the BreaKHis dataset showed improved accuracy obtained by CNN when compared to traditional machine learning models trained on the same dataset but with state of the art texture descriptors. Future work can explore different CNN architectures and the optimization of the hyperparameters. Also, strategies to select representative patches in order to improve the accuracy can be explored.

REFERENCES

[1] P. Boyle and B. Levin, Eds., *World Cancer Report 2008*. Lyon: IARC, 2008. [Online]. Available: http://www.iarc.fr/en/publications/pdfs-online/wcr/2008/wcr_2008.pdf

[2] S. R. Lakhani, E. I.O., S. Schnitt, P. Tan, and M. van de Vijver, *WHO classification of tumours of the breast*, 4th ed. Lyon: WHO Press, 2012.

[3] B. Stenkvist, S. Westman-Naeser, J. Holmquist, B. Nordin, E. Bengtsson, J. Vegelius, O. Eriksson, and C. H. Fox, "Computerized nuclear morphometry as an objective method for characterizing human cancer cell populations," *Cancer Research*, vol. 38, no. 12, pp. 4688–4697, 1978.

[4] M. Kowal, P. Filipczuk, A. Obuchowicz, J. Korbicz, and R. Monczak, "Computer-aided diagnosis of breast cancer based on fine needle biopsy microscopic images," *Computers in Biology and Medicine*, vol. 43, no. 10, pp. 1563–1572, 2013.

[5] P. Filipczuk, T. Fevens, A. Krzyżak, and R. Monczak, "Computer-aided breast cancer diagnosis based on the analysis of cytological images of fine needle biopsies," *IEEE Transactions on Medical Imaging*, vol. 32, no. 12, pp. 2169–2178, 2013.

[6] Y. M. George, H. L. Zayed, M. I. Roushdy, and B. M. Elbagoury, "Remote computer-aided breast cancer detection and diagnosis system based on cytological images," *IEEE Systems Journal*, vol. 8, no. 3, pp. 949–964, 2014.

[7] Y. Zhang, B. Zhang, F. Coenen, and W. Lu, "Breast cancer diagnosis from biopsy images with highly reliable random subspace classifier ensembles," *Machine Vision and Applications*, vol. 24, no. 7, pp. 1405–1420, 2013.

[8] Y. Zhang, B. Zhang, F. Coenen, J. Xiau, and W. Lu, "One-class kernel subspace ensemble for medical image classification," *EURASIP Journal on Advances in Signal Processing*, vol. 2014, no. 17, pp. 1–13, 2014.

[9] S. Doyle, S. Agner, A. Madabhushi, M. Feldman, and J. Tomaszewski, "Automated grading of breast cancer histopathology using spectral clustering with textural and architectural image features," in *Proceedings of the 5th IEEE International Symposium on Biomedical Imageging (ISBI): From Nano to Macro*, vol. 61. IEEE, May 2008, pp. 496–499.

[10] A. J. Evans, E. A. Krupinski, R. S. Weinstein, and L. Pantanowitz, "2014 american telemedicine association clinical guidelines for telepathology: Another important step in support of increased adoption of telepathology for patient care," *Journal of Pathology Informatics*, vol. 6, 2015.

[11] F. Spanhol, L. S. Oliveira, C. Petitjean, and L. Heutte, "A dataset for breast cancer histopathological image classification," *IEEE Transactions of Biomedical Engineering*, 2016.

[12] Y. Bengio, A. Courville, and P. Vincent, "Representation learning: A review and new perspectives," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, pp. 1798–1828, 2013.

[13] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, pp. 436–444, 2015.

[14] Y. LeCun, B. Boser, J. S. Denker, D. Henderson, R. E. Howard, W. Hubbard, and L. D. Jackel, "Backpropagation applied to handwritten zip code recognition," *Neural Computation*, vol. 1, no. 4, pp. 541–551, 1989.

[15] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Advances in Neural Information Processing Systems 25*, 2012, pp. 1097–1105.

[16] X. X. Niu and C. Y. Suen, "A novel hybrid cnn–svm classifier for recognizing handwritten digits," *Pattern Recognition*, vol. 45, no. 1318-1325, 2012.

[17] L. G. Hafemann, L. S. Oliveira, and P. Cavalin, "Forest species recognition using deep convolutional neural networks," in *International Conference on Pattern Recognition*, 2014, pp. 1103–1107.

[18] T. H. Vu, H. S. Mousavi, V. Monga, U. A. Rao, and G. Rao, "Dfdl: Discriminative feature-oriented dictionary learning for histopathological image classification," in *Proceedings of the IEEE 12th International Symposium on Biomedical Imaging (ISBI)*. IEEE, Apr. 2015, pp. 990–994.

[19] A. L. Mescher, *Junqueira's basic histology: text and atlas*. New York: McGraw-Hill Lange, 2013.

[20] F. Xing, Y. Xie, and L. Yang, "An automatic learning-based framework for robust nucleus segmentation," *IEEE Transactions on Biomedical Imaging*, 2015.

[21] A. Prasoon, K. Petersen, C. Igel, F. Lauze, E. Dam, and M. Nielsen, "Deep feature learning for knee cartilage segmentation using a triplanar convolutional neural network," in *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2013*, ser. Lecture Notes in Computer Science, K. Mori, I. Sakuma, Y. Sato, C. Barillot, and N. Navab, Eds. Springer Berlin Heidelberg, 2013, vol. 8150, pp. 246–253.

[22] A. Cruz-Roa, J. Arevalo Ovalle, A. Madabhushi, and F. A. González Osorio, "A deep learning architecture for image representation, visual interpretability and automated basal-cell carcinoma cancer detection," in *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2013*, ser. Lecture Notes in Computer Science, K. Mori, I. Sakuma, Y. Sato, C. Barillot, and N. Navab, Eds. Springer Berlin Heidelberg, 2013, vol. 8150, pp. 403–410.

[23] D. C. Ciresan, A. Giusti, L. M. Gambardella, and J. Schmidhuber, "Deep neural networks segment neuronal membranes in electron microscopy images," in *Proceedings of 26th Annual Conference on Neural Information Processing Systems 2012 (NIPS)*, P. L. Bartlett, F. C. N. Pereira, C. J. C. Burges, L. Bottou, and K. Q. Weinberger, Eds., Dec. 2012, pp. 2852–2860. [Online]. Available: http://papers.nips.cc/paper/4824-imagenet-classification-with-deep-convolutional-neural-networks

[24] Y. LeCun, K. Kavukcuoglu, and C. Farabet, "Convolutional networks and applications in vision," in *Proceedings of 2010 IEEE International Symposium on Circuits and Systems (ISCAS)*. Springer-Verlag, Jun. 2010, pp. 253–256.

[25] D. Scherer, A. Müller, and S. Behnke, "Evaluation of pooling operations in convolutional architectures for object recognition," in *Proceedings of the 20th International Conference on Artificial Neural Networks: Part III*, K. Diamantaras, W. D. Duch, and L. S. Iliadis, Eds. Springer-Verlag, Sep. 2010, pp. 92–101.

[26] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Proceedings of 26th Annual Conference on Neural Information Processing Systems 2012 (NIPS)*, P. L. Bartlett, F. C. N. Pereira, C. J. C. Burges, L. Bottou, and K. Q. Weinberger, Eds., Dec. 2012, pp. 1106–1114. [Online]. Available: http://papers.nips.cc/paper/4824-imagenet-classification-with-deep-convolutional-neural-networks

[27] C. Farabet, C. Couprie, L. Najman, and Y. LeCun, "Learning hierarchical features for scene labeling," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, pp. 1915–1929, 2013.

[28] Y. Lecun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proceedings of the IEEE*, vol. 86, pp. 2278–2324, 1998.

[29] L. Bottou, "Stochastic gradient tricks," in *Neural Networks, Tricks of the Trade, Reloaded*, ser. Lecture Notes in Computer Science (LNCS 7700), G. Montavon, G. B. Orr, and K.-R. Müller, Eds. Springer, 2012, pp. 430–445. [Online]. Available: http://leon.bottou.org/papers/bottou-tricks-2012

[30] J. Kittler, M. Hatef, R. P. W. Duin, and J. Matas, "On combining classifiers," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 20, no. 3, pp. 226–239, 1998.

[31] NVIDIA Corporation. (2015) Nvidia tesla product literature. [Online]. Available: http://www.nvidia.com/object/tesla_product_literature.html

[32] Y. Jia, E. Shelhamer, J. Donahue, S. Karayev, J. Long, R. Girshick, S. Guadarrama, and T. Darrell, "Caffe: Convolutional architecture for fast feature embedding," *arXiv preprint arXiv:1408.5093*, 2014.

[33] I. Fogel and D. Sagi, "Gabor filters as texture discriminator," *Biological Cybernetics*, vol. 61, pp. 103–113, 1989.

[34] C. Bovik, M. Clark, and W. S. Geisler, "Multichannel texture analysis using localized spatial filters," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 12, pp. 55–73, 1990.

[35] N. A. Hamilton, R. S. Pantelic, K. Hanson, and R. D. Teasdale, "Fast automated cell phenotype image classification," *BMC Bioinformatics*, vol. 8, 2007. [Online]. Available: http://www.biomedcentral.com/1471-2105/8/110

[36] L. P. Coelho, A. Ahmed, A. Arnold, J. Kangas, A. S.Sheikh, E. P. Xing, W. Cohen, and R. F. Murphy, "Structured literature image finder: extracting information from text and images in biomedical literature," in *Linking Literature, Information, and Knowledge for Biology*, ser. LNCS, C. Blaschke and H. Shatkay, Eds., 2010, vol. 6004, pp. 23–32.

[37] P. L. Bartlett, F. C. N. Pereira, C. J. C. Burges, L. Bottou, and K. Q. Weinberger, Eds., *Advances in Neural Information Processing Systems 25: 26th Annual Conference on Neural Information Processing Systems 2012.*, Dec. 2012. [Online]. Available: http://papers.nips.cc/book/advances-in-neural-information-processing-systems-25-2012