Classification of Breast Cancer Histology Images Through Transfer Learning Using a Pre-trained Inception Resnet V2

Carlos A. Ferreira[1(✉)], Tânia Melo[1], Patrick Sousa[1], Maria Inês Meyer[1], Elham Shakibapour[1], Pedro Costa[1], and Aurélio Campilho[1,2]

[1] INESC-TEC - Institute for Systems and Computer Engineering, Technology and Science, Porto, Portugal
{carlos.a.ferreira,tania.f.melo,patrick.sousa,maria.i.meyer, elham.shakibapour,pvcosta}@inesctec.pt
[2] Faculty of Engineering, University of Porto, Porto, Portugal
campilho@fe.up.pt

Abstract. Breast cancer is one of the leading causes of female death worldwide. The histological analysis of breast tissue allows for the differentiation of the tissue suspected to be abnormal into four classes: normal tissue, benign tumor, *in situ* carcinoma and invasive carcinoma. Automatic diagnostic systems can help in that task. In this sense, this work propose a deep neural network approach using transfer learning to classify breast cancer histology images. First, the added top layers are trained and a second fine-tunning is done on some feature extraction layers that are frozen previously. The used network is an Inception Resnet V2. In order to overcome the lack of data, data augmentation is performed too. This work is a suggested solution for the ICIAR 2018 BACH-Challenge and the accuracy is 0.76 in the blind test set.

Keywords: Breast cancer diagnosis
Breast histology images classification · Convolutional neural network
Inception resnet v2 · Transfer learning · Data augmentation
Fine-tuning

1 Introduction

Breast cancer consists in a malignant proliferation of the epithelial cells present in the ducts or lobes of the breast [1]. It is the most diagnosed cancer in women worldwide and the second leading cause of death among them. Fortunately, in the last years, the morbidity and mortality rates have decreased due to better screening procedures, which allow to do an early detection of the disease, and the improvement of treatment options [2].

Breast cancer is commonly detected by clinical examination of the breast via palpation or mammograms. Most masses detected in the screening examinations

A. Campilho et al. (Eds.): ICIAR 2018, LNCS 10882, pp. 763–770, 2018.
https://doi.org/10.1007/978-3-319-93000-8_86

are benign. Unlike the malignant tissue, the benign lesions do not grow uncontrollably and are not life-threatening [3]. When the detected mass is suspicious, a pathologist performs a histological examination of the tissue obtained from a needle biopsy. This examination allows to distinguish between normal, benign and malignant tissues based on the analysis of tissue architecture and nuclei organization [4]. However, this analysis is not trivial and different pathologists can classify the same sample in a different way, which leads to overinterpreted and underinterpreted diagnoses [5]. Furthermore, it is a time-consuming task.

In order to reduce the cost and increase the efficiency of the classification, in this work, we present a methodology for automatic classification of hematoxylin and eosin (H & E) stained breast microscopy images into four classes: normal tissue, benign abnormality, malignant *in situ* carcinoma and malignant invasive carcinoma (Fig. 1). A benign tumor is characterized by the accumulation of a significant number of cells in the ducts and lobes. Unlike the malignant tissues, the variation of cells' size and shape in this type of lesions is minimal. In turn, the classification of a malignant tissue as invasive carcinoma or *in situ* carcinoma depends on the presence of abnormal cells in other breast regions beyond the ducts and lobes. The staining enhances nuclei (purple) and cytoplasm (pinkish). The lumen appears white.

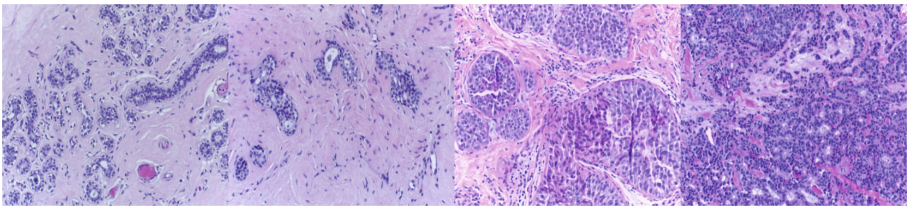

Fig. 1. Hematoxylin and eosin stained breast histology microscopy images (from left to right): normal tissue, benign lesion, *in situ* and invasive carcinoma (Color figure online)

2 Related Work

Taking into account that a benign tissue can be distinguished from a malignant tissue based on the analysis of nuclei morphology, several authors extract nuclei-based features from stained microscopy images and use them for classifying the tissues into benign and malignant (binary classification) [6–8]. Filipczuk et al. [7] use the circular Hough transform for an initial detection of the nuclei candidates and a support vector machine (SVM) to discriminate "correct candidates" from "incorrect candidates". Morphological and texture features are then extracted from the candidates that were classified as "correct" in the previous step. In [8] the authors refine the nuclei contour using a watershed and in [6] four clustering algorithms for nuclei segmentation are tested. Belsare et al. [9] segment the epithelial cells surrounding the lumen using a spatio-color-texture graph segmentation method,

extract texture features and then compare the performance of a linear discriminant analyzer classifier with k-Nearest Neighbors classifier and SVM.

Although all works mentioned above present methodologies to classify the breast histology images into benign or malignant, other authors considered the classification of these images as a multiclass problem. For instance, in [10] the author used a fusion of two Random Subspace classifier ensembles in order to classify the images into 3 classes: normal, *in situ* carcinoma and invasive carcinoma. In this approach, one of the ensembles consists of SVM classifiers and the other is based on multiple layer perceptrons. Regarding to the feature extraction step, Zhang [10] combines a local binary pattern texture analysis with Curvelet Transform.

In the last years, the deep learning models have become more and more popular. Therefore, several authors also extended the use of Convolutional Neural Networks (CNNs) for classifying breast histology images. Cruz-Roa et al. [11] presented a deep learning approach for the detection of invasive ductal carcinoma in whole slide images. In this approach, a CNN is trained over a large amount of tissue regions (patches) with size 100×100 pixels. Since the features extracted contain information about not only the nuclei but also the tissue organization, this approach outperforms the other state-of-the-art methods. Recently, Araújo et al. [4] also proposed a deep learning approach for classifying breast histology images into four classes: normal tissue, benign lesion, *in situ* carcinoma and invasive carcinoma. Similarly to [11], the architecture of the CNN used in [4] also retrieves information at different scales. In this work, a data augmentation method was also applied to increase the amount of training data. For comparative purposes, Araújo et al. [4] trained a SVM with the features detected by the CNN. The accuracy of the approach presented in [4] is 77.8% and the sensitivity for cancer cases is 95.6%.

2.1 Deep Convolutional Neural Networks

Deep convolutional neural networks (DCNN) have achieved unprecedented performance in the field of image classification and recognition. A DCNN recognizes an object by looking for low level features such as edges, lines and curves, and then building up more abstract features. The use of these methods becomes particularly interesting and useful in approaches where feature extraction is required. The input image passes through a series of filters before entering into a neural network. Instead of having fixed numbers in kernels, they are trained on the data. While the convolutional network is trained, the kernel will get better and better at filtering a given image for relevant information. Therefore, the model parameters are optimized in order to minimize the loss function.

There is often a direct relationship between the amount of data needed and the performance of the model. The larger the training dataset, the greater the diversity of data, and the better the generalization achieved. However, most of the times, the amount of data needed to build models is impossible to find, especially in the medical field. One usual solution to overcome this issue is the transfer learning. Models trained on one task capture relations in the data type

and can easily be reused for different problems in the same domain. Output reusable features of a pre-trained model in a large dataset are used to allow the learning process to be closer to the optimal solution parameters for the problem. The data augmentation is another methodology that allows to overcome the limitation of data. Often rotations, flips, zooms, shears do not change the classes of the images and allow to have a high amount of data and obtain models with a greater generalization.

Regarding the architecture, VGG networks, ResNets or Inception Networks are usually used for classification with transfer learning. The VGG networks follow the traditional layout of basic convolutional neuronal networks: a series of convolutional, max-pooling, and activation layers before some fully-connected classification layers which are in the end. On the other hand, the ResNet has the essential structure of a shortcut connection which is flexible and dependent on the tasks. Shortcut connections can skip one or more layers. Finally, an inception network uses convolution kernels of multiple sizes as well as pooling within one layer.

3 Methods

3.1 Material

This paper describes an approach to solve the part I of the challenge held of ICIAR 2018 conference: the classification of breast histology images. The dataset is composed by high-resolution (2048×1536 pixels) RGB images. All images were digitized in the same acquisition conditions, with magnification of 200x and pixel size of $0.42\,\mu m \times 0.42\,\mu m$. The training dataset contains a total of 400 images (100 per each class). Of the 100 images per class, 70, 20 and 10 were sub-selected for training, testing and validation before the release of the evaluation dataset by the challenge organizers. The division was random. The images that are part of the validation and test set are indicated in the appendix. The blind test set contains 100 images.

The processor used was Intel (R) Core TM i7-5829K CPU @ 3.30 GHz, the RAM was 32 GB and the GPU was 8 GB.

3.2 DCNN Model

Given the reduced dataset size, the use of transfer learning and data augmentation methods is indispensable. In this case a pre-trained Inception-Resnet-v2 network [12] for the ImageNet without the fully-connected layers is used. The ImageNet database is often used for pre-trained models of transfer learning as it contains a great diversity of intra and inter classes and it is publicly available. The pre-trained top layers are removed previously because they are very specific to the training occurrence. This architecture uses the tricks and decisions of an inception network with residual connection variants. This was used since it is one of the architectures that obtained better results for the ImageNet

dataset. It was expected that most of the filters used in the extraction of features were already adjusted to features suitable for this classification problem. No pre-processing is done. The images are first reshaped to 244×244 pixels, a very common input size for DCNN and rescaled to [0-1]. DCNN usually work better with square images. The reshape of the images does not greatly affect the accuracy and the shape of the cellular structures and it allows to reduce the computational cost.

The top layers consist of a global average pooling layer, a fully connected layer of 256 neurons (with activation of rectified linear unit) and finally the neurons that allow classification in each of the four classes (with softmax activation). The global average pooling and dropout of 0.5 after the fully connected layer help to reduce overfitting.

In a first phase, only the fully-connected layers are trained. In a second phase, the DCNN is retrained on top layer but also there is fine-tuning of the weights of some pretrained network layers. It is common to keep the weights of some bottom layers (due to overfitting issues) and only perform fine-tunning of high-level features. The most generic features (e.g. edges and blobs) are maintained. In this case the weights are frozen until the layer 678. Figure 2 represents a schematic of the proposed architecture.

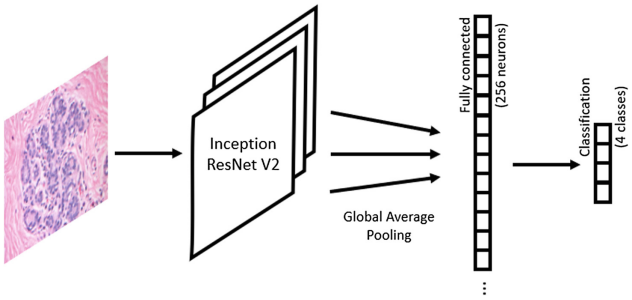

Fig. 2. Implemented CNN architecture

Data augmentation increases the amount and generality of the data. In this case it is built through the application in the training set of 18 rotations, vertical and horizontal flips, zooms of 10% and vertical and horizontal shifts of 10%. The shifts and the zoom range is not too large, otherwise the anatomical structures could be lost in the transformation.

Early stopping is used in both phases, i. e., no more epochs are ran when the validation loss does not reach a minimum over 20 consecutive epochs. The weights used for the evaluation are from the epoch in which the loss of the validation set is minimum. The loss is categorical cross-entropy. The optimizer used is the SGD. SGD tends to converge to better solutions but in return takes longer time. In the training of the top layers, a high learning rate of 0.2 is used,

while with fine-tuning of the bottom layer, a low learning rate of 0.0001 is used. The value is smaller since the solution was already close to the intended one.

4 Results

The performance of the method was evaluated based on accuracy. Besides that, a confusion matrix was computed to understand which classes are more misclassified and the log loss was considered since it takes into account the uncertainty of prediction. Table 1 presents the obtained results.

Table 1. Accuracy and loss of the datasets

	Accuracy	Loss
Training	0.99	0.23
Validation	0.93	0.23
Test	0.90	0.59
Evaluation	0.76	Not available

In the top layer training, the minimum loss of the evaluation set is reached at approximately 65 epochs while in the fine-tuning phase it happens at approximately 35 epochs. Overall, accuracy improved by 0.15 with fine-tuning.

Table 2 shows the confusion matrix for the selected test set. False negatives are spread across the different classes. However, some images have very white areas and classification generally fails for these situations. These are damaged tissues or tissues that did not attach properly to the blade preparation.

Table 2. Confusion matrix of this test set

		Real			
		Normal	Benign	*In situ*	Invasive
Predict	Normal	19	0	1	0
	Benign	0	18	1	2
	In situ	1	2	17	0
	Invasive	0	0	1	18

Throughout the training, there is a little overfitting. The existence of overfitting leads to a large decrease in accuracy for the blind test set over previously known and evaluated datasets. For future work, some strategies can be adopted in order to avoid overfitting: use the lower unfreeze layers, apply few neurons in the top layers, increase the number of images for validation and testing, promote greater similarity between datasets and/or introduce cross-validation and

regularization techniques. Figure 3 shows an example in which the proposed algorithm works well and another one in which the result does not correspond to the entire image. Both are invasive carcinoma situations. In the first one (Fig. 3 (a)), it is possible to verify a spreading of cells along all the image. In the second example (Fig. 3 (c)), lumen structures (white) with a small amount of cells are distinguishable. This has the appearance of a benign tumor and the algorithm is not robust enough to perceive the particulars of the invasive carcinoma. This conclusion can be achieved through saliency maps (the hottest colors represent the most relevant locations for classification).

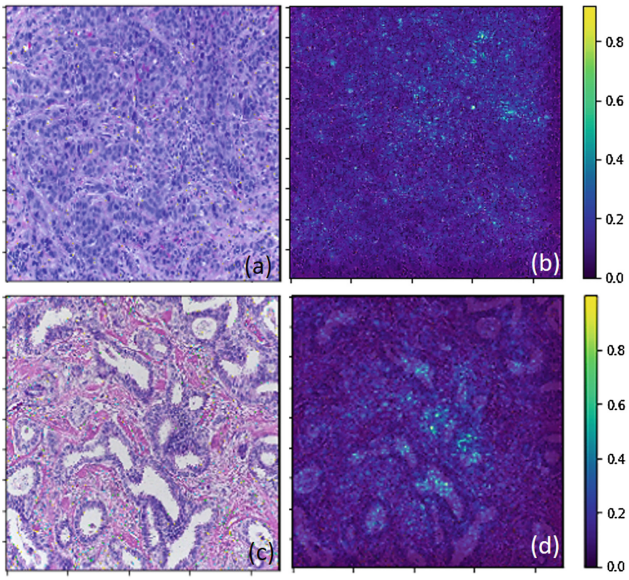

Fig. 3. Classification results for two invasive carcinomas: (a and c) original images; (b and d) saliency maps

5 Conclusion

This work proposes a Inception ResNet V2 which through transfer learning, fine-tunning and data augmentation allows to classify histological images of breast cancer stained with H & E. The network learns the relevant features for classification which becomes particularly important since it is not always easy to understand which features should be extracted using traditional methods. The classification of each individual class produce reliable results that overcomes some state-of-art works. In order to improve the results, an approach with image patches, some pre and post-processing methods will be tested. Other strategies will also be applied to reduce the overfitting.

Acknowledgments. This work is financed by the North Portugal Regional Operational Programme (NORTE 2020), under the PORTUGAL 2020 Partnership Agreement, and the European Regional Development Fund (ERDF), within the project "NanoSTIMA: Macro-to-Nano Human Sensing: Towards Integrated Multimodal Health Monitoring and Analytics/NORTE-01-0145-FEDER-000016".

A Appendix

Each class has images numbered from 1 to 100. In each class, the following images were used to construct the datasets: Test set - 2, 3, 4, 15, 20, 27, 38, 39, 42, 44, 47, 54, 60, 61, 67, 69, 75, 80, 93 and 96; Validation set - 11, 18, 21, 23, 45, 49, 55, 87, 89 and 99; Training set - remaining images.

References

1. Kasper, D., Fauci, A., Hauser, S., Longo, D., Jameson, J.: Harrison's Principles of Internal Medicine. McGraw-Hill Education, New York (2015)
2. National Breast Cancer Foundation Inc: Breast Cancer Facts. http://www.nationalbreastcancer.org/breast-cancer-facts
3. American Cancer Society: Breast Cancer Facts&Figures 2017–2018. American Cancer Society, Inc., Atlanta (2017)
4. Araújo, T., Aresta, G., Castro, E., Rouco, J., Aguiar, P., Eloy, C., Polónia, A., Campilho, A.: Classification of breast cancer histology images using Convolutional Neural Networks. PLOS ONE **12**(6), 1–14 (2017)
5. Elmore, J., Longton, G., Carney, P., Geller, B., Onega, T., Tosteson, A., Nelson, H., Pepe, M., Allison, K., Schnitt, S., O'Malley, F., Weaver, D.: Diagnostic concordance among pathologists interpreting breast biopsy specimens. JAMA **313**(11), 1122–1132 (2015)
6. Kowal, M., Filipczuk, P., Obuchowicz, A., Korbicz, J., Monczak, R.: Computer-aided diagnosis of breast cancer based on fine needle biopsy microscopic images. Comput. Biol. Med. **43**(10), 1563–1572 (2013)
7. Filipczuk, P., Fevens, T., Krzyzak, A., Monczak, R.: Computer-aided breast cancer diagnosis based on the analysis of cytological images of fine needle biopsies. IEEE Trans. Med. Imaging **32**(12), 2169–2178 (2013)
8. George, Y., Zayed, H., Roushdy, M., Elbagoury, B.: Remote computer-aided breast cancer detection and diagnosis system based on cytological images. IEEE Syst. J. **8**(3), 949–2178 (2013)
9. Belsare, A., Mushrif, M., Pangarkar, M., Meshram, M.: Classification of breast cancer histopathology images using texture feature analysis. In: TENCON 2015–2015 IEEE Region 10 Conference, Macau (2015)
10. Zhang, B.: Breast cancer diagnosis from biopsy images by serial fusion of Random Subspace ensembles. In: 4th International Conference on Biomedical Engineering and Informatics, Shanghai (2011)
11. Cruz-Roa, A., Basavanhally, A., González, F., Gilmore, H., Feldman, M., Ganesan, S., Shih, N., Tomaszewsk, J., Madabhushi, A.: Automatic detection of invasive ductal carcinoma in whole slide images with convolutional neural networks. In: Proceedings of the Medical Imaging 2014: Digital Pathology, San Diego, vol. 9041 (2014)
12. Szegedy, C., Ioffe, S., Vanhoucke, V.: Inception-v4, inception-resnet and the impact of residual connections on learning (2016). arXiv:1602.07261