Efficient Breast Cancer Classification Using Histopathological Images and a Simpler VGG

Marcelo Luis Rodrigues Filho, Omar Andres Carmona Cortes

[1]Departamento de Computação – Instituto Federal do Maranhão (IFMA)
Av. Getulio Vargas, 04 – São Luis - MA - Brazil

marceloluis@acad.ifma.edu.br, omar@ifma.edu.br

Abstract. *Breast cancer is the second most deadly disease worldwide. This severe condition led to 627,000 people dying in 2018. Thus, early detection is critical for improving the patients' lifetime or even cure them. In this context, we can appeal to Medicine 4.0 that exploits the machine learning capabilities to obtain a faster and more efficient diagnosis. Therefore, this work aims to apply a simpler convolutional neural network, called VGG-7, for classifying breast cancer in histopathological images. Results have shown that VGG-7 overcomes the performance of VGG-16 and VGG-19, showing an accuracy of 98%, a precision of 99%, a recall of 98%, and an F1 score of 98%.*

1. Introduction

Breast cancer is a severe disease that attacks women primarily. However, about 1% of men are also affected by it [INCA 2016]. In fact, breast cancer is the most common cancer among women and the second one in general [AICR 2020]. The World Health Organization estimates that this kind of cancer impacts 2.1 million women per year with 627,000 deaths, representing about 15% of all deaths caused by cancer [WHO 2020]. Thus, it is critical to perform the early diagnoses to start the treatment as fast as possible, increasing the lifetime expectation and maybe getting the patient's cure. Thus, Medicine 4.0 [Wolf and Scholze 2017] arises proposing using technology such as computer vision and artificial intelligence to help physicians in this task.

In this work, we are engaged in evaluating how convolutional neural networks (CNNs) classify cancer in histopathological images. During the assessment of images, the pathologist looks for signs of malignancies to determine if a tumor is growing as a malignant one [Titoriya and Sachdeva 2019]. On the one hand, the pathologist is the expert who can confirm the diagnose. On the other hand, the expert is human; therefore, he is subject to physical and visual distresses. In this context, computational tools become essential for a precise and fast diagnosis.

The main problem with using CNNs is that these architectures usually involve many layers, demanding considerable time to train them, especially when using large datasets such as ImageNet. Thus, this work proposes a simpler CNN architecture that can perform efficiently in classifying histopathological images. We used the VGG-16 as a baseline because it has presented good results, then tested a smaller configuration named VGG-7 with no transfer learning, *i.e.*, we trained our VGG from scratch using only the BreakHis [Spanhol et al. 2016b] dataset. In this context, the work is divided as follows: Section 2 shows some related works; Section 3 details the VGG architectures and our proposal (VGG-7); Section 4 describes the configuration, the dataset, and the results

of our computational experiments; finally, Section 5 presents the conclusions and future work.

2. Related Works

Convolutional Neural Networks have played an important role in image segmentation and classification, especially in medical and biomedical applications. Particularly, the use of CNNs in the field of medical research has been demonstrated in several works such as [Gulshan et al. 2016], [Bej], [Ismail and Sovuthy 2019], [Singh et al. 2020], and many more. In the field of breast cancer detection, [Spanhol et al. 2016b] introduces a public dataset [1] composed of 7909 images of 82 patients. This dataset has been the main dataset in testing machine learning algorithms, including CNN architectures presented by the same author in [Spanhol et al. 2016a].

Regarding the VGGs, several works deal with the architectures' classification ability or try to improve it, especially in biomedical applications. For instance, [Shallu and Mehra 2018] compares VGG-16 and VGG-19 against ResNet50 with and without transfer learning, showing that VGGs are more efficient than the ResNet. In [Saikia et al. 2019], a study on the performance of VGG16, VGG19, ResNet-50, and GoogLeNet-V3 is carried out in fine-needle aspiration cytology images, in which GoogLeNet-V3 reached the best results after a fine-tuning. Further, a novel attention-based deep learning model using VGG-16 is proposed by [Sitaula and Hossain 2020] to improve COVID-19 classification using x-ray images getting the best results. Furthermore, a modified VGG, called MVGG, is proposed and implemented in [Khamparia et al. 2021] to increase the detection ability on 2D and 3D mammogram image datasets.

As we can see, the field of using and studying deep learning models in the classification of biomedical images, especially the VGG architecture, is vast. Thus, in this paper, we propose a simpler VGG architecture, called VGG-7, to provide a faster training deep learning model that also overcomes the efficiency of the classical VGGs.

3. The VGG Architecture

A CNN is a deep neural network that depends on the correlation of neighboring pixels [Dabeer et al. 2019]. Thus, being devised to work with images made it achieved significant success in image classification problems. In other words, working with images is the nature of CNNs that makes them efficient in image classification.

The CNN architecture uses three kinds of layers. The first one is convolution, in which a filter looks for particular features. That is why this layer is also called a feature map. The second one is the pooling layer that re-samples the image, *i.e.*, reduces its dimensionality, consequently tending to reduce the overfitting. After using a combination of convolutional and pooling layers, the output can be passed to a fully connected layer for classification.

The VGG [Simonyan and Zisserman 2015], a popular CNN network, was introduced in 2015. VGG-16 and VGG-19 are the most famous ones and are commonly used for image detection. The number in front of the name stands for the number of weight

[1]The dataset is available in: `http://web.inf.ufpr.br/vri/breast-cancer-database`

layers in the network. In this work, we propose a simpler version of VGG called VGG-7 that comprises the layers presented in Figure 1.

Figure 1. VGG-7 Architecture

The VGG-7 model consists of four convolution layers divided into two blocks, followed by max-pooling layers. Max-pooling can divide the images into several blocks of the same size and only take each block's higher value. Also, the global contextual information with embedded channel-wise statistics was gathered with a global average pooling layer. The three last layers are Fully-Connected (FC) ones: the first two ones have 128 and 64 units, respectively, and the third one performs binary classification with a sigmoid function.

We use 3x3 zero-padding convolutions layers with stride 1, each followed by a rectified linear unit (ReLU) and 2×2 max pooling operation with stride equals 2 for feature extraction. Either, we double the number of feature channels on each pooling operation step. Then, during the training process, the input to VGG-7 is fixed-size with a 150×150 RGB image. The input images and their corresponding labels are used to train the network with Adam optimizer [Kingma and Ba 2014] and Binary Cross-Entropy loss function. All the kernels are initialized with a random uniform distribution procedure of Xavier[Glorot and Bengio 2010]. We use class weighting to create a model where that loss values for classes 'benign' and 'malign' will be multiplied by their corresponding weight values to avoid unbalanced classes in the dataset.

Furthermore, the width of convolutions layers (the number of channels) started from 32 in the first layer and then increased by a 2-factor after each max-pooling layer until it reaches 64. The units of the last layers are relatively small to reduce computational cost. The first Fully-Connected (FC) layer has 128 units that correspond to double the filter from the previous max-pooling layer, and the second layer has 128/2 units. The final layer is the Sigmoid layer responsible for the binary classification.

4. Computer Experiments

The application was implemented in Python 3.0 using recent versions of TensorFlow[Abadi et al. 2015] and Scikit-learn [Pedregosa et al. 2011] library. The code and the training step have been done in Google Colaboratory [Google 2020], which was essential to this work because we can use GPU computing for training the CNN. The virtual machine is a two CPUs Intel® Xeon 2.30 GHz, 14 GB of RAM, and 37.11 GB of HD. Even though we used GPUs, the training step takes about 1 hour for each network configuration.

The learning rate is Adam's default, 0.001, the momentum equals 0.9, and gamma equals 0.1. We train the neural network by slicing the data into batches of size 32 and

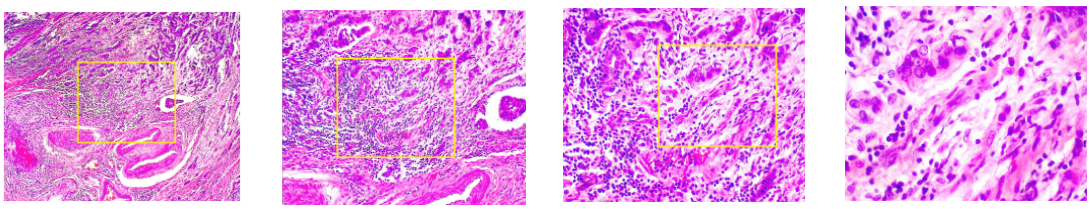
repeatedly iterating over the entire dataset for 50 epochs. Moreover, we tested two types of sampling: hold-out (80/20 and 90/10) and k-fold ($k = 10$) cross-validation.

4.1. Dataset

The dataset comes from Breast Cancer Histopathological Dataset [Spanhol et al. 2016b], a public domain dataset made available by the Laboratory of Vision, Robotics, and Images from Universidade Federal do Paraná (UFPR). The dataset comprises 7909 images of tumoral tissues from 82 different patients. The images have different zoom magnitudes devised by 40x, 100x, 200x, and 400x as presented in Fig. 2, which shows a malign tumor. Either, images are in the "png" format, having a resolution of $700 \times 460 pixels$, three RGB channels, and 8 bits depth in each one. Table 1 presents the number of benignant and malignant tumors according to the magnitude.

Table 1. Dataset Structure

Magnitude	Benign	Malign	Sub-Total
40X	652	1370	1995
100X	644	1437	2081
200X	623	1390	2013
400X	588	1232	1820
Total	2480	5429	**7909**



(a) 40X (b) 100X (c) 200X (d) 400X

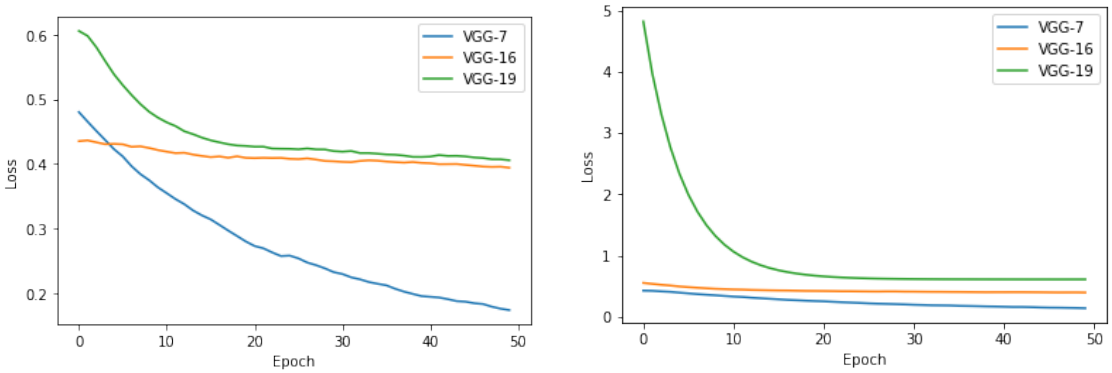
Figure 2. Malignant cancer in different magnitudes [Spanhol et al. 2016b]

4.2. Data augmentation

Data augmentation is essential to teach the network the desired invariance and robustness properties when only a few training samples are available. We perform standard in-place data augmentation techniques such as random rotation, random horizontal and vertical flip, width and height shift range. We also pre-processed the inputs in the same way as the original VGG [Simonyan and Zisserman 2015], in which we subtracted the mean RGB value computed on the training set from each pixel. The complete set of operations are: randomly flip horizontally, randomly flip vertically, random rotations of 50 degrees, width and height shift.

4.3. Results

The first result regards the losses in the training stage using only hold-out sampling. Figure 3 shows the loss as epochs increase in 80/20 (a) and 90/10 (b). As we can see, the error decreases faster and more than the other ones in VGG-7, reaching a value closer to zero.

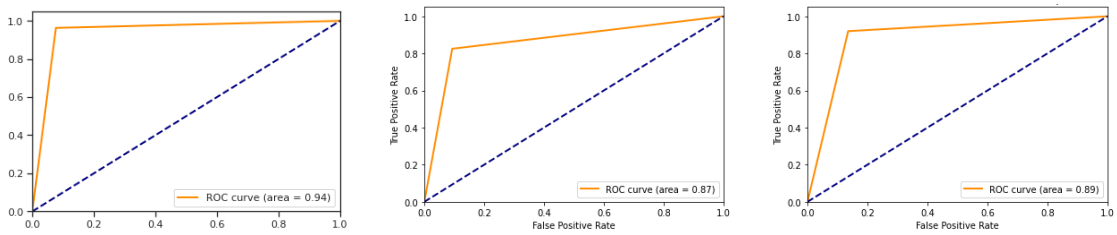


(a) Hold out 80/20 (b) Hold out 90/10

Figure 3. Training loss as the epoch count increases

Furthermore, our proposal presents a significant advantage over the other VGGs because our approach needs to search for much lesser parameters than the other ones. The total number of parameters is 82,209, 15,765,313, and 21,075,009 for VGG-7, VGG-16, and VGG-19. Consequently, the training stage of VGG-7 is faster than the other ones. Additionally, our proposal requires much less memory in both the training stage and deployment or embedment.

Figure 4 shows the ROC curve of the three CNN architectures with the areas of 0.94, 0.87, and 0.89 for VGG-7, VGG-19, and VGG-16, respectively in 80/20 hold-out. Then, considering that the ROC curve presents the probability of confirming the illness's presence, the figure validates the efficiency of our proposal.



(a) VGG-7 (b) VGG-19 (c) VGG-16

Figure 4. ROC Curves

Concerning the efficiency of hold-out sampling, Table 2 shows the confusion matrix for the three VGG architectures, in which we can see that our proposal, the VGG-7, classified 1540 images correctly, whereas VGG-19 and VGG-16 classified 1280 and 1290 correctly, respectively. All in all, the VGG-7 correctly identifies 260 and 250 more images than VGG-19 and VGG-16, respectively. This means that at least 30 more patients will undergo treatment.

In order to determine the efficiency of the CNN networks, Tables 3 presents all metrics obtained by the experiment using the hold-out sampling using the rate of 80/20 and 90/10, respectively. As we can see, the VGG-7 overcame the classical VGGs in all metrics, meaning that VGG-7 is more efficient than the other networks with fewer

Table 2. Confusion Matrix - VGG-7, VGG-19, and VGG-16 - Hold-Out: 80/20

	VGG-7		VGG-19		VGG-16	
	Malignant	Benignant	Malignant	Benignant	Malignant	Benignant
Malignant	**1100**	32	**1000**	44	**1000**	39
Benignant	60	**440**	220	**280**	200	**290**

misclassifications. As expected, the generalization error tends to decrease as we increase the training set size in all architectures. Nonetheless, VGG-7 overcomes the other ones in all metric.

Table 3. Metrics - Hold-Out: 80/20 and 90/10

	80/20			90/10		
Metric	VGG-7	VGG-19	VGG-16	VGG-7	VGG-19	VGG-16
Accuracy	**0.94**	0.83	0.85	**0.95**	0.69	0.86
Precision	**0.95**	0.83	0.84	**0.96**	0.69	0.86
Recall	**0.97**	0.96	0.96	**0.96**	1.00	0.96
F1 Score	**0.96**	0.89	0.90	**0.96**	0.81	0.91

Concerning k-fold cross-validation sampling, Table 4 shows the results of the stratified cross-validation, *i.e.*, the stratified re-sampling guarantees that the same distribution is used in training and test sets. As we can see, the VGG-7 outperforms the other VGG version in all metrics.

Table 4. Metrics - K-fold (k=10) cross-validation

	VGG-7	VGG-16	VGG-19
Average folder Accuracy	$\mathbf{0.98 \pm 0.02}$	0.97 ± 0.05	0.96 ± 0.04
Average folder F1 score	$\mathbf{0.98 \pm 0.02}$	0.98 ± 0.03	0.97 ± 0.02
Average folder Precision	$\mathbf{0.99 \pm 0.01}$	0.98 ± 0.02	0.97 ± 0.03
Average folder Recall	$\mathbf{0.98 \pm 0.03}$	0.98 ± 0.05	0.98 ± 0.02

5. Conclusions

This paper proposed a simpler VGG neural network called VGG-7 for classifying breast cancer in histopathological images. We trained the VGG-7 from scratch with no transfer learning. Our proposal was compared against the classical VGGs (VGG16 and VGG19) using transfer learning.

We also tested the classical VGGs with no transfer learning, achieving a lower performance than using it; thus, we omitted those results in this work. Furthermore, our proposal achieved 95% of accuracy, 96% of precision, 96% of recall, and 96% of F1 Score in 90/10 hold-out sampling. Moreover, the VGG-7 reached 98% of accuracy, 98% of precision, 99% of recall, and 98% in the k-fold cross-validation test, overcoming the other architectures as well.

Future work includes: (i) compare the VGG-7 against other CNNs architectures; (ii) test the VGG-7 using other fully connected networks; (iii) add different classifiers, such as SVM replacing the VGG-7 fully connected layer; and (iv) use VGG-7 with other

algorithms creating ensemble classifiers in order to improve the efficiency of the classification.

References

Abadi, M., Agarwal, A., Barham, P., Brevdo, E., Chen, Z., Citro, C., Corrado, G. S., Davis, A., Dean, J., Devin, M., Ghemawat, S., Goodfellow, I., Harp, A., Irving, G., Isard, M., Jia, Y., Jozefowicz, R., Kaiser, L., Kudlur, M., Levenberg, J., Mané, D., Monga, R., Moore, S., Murray, D., Olah, C., Schuster, M., Shlens, J., Steiner, B., Sutskever, I., Talwar, K., Tucker, P., Vanhoucke, V., Vasudevan, V., Viégas, F., Vinyals, O., Warden, P., Wattenberg, M., Wicke, M., Yu, Y., and Zheng, X. (2015). TensorFlow: Large-scale machine learning on heterogeneous systems. Software available from tensorflow.org.

AICR (2020). *American Institute for Cancer Research, Breast cancer statistics*. https://www.wcrf.org/dietandcancer/cancer-trends/breast-cancer-statistics, Visit on 05-31-2020.

Dabeer, S., Khan, M. M., and Islam, S. (2019). Cancer diagnosis in histopathological image: CNN based approach. *Informatics in Medicine Unlocked*, 16:100231.

Glorot, X. and Bengio, Y. (2010). Understanding the difficulty of training deep feedforward neural networks. *Journal of Machine Learning Research - Proceedings Track*, 9:249–256.

Google (2020). *Google Colaboratory*. https://colab.research.google.com/notebooks/intro.ipynb, Visit on 06-15-2020.

Gulshan, V., Peng, L., Coram, M., Stumpe, M. C., Wu, D., Narayanaswamy, A., Venugopalan, S., Widner, K., Madams, T., Cuadros, J., Kim, R., Raman, R., Nelson, P. C., Mega, J. L., and Webster, D. R. (2016). Development and Validation of a Deep Learning Algorithm for Detection of Diabetic Retinopathy in Retinal Fundus Photographs. *JAMA*, 316(22):2402–2410.

INCA (2016). *Ministério da Saúde - Instituto Nacional de Câncer, Câncer de mama: vamos falar sobre isso?* https://www.inca.gov.br/sites/ufu.sti.inca.local/files//media/document//cartilha-cancer-de-mama-vamos-falar-sobre -isso2016.pdf, Visit on 05-31-2020.

Ismail, N. S. and Sovuthy, C. (2019). Breast cancer detection based on deep learning technique. In *2019 International UNIMAS STEM 12th Engineering Conference (EnCon)*, pages 89–92.

Khamparia, A., Bharati, A., Podder, P., Gupta, D., Khanna, A., K., P. T., and Thanh, D. N. H. (2021). Diagnosis of breast cancer based on modern mammography using hybrid transfer learning. *Multidimensional Systems and Signal Processing*, 32:747–765.

Kingma, D. and Ba, J. (2014). Adam: A method for stochastic optimization. *International Conference on Learning Representations*.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau,

D., Brucher, M., Perrot, M., and Duchesnay, E. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830.

Saikia, A. R., Bora, K., Mahanta, L. B., and Kumar Das, A. (2019). Comparative assessment of CNN architectures for classification of breast fnac images. *Tissue and Cell*, 57:8–14. EM in cell and tissues.

Shallu and Mehra, R. (2018). Breast cancer histology images classification: Training from scratch or transfer learning? *ICT Express*, 4(4):247–254.

Simonyan, K. and Zisserman, A. (2015). Very deep convolutional networks for large-scale image recognition. In *International Conference on Learning Representations*.

Singh, R., Ahmed, T., Kumar, A., Singh, A. K., Pandey, A. K., and Singh, S. K. (2020). Imbalanced breast cancer classification using transfer learning. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, pages 83–93.

Sitaula, C. and Hossain, M. (2020). Attention-based VGG-16 model for covid-19 chest x-ray image classification. *Applied Intelligence*, 51:2850–2863.

Spanhol, F. A., Oliveira, L. S., Petitjean, C., and Heutte, L. (2016a). Breast cancer histopathological image classification using convolutional neural networks. In *2016 International Joint Conference on Neural Networks (IJCNN)*, pages 2560–2567.

Spanhol, F. A., Oliveira, L. S., Petitjean, C., and Heutte, L. (2016b). A dataset for breast cancer histopathological image classification. *IEEE Transactions on Biomedical Engineering*, 63(7):1455–1462.

Titoriya, A. and Sachdeva, S. (2019). Breast cancer histopathology image classification using AlexNet. In *4th International Conference on Information Systems and Computer Networks (ISCON)*, pages 708–712.

WHO (2020). *World Health Organization, Breast cancer*. https://www.who.int/cancer/prevention/diagnosis-screening/breast-cancer/en/, Visit on 05-31-2020.

Wolf, B. and Scholze, C. (2017). "medicine 4.0". *Current Directions in Biomedical Engineering*, 3(2):183–186.