

Breast Nodule Detection and Classification Using Convolutional Neural Network with Transfer Learning

Flávia Telles[1], Luiza de Macedo Mourelle[2(✉)], and Nadia Nedjah[3]

[1] Postgraduate Program in Electronics Engineering, State University of Rio de Janeiro, Rio de Janeiro, Brazil
flaviaesanto@gmail.com

[2] Postgraduate Program in Electronics Engineering, Department of Systems Engineering and Computation, State University of Rio de Janeiro, Rio de Janeiro, Brazil
ldmm@eng.uerj.br

[3] Postgraduate Program in Electronics Engineering, Department of Electronics Engineering, State University of Rio de Janeiro, Rio de Janeiro, Brazil
nadia@eng.uerj.br

Abstract. Diagnosing breast cancer using mammograms is not always easy, since it depends on human interpretation and the images can be complex. In this sense, early detection is still a challenge. In this paper, we propose and evaluate two methods based on Convolutional Neural Networks. We use U-Net and VGG16 architectures to analyze images from the CBIS-DDSM database. We compare semantic segmentation with direct classification using Transfer Learning. The goal is to check whether it is really necessary to identify the exact lesion area for diagnosis. The results show that the classification approach performed better than segmentation. The U-Net model had convergence problems with this dataset. The VGG16 model achieved the best results after applying regularization techniques, reaching an accuracy of 65.3% and a sensitivity of 66.0% in detecting malignant cases.

Keywords: Breast cancer · Deep learning · Transfer learning · Convolutional neural networks · Computer-aided diagnosis

1 Introduction

Cancer is a serious disease and a big health problem around the world. It causes many deaths and affects many families. Breast cancer is one of the most common types. It significantly affects the longevity and quality of life of women worldwide [1]. One problem is that it can grow without clear symptoms. Often, the diagnosis happens only when the disease has already reached advanced stages. This reduces treatment options and lowers survival rates. So, early detection is very important. We also need better ways to detect it earlier [2].

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2027
O. Gervasi et al. (Eds.): ICCSA 2026 Workshops, LNCS 16761, pp. 22–35, 2027.
https://doi.org/10.1007/978-3-032-30527-5_2

Cancer is one of the most complex diseases today and a leading cause of early death worldwide [1]. In Brazil, as shown in Fig. 1, breast cancer has the highest mortality rates among women [5]. This highlights the need for more effective diagnosis methods [3]. The chance of survival is directly related to the detection stage, showing a 99% survival rate for early cases, compared to only 28% for advanced cases with distant metastasis [4].

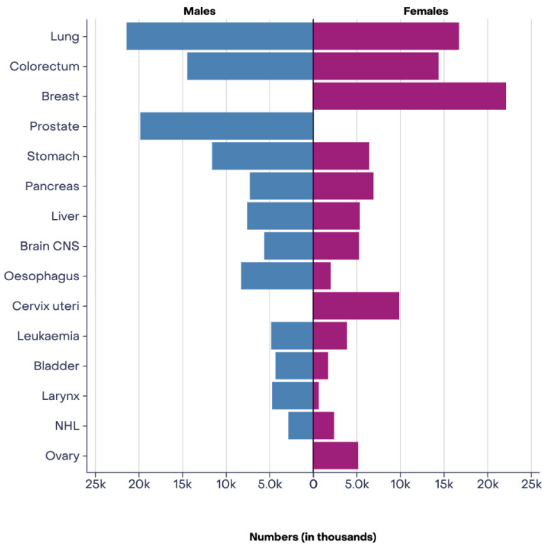

Fig. 1. Mortality according to primary tumor location and gender in Brazil in 2022 [5].

The first exams for cancer nodule detection are self-examination and physical exam by a doctor. Both are simple methods and non invasive, but their sensitivity and specificity are limited, which could not detect small tumors or confuse benign and malignant alterations. Mammogram is an x-ray exam, usually indicated as the first one, but it can be uncomfortable or has some limitations in dense breasts, where the glandular tissue can obscure lesions. Ultrasound is a complementary exam to mammogram, useful to differentiate solid nodules from cysts and evaluate lesions in dense breasts. Magnetic resonance (MR) is a more sensitive image exam, but less specific, of high cost, indicated for women with high risk of breast cancer, family history or genetic predisposition. Biopsy is the only exam that provides a definitive diagnosis for breast cancer, but it is very invasive.

Today, doctors usually analyze mammograms visually [15]. Although this is the standard method, it depends entirely on human visual accuracy. It is also subject to variables such as fatigue and professional experience, which can contribute to incorrect diagnosis [16]. Because of these problems, computer systems

can help doctors, acting as a second opinion. It can process large amounts of data and identify subtle patterns that the human eye cannot easily see [17].

To help clinical practice, computer-aided diagnosis systems use Artificial Neural Networks (ANNs), which are models inspired by the human brain [28]. The evolution of these architectures led to Convolutional Neural Networks (CNNs), which were designed to process grid-like data, such as mammogram images. They use convolution layers to detect patterns in images and pooling layers, reducing data dimensions and ensuring resistance to small movements [30]. In the context of breast diagnosis, this helps the model detect simple and complex patterns, finding everything from simple edges to complex morphological patterns of masses and calcifications.

In this work, we explore the potential of two CNN architectures to improve detection and diagnosis of breast cancer: U-Net and VGG16. The main objective is to investigate and compare the effectiveness of semantic segmentation versus direct image classification. This comparison helps us understand which method works best in practice.

This paper is organized as follows. Section 2 reviews related works and places this research in the current state of the art by analyzing similar approaches in the scientific literature. Section 3 introduces the theoretical background in artificial neural networks in the context of this work and the architectures applied. Section 4 shows the results obtained, the evaluation metrics, and the comparative performance analysis of the trained models. Finally, Sect. 5 summarizes the study findings, points out conclusions and directions for future work.

2 Related Works

Academic research on breast cancer diagnosis has evolved by integrating standardized databases and deep neural network architectures. Pinheiro [14] proposed a method using U-Net [33] and optimization algorithms. It approaches nodule detection through an innovative combination of the U-Net architecture and several swarm intelligence methodologies. By exploring a wide range of bio-inspired algorithms to optimize the segmentation process, the author tested different methods and compared the results. These consolidated results can be seen in detail in Table 1. Among the conducted experiments, particle swarm optimization achieved the best performance. It reached an average accuracy of 93.78% and a recall of 93.04%, showing that U-Net performs quite well for this task.

Regarding the classification of mammogram abnormalities, Falconi and his colleagues [13] conducted an extensive research. They specifically used the CBIS-DDSM database [32] to evaluate twenty pre-trained neural network models. The results showed that the VGG16 [34] architecture achieved an accuracy of 84.4% when subjected to fine-tuning. This consolidated this model as a robust feature extractor for diagnosing tumor masses. This evidence supports the choice of VGG16 in this research as a fundamental paradigm for the computer-aided diagnosis task.

In addition, Baffa *et al* [20] focused on ROI (Region of Interest) segmentation in mammary thermography, using the public database DMR-UFF [6] composed

Table 1. Best-performing comparison [14]

Algorithm	Accuracy (ReLU 2×2)	Recall (ReLU 2×2)
Particle Swarm Optimization (PSO)	93.78%	93.04%
Gravitational Search Algorithm (GSA)	93.52%	92.60%
Harmony Search (HS)	93.17%	92.39%
Artificial Bee Colony (ABC)	92.95%	91.95%
Gray Wolf Optimization (GWO)	92.80%	91.83%
Social Spider Algorithm (SSA)	92.62%	92.03%
Bacterial Foraging Optimization (BFO)	92.57%	90.49%

of 283 infrared images. The segmentation methodology used a variant of the U-Net aiming at segmenting the breast external region, defined from inflammatory pleats to the neck. The U-Net received input images in grey scale of 256×256 pixels and used 12 convolutional layers with ReLU activation function. The output layer was composed of a 1×1 convolution with Sigmoid as activation function. The proposed U-Net provided a better performance than other methods based on VGG, achieving a mean accuracy of 98.24% and precision of 98.86%. This last metric indicates the reliability in positive detections or how rarely it indicates a false tumor. Besides this, it obtained an Intersection over Union (IoU) of 94.38%, demonstrating high efficiency in precise segmentation of ROI. The research emphasized the importance of image pre-processing to improve accuracy. It demonstrated that deep models outperform traditional computer vision methods in distinguishing between benign and malignant cases.

Furthermore, the study by Mendez and Muñoz [23] explored the identification of the Region of Interest in malignant mammograms. They compared the U-Net and YOLO (You Only Look Once) [7] architectures. The research used two data sets from The Cancer Imaging Archive (TCIA) [8]. The first was CBIS-DDSM, with 1318 mammography images and the second was Brain MRI Segmentation, based on TCGA-LGG [9], with 2090 magnetic resonance images. The work implemented the YOLO architecture and the U-Net, specialized in semantic segmentation. The learning rate, that provided the best results, were 2×10^{-4} for YOLO and 0.0001 for U-Net, both trained with ADAM optimizer [10]. The authors concluded that U-Net showed superior performance in spatial localization accuracy, reaching an Intersection over Union (IoU) index of 0.67 for malignant mammograms.

Finally, the work by Silva [24] used the transfer learning technique for attribute extraction, based CNNs pre-trained in ImageNet [18], such as LeNet [19], ResNet50 [12], Inception V3 [28], SqueezeNet as feature extractors. Several traditional classifiers were tested, including Naive Bayes [21] Random Forest [22], Suppport Vector Machine (SVM) [25]. The best combination without attribute selection was from Inception V3 plus SVM polynomial (degree 3), achieving an accuracy of 79.92%.

The analysis of these five works supports the strategy adopted in this research. It validates the use of the CBIS-DDSM database and the focus on the potential of U-Net and VGG16 networks to overcome the limitations of conventional visual diagnosis.

3 Neural Network-Assisted Diagnosis for Cancer

Several architectures for CNNs were used with success in classification of cancer medical images, such as AlexNet [11], VGG Net [34], GoogleNet/Inception [28], ResNet. Among these, the VGG Net family, in particular VGG16, provides an efficient feature extractor for Transfer Learning technique. It stands out for using sequential convolution blocks with small filters of 3×3 pixels and a stride of one pixel. This setup allows us to increase the network depth while keeping a controlled number of parameters. It also makes learning discriminative features easier through the rectified linear activation function [31]. The implementation uses Transfer Learning with weights previously adjusted on the ImageNet dataset. This technique allows the model to reach convergence more efficiently when dealing with the mammogram images database [35]. Figure 2 displays the architecture of the VGG16 network used in this research.

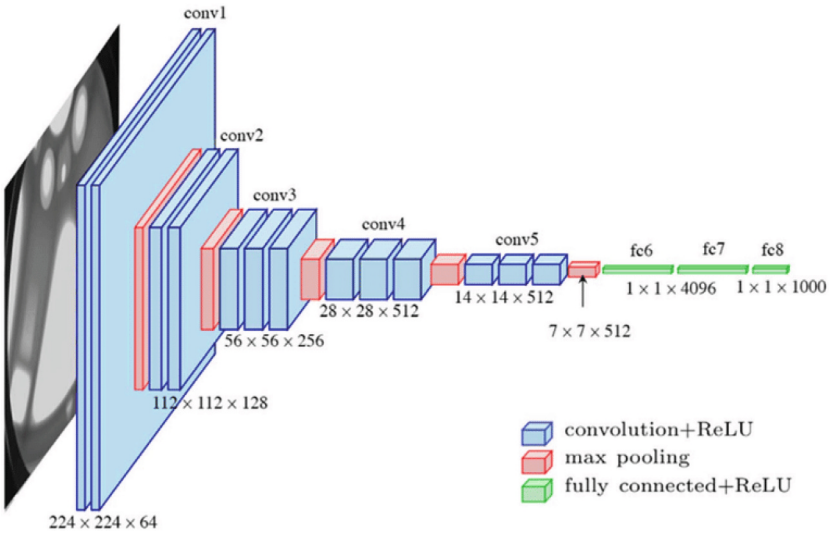

Fig. 2. Traditional VGG16 architecture.

Medical image segmentation is a challenging task that requires a class for each image pixel. CNN architectures were designed specifically for this aim, such as Fully Convolutional Networks (FCN) [26], SegNet [27], U-Net. In our work, we used U-Net for the segmentation process, featuring a contraction path

to capture context and a symmetric expansion path for precise localization [33]. The network uses skip connections to link feature maps from the contraction path with those from the expansion path. This preserves high-resolution information, which is essential to define the edges of a tumor. The complete structure of the U-Net architecture is presented in Fig. 3.

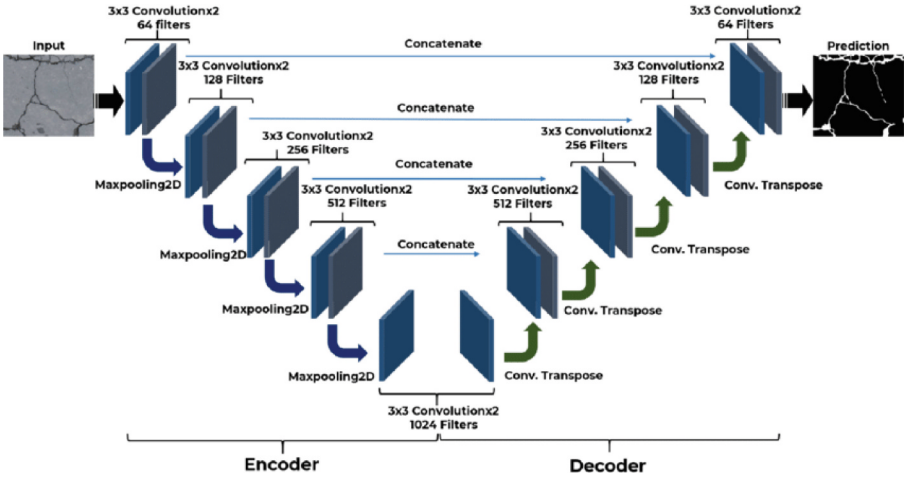

Fig. 3. Traditional U-Net architecture.

The database used for training and validation was the Curated Breast Imaging Subset of DDSM [32]. This repository provides digitized mammograms with precise annotations and segmentation masks. These masks serve as the gold standard for the supervised training of the proposed neural networks.

4 Results and Discussion

The experimental evaluation of this study is based on a structured methodology. It covers everything from data acquisition to the final performance analysis. The development environment was Google Colaboratory Platform [29], providing the necessary computational resources, like the NVIDIA T4 graphics processing unit [36], to train the deep learning models.

First, we prepared the dataset. The Curated Breast Imaging Subset of DDSM dataset contains mammogram images categorized into benign and malignant cases according to the breast imaging reporting and data system [32,38]. The experiment followed sequential pre-processing steps to ensure input data quality. First, we filtered the original database to select only images with mass annotations. We discarded calcification cases to keep the sample uniform. Then, we resized the images to a resolution of 224×224 pixels and converted them to the

RGB color space. Finally, we divided the images into training, validation, and test groups, in order to better evaluate the model.

For the U-Net segmentation architecture, we conducted four different experiments. These tests evaluated baseline settings, data augmentation, learning rate reduction and a hybrid loss function, consisting of Dice coefficient with binary cross entropy. For Experiment 1, we applied the standard learning rate of 1×10^{-4}, usually used as a starting point for computational vision tasks. For Experiment 2, data augmentation was introduced, while keeping the standard learning rate. For Experiment 3, the learning rate was reduced to 1×10^{-5}. For Experiment 4, a hybrid loss function was introduced, combining the Dice coefficient with binary cross entropy, while keeping data augmentation. Since this network did not perform well in identifying nodules, none of the optimization techniques outperformed the initial model.

The best result remained in the baseline, achieving an accuracy of 55%, a sensitivity of 60%, and a maximum regional similarity coefficient of about 64%. This shows that the model did not learn well (underfitting), highlighting its inability to capture the complex nature of the data for the tumor boundary detection task. Table 2 details the results of all attempts with the U-Net.

Table 2. U-Net results based on four models: baseline, data augmentation, learning rate and hybrid loss function

	Experiment 1	Experiment 2	Experiment 3	Experiment 4
Best Dice	0.6396	0.6349	0.6341	0.6290
Best IoU	0.4871	0.4819	0.4825	0.4749
Best Epoch	10	14	19	18
Accuracy	55%	55%	55%	55%
Precision	54%	54%	55%	55%
Recall	60%	49%	52%	45%
F1-Score	0.57	0.52	0.53	0.49

Due to the segmentation limits, the research moved to a second phase focused on direct classification using the VGG16 architecture. We subjected this model to four rigorous optimization experiments. The first experiment established the baseline through transfer learning with a fully frozen feature extractor, which means that the weights were fixed and only the final classification layers were trained with mammogram data. This approach allows the model to use the generalized knowledge achieved in ImageNet and adapt to the new specific task.

The graphic Accuracy per Epoch, in Fig. 4, shows that the training accuracy rises consistently, reaching values greater than 80%, indicating that the model

has the capacity to learn and memorize the training data. However, the Validation Accuracy stagnates prematurely, which is a symptom of overfitting. This diagnosis is confirmed by the graphic Loss per Epoch, in Fig. 4, in which we can see that, after Epoch 7, the model stopped learning patterns, losing performance with new data. The results on the test set showed an overall accuracy of 63.8% and a sensitivity of 65.3%.



Fig. 4. Accuracy and Loss per Epoch for the baseline experiment with VGG16.

However, the model produced a high volume of false positives, reducing precision to 52.7%. This indicates that, after 182 times, the model predicted malignant, while 47.3% of them were incorrect, which, in a real scenario, would generate a high volume of unnecessary biopsies. The clinical diagnosis pointed to severe overfitting. The network memorized the training data, but failed to generalize. This behavior can be seen in detail in Fig. 5 from the baseline confusion matrix.

The second experiment tried to increase the learning capacity by applying fine tuning. We changed the last convolutional block and introduced an automated scheduler to progressively reduce the learning rate. This intervention, without proper compensation, made the model unstable and uncontrollable. The graphic Accuracy and Loss per Epoch, seen in Fig. 6, shows that a Training Accuracy rising to more than 90%, while the Training Loss converges to values near zero. This demonstrates that the model, now more complex, learned the training data with extreme efficiency.

On the other hand, the Validation Loss is volatile, with a tendency to increase almost immediate, with high peaks. The Validation Accuracy stagnates and becomes erratic, confirming that the model was overfitting more aggressively and unstable than the baseline. The sensitivity dropped drastically to 54% because the network failed to identify almost half of the malignant cases. Although with 53% of precision, F1-score dropped to 54%, confirming loss in general performance. As illustrated by the confusion matrix in Fig. 7, this generated the worst

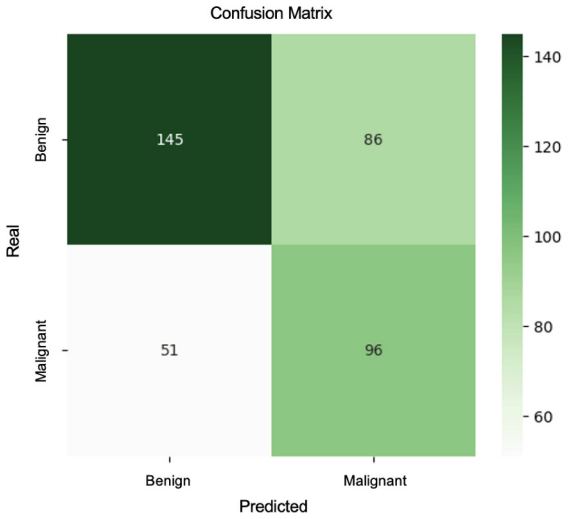

Fig. 5. Confusion Matrix for VGG16 baseline experiment.



Fig. 6. Accuracy and Loss per Epoch for VGG16 with fine tuning.

clinical result of the phase. The application of fine tuning, without a strong and increase in regularization, was prejudicial. The model became more powerful, but also unstable, compromising sensibility.

To control the instability of the second experiment, the third one implemented a strong regularization strategy. We kept the fine tuning, but added data augmentation, to increase training data variability, and raised the neuron dropout rate to the extreme value of 0.7, forcing the network to be less dependent on specific neurons. The validation curves stopped diverging, proving that the aggressive regularization was successful in stabilizing the training and preventing learning collapse. This evolution is demonstrated in Fig. 8 through the accuracy and loss curves per epoch. The Validation curves are extremely volatile,

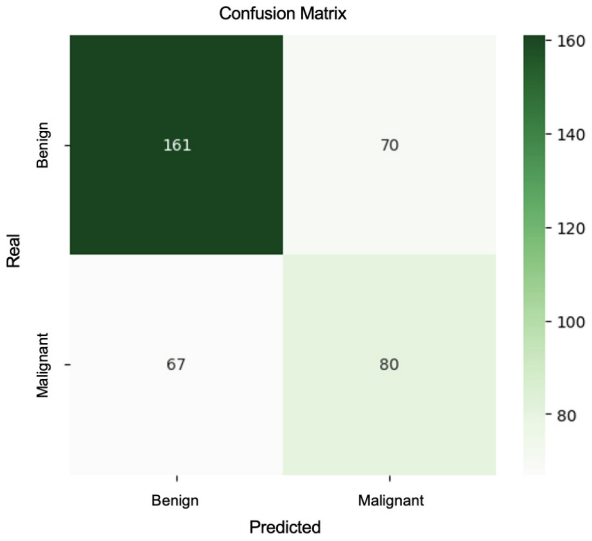

Fig. 7. Confusion Matrix for VGG16 fine tuning experiment.

but they stopped diverging, indicating that the regularization was successful in controlling the model and avoiding the learning collapse.

The fourth and final experiment evaluated whether introducing class balancing through weighted sampling would optimize the system even more [39]. This technique forced the data loader to select samples from the minority class more frequently in order to ensure balanced training batches. The training and validation curves, in Fig. 9, are quite similar to those of experiment three. Both experiments three and four converged to the same performance, establishing the

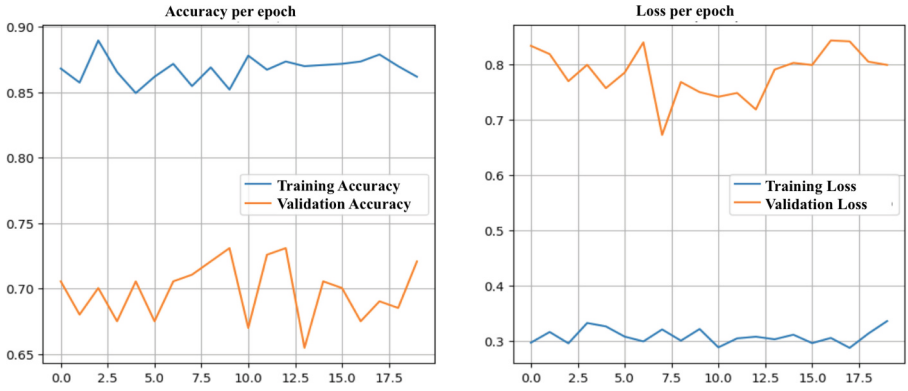

Fig. 8. Accuracy and Loss curves for training and validation per epoch for VGG16 with strong regularization.

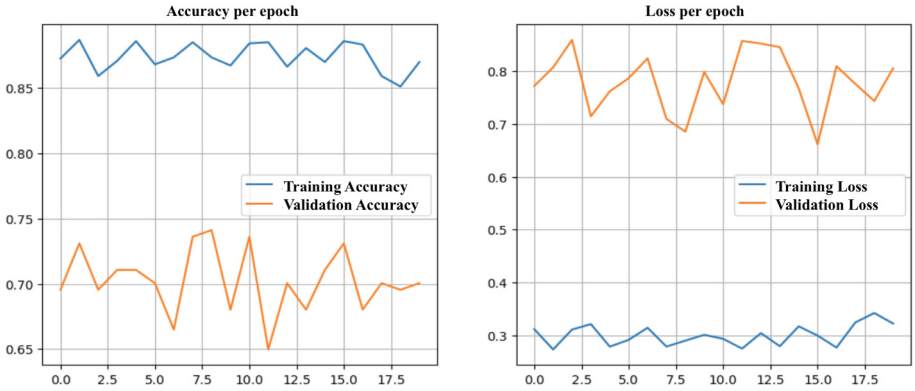

Fig. 9. Accuracy and Loss per Epoch for VGG16 with class balancing.

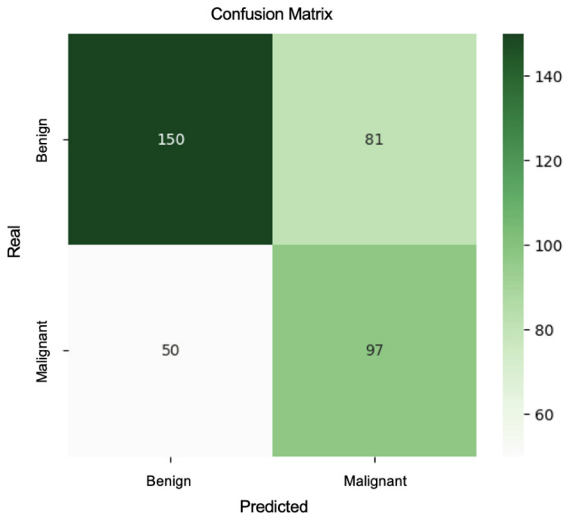

Fig. 10. Confusion Matrix for VGG16 with strong regularization and class balancing.

best results in this work. The significant finding of this experiment was that the methodological addition did not bring isolated improvements. The model converged to the exact same performance as experiment three on the test set. This combined result established the best performance of the work, with an overall accuracy of 65.3% and a sensitivity of 66.0% in detecting malignant cases. Figure 10 displays the resulting confusion matrix from this last step and conclusively illustrates the robustness of the model in correctly identifying pathologies.

5 Conclusion

This study compared semantic segmentation and direct classification strategies to help diagnose breast cancer. Experimental evidence showed that the U-Net architecture has theoretical potential for finding exact regions of interest. However, convergence was complex. The network also required massive amounts of data for pixel-by-pixel annotation. Under the evaluated conditions, this resulted in severe underfitting.

On the other hand, the VGG16 architecture worked better in our experiments. This model was based on the transfer learning paradigm. We implemented a strong regularization strategy using a neuron dropout rate of 0.7. This was the necessary methodological difference to reduce overfitting and ensure knowledge generalization. The system achieved a relevant diagnosis balance, with a 66.0% sensitivity in identifying malignant cases. This suggests that global features may be enough for this task.

In future work, it would be interesting to explore new architectural and methodological limits. Regarding classification, we suggest evaluating more modern and efficient networks, such as the EfficientNet family or Vision Transformers architectures. These models can capture long-range global connections in the image. For the segmentation task, research should move to more advanced U-Net versions, such as Attention U-Net or U-Net++ add attention mechanisms. These mechanisms focus selectively on the most relevant features of the lesion, which could potentially solve the observed underfitting problem.

Acknowledgments. This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior, Brasil (CAPES) – Finance Code 001.

References

1. World Health Organization (WHO). https://www.who.int/news-room/fact-sheets/detail/noncommunicable-diseases. Accessed 03 Apr 2025
2. World Health Organization. https://gco.iarc.fr/tomorrow/en/dataviz. Accessed 03 Apr 2025
3. World Health Organization (WHO). https://www.who.int/news-room/fact-sheets/detail/cancer. Accessed 03 Apr 2025
4. National Cancer Institute. https://www.cancer.gov/types/breast/patient/breast_treatment-pdq_185. Accessed 03 Apr 2025
5. World Health Organization. https://gco.iarc.who.int. Accessed 03 Apr 2025
6. Silva, L.F.D., et al.: A new database for breast research with infrared image. J. Med. Imaging Health Inform. **4**(1), 92–100 (2014)
7. Redmon, J., Divvala, S., Girshick, R., Farhadi, A.: You only look once: unified, real-time object detection. In: Computer Vision Foundation, pp. 779–788 (2016)
8. The Cancer Imaging Archive. About - TCIA. https://www.cancerimagingarchive.net/about-the-cancer-imaging-archive-tcia/. Accessed 23 Nov 2025
9. BUDA, M. MRI brain segmentation. https://www.kaggle.com/mateuszbuda/lgg-mri-segmentation. Accessed 25 Nov 2025

10. Kingma, D.P., Ba, J.: Adam: a method for stochastic optimization. In: Proceedings of the 3rd International Conference on Learning Representations (ICLR), San Diego, CA, USA (2015). https://arxiv.org/abs/1412.6980. Accessed 25 Nov 2025
11. Krizhevsky, A., Sutskever, I., Hinton, G.E.: ImageNet classification with deep convolutional neural networks. In: Advances in Neural Information Processing Systems 25 (NIPS 2012), pp. 1097–1105 (2012)
12. He, K., et al.: Deep residual learning for image recognition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, pp. 770–778 (2016)
13. Falconi, L.G., et al.: Transfer learning and fine tuning in breast mammogram abnormalities classification on CBIS-DDSM database. Adv. Sci. Technol. Eng. Syst. J. **5**(2), 154–165 (2020)
14. Pinheiro, C.A.D.P.: Detecção e Classificação de Nódulos Pulmonares Utilizando Transferência de Aprendizagem Profunda e Inteligência de Enxame. Master's thesis. Universidade do Estado do Rio de Janeiro (UERJ), Rio de Janeiro, RJ (2019)
15. Conselho Nacional de Saúde. https://www.gov.br/conselho-nacional-de-saude/pt-br/atos-normativos/recomendacoes/2014/recomendacao-no-007.pdf/view. Accessed 03 Apr 2025
16. Rodríguez-Ruiz, A., et al.: Can we reduce the workload of mammographic screening by automatic identification of normal exams with artificial intelligence? A feasibility study. Eur. Radiol. **29**(9), 4825–4832 (2019)
17. Lundervold, A.S., Lundervold, A.: An overview of deep learning in medical imaging focusing on MRI. Z. Med. Phys. **29**(2), 102–127 (2019)
18. Russakovsky, O., et al.: ImageNet large scale visual recognition challenge. Int. J. Comput. Vision (IJCV) **115**, 211–252 (2015)
19. Lecun, Y., et al.: Gradient-based learning applied to document recognition. Proc. IEEE **86**(11), 1–46 (1998). ISSN 00189219
20. Baffa, M., Coelho, A.M., Conci, A.: Segmentação de imagens infravermelhas para detecção do câncer de mama utilizando U-net CNN. In: Proceedings of the XXXIV Conference on Graphics, Patterns and Images (SIBGRAPI), pp. 361–368 (2021)
21. John, G.H., Langley, P.: Estimating continuous distributions in Bayesian classifiers. In: Proceedings of the Eleventh Conference on Uncertainty in Artificial Intelligence (UAI), pp. 338–345. Morgan Kaufmann (1995)
22. Breiman, L.: Random forests. Mach. Learn. **45**(1), 5–32 (2001)
23. Méndez, M., Muñoz, J.P.: Identificación de la Región de Interés (ROI) en imágenes médicas, utilizando técnicas de Deep Learning. Master's thesis (2021)
24. Silva, A.F.D.: Sistema Inteligente para o Apoio ao Diagnóstico do Câncer de Mama através da análise de imagens termográficas e do uso de Redes Neurais Artificiais Profundas. Master's thesis. Universidade Federal de Pernambuco (UFPE), Recife, PE (2022)
25. Cortes, C., Vapnik, V.: Support-vector networks. Mach. Learn. **20**(3), 273–297 (1995)
26. Long, J., Shelhamer, E., Darrell, T.: Fully convolutional networks for semantic segmentation. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Boston, MA, USA, pp. 3431–3440 (2015)
27. Badrinarayanan, V., Kendall, A., Cipolla, R.: SegNet: a deep convolutional encoder-decoder architecture for image segmentation. IEEE Trans. Pattern Anal. Mach. Intell. **39**(12), 2481–2495 (2017)
28. Szegedy, C., et al.: Going deeper with convolutions. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Boston, MA, USA, pp. 1–9 (2015)

29. Bisong, E.: Building Machine Learning and Deep Learning Models on Google Cloud Platform: A Comprehensive Guide for Beginners. Apress (2019)
30. Nielsen, M.A.: Neural Networks and Deep Learning. Determination Press (2015). http://neuralnetworksanddeeplearning.com/
31. Goodfellow, I., Bengio, Y., Courville, A.: Deep Learning. MIT Press (2016). https://www.deeplearningbook.org/
32. Lee, R.S., et al.: A curated mammography data set for use in computer-aided detection and diagnosis research. Sci. Data **4**, 170177 (2017)
33. Ronneberger, O., Fischer, P., Brox, T.: U-Net: convolutional networks for biomedical image segmentation. In: Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI), Munich, Germany, pp. 234–241 (2015)
34. Simonyan, K., Zisserman, A.: Very Deep Convolutional Networks for Large-Scale Image Recognition. arXiv preprint, pp. 1097–1105 (2012)
35. Hosny, A., et al.: Artificial intelligence for precision oncology. Nat. Rev. Cancer **22**(10), 563–578 (2022)
36. Nvidia. A new level of acceleration has arrived. https://www.nvidia.com/pt-br/data-center/tesla-t4/. Accessed 25 Mar 2025
37. Bisong, E.: Google Colaboratory, pp. 59–64. Apress (2019). Cloud development environment
38. American College of Radiology: ACR BI-RADS Atlas: Breast Imaging Reporting and Data System, 5th edn. American College of Radiology, Reston, VA (2013)
39. Paszke, A., et al.: Pytorch: an imperative style, high-performance deep learning library. In: Advances in Neural Information Processing Systems (NeurIPS), vol. 32, pp. 8024–8035. Curran Associates, Inc. (2019)