Metastasis Detection of Breast Cancer using Ensemble Deep Learning

Danyllo Carlos Silva e Silva[1], Omar Andres Carmona Cortes[2]

[1]Programa de Mestrado em Engenharia da Computação e Sistemas (PECS)
Universidade Estadual do Maranhão (UEMA)
Cidade Universitária Paulo VI – Caixa Postal 09 – São Luís, MA.

[2]Departamento de Computação (DComp)
Instituto Federal do Maranhão (IFMA)
Av. Getulio Vargas, 04 – 65030-005 – São Luis, MA – Brazil

{danyllosilva@aluno.uema.br, omar@ifma.edu.br}

Abstract. *Breast cancer is one of the diseases which mainly affects women and is responsible for most of the deaths in Brazil, followed by skin and lung cancer. Among the consequences of occurrence, there are genetic predisposition, sedentarism, and late menopause, for example. The metastatic stage of this illness has a low survival rate because the disease spreads from the breast to other parts of the body, and the patients need the diagnosis as fast as possible to start the treatment. Moreover, state-of-art works claim that pathologists can reach 0.72 AUC in analyzing an exam composed of thousands of histopathologic images of lymph node sections. In this context, this work presents an Ensemble Convolutional Neural Network with Transfer Learning, called U-net_VGG19, for detection using the PatchCamelyon dataset. Results indicate that the proposal reached an AUC of 0.9565 and a loss of 0.2869, reaching better results than state-of-the-art CNNs such as VGG16, VGG19, MobileNetV3Large, ConcatNet, and a custom-made CNN.*

1. Introduction

Breast cancer is the disordered development of cells in the breast region with the potential to invade other organs [PREVENTION 2023a]. The diagnosis of this disease is hard to conclude even for an experienced pathologist professional because so many characteristics need to be verified.

Moreover, not only one single cause provokes cancer occurrence. Genetic predisposition and hereditary or unhealthy life habits are important factors contributing to breast cancer incidence. The statistics in Brazil reveal that this disease affects mainly women and 1% of men, with a registry of 18,295 deaths in 2019: 18,068 women and 227 men. For 2023, the estimation is 74,000 new occurrences of breast cancer. For prevention, the individual can choose two strategies: tracking and early diagnosis [PREVENTION 2023b].

Early diagnosis is one of the best ways to obtain the cure for any cancer, mainly when it is in the first stages. The tracking strategy [PREVENTION 2023a] allows the individual to track the health stage of breast cells to identify any abnormalities, such as physical differences in organ structure. Beyond the strategies [Younis et al. 2022], some diagnostic methods complement them, such as mammography (most recommended for tracking stage), magnetic resonance imaging, and self-examination of the breasts (commonly to prevent by the physical touch around the region). Mammography is considered the first option regarding exams to detect earlier cancer because it can anticipate the diagnosis up to three years before the first symptoms.

Beyond strategies and methods, pathologists have confidence in their expertise [Litjens et al. 2022] to reach the most precise diagnosis by looking thoroughly for histopathologic features to find specific cellular structures to justify the conclusion. This task requires an individual effort that can be impacted by external factors such as exhaustion, visual fatigue, and work pressure, among many others; thus, providing the pathologist with tools to support the final decision is crucial. Additionally, pathologists have shown a poor performance of 38% in micro-metastases detection under simulated time constraints [Bejnordi et al. 2017]. Therefore, machine learning algorithms, such as neural networks, have become essential tools for helping pathologists with this diagnosis.

In this context, this work aims to help pathologists by using an ensemble Convolutional Neural Network and applying concepts of Transfer Learning to detect metastatic breast cancer. The results are also compared against the following architectures: VGG-16, VGG-19, CustomNet (network created from scratch), and MobileNetV3Large. Therefore, this work is divided as follows: Section 2 presents some related works; Section 3 shows some literature review and our proposal; Section 4 shows the computational experiment, its setup, and results; finally, Section 5 presents the conclusions of this investigation and future work.

2. Related Works

Fukushima [Fukushima 1980] inspired the first CNN [Lecun et al. 1998] in 1998. Since then, CNNs have played an important role in image segmentation and classification tasks. In other words, this model represents a milestone in image-based machine learning applications [Filho and Cortes 2022]. Since LeCun's publication, intensive research has been done in the area, especially in medical and biomedical fields, as we can see in research such as [Gulshan et al. 2016], [Gayathri et al. 2020], and [Su et al. 2021], between many others.

The VGGs architectures, a popular CNN, were proposed by [Simonyan and Zisserman 2015a]. Thereafter, several works have been proposed. For instance, [Shallu and Mehra 2018] compares VGG16 and VGG19 against ResNet50 using transfer learning to breast cancer classification. In [Saikia et al. 2019], an investigation of the performance of VGG16, VGG19, ResNet-50, and GoogLeNet-V3 is carried out in fine-needle aspiration cytology images.

Regarding breast cancer detection, the architectures ResNet-18, ResNet-152, and GoogLeNet were evaluated for classifying breast cancer using histopathological images in [Silva and Cortes 2020]. Also, [Ismail and Sovuthy 2019] compared ResNet-50 and VGG16 for breast cancer detection using mammograms. Furthermore, [Singh et al. 2020] investigated the issue of imbalanced data in datasets using VGG19.

In the particular field of breast cancer using a VGG architecture and PatchCamelyon dataset, [Bejnordi et al. 2017] proposed the ConcatNet comparing its results with GoogLeNet, ResNet, and VGG16, reaching an AUC of 0.924. Thus, we used this result as a comparison of ensemble models. It is crucial to notice that using PatchCamelyon offers an additional challenge to this investigation due to the large number of images in the dataset.

3. Material and Method

3.1. Convolutional Neural Networks

The first Artificial Neural Network (ANN) was developed in 1940 by [McCulloch and Pitts 1943] to mimic the learning capacity of the human brain through perception and context interpretation as input and get one conclusion according to context recognition. Next, [Rumelhart et al. 1986] introduced the Multilayer Perceptron (MLP), in which between the input and the output, one or more hidden layers can exist for processing data considering specific learning rules and parameters adjusted through epochs (a period that occurs one stage of measure about the output from the network).

Nowadays, the ANNs have evolved immensely, *i.e.*, ANNs grew beyond the used two hidden layers from traditional MLPs. Since 2010, thanks also to the growth in computational power, the processing stage is the focus of development[Aggarwal 2018], and all the efforts were destinated on understanding the complexity of how to get the full and essential features of an object, which can be an image or video, for instance. With this purpose, the Convolutional Neural Networks (CNN) came into the field to maximize the learning power and be nearer to mimicking the behavior of the human brain.

A regular CNN is devised by the following layers: an input, convolutional, pooling, and a fully connected layer with an output. The input receives the images and passes them to the convolutional layers, which use filters to scan over the image and extract features from it. Usually, these features are colors, shapes, and edges that ultimately define a specific image [Beysolow II 2017]. A model can have as many convolutional layers as necessary. Figure 1 shows a CNN with 8 layers (input, a pair of convolutional + pooling, a fully connected with 2 layers, and the output).

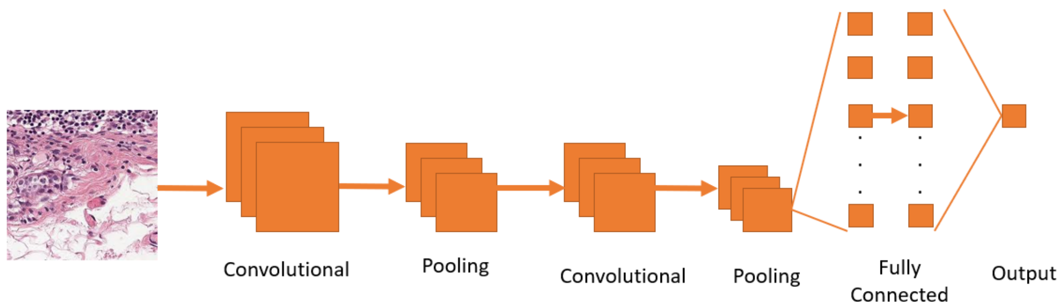

Figure 1. A CNN architecture with 8 layers

Between the convolutional layers, it is possible to use a pooling layer, which takes the feature maps produced in the convolution layer and "pools" them into an image [Beysolow II 2017], performing a dimensionality reduction by using a maximum or average operation. Then, the fully connected input receives a much smaller image than that presented to the network, processing the resultant images and providing the outcome, usually a classification.

Among the various available architectures for deep learning, such as ResNet [He 2015], DenseNet [H. 2016], Inception, Xception, Yolo, VGG, Mobilenet, and many others, the VGG has attracted researchers' attention due to its ability for biomédical applications, especially in image-based exam diagnosis. Figure 2 shows the architecture of the VGG19.

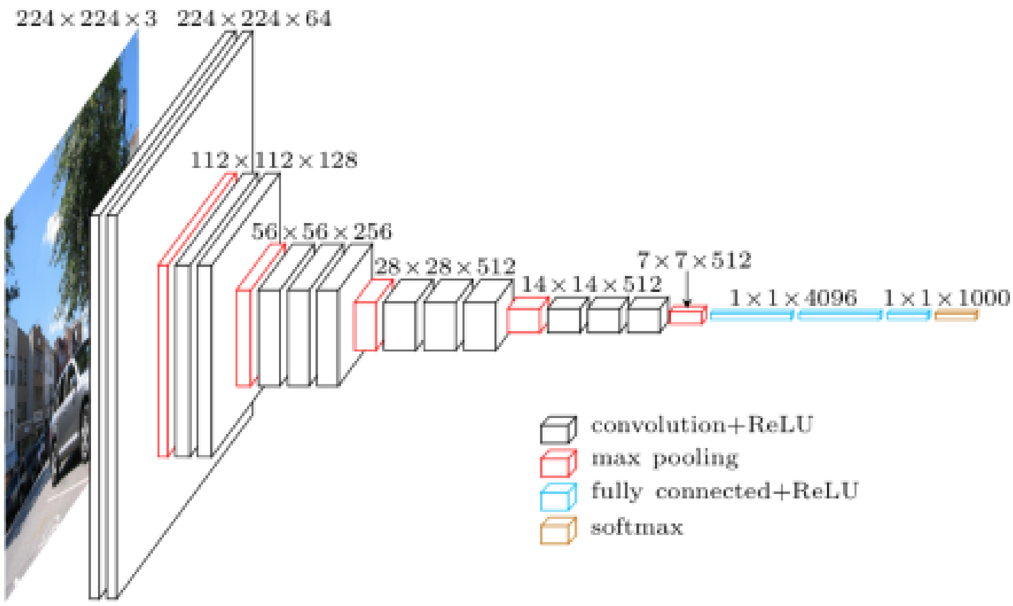

Figure 2. VGG19 architecture

The VGG19, unlike the VGG16 [Simonyan and Zisserman 2015b] with 16 layers deep, presents nineteen layers deep with the following structure: 16 Convolutional layers, 3 Fully-connected (FC) layers (all layers with, at least, one connection between each other), 5 Maxpool layers, and 1 Softmax Layer. Both architectures are available and already pre-trained with ImageNet. Next, we present the proposed ensemble network called U-net_VGG19, which takes advantage of the synaptic weights from the previous training stage, the so-called fine-tuning from Transfer Learning.

3.2. Proposal: U-net_VGG19

The combination of different architectures [Zhang et al. 2022], the so-called ensemble approaches, allows exploring the combination of different models creating new ones, and taking advantage of the best of each network combined to get better results. In this work, we investigate the combination of two distinct architectures: the U-Net [Ronneberger et al. 2022] and the VGG19.

In our proposal, we used the VGG19 because it is commonly used in biomedical applications. In contrast, U-Net captures graphical elements more accurately than other architectures [Ronneberger et al. 2022], demonstrating proper performance segmenting interested areas. Figure 3 shows our architecture proposal. It receives a histopathological image as input with dimensions PatchCamelyonxPatchCamelyonx3 (RGB), uses U-Net as an encoder to feature extraction and learns with abstract representations in convolution blocks. In the middle of the model, a bridge concatenates both architectures, U-Net and

VGG-19. VGG-19 is responsible for finding semantic meaning using 3x3 convolution blocks and the ReLu activation function to get the output. The final result is a dimension output indicating the presence or absence of metastatic breast cancer.

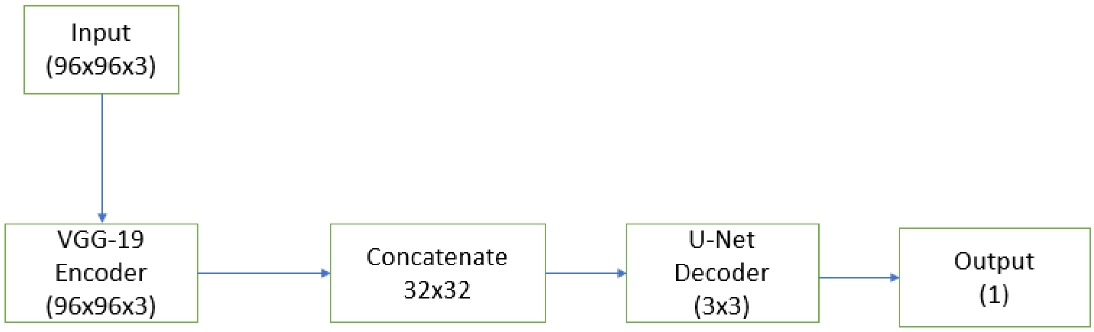

Figure 3. U-net_VGG19 architecture

4. Computational Experiments

4.1. Setup

The work was coded in a professional Google Collaboratory environment with the following configurations: 166 GB HDD, 26 GB memory, and Py3 GPU using Keras and TensorFlow frameworks. Keras ImageDataGenerator was used to generate training and testing subsets. In the end, the model was trained using the validation approach with 110,593 images for training and 27,648 for testing in the ratio of 80/20, similarly to [Jin et al. 2020] Furthermore, the Keras parameter ReduceLROnPlateau was used to reduce the learning when the model tends not to improve in the subsequent epochs and checkpoints.

The model was trained with ten epochs, with batch size equal to 32, a seed of 1,337, a learning rate equal to 0.0001 (empirically chosen based on the state-of-the-art works) with Adam optimizer and activation function sigmoid and loss function binary_crossentropy. The U-NET_VGG19 used Focal Twersky Loss to calculate the loss rate because the U-NET architecture was designed to work better with this function. Figure 4 shows the CustomNet architecture, a simple CNN composed of seven CNN layers and four fully-connected layers with 4608, 256, 256, and 1 neuron, respectively. The custom net was empirically implemented as a comparison baseline with no transfer learning.

4.2. Database

The dataset is a reference for testing algorithms of metastatic tissue detection on histopathological images, called PatchCamelyon [Veeling et al. 2018]. This database contains 327.800 microscopic images of binary tissues classified indicating metastasis's presence or absence. The base ratio is 50/50 for the two classes (with or without metastasis). The image resolution is 96×96, with 3-channel color and TIF file format. Figure 5 shows an example of the dataset marked with positive images by experienced professional pathologists from the department of pathology at Radboud University Medical Center (a specialized center in nephrology, neuro-oncology, and molecular pathologies).

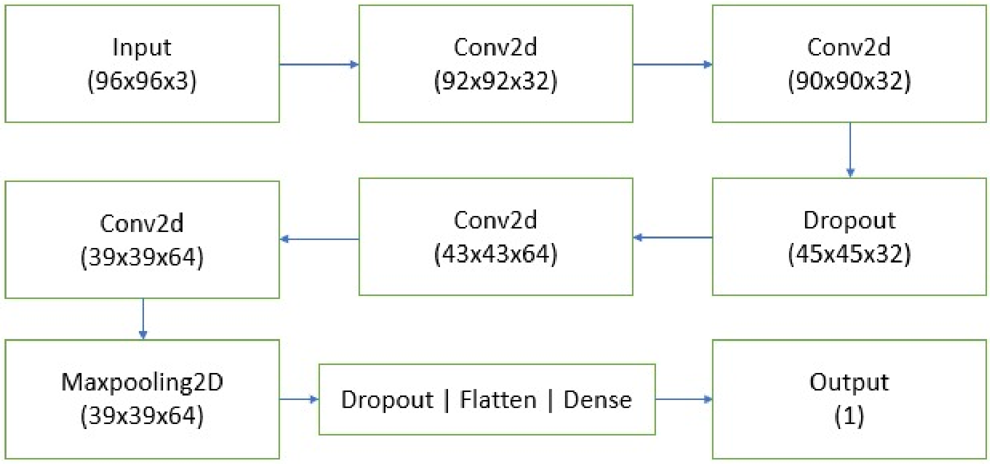

Figure 4. Custom CNN

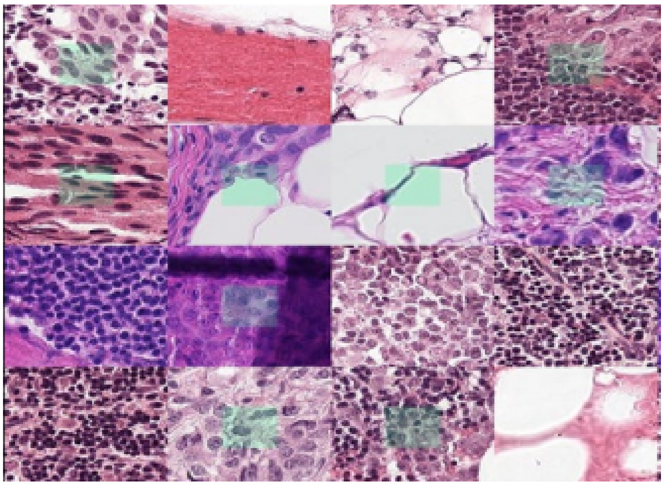

Figure 5. Labeled images from PatchCamelyon dataset.

4.3. Pre-Processing

In pre-processing, the changes were carried out to ensure effective learning. During processing input data, it was necessary to remove some images for the following reasons:

- Duplicate images;
- Illegible with the presence of noise (white/black spots);
- Excessively light or dark, covering regions of interest;
- Corrupted and unreadable images.

As a result, only 138,241 images were readable for the training stage, which changes the balance of the database to the ratio 0.6/0.4 for classes 0 and 1, respectively. All in all, the algorithm follows the steps described in Figure 6.

4.4. Results

The results take into account the metrics Loss, Accuracy, and AUC in five different architectures (VGG16, VGG19, a custom CNN, MobileNetV3Large, and U-NET_VGG19) can be seen from Figures 7 to 11.

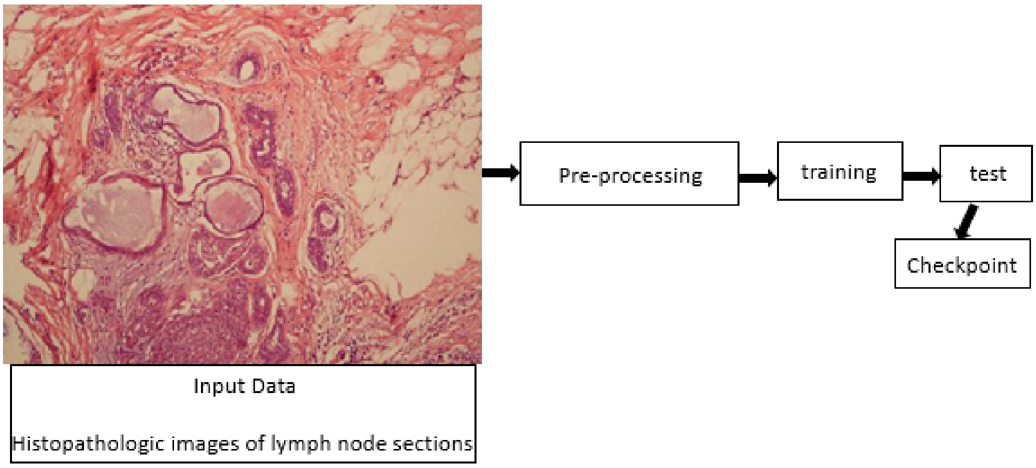

Figure 6. Execution flow of the algorithm

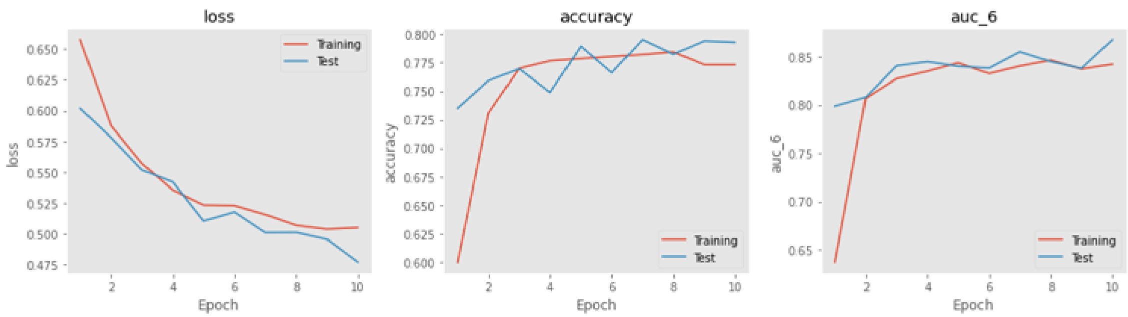

Figure 7. Loss, ACC, and AUC of VGG16

Looking at metrics in Figures 7 to 11, it is impossible to identify the best results at a glance. Excepting in MobileNetV3Large, all architectures tend to converge. However, looking at the y-axis, it is possible to realize that our proposal reaches the best AUC, confirmed by Table 1 with a loss of 0.2869, an accuracy of 0.8750, and an AUC of 0.9565. Surprisingly, CustomNet presented an AUC close to VGG16 and VGG19 but using a reduced architecture.

Table 1. Comparison table of the results.

Algorithm	Test_Loss	Test_Acc	AUC
VGG16	0.4772	0.7928	0.8677
VGG19	0.4776	0.7940	0.8632
CustomNet	0.4703	0.7743	0.8527
MobileNetV3Large	0.6518	0.5926	0.6392
U-Net_VGG-19	**0.2869**	**0.8750**	**0.9565**

In Table 2, we compare our proposal with ConcatNet [Jin et al. 2020], in which we can see that our proposal reached a better AUC, test accuracy, and test loss. Moreover, in ConcatNet, the authors used four U-Net models, training them from 50 to 1000 epochs, while U-Net_VVG 19 was trained using only ten epochs.

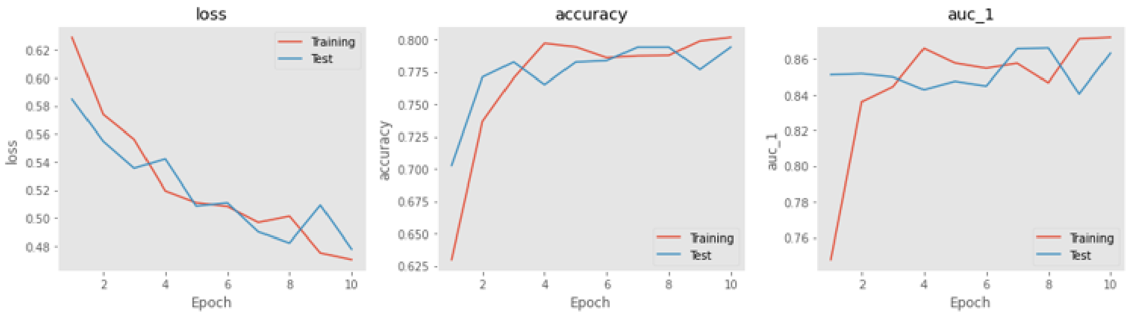

Figure 8. Loss, ACC, and AUC of VGG19

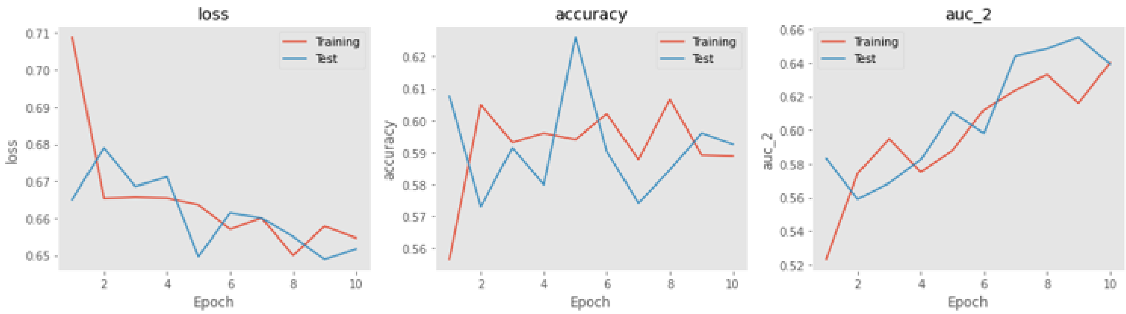

Figure 9. Loss, ACC, and AUC of MobileNetV3Large

Table 2. Comparison against ConcatNet

	AUC	Test_Acc	Test_Loss
This Work	0.9565	0.875	0.2869
ConcatNet	0.92	0.862	0.4357

5. Conclusions and Future Work

This work proposed an ensemble model composed of two well-established CNN architectures: U-NET and VGG19. Results have shown that our proposal reached a 0.9565 AUC, a test accuracy of 0.8750, and a test loss of 0.2869, performing better than state-of-the-art architectures such as VGG16, VGG19, and MobileNetV3Large. Moreover, we devised a custom CNN as a baseline comparison, which reached promising results close to those from VGG architectures but using fewer layers and no transfer learning.

As we can see by the results, the proposal presented promising outcomes indicating that the proposed approach can be helpful for pathologists for more precise and faster diagnosis, which encourages us to develop an application that can embed the U-NET_VGG model in a mobile application or a web service.

Additionally, future work includes applying double transfer learning (DTL) and adding a training stage using the BreaKHis dataset as well. Furthermore, it is possible to explore data augmentation, which we did not analyze in this investigation. Furthermore, the number of training epochs can be reconsidered at a reasonably higher value, such as 100.

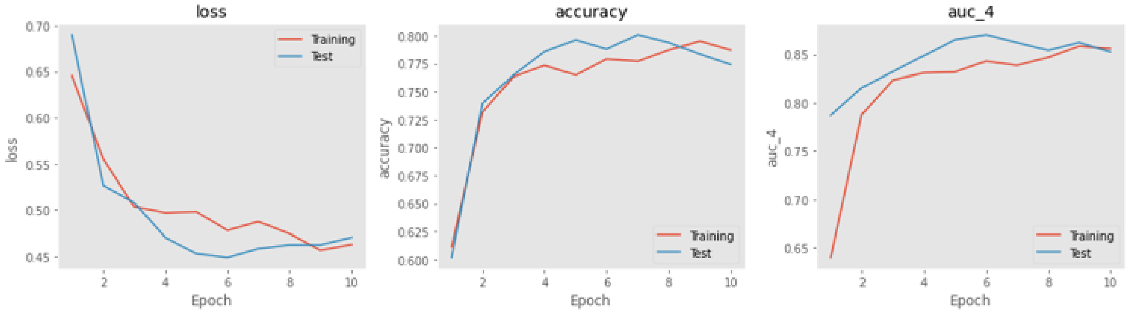

Figure 10. Loss, ACC, and AUC of the Costume CNN

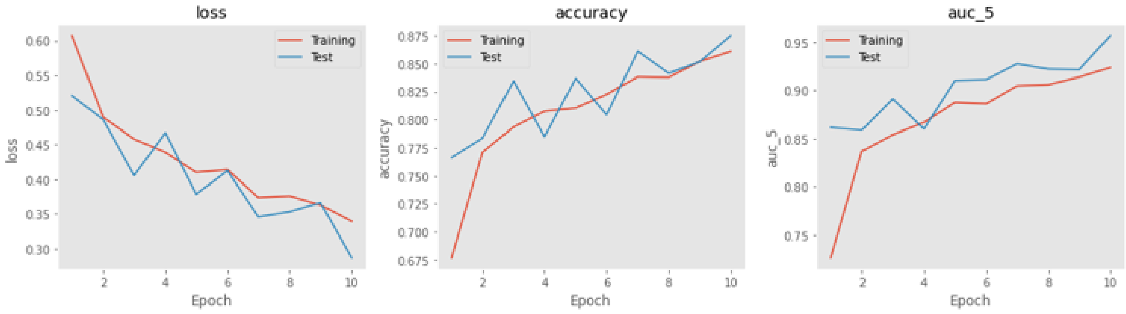

Figure 11. TLoss, ACC, and AUC of U-NET_VGG19

Acknowledgment

We would like to thank the Fundação de Amparo a Pesquisa do Maranhão (FAPEMA) for supporting this work.

References

Aggarwal, C. C. (2018). Neural networks and deep learning. In Springer, editor, *Book on neural networks and deep learning*, pages 9–53. Springer.

Bejnordi, B. E., Veta, M., van Diest, P. J., van Ginneken, B., Karssemeijer, N., Litjens, G., van der Laak, J. A. W. M., and the CAMELYON16 Consortium (2017). Diagnostic Assessment of Deep Learning Algorithms for Detection of Lymph Node Metastases in Women With Breast Cancer. *JAMA*, 318(22):2199–2210.

Beysolow II, T. (2017). *Introduction to Deep Learning Using R: a step-by-step guide to learning and implementing Deep Learning Models Using R*. Apress.

Filho, M. L. R. and Cortes, O. A. C. (2022). Efficient breast cancer classification using histopathological images and a simple vgg. *Revista de Informática Teórica e Aplicada*, 29(1):102—-114.

Fukushima, K. (1980). Neocognitron: A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position. In Springer, editor, *Biological Cybernetics*, pages 193–202. Springer.

Gayathri, S., Gopi, V. P., and Palanisami, P. (2020). A lightweight cnn for diabetic retinopathy classification from fundus images. *Biomedical Signal Processing and Control*, 62:102115.

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.

H., Gao, L. Z. R. S. M. L. V. D. (2016). Densely connected convolutional networks. In University, C., editor, *Cornell University (Arvix.org)*, pages 1–9. Cornell University.

He, K., Z. X. R. S. S. J. (2015). Deep residual learning for image recognition. In University, C., editor, *Cornell University (Arvix.org)*, pages 1–12. Cornell University.

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.

Jin, Y. W., Jia, S., B., A. A., and Hu, P. (2020). Integrative Data Augmentation with U-Net Segmentation Masks Improves Detection of Lymph Node Metastases in Breast Cancer Patients. *Cancers (Basel)*, 12(10):2934.

Lecun, Y., Bottou, L., Bengio, Y., and Haffner, P. (1998). Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86:2278 – 2324.

Litjens, G., Bandi, P., Bejnordi, B. E., Geessink, O., Balkenhol, M., Bult, P., Halilovic, A., Hermsen, M., van de Loo, R., Vogels, R., Manson, Q. F., Stathonikos, N., Baidoshvili, A., van Diest, P., Wauters, C., van Dijks, M., and van der Laak, J. (2022). 1399 hestained sentinel lymph node sections of breast cancer patients: the camelyon dataset. In GigaScience, O., editor, *the CAMELYON dataset*, pages 2–3. Oxford.

McCulloch, W. S. and Pitts, W. (1943). A logical calculus of the ideas immanent in nervous activity. In Springer, editor, *Bulletin of mathematical biophysics*, pages 115–133. Springer.

PREVENTION, B. N. I. O. C. (2023a). Estimate — 2023. In of Health, B. M., editor, *Cancer Incidence in Brazil*, pages 39–40. NATIONAL INSTITUTE OF CANCER PREVENTION (BRAZIL).

PREVENTION, B. N. I. O. C. (2023b). Estimate — 2023. In of Health, B. M., editor, *Cancer Incidence in Brazil*, page 31. NATIONAL INSTITUTE OF CANCER PREVENTION (BRAZIL).

Ronneberger, O., Fischer, P., and Brox, T. (2022). Convolutional networks for biomedical image segmentation. In Springer, editor, *Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, pages 234–241. Springer.

Rumelhart, D. E., Hinton, G. E., and Williams, R. J. (1986). Learning representations by back-propagating errors. *nature*, 323(6088):533–536.

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.

Silva, D. C. S. e. and Cortes, O. A. C. (2020). On convolutiona neural networks and transfer learning for classifying breast cancer on histopahological images using gpu. In *XXVII Brazilian COngress on Biomedical Engineering*.

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

Simonyan, K. and Zisserman, A. (2015b). Very deep convolutional networks for large-scale image recognition. In Oxford, editor, *Computer Vision and Pattern Recognition*, pages 1–14. Oxford.

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.

Su, Y., Li, D., and Chen, X. (2021). Lung nodule detection based on faster r-cnn framework. *Computer Methods and Programs in Biomedicine*, 200:105866.

Veeling, B. S., Linmans, J., Winkens, J., Cohen, T., and Welling, M. (2018). Rotation equivariant cnns for digital pathology. In Springer, editor, *Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, pages 234–241. Springer.

Younis, Y. S., Ali, A. H., Alhafidhb, O. K. S., Yahia, W. B., Alazzam, M. B., Hamad, A. A., and Meraf, Z. (2022). Early diagnosis of breast cancer using image processing techniques. In Velmurugan, P., editor, *Applications of Nanomaterials and Nanotechnology in Engineering, Environment and Life Sciences*, pages 2–5. Hindawi.

Zhang, Y., Liu, J., and Shen, W. (2022). A review of ensemble learning algorithms used in remote sensing applications. In Journals, M. O. A., editor, *Applied Science*, pages 1–14. MDPI Open Access Journals.