

Classificação de Imagens Histopatológicas Usando uma VGG e Mecanismo de Atenção para Detecção de Câncer de Mama

Marcelo Luis Rodrigues Filho¹, Omar Andres Carmona Cortes²

¹Bacharelado em Sistemas de Informação – Instituto Federal do Maranhão (IFMA)
Av. Getúlio Vargas, 04 – Monte Castelo – 65030-005 – São Luis – MA.

²Departamento de Computação (DComp) – Instituto Federal do Maranhão (IFMA)

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

Abstract. *Breast cancer is a severe illness caused by uncontrolled cell division. Thus, early detection is critical to improve the patient's life or even cure the disease. In this context, this work proposes to use a convolutional neural network, the so-called miniVGG, devised by seven layers and derived from traditional VGG architectures. Results have shown that a less deep network combined with the proper choice of learning rate and the use of attention mechanism overcomes the performance of traditional VGGs in both time and classification performance, reaching an accuracy of 93%, precision of 91%, Recall of 92%, F1_Score of 92%, and AUC of 99%.*

Resumo. *O câncer de mama é causada pela divisão celular descontrolada. Dessa forma, a detecção precoce da doença é fundamental para melhorar a vida dos pacientes ou até mesmo curá-los. Nesse contexto, o presente trabalho propõe o uso de uma arquitetura de rede convolucional (CNN) chamada miniVGG, formada por sete camadas e derivada das arquiteturas VGG tradicionais. Os resultados mostraram que uma variação menos profunda combinada com uma estratégia de escolha da taxa de aprendizado e mecanismo de atenção supera o desempenho em tempo de treinamento e classificação das arquiteturas VGG tradicionais alcançando uma acurácia de 93%, precisão de 91%, recall de 92%, F1_Score de 92% e um AUC de 99%.*

1. Introdução

A Organização Mundial da Saúde (OMS) estimou que 2,3 milhões de mulheres foram diagnosticadas com câncer de mama em 2020 e 685 mil mortes foram registrada em todo o mundo. No mesmo ano, havia 7,8 milhões de mulheres vivendo com câncer de mama nos últimos 5 anos, sendo o tipo da doença mais prevalente do mundo [OMS 2021].

Uma das formas clínicas de se detectar a presença de câncer de mama é através de exames histopatológicos, na qual são analisados tecidos afetados através do microscópio. O principal problema da inspeção visual é que ela é suscetível a condições externas como estresse, fadiga física ou visual, dentre outras. Assim, o uso de ferramentas de inteligência computacional como as redes neurais convolucionais (CNN) tornam-se essencial para contornar as condições externas e fornecer um diagnóstico rápido e eficaz. Não obstante, as imagens histopatológicas são também um desafio para os classificadores devido a similaridade entre imagens de diferentes tipos de tumores assim como a alta variabilidade de imagens para um único tipo de classe [Han et al. 2017].

O trabalho de [Spanhol et al. 2016] introduziu uma base de dados pública denominada BreakHis contendo 7909 imagens de 82 pacientes dentre 8 tipos de câncer de mama, se tornando a principal base de dados para o teste de algoritmos de aprendizagem de máquina. As CNNs tem sido especialmente aplicadas na classificação do BreakHis usando as 8 classes de câncer de mama, pois tem se mostrado eficientes em diversos trabalhos, como por exemplo, [Han et al. 2017] e [Alom et al. 2019] que usam CNNs na classificação em múltiplas classes, porém a limitação destes métodos consiste na complexidade na abordagem para acumulação de *features* ou combinação de diferentes Funções de Perda usando várias imagens do BreakHis como múltiplas entradas da função. Não obstante, destacam-se um CNN chamada VGG, proposta em 2015 [Simonyan and Zisserman 2015]. Essa rede tem sido usada em trabalhos como, por exemplo, [Sitaula and Hossain 2020] e [Younis et al. 2022] que usam a VGG com mecanismos de atenção.

Dessa forma, neste trabalho propõe-se o uso de uma miniVGG com Mecanismo de Atenção, visando melhorar a capacidade do modelo para reconhecimento de tumores, aumentando a capacidade da rede em focar em *features* importantes da imagem, muito embora o mecanismo afete o tempo de treinamento da rede. Adicionalmente, o método *Gradient Class Activation Map(Grad-CAM)* foi utilizado para visualizar as ativações da última camada das redes e obter melhores observações sobre as partes das imagem histopatológica, conseqüentemente melhorando também a interpretabilidade da classificação. Nesse contexto, este trabalho contribui com a criação de uma arquitetura VGG menos profunda usando o Mecanismo de Atenção em Canal de Cor e com treinamento baseado *Cyclical Learning Rate* [Smith 2015, Smith and Topin 2017] que determina a melhor taxa de aprendizagem.

2. MiniVGG com Mecanismo de Atenção

As redes VGG são formadas por blocos de convolução com pequenos kernels (filtros) de tamanho 3x3 e 1x1 com camadas densas para classificação de imagens. Os filtros são usados para extrair características (mais conhecida como *features*) da imagem digital, como bordas, cores e formatos, o que é importante no diagnóstico patológico, sendo as camadas densas utilizadas para realizar a inferência sobre qual o tipo de câncer presente na imagem. Assim, neste trabalho propõe-se o uso de mecanismos de atenção, um mecanismo de aprendizagem que simula a atenção cognitiva humana, em uma miniVGG, sendo que o mecanismo de atenção é complementar às convoluções, modelando com eficiência as dependências de vários níveis e de longo alcance em toda as regiões da imagem [Chaudhari et al. 2019].

Introduzido por [Hu et al. 2017], o Mecanismo de Atenção no Canal é usado para classificação de imagens. Neste trabalho, o mecanismo usa *Global Average Pooling*, Convolução 1x1 com Função ReLU no lugar das Rede Neurais Densas usadas por [Hu et al. 2017] para evitar aumento de complexidade da rede e buscar um mecanismo mais nativo para as *features* convolucionais. Essa abordagem formula a relação entre as *features*, entendida como estado intermediário da imagem de entrada, com os pesos de atenção resultantes no mecanismo de atenção. A Figura 1 apresenta a VGG e a modificação dela com o mecanismo de atenção, que é colocada após cada camada convolucional + *pooling*, porém gerando apenas um pequeno aumento na quantidade de parâmetros.

conv3-64 + BN conv3-128 + BN conv3-256 + BN conv3-512 + BN GAP Rede Neural Densa-512 Rede Neural Densa -256 Rede Neural Densa -128 soft-max TOTAL DE PARÂMETROS: 1,718,152	conv3-64 + BN Atenção conv3-128 + BN Atenção conv3-256 + BN Atenção conv3-512 + BN Atenção GAP Rede Neural Densa-512 Rede Neural Densa -256 Rede Neural Densa -128 soft-max TOTAL DE PARÂMETROS: 1,761,672
---	---

Figura 1. MiniVGG proposta sem e com atenção

Nos experimentos foram realizadas operações de Aumento de Dados em Tempo de Execução (*Online Data Augmentation*) usadas foram Rotação Horizontal e Rotação Vertical da imagem com probabilidade de 50%. As imagens foram redimensionadas para 256x256 para reduzir o consumo de memória RAM. As redes VGG11, VGG13, VGG16 e VGG19 foram otimizadas com o otimizador Adam configurado com taxa de aprendizado 0,001. Os *pixels* das imagens foram normalizados dividindo cada valor de pixel por 255 para deixar seus valores no intervalo [0, 1]. O tamanho do lote é fixado em 32.

3. Experimentos e Resultados

As CNNs foram treinadas no Google Colab Pro, sendo o aumento de dados, feito baseado em *Data Warping* em tempo de treinamento para suprir a necessidade de alta demanda computacional. A base de dados utilizada é a BreakHis com as 8 classes, com a resolução ajustada em 256x256 que corresponde à dimensão das imagens do conjunto de dados ImageNet onde as redes VGGs foram pré-treinadas. Para o treinamento utilizou-se a abordagem de validação na proporção 80/20. Já no experimento para verificar o efeito das magnitudes do microscópio, utilizou-se o *k-fold*.

A Tabela 1 mostra as métricas de classificação para as 4 configurações usuais da VGG, para a miniVGG e para a miniVGG com mecanismo de atenção e o uso da técnica de CLR (*Cyclical Learning Rate*), que aumentou o valor das métricas com uma taxa de aprendizado de 0,0002 para a miniVGG sem atenção e 0,003 com atenção. Dentre as tradicionais, a VGG11 obteve os melhores resultados, indicando que menos camadas pode gerar melhores resultados, o que acaba sendo confirmado pela miniVGG. Além disso, é possível observar que o tempo de execução é proporcional ao número de parâmetros das redes, e como resultado a miniVGG obteve o menor tempo de treinamento.

O Mecanismo de Atenção em Canal de cor deixou a miniVGG mais precisa na classificação, mas aumentou o tempo de treinamento. Isso provavelmente é devido o alto requisito computacional das operações de multiplicação das máscaras ou pesos de atenção com os mapas de características *feature maps* extraídos das camadas convolucionais, apesar de que os parâmetros da miniVGG e da miniVGG com Mecanismo de Atenção não possuem grande diferença de quantidade. Além disso, a atenção sobre os canais das *features* indicam que a cor da imagem é uma importante *feature* para melhorar a classificação.

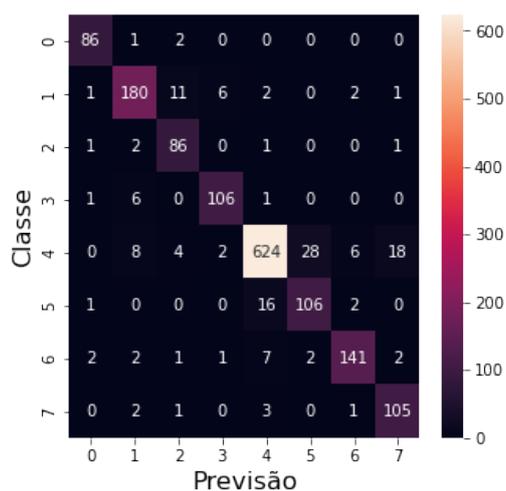
A matriz de confusão na Figura 2 mostra que o modelo proposto classificou corretamente 1434 imagens de 1582 nos dados de teste considerando as oito classes de tumores,

Tabela 1. Resultados das VGGs

MODELO	PRC	RECALL	F1	ACC	ROC/AUC	Tempo
VGG11	0.75	0.71	0.71	0.71	0.93	52min 8s
VGG13	0.70	0.68	0.68	0.65	0.93	1h 8min 11s
VGG16	0.69	0.67	0.68	0.66	0.92	1h 22min 48s
VGG19	0.68	0.66	0.66	0.61	0.91	1h 38min 24s
miniVGG	0.91	0.91	0.91	0.91	0.99	50min 19s
miniVGG-Atenção	0.93	0.92	0.92	0.92	0.99	1h 1min 31s

o que representa 90,6% de acertos. Além disso, é possível observar que o modelo tendeu a confundir as classe de código 4 (*ductal carcinoma*) e código 5 (*lobular carcinoma*) com um total de 44 erros, o que indica que as imagens destas classes de câncer maligno possuem alta similaridade em comparação com outras classes.

Figura 2. Matriz de confusão - miniVGG



Considerando que a miniVGG apresentou os melhores resultados, investigou-se como a rede se comporta para cada magnitude de zoom da imagem. A Tabela 2 mostra a média das métricas de avaliação por fator de ampliação para 10 subdivisões (*folds*) estratificadas. Os resultados mostram que a miniVGG possui bom desempenho de classificação das Imagens Histopatológicas no fator de ampliação 40X, correspondendo à primeira opção na qual os patologistas escolhem as regiões de interesse, já que nesta ampliação aparecem mais estruturas microscópicas por toda a imagem que podem indicar a presença de câncer. Portanto, o modelo é similar ao patologista quando tem a necessidade de encontrar mais estruturas ou *features* que estão sempre presentes na imagem com menor ampliação.

A Figura 3 apresenta através do Grad-CAM como a miniVGG “enxerga” uma adenosis sem atenção e com atenção, sendo a imagem com atenção bem mais definida. A Figura 4 mostra um comparativo entre a miniVGG sem e com atenção para cada classe do BreakHis. Observa-se que diferentes partes da imagem são consideradas como regiões de interesse dependendo do uso do Mecanismo de Atenção.

Tabela 2. Resultados para 10 folds de Validação Cruzada (k = 10)

Métrica	40X	100X	200X	400X
Acurácia	0,92 ± 0,02	0,90 ± 0,03	0,87 ± 0,03	0,87 ± 0,02
F1 score	0,89 ± 0,02	0,89 ± 0,03	0,86 ± 0,02	0,85 ± 0,02
Precisão	0,90 ± 0,01	0,90 ± 0,03	0,87 ± 0,02	0,87 ± 0,02
Recall	0,89 ± 0,02	0,89 ± 0,03	0,85 ± 0,02	0,85 ± 0,02
ROC/AUC	0,89 ± 0,02	0,89 ± 0,03	0,85 ± 0,02	0,85 ± 0,02

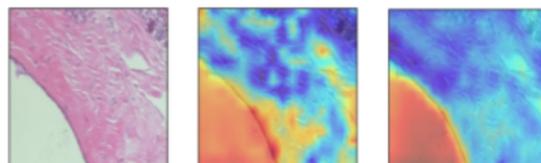


Figura 3. Ativação da Adenosis (esquerda) sem Atenção (meio) e com Atenção (direita).

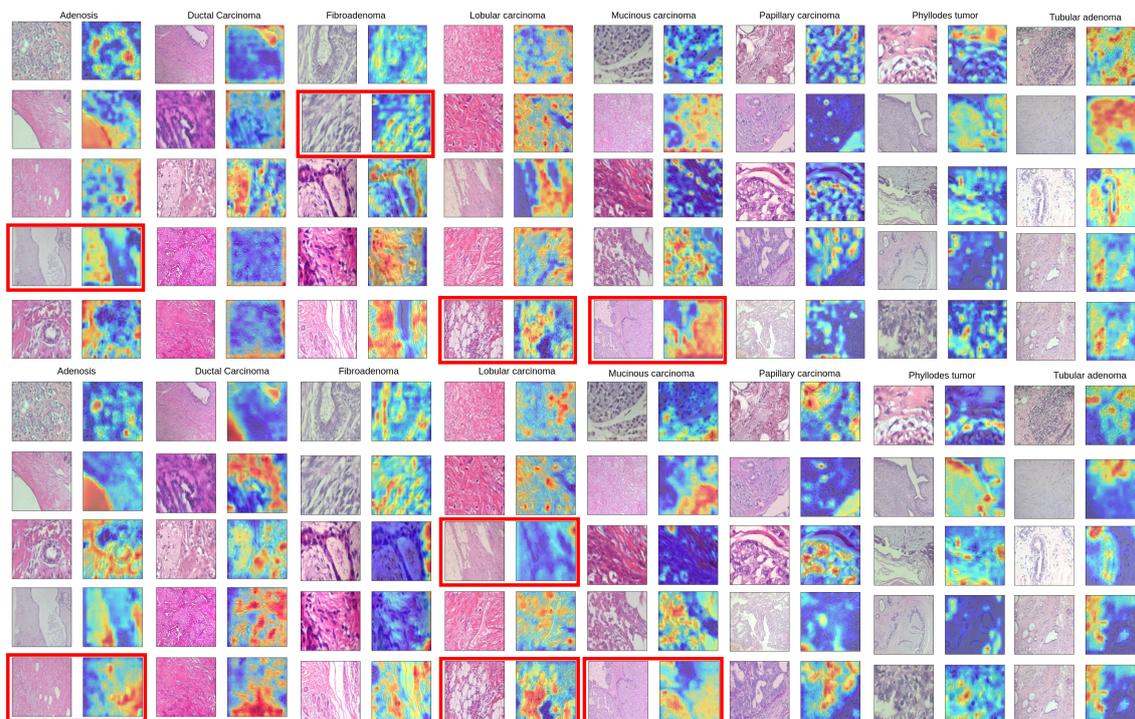


Figura 4. miniVGG sem e com Atenção (topo e base da figura): Ativações por Classe.

Na comparação entre classes, em algumas imagens o modelo se concentra nas bordas para as classe o *Ductal Carcinoma* e *Lobular carcinoma*. Surpreendentemente, o modelo nem sempre foca a atenção nas estruturas celulares que o patologista considera antes de realizar o diagnóstico, pois em algumas delas são encontradas evidências de que o modelo considera a parte branca da imagem - tecido adiposo ou matriz extracelular - que não é uma parte discriminativa na imagem para indicação de Câncer de Mama por um patologista. Adicionalmente, as ativações parecem mais delineadas e menos dispersas sobre certas partes das imagem nas ativações com atenção, portanto, o melhor delineamento

das estruturas e foco em aspectos minuciosos das imagens histopatológicas resultaram na melhoria dos resultados do modelo.

4. Conclusões

Este trabalho apresentou um miniVGG com mecanismo de atenção para classificação multiclasse de câncer de mama. Os resultados mostraram que redes menos profundas podem obter bons resultados chegando próximo a 93% nas métricas de desempenho. Além disso, a utilização do Grad-CAM mostrou como o mecanismo de atenção gera imagens com melhor definição de bordas, levando a melhores resultados sem aumentar significativamente o tempo de treinamento. Os trabalhos futuros incluem a classificação do BreakHis com dados separados por pacientes e avaliação das arquiteturas de *Mobile Transformer* para capturar o contexto global das *features* sem convolução.

Referências

- Alom, M. Z., Yakopcic, C., Taha, T., and Asari, V. (2019). Breast cancer classification from histopathological images with inception recurrent residual convolutional neural network. *Journal of Digital Imaging*, 32.
- Chaudhari, S., Polatkan, G., Ramanath, R., and Mithal, V. (2019). *An Attentive Survey of Attention Models*. *CoRR*, abs/1904.02874.
- Han, Z., Wei, B., Zheng, Y., Yin, Y., Li, K., and Li, S. (2017). Breast cancer multi-classification from histopathological images with structured deep learning model. *Scientific Reports*, 7.
- Hu, J., Shen, L., and Sun, G. (2017). Squeeze-and-excitation networks. *CoRR*, abs/1709.01507.
- OMS (2021). *Breast Cancer*. Disponível em: <https://www.who.int/news-room/fact-sheets/detail/breast-cancer>.
- Simonyan, K. and Zisserman, A. (2015). Very deep convolutional networks for large-scale image recognition. In *International Conference on Learning Representations*.
- Sitaula, C. and Hossain, M. (2020). Attention-based VGG-16 model for covid-19 chest x-ray image classification. *Applied Intelligence*, 51:2850–2863.
- Smith, L. N. (2015). Cyclical learning rates for training neural networks. *CoRR*, abs/1506.01186.
- Smith, L. N. and Topin, N. (2017). Super-convergence: Very fast training of networks using large learning rates. *CoRR*, abs/1708.07120.
- Spanhol, F. A., Oliveira, L. S., Petitjean, C., and Heutte, L. (2016). *A Dataset for Breast Cancer Histopathological Image Classification*. *IEEE Transactions on Biomedical Engineering*, 63(7):1455–1462.
- Younis, A., Qiang, L., Nyatega, C. O., Adamu, M. J., and Kawuwa, H. B. (2022). Brain tumor analysis using deep learning and vgg-16 ensembling learning approaches. *Applied Sciences*, 12(14).