

Análise comparativa da influência de otimizadores no desempenho de uma CNN para detecção do câncer de mama

Stefane A. Santos¹, Andressa G. Moreira¹, Ialis C. P. Junior¹

¹Universidade Federal do Ceará (UFC)
Engenharia da Computação – Sobral – CE – Brasil

{stefaneadna, andressagomes}@alu.ufc.br, ialis@sobral.ufc.br

Abstract. *The field of artificial intelligence (AI) presents remarkable advances in medicine. Studies analyzed on the application of Convolutional Neural Networks for the detection of breast cancer. In this article, a comparative analysis is performed between the optimization methods (Adam, Adadelta, Adagrad, Adamax, Nadam, RMSprop) related to a VggNet16 architecture for the classification of neoplasms in histopathological images. The experiments were carried out by creating models for the magnification factors (40x, 100x, 200x, 400x) of the images extracted from the BreakHis dataset. The Adam optimizer got the best result for the image set, specifically on base 400x.*

Resumo. *O campo da inteligência artificial (IA) apresenta notáveis avanços na medicina. Estudos analisam a aplicação de Redes Neurais Convolucionais para a detecção de câncer de mama. Neste artigo, é realizada uma análise comparativa entre os métodos de otimização (Adam, Adadelta, Adagrad, Adamax, Nadam, RMSprop) aplicados a uma arquitetura VggNet16 para a classificação de neoplasias em imagens histopatológicas. Os experimentos foram realizados com a criação de modelos para os fatores de ampliação (40x, 100x, 200x, 400x) das imagens extraídas do dataset BreakHis. O otimizador Adam obteve o melhor resultado para o conjunto de imagens, especificamente na base 400x.*

1. Introdução

O câncer de mama é o tipo de câncer que mais ameaça a vida das mulheres, após o câncer de pulmão [Alanazi SA 2021]. Segundo a American Cancer Society, o câncer de mama é caracterizado pelo crescimento das células mamárias, que formam nódulos, que podem ser benignos ou malignos [American Cancer Society 2019]. As causas do câncer de mama continuam desconhecidas, logo, a detecção precoce é o meio mais eficaz para reduzir a taxa de mortalidade [Osareh and Shadgar 2011]. Diferentes técnicas são usadas para diagnosticar a doença, como a mamografia, biópsia e ultrassom. Entretanto, o diagnóstico manual pode atrasar a detecção. Dessa forma, técnicas de aprendizado de máquina estão sendo amplamente utilizadas para realizar a detecção do câncer de mama [Gayathri and Sumathi 2016].

As Redes Neurais Convolucionais (CNNs) abrangem diversas aplicações referentes ao reconhecimento e à classificação de imagens. Ademais, estudos recentes mostram resultados significativos na utilização do algoritmo de aprendizado profundo CNN em aplicações médicas. O trabalho proposto por [Horie et al. 2018] analisa os resultados de diagnóstico de câncer de esôfago usando CNN. [Tschandl et al. 2018] fizeram uso da

técnica em um estudo para realizar o diagnóstico de câncer de pele não pigmentado. Por fim, [Spanhol et al. 2016a], propõe um método baseado em CNN para a classificação de câncer de mama.

Dessa forma, a CNN é composta por múltiplas funções e camadas de convolução e apresenta ampla quantidade de parâmetros e diversas arquiteturas [Vargas et al. 2016]. Dentre os parâmetros que contribuem para o funcionamento do classificador, os otimizadores possuem um papel de total relevância. O otimizador é responsável pela atualização dos pesos para minimizar a função de perda. Em suma, o objetivo do otimizador é diminuir os erros da rede neural, e conseqüentemente, melhorar o desempenho da rede [Ribeiro and Junior 2020, Yaqub et al. 2020].

Neste artigo é proposto um estudo para comparar a influência de métodos de otimização no desempenho de uma CNN para a detecção do câncer de mama. O desempenho dos modelos foi avaliado por meio de três métricas de avaliação. Portanto, para o desenvolvimento do trabalho foram aplicados a uma arquitetura VggNet16 os otimizadores Adam, Adadelta, Adagrad, Adamax, Nadam e RMSprop.

Este trabalho é apresentado nas seções que seguem: um embasamento teórico sobre a CNN e os otimizadores utilizados nesse artigo é apresentado na seção 2, a seção 3 descreve a metodologia usada para o desenvolvimento do trabalho, na seção 4 são discutidos os resultados obtidos, e por fim, na seção 5 conclui-se o artigo e delimitam-se as pesquisas futuras.

2. Referencial Teórico

Nesta seção são destacados os principais conceitos utilizados para fundamentar a pesquisa. Destacam-se aqui o algoritmo Rede Neural Convolutacional, a arquitetura VGG16 e os métodos de otimização, Adam, Adadelta, Adagrad, Adamax, Nadam e RMSprop.

2.1. Rede Neural Convolutacional

O conceito de Rede Neural Convolutacional (CNN) foi introduzido pela primeira vez em um artigo de Kunihiko Fukushima, definido como um modelo de rede neural para mecanismo de reconhecimento de padrões visuais [Fukushima 1980]. O algoritmo de aprendizado profundo CNN vem sendo amplamente utilizado na classificação e reconhecimento de imagens. A rede é constituída por múltiplas camadas e funções capazes de atribuir importância aos dados de entrada e definir a probabilidade de um dado pertencer a uma das classes disponíveis.

As camadas de convolução são responsáveis pelo processo de aplicação de um filtro a uma entrada para produzir um mapa de característica. Ademais, a camada de ativação transforma os dados recebidos, para apresentar não linearidade. Em seguida, é comumente utilizada a camada de agrupamento (*pooling*) a fim de reduzir a dimensionalidade dos dados. Por fim, ao final da rede seguem as camadas totalmente conectadas, cujo objetivo é traçar os recursos extraídos das camadas anteriores para classificar a saída. Logo, após a última camada adiciona-se uma função de classificação para definir a probabilidade alvo, e conseqüentemente, o resultado da rede [Vargas et al. 2016, Kandel et al. 2020].

2.2. Arquitetura VGG16

O modelo VGG16, proposto em [Simonyan and Zisserman 2015], consiste em uma pilha de camadas, sendo 13 camadas de convolução e 3 camadas totalmente conectadas. O

filtro de convolução utilizado possui tamanho 3×3 e passo de 1 pixel. Ademais, para reduzir o tamanho das matrizes resultantes das camadas de convolução, é realizado o *pool* máximo com uma janela de tamanho 2×2 e passo de 2 pixels. Ao final das camadas de convolução, seguem três camadas *Fully-Connected* (FC), seguidas pela camada final, com a função de classificação, Softmax, que garante que a probabilidade de saída é igual um. [Kandel et al. 2020, Simonyan and Zisserman 2015].

2.3. Otimizadores

A técnica de aprendizagem profundo, CNN, consiste em treinar um conjunto de dados e comparar o rótulo da entrada com o rótulo previsto, a fim de minimizar a função de custo. Desse modo, é mister a utilização de métodos de otimização para obter uma função de custo mínima. Portanto, nesta seção são apresentados algoritmos de otimização responsáveis por minimizar a taxa de erro e melhorar o desempenho da rede neural, buscando evitar o sobreajuste dos dados.

- O otimizador Adam, proposto por [Kingma and Ba 2017], é um método de otimização baseado em descida gradiente estocástica com poucos requisitos de memória, sendo um algoritmo de simples implementação e amplamente utilizado em problemas que demanda grandes quantidades em termos de dados e parâmetros. O otimizador calcula as taxas de aprendizagem adaptativa para diversos parâmetros que abrangem momentos de primeira e segunda ordem.
- O método otimizador Adadelta, introduzido por [Zeiler 2012], é um algoritmo de descida gradiente estocástico baseado em taxa de aprendizagem por dimensão. Trata-se de um método com baixa sobrecarga computacional e que não requer ajuste manual da taxa de aprendizagem. Ademais, é um algoritmo capaz de resolver a queda contínua das taxas de aprendizagem durante o treinamento.
- O Adagrad, proposto por [Duchi et al. 2011], é um otimizador baseado em taxas de aprendizagem específica, que são atualizadas de acordo com os parâmetros durante o treinamento. Dessa forma, trata-se de uma nova subdivisão de métodos de subgradientes que abrangem conhecimento sobre os dados observados durante as interações para que seja possível a realização de um aprendizado mais informativo baseado em gradiente. Portanto, o método provém de aplicações estocástica e online que empregam funções proximais para o controle das etapas de gradiente.
- O otimizador Adamax é uma variação atualizada do método Adam, proposto por [Kingma and Ba 2017]. O algoritmo apresenta uma versão de gradiente descendente baseada na norma do infinito, eficaz para acelerar o processo de otimização.
- O método Nadam introduzido por [Dozat 2016] é uma extensão do algoritmo de otimização, Adam. A proposta do algoritmo é combinar os principais componentes do Adam, que seriam o momentum e a taxa de aprendizagem adaptativa, com a descida do gradiente de *momentum* de Nesterov. Logo, é possível melhorar a velocidade de convergência e a qualidade dos modelos.
- RMSprop foi introduzido por [Hinton et al. 2020] como uma forma para reduzir o empecilho da aprendizagem monotonicamente decrescente, desvantagem encontrada no método Adagrad. Desse modo, trata-se de um método de taxa de aprendizagem adaptativa que realiza a divisão da taxa de aprendizado pela média exponencialmente decrescente de gradientes quadrados.

3. Metodologia

Esta seção apresenta a metodologia adotada para realizar a comparação da influência dos métodos otimizadores no desempenho de uma CNN para a detecção do câncer de mama. Como linguagem de programação foi assumido Python e biblioteca principal TensorFlow.

3.1. Dataset

Os experimentos foram realizados por meio da utilização do conjunto de dados para a classificação de imagens histopatológicas do câncer de mama, BreakHis¹. O dataset possui amostras de tamanho igual a 700 X 460 pixels, na escala RGB de 3 canais, profundidade de 8 bits em cada canal e formato PNG. Além disso, as imagens estão separadas em dois grupos principais, tumores benignos e tumores malignos, na qual contém 2.480 amostras benignas e 5.429 amostras malignas e estão distribuídas por diferentes fatores de aumento (40x, 100x, 200x, 400x) [Spanhol et al. 2016b].

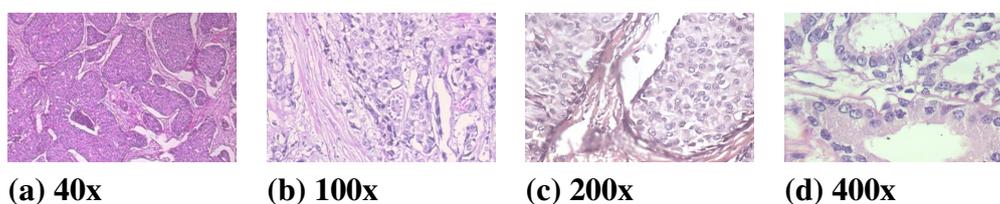


Figura 1. Imagens do grupo de tumor maligno de acordo com os seus fatores de aumento.

3.2. Pré-processamento

A rede VggNet16 possui dimensões de entrada pré-estabelecidas por padrão. Com isso, as imagens do dataset BreakHis foram redimensionadas para o tamanho 224x224 pixels nos 3 canais de cores RGB, em seguida, os valores de intensidade foram normalizados, resultando em valores entre 0 e 1.

3.3. Métricas de Avaliação

As métricas de avaliação são utilizadas com o intuito de analisar o desempenho dos modelos em dados desconhecidos. Os modelos utilizados para a análise das métricas de avaliação são os obtidos através do treinamento da rede VggNet16 com os otimizadores estudados neste trabalho.

As métricas de avaliação utilizadas neste trabalho são a sensibilidade (SEN), especificidade (ESP) e acurácia (ACC). A acurácia diz respeito a quantidade de previsões corretas do modelo, a sensibilidade é responsável por avaliar a capacidade do modelo na classificação das neoplasias malignas enquanto especificidade avalia a capacidade do modelo na classificação das neoplasias benignas.

¹Disponível em: <https://web.inf.ufpr.br/vri/databases/breast-cancer-histopathological-database-breakhis/>

4. Resultados

O propósito principal deste trabalho é utilizar a rede VggNet16 para realizar uma avaliação comparativa de diversas funções otimizadoras, com o intuito de verificar qual função possui o melhor desempenho na classificação das neoplasias benignas e malignas do câncer de mama.

O processo de avaliação consiste em treinar a rede diversas vezes, sendo que em cada treino uma função otimizadora diferente será utilizada. Para isso, foram utilizadas 300 épocas de treinamento para a rede, um lote de tamanho 32, com a função de perda *categorical cross entropy* e o valor da taxa de aprendizagem(lr) de cada função otimizadora foi escolhido empiricamente. Para cada função otimizadora foram realizados quatro testes, onde lr tinha o valor de 1×10^{-n} , sendo n o número do teste. Os valores escolhidos podem ser vistos na Tabela 1.

Tabela 1. Taxa de aprendizagem utilizada em cada função otimizadora.

	Adadelta	Adagrad	Adam	Adamax	Nadam	RMSprop
Taxa de Aprendizagem	0,1	0,01	0,0001	0,0001	0,0001	0,0001

As imagens do dataset foram divididas aleatoriamente em 70% para treino, 20% para validação e 10% para os testes. [Singh et al. 2021, Attallah et al. 2021] Além disso, para cada fator de aumento das imagens do dataset, foram gerados modelos de uma combinação de treino da rede VggNet16 com as funções otimizadoras abordadas neste trabalho.

As métricas de avaliação foram utilizadas com o intuito de verificar a eficácia de classificação de cada modelo gerado. Com isso, um modelo com boa taxa de sensibilidade possui alta capacidade de categorização das neoplasias malignas, da mesma forma que um modelo com boa taxa de especificidade possui alta capacidade de categorização das neoplasias benignas. Além disso, um modelo com alto índice de acurácia possui eficiência tanto na classificação das neoplasias benignas como também das neoplasias malignas.

Para as imagens com o fator de aumento de 40x, a função otimizadora RMSprop obteve a melhor taxa sensibilidade significando que essa função possui melhor eficácia para classificação das neoplasias malignas. Já o otimizador Adam alcançou melhores taxas de especificação e acurácia, chegando a alcançar 98,24% de especificidade, demonstrando ser a melhor função na classificação das neoplasias benignas e também a função que obteve o maior número de previsões corretas.

Com relação às imagens que contém o fator de aumento de 100x, as funções otimizadoras Adam e Nadam contiveram os melhores resultados na taxa de sensibilidade, demonstrando serem as melhores funções para classificação das neoplasias malignas. Além disso, a função Adadelta atingiu as melhores taxas de especificação e acurácia. Esta obteve uma acurácia de 96,63% e, por consequência, a com maior número de previsões corretas na classificação das neoplasias benignas.

Para as imagens com o fator de aumento 200x, a função otimizadora Adam obteve os melhores resultados em todas as métricas de avaliação, tendo uma acurácia de 97,02%. Com isso, ela se demonstrou a melhor função na classificação das neoplasias benignas e malignas, sendo também a função que obteve o maior número de previsões corretas.

Tabela 2. Métricas avaliativas dos modelos treinados utilizando as imagens com o fator de aumento de 40x e 100x

Otimizadores	40x			100x		
	SEN	ESP	ACC	SEN	ESP	ACC
Adam	97,20%	98,24%	97,5%	97,84%	92,75%	96,15%
Adadelta	93,00%	85,96%	91,00%	97,12%	95,65%	96,63%
Adagrad	95,10%	56,14%	84,00%	97,12%	71,01%	88,46%
Adamax	96,50%	87,19%	94,00%	97,84%	88,40%	94,71%
Nadam	93,00%	94,73%	93,5%	97,84%	92,75%	96,15%
RMSprop	97,90%	92,98%	96,5%	95,68%	89,85%	93,75%

Para as imagens com o fator de aumento de 400x, a função Adam atingiu a melhor porcentagem de sensibilidade, tendo uma taxa de sensibilidade de 98,37% e sendo por consequência a função mais eficaz na categorização das neoplasias malignas. A função otimizadora Adamax conteve as melhores taxas de especificação e acurácia, sendo a função com o maior número de previsões corretas e com a melhor capacidade classificatória das neoplasias benignas.

Tabela 3. Métricas avaliativas dos modelos treinados utilizando as imagens com o fator de aumento de 200x e 400x

Otimizadores	200x			400x		
	SEN	ESP	ACC	SEN	ESP	ACC
Adam	96,47%	98,33%	97,02%	98,37%	72,27%	91,20%
Adadelta	94,36%	95,00%	94,55%	95,93%	81,35%	91,20%
Adagrad	88,02%	90,0%	88,61%	97,56%	67,79%	87,91%
Adamax	92,95%	98,33%	94,55%	97,56%	84,74%	93,40%
Nadam	90,14%	88,33%	89,60%	95,12%	81,35%	90,65%
RMSprop	92,95%	93,33%	93,06%	96,74%	83,05%	92,30%

A função otimizadora Adagrad, em todos os fatores de aumento de imagens, possui as piores taxas de acerto nas previsões da classificação das neoplasias, se demonstrando pouco confiável para ser utilizada em conjunto com a arquitetura VggNet16 para resolução desse problema. Ademais, as funções otimizadoras RMSprop e Nadam, por mais que possuam resultados satisfatórios na predição correta das neoplasias, podem não ser suficientes para esse problema, uma vez que ambas têm resultados intermediários e não possuem as melhores acurácias em nenhum tipo de fator de aumento de imagem.

As funções Adamax e Adadelta possuem um bom número de previsões corretas na classificação das neoplasias. No entanto, para as imagens com o fator de aumento de 40x elas obtiveram taxas baixas de especificidade, podendo não ser confiável na classificação das neoplasias benignas desse fator de aumento. Além disso, a função otimizadora Adam é a que se demonstra mais confiável para as imagens de todos os fatores de aumento, uma vez que ela está sempre entre as três melhores taxas de acurácias.

5. Conclusão

Esse trabalho realizou uma análise comparativa do desempenho de seis funções otimizadoras aplicadas a rede VggNet16, com o objetivo de identificar as melhores funções para a classificação das neoplasias benignas e malignas do câncer de mama.

Entre todas as funções otimizadoras avaliadas, a função Adam é a que se demonstra mais confiável para realizar a classificação dessas neoplasias. Entre os benefícios dessa função, estão a fácil implementação e a eficiência computacional. Além disso, o que contribui para sua confiabilidade é que a base de dados utilizada possui imagens com uma larga escala de pixels e essa função é muito adequada para o treinamento de bases com grandes dados e com muitos parâmetros [Kingma and Ba 2017].

Com a análise abordada nesta pesquisa, fica claro que é de suma importância a aplicação de testes com o intuito de definir os melhores parâmetros para serem aplicados em uma arquitetura de rede neural. Uma vez que algumas funções otimizadoras demonstram ser mais confiáveis e com resultados mais promissores para serem utilizados na classificação das neoplasias.

Para pesquisas futuras, pretende-se analisar outras funções otimizadoras, com o intuito de definir qual a melhor função para ser utilizada para a classificação dessas neoplasias. Além disso, existe a intenção de analisar métodos de aumento de dados, como as Redes Adversárias Generativas, com o objetivo de aumentar as imagens do dataset utilizado e por consequência melhorar a acurácia dos modelos.

Referências

- Alanazi SA, Kamruzzaman MM, I. S. M. A. M. A. Y. A. N. S. M. (2021). 3 boosting breast cancer detection using convolutional neural network. In Magnenat-Thalmann, N. and Thalmann, D., editors, *J Healthc Eng*. John Wiley & Sons Ltd.
- American Cancer Society (2019). What is breast cancer? Disponível em: <https://www.cancer.org/cancer/breast-cancer/about/what-is-breast-cancer.html>. Acessado em: 11 jul. 2021.
- Attallah, O., Anwar, F., Ghanem, N., and Ismail, M. (2021). Histo-cadx: duo cascaded fusion stages for breast cancer diagnosis from histopathological images. *PeerJ Computer Science*, 7:e493.
- Dozat, T. (2016). Incorporating Nesterov Momentum into Adam. In *Proceedings of the 4th International Conference on Learning Representations*, San Juan, Puerto Rico.
- Duchi, J., Hazan, E., and Singer, Y. (2011). Adaptive subgradient methods for online learning and stochastic optimization. *J. Mach. Learn. Res.*, 12(null):2121–2159.
- Fukushima, K. (1980). Neocognitron: a self organizing neural network model for a mechanism of pattern recognition unaffected by shift in position. *Biological cybernetics*, 36(4):193—202.
- Gayathri, B. M. and Sumathi, C. P. (2016). Comparative study of relevance vector machine with various machine learning techniques used for detecting breast cancer. In *2016 IEEE International Conference on Computational Intelligence and Computing Research (ICCIC)*, pages 1–5.

- Hinton, G., Srivastava, N., and Swersky, K. (2020). Neural networks for machine learning, lecture 6a overview of mini-batch gradient descent. Disponível em: <http://www.cs.toronto.edu/~hinton/coursera/lecture6/lec6.pdf>. Acessado em: 12 jul. 2021.
- Horie, Y., Yoshio, T., Aoyama, K., Yoshimizu, S., Horiuchi, Y., Ishiyama, A., Hirasawa, T., Tuchida, T., Ozawa, T., Ishihara, S., Kumagai, Y., Fujishiro, M., Maetani, I., Fujisaki, J., and Tada, T. (2018). The diagnostic outcomes of esophageal cancer by artificial intelligence using convolutional neural networks. *Gastrointestinal Endoscopy*, 89.
- Kandel, I., Castelli, M., and Popovič, A. (2020). Comparative study of first order optimizers for image classification using convolutional neural networks on histopathology images. *Journal of Imaging*, 6(9).
- Kingma, D. P. and Ba, J. (2017). Adam: A method for stochastic optimization.
- Osareh, A. and Shadgar, B. (2011). A computer aided diagnosis system for breast cancer.
- Ribeiro, A. M. and Junior, F. P. S. A. (2020). Um estudo comparativo entre cinco métodos de otimização aplicados em uma RNC voltada ao diagnóstico do glaucoma. *Revista de Sistemas e Computação*, 10(1):122–130.
- Simonyan, K. and Zisserman, A. (2015). Very deep convolutional networks for large-scale image recognition.
- Singh, V., Pencina, M., Einstein, A. J., Liang, J. X., Berman, D. S., and Slomka, P. (2021). Impact of train/test sample regimen on performance estimate stability of machine learning in cardiovascular imaging. *Scientific Reports*, 11.
- 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.
- Tschandl, P., Rosendahl, C., Akay, B., Argenziano, G., Blum, A., Braun, R., Cabo, H., Gourhant, J.-Y., Kreusch, J., Lallas, A., Lapins, J., Marghoob, A., Menzies, S., Neuber, N., Paoli, J., Rabinovitz, H., Rinner, C., Scope, A., Soyer, P., and Kittler, H. (2018). Expert-level diagnosis of nonpigmented skin cancer by combined convolutional neural networks. *JAMA Dermatology*, 155.
- Vargas, Paes, and Vasconcelos (2016). Um estudo sobre redes neurais convolucionais e sua aplicação em detecção de pedestres. In *CONFERENCE ON GRAPHICS, PATTERNS AND IMAGES*, volume 29.
- Yaqub, M., Feng, J., Zia, M. S., Arshid, K., Jia, K., Rehman, Z. U., and Mehmood, A. (2020). State-of-the-art CNN optimizer for brain tumor segmentation in magnetic resonance images. *Brain Sciences*, 10(7).
- Zeiler, M. D. (2012). Adadelta: An adaptive learning rate method.