

# Classificação da densidade mamária em mamografias utilizando redes neurais convolucionais

Marcos Felipe Vendramini Carvalho  
Departamento de Ciência da Computação, PUC Minas  
Belo Horizonte - MG  
Email: marcos.vendramini@sga.pucminas.br

Alexei Manso Correa Machado  
Departamento de Ciência da Computação, PUC Minas  
Faculdade de Medicina, UFMG  
Belo Horizonte - MG  
Email: alexeimcmachado@gmail.com

**Resumo**—Neste estudo, avaliou-se o potencial das redes neurais convolucionais na classificação de texturas para diagnóstico de câncer de mama na escala BI-RADS de quatro níveis. A base de dados foi constituída de 5024 exames mamográficos, com recortes de 128x128 pixels. As escalas foram avaliadas em dois conjuntos, o primeiro agrupando as escalas não-densas e densas, e o segundo avaliando os níveis individualmente. Os métodos apresentaram acurácia de 89% e 70%, para o primeiro e o segundo conjuntos, respectivamente, mostrando-se competitivos com trabalhos da literatura.

**Abstract**—In this study, the potential of convolutional neural networks in the classification of textures for the diagnosis of breast cancer in the four-level BI-RADS scale was evaluated. The database consisted of 5024 mammographic exams, with 128x128 pixel patches. The scales were evaluated in two sets, the first grouping the non-dense and dense scales, and the second scaling the four levels individually. The methods were shown to be competitive, with accuracy of 89% and 70%, for the first and second sets, respectively.

## I. INTRODUÇÃO

O câncer de mama representa 29,5% dos casos de câncer em mulheres no Brasil e, em 2018, foi responsável por 15.593 mortes [1]. Sabe-se que alta taxa de mortalidade do câncer de mama está associada com diagnóstico tardio e consequente atraso no tratamento, o que justifica a realização de exames em mulheres a partir de 50 anos de idade, a cada dois anos [1]. A mamografia continua a ser a mais importante técnica de imagem para as mamas, onde é possível localizar nódulos e calcificações que representam risco da doença. A escala mais usada para classificação da densidade mamária é a BI-RADS (*Breast Imaging Reporting and Data System*) [2]. Ela propõe uma classificação qualitativa em quatro grupos, baseado na proporção entre tecido gorduroso e fibroglandular: BI-RADS I (CI), predominantemente gordurosa; BI-RADS II (CII), contendo áreas dispersas com tecido fibroglandular; BI-RADS III (CIII), heterogeneamente denso; e BI-RADS IV (CIV), extremamente denso (Fig. 1) [3]. A classificação das densidades é de extrema importância, uma vez que mamas das classes CIII e CIV podem esconder pequenas lesões e levar à detecção tardia da doença [4].

O principal desafio para a classificação automática de densidades é que as mamas podem ser compostas de tecidos heterogêneos. A Fig. 2 mostra exames nas quatro escalas de densidade que apresentam texturas semelhantes. Os exemplos mostrados não são considerados exceções. Na verdade, um

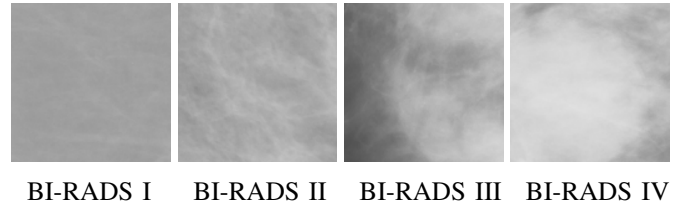


Figura 1. Exemplos de texturas mamárias típicas para cada escala.



Figura 2. Exemplos de texturas mamárias semelhantes com diferentes escalas.

banco de dados pode ser muito heterogêneo, dificultando a tarefa de classificação para sistemas de diagnóstico automático, pois os limites de cada classe não são claros. A dificuldade de se obter resultado consistente na classificação de uma mesma imagem por diferentes médicos, faz desse problema um potencial alvo para estudo com auxílio de técnicas de aprendizado de máquina [5].

O objetivo deste estudo é propor métodos de classificação automática da densidade mamária pela análise de imagens de raio X usando redes neurais convolucionais (CNN). Essa classificação é de extrema importância na prevenção do câncer, pois a categoria diagnosticada determina a recomendação médica quanto à necessidade de exames complementares. Embora o estudo esteja voltado para uma solução de um problema médico, o objetivo secundário é testar a capacidade das CNN em diferenciar texturas complexas, principalmente aquelas cujos elementos constituintes são de difícil percepção pelo olho humano [6].

## II. TRABALHOS RELACIONADOS

Métodos baseados em aprendizado de máquina têm apresentado grande eficácia na classificação de imagens, como o estudo proposto por Mohamed et al. [7] que demonstrou o uso de CNN com base no modelo AlexNet [8] para distinção das

categorias de densidade densa e não densa. Utilizando uma extensa base de dados privada, composta de 15415 imagens completas, o método obteve acurácia de 92%.

O estudo proposto por Rungue [9] aplicou máquinas de vetores de suporte (SVM) na classificação de densidades mamária nos quatro níveis BI-RADS utilizando diversos descritores de textura. A classificação foi feita usando recortes dos exames de uma base pública de 5024 imagens [10]. O estudo mostrou que as classes intermediárias (CII e CIII) são as mais complexas para serem classificadas.

Outro estudo mais recente, proposto por Lehman et al. [11], apresenta um método de classificação de densidade mamária com CNNs, utilizando como base a rede ResNet-18 [12]. No processo, foi feita a classificação em dois e quatro grupos, usando 58894 imagens completas de um banco de dados privado não balanceado onde 75% das imagens estão concentrados nas classes que apresentam as maiores taxas de acerto.

Os trabalhos citados acima que utilizam CNNs, classificam as densidades com base em imagens de mamografias completas. A maior parte deles adota a classificação binária que é mais simples que a realizada em quatro níveis. O presente estudo difere-se dos demais pois utiliza CNNs para classificar somente um recorte do exame, onde apenas a textura está representada. Outro diferencial do trabalho é utilização de uma base pública [10], enquanto a maioria dos trabalhos da literatura se baseiam em imagens de acesso restrito, o que dificulta a comparação e reprodução dos resultados.

### III. MATERIAIS E MÉTODOS

O conjunto de dados utilizado no estudo consiste em 5024 imagens de mamografias gentilmente oferecidas pelo projeto IRMA da Universidade de Tecnologia Aachen e publicado por Oliveira et al. [10]. As imagens estão distribuídas de forma balanceada entre as classes com 1256 amostras para cada uma, das quais foram extraídas regiões de interesse de 128x128 *pixels*, armazenadas no formato PNG (*Portable Network Graphics*). As regiões de interesse extraídas representam a parte mais significativa do tecido das mamas, removendo-se artefatos como anotações e marcas presentes nas imagens. Para aumentar o número de imagens foi realizada a inversão horizontal, vertical e rotação das imagens para a geração de novas entradas com o objetivo de redução de *overfitting*. As imagens foram separadas em três grupos, mantidas as proporções das classes: treinamento, teste e validação, na proporção 80%, 5% e 15% da base, respectivamente.

Foram implementados os modelos de redes residuais (ResNet), AlexNet e VGG. O modelo Residual não permitiu a diferenciação entre as classes, uma vez que não houve convergência durante o treino. O AlexNet obteve um resultado próximo ao modelo VGG. Porém, por possuir camadas com filtros maiores, o AlexNet apresentou maior custo de processamento. Como as imagens usadas são pequenas, a VGG mostrou ser a melhor arquitetura para o problema, evitando que as imagens tivessem seu tamanho muito reduzido ao passarem através das camadas da rede. Dessa forma, a rede

VGG-16 foi usada como a base do classificador e a ela foram adicionadas normalização de *batch* após as camadas convolucionais e *dropout* de 50% após cada camada densa. Todas as camadas convolucionais apresentam janelas de 3x3 *pixels*, com espaçamento entre janelas de 3 *pixels*, borda de 1 *pixel* e a função de ativação ReLU. As camadas densas intermediárias utilizaram as funções de ativação, ReLU, e a última camada responsável pela classificação utilizou função sigmoide para a classificação binária e *softMax* na multiclasse.

Adicionalmente, foi utilizada a técnica de extração de características inserindo-se, no início do modelo, outra rede VGG-16 previamente treinada com a base de dados ImageNet [8]. Essa rede foi inicializada com os pesos e valores fixos, para que estes não mudassem durante o treinamento. A saída da rede foi retirada após a última camada de *pooling*, uma vez que as camadas seguintes são densas e não possuem como saída tensores 3D, e sim vetores. Essa saída foi direcionada ao restante da rede convolucional responsável pelo treinamento específico do problema. Os modelos AlexNet e ResNet também foram testados como redes extratoras, mas não apresentaram resultados superiores à VGG16. A topologia da rede utilizada pode ser visualizada na Fig. 3.

Com relação aos hiperparâmetros, foram testados 100 a 400 épocas, e 16, 32 e 64 imagens por lote. Em todos os casos, 400 épocas e 32 imagens por lote proporcionaram resultados melhores e, por isso, esse padrão também foi seguido durante todo o processo. Em todos os testes foram usados entropia cruzada como função de perda e RMSProp como otimizador.

Uma vez que o modelo de rede proposto apresentou resultados superiores para a classificação binária, foi idealizada a construção de uma cascata de redes para a classificação dos quatro níveis BI-RADS. O modelo cascata consiste em fazer a concatenação de redes de classificação binária para resolver um problema multiclasse. Para o desenvolvimento desse modelo, foram implementadas todas as possibilidades de redes binárias, num total de 23 configurações, sendo as mais eficazes representadas na Fig. 4.

Todos os algoritmos foram implementados usando a plataforma *Anaconda* versão 4.5.4 com *Python* 3.6 e a biblioteca *Keras* 2.2.4. Para o processamento da rede foi utilizado um computador *Windows* 10 Pro 64x, Intel Core i7-2600 CPU 3.40GHz, com 16GB RAM e *Nvidia GeForce GTX* 1050 TI 4GB. Cada rede levou entre quatro e oito horas para ser treinada, com processamento paralelo em GPU usando o CUDA.

### IV. RESULTADOS

A partir dos modelos desenvolvidos, foram escolhidas quatro configurações, sendo um classificador binário entre mamas densas e não densas, um classificador multiclasse de CNN única e os dois modelos cascatas representados na Fig. 4.

O classificador binário apresentou acurácia de 89,1%, com precisão de 0.92 para o tecido não denso e 0.87 para denso, e sensibilidade de 0.85 para tecido não denso e 0.93 para denso.

O modelo multiclasse de rede única apresentou acurácia de 56,2% e baixa taxa de acerto na diferenciação de classes

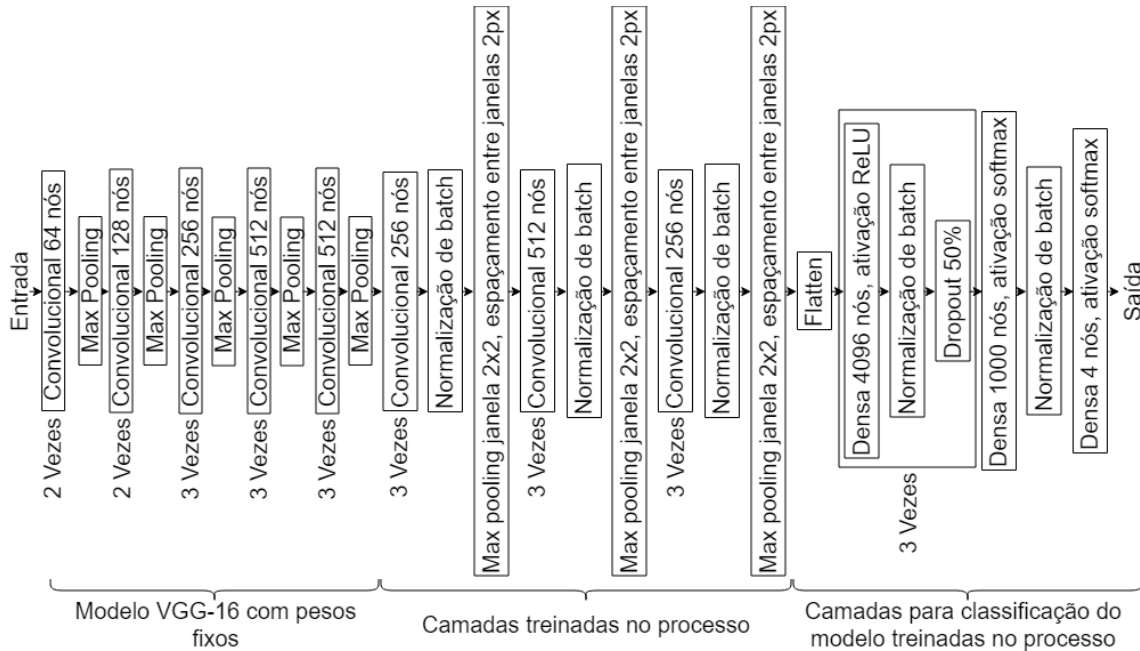


Figura 3. Rede usada na classificação multiclasse.

intermediárias (CII e CIII), embora os níveis CI e CIV tenham sido classificados corretamente na maioria dos casos. Dessa forma, pode-se concluir que este modelo não foi o ideal para a solução do problema.

Usando a estrutura em cascata do Modelo 1, dentre as possíveis combinações de redes, o melhor resultado obteve acurácia de 66,2%. Para alcançar esse resultado, na Rede 1, foi feita a classificação entre imagens de classe CI e as demais classes. Não sendo classificada como CI, a imagem foi usada como entrada para a Rede 2, onde foi feita a classificação entre imagens de classe CII e as restantes. Por último, na Rede 3, a imagem é classificada entre CIII e CIV. Considerando-se os resultados obtidos nesse modelo, é possível observar que o uso de múltiplas classificações binárias para resolver um problema multiclasse pode tornar a classificação mais precisa.

Para o classificador em cascata do Modelo 2, dentre as possíveis combinações de redes, o melhor resultado obteve acurácia de 69,8%. Para isso, na Rede 1 foi feita a classificação entre tecido denso e não denso. As imagens classificadas como não densas foram usadas como entrada para a Rede 2 e as densas como entrada para a 3. Na Rede 2, foi feita a

classificação entre CI e CII e na 3, foi feita a classificação entre CIII e CIV. Considerando-se os resultados obtidos no segundo modelo cascata, pode-se observar uma melhora significativa em relação aos modelos anteriores, embora o nível CII ainda apresentasse uma taxa de erro maior quando comparado aos obtidos nas demais classes.

## V. DISCUSSÃO

Os resultados obtidos no presente trabalho foram comparados com três abordagens. A primeira, proposta por Rungue [9], utiliza SVM e diversos descritores de textura. Os experimentos foram realizados com a mesma base de imagens pública usada neste trabalho. O segundo, proposto por Mohamed et al. [7], apresentou resultados para classificação binária em imagens nas posições mediolateral oblíqua (MLO) e craniocaudal (CC). O terceiro, proposto por Lehman et al. [11], apresentou resultados para classificação binária e multiclasse para o problema, avaliadas usando-se imagens de teste e em implantação clínica, apenas na posição MLO.

Mohamed et al. e Lehman et al. apresentam soluções semelhantes, também envolvendo CNNs, porém ambos diferem-se do presente trabalho, uma vez que utilizam imagens completas de mamografias, enquanto este usa apenas regiões de interesse. As bases privadas possuem três e oito vezes mais imagens, respectivamente, enquanto o método proposto neste artigo usa uma base pública [10].

Para a classificação binária o melhor resultado obtido foi a implementação clínica de Lehman et al., com acurácia de 94%, seguido pelo classificador de Mohamed et al., com acurácia de 92%, como pode ser observado na Fig. 5. A comparação de acurácia é dificultada pelo fato dos 2 trabalhos terem utilizado bases de imagens privadas, além de analisarem todo

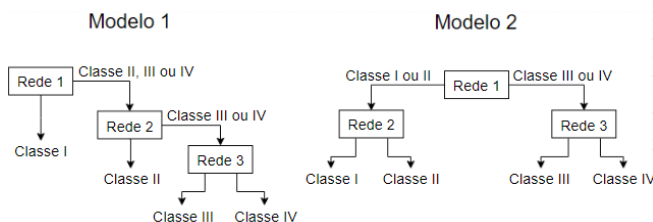


Figura 4. Modelos cascata.

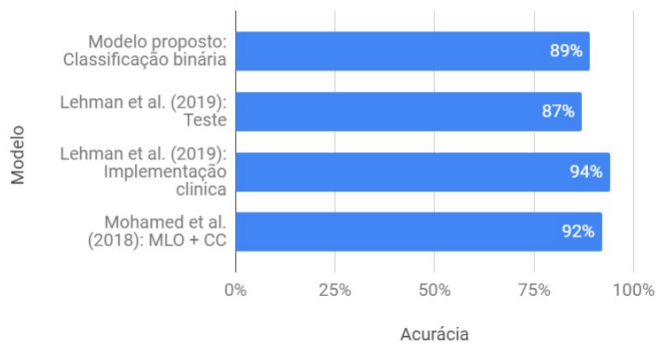


Figura 5. Comparação dos resultados para a classificação binária.

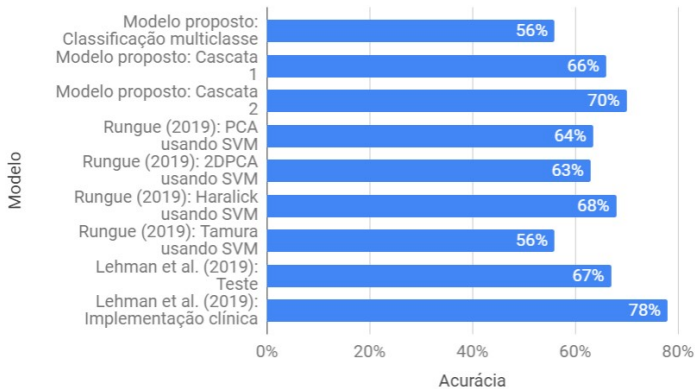


Figura 6. Comparação dos resultados para a classificação multiclasse.

o contexto da mamografia. No caso da implementação clínica de Lehman et al., o número de pacientes com cada nível BI-RADS estava desbalanceado, o que pode ter favorecido uma maior taxa de acerto. De fato, o índice de acurácia verificado nos testes balanceados ficou 7 pontos percentuais menor que a implementação clínica, valor esse abaixo do obtido pelo método proposto neste trabalho.

Na Fig. 6 é possível comparar os melhores resultados para classificação multiclasse. Rungue propôs soluções com descritores de textura PCA, DPCA, Haralick e Tamura. Sua classificação utilizando Haralick e SMV atingiu a acurácia de 68%. Lehman et al. apresentaram a melhor acurácia dentre os resultados obtidos, atingindo o valor de 78%, mas apenas na implementação clínica com base desbalanceada, possuindo 732 CI, 5702 CII, 4039 CIII e 290 CIV. Dessa forma, a rede gerou acerto de 91% e 94% nos níveis II e III, e 69% e 72% nos níveis I e IV, respectivamente, ou seja, a rede acertou mais nas imagens mais frequentes. Nos testes balanceados, os resultados de Lehman et al. ficaram em apenas 67%. Somente Rungue apresenta o mesmo cenário que o utilizado no presente trabalho, com 68% de acurácia. Dessa forma, com a acurácia de 70%, o modelo proposto apresentou uma melhora nos resultados disponíveis na literatura.

## VI. CONCLUSÃO

Este trabalho propôs uma solução utilizando cascatas de redes neurais convolucionais para o problema de classificação

de texturas mamográficas. Após a análise de diversas estruturas de redes e técnicas de extração de características, o método alcançou acurácia de 89% para classificação binária e 70% para classificação multiclasse. O classificador binário, apesar de não possuir a melhor acurácia, mostrou-se competitivo em relação aos demais modelos, mesmo utilizando imagens menores, fator que reduz o tempo de processamento e as limitações de hardware. O modelo cascata multiclasse, no entanto, apresentou um ganho de acurácia quando comparado aos demais métodos da literatura, além de estar baseado em testes com dados públicos.

Os resultados publicados para o problema parecem modestos para classificações utilizando CNNs mas, até o momento, os modelos existentes não resolvem de forma eficaz o problema de classificação de texturas mamográficas, principalmente quando se trabalha somente com regiões de interesse. Apesar disso, eles se mostraram promissores para estudos de classificação de densidade de imagens que podem auxiliar na detecção do câncer de mama.

A partir dos resultados obtidos foi possível entender a importância da correta parametrização da rede, assim como a utilização combinada de métodos para solução do problema. Dentre os métodos analisados, a extração de características se mostrou eficaz para diferenciar imagens semelhantes em classes diferentes, e o modelo cascata provou conseguir ganho de acurácia quando comparado a redes únicas.

## REFERÊNCIAS

- [1] INCA, "Instituto nacional de câncer José Alencar Gomes da Silva," 2018, acesso em: 02/03/2019. [Online]. Available: <https://www.inca.gov.br>
- [2] C. D'Orsi and Acr, *2013 ACR BI-RADS Atlas: Breast Imaging Reporting and Data System*. American College of Radiology, 2014.
- [3] D. Spak, J. Plaxco, L. Santiago, M. Dryden, and B. Dogan, "Bi-rads® fifth edition: A summary of changes," *Diagnostic and Interventional Imaging*, vol. 98, 01 2017.
- [4] J. N. Wolfe, "Breast patterns as an index of risk for developing breast cancer," *Am J Roentgenology*, vol. 126, pp. 1130–7, 1976.
- [5] M. Muštra, P. M. Grgić, and K. Delač, "Breast density classification using multiple feature selection," *Automatika*, vol. 53, no. 4, pp. 362–372, 2012.
- [6] F. Chollet, *Deep Learning with Python*. New York: Manning Publications Co., 2018.
- [7] A. A. Mohamed, Y. Luo, H. Peng, R. C. Jankowitz, and S. Wu, "Understanding clinical mammographic breast density assessment: a deep learning perspective," *Journal of Digital Imaging*, vol. 31, no. 4, pp. 387–392, 2018.
- [8] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, A. C. Berg, and L. Fei-Fei, "ImageNet Large Scale Visual Recognition Challenge," *International Journal of Computer Vision (IJCV)*, vol. 115, no. 3, pp. 211–252, 2015.
- [9] H. A. Rungue, "Classificação de imagens mamográficas na escala de densidade bi-rads utilizando descritores de textura," Master's thesis, Pontifícia Universidade Católica de Minas Gerais, 2018.
- [10] J. E. Oliveira, M. O. Gueld, A. d. A. Araújo, B. Ott, and T. M. Deserno, "Toward a standard reference database for computer-aided mammography," in *Medical Imaging 2008: Computer-Aided Diagnosis*, vol. 6915, 2008, p. 69151Y.
- [11] C. D. Lehman, A. Yala, T. Schuster, B. Dontchos, M. Bahl, K. Swanson, and R. Barzilay, "Mammographic breast density assessment using deep learning: Clinical implementation," *Radiology*, vol. 290, no. 1, pp. 52–58, 2019.
- [12] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *2016 IEEE CVPR*, 2016, pp. 770–778.