

Classificação de lesões mamárias em imagens de exames histopatológicos utilizando *Deep Learning Features* e *Support Vector Machines*

José Igor de Carvalho¹, Igor Rafael Silva Valente¹

¹Instituto Federal de Educação, Ciência e Tecnologia do Ceará (IFCE)
Eixo Tecnológico da Computação

jose.igor.carvalho05@aluno.ifce.edu.br, igor@ifce.edu.br

Abstract. *Breast cancer is one of the most common forms of the disease in the world. It is estimated that about 8% of the female population will be affected at some point in their lives by this pathology. Despite all advances in the treatment of this disease, early diagnosis remains essential for the effectiveness of any applied technique. Allied to this purpose, several techniques and algorithms have been developed in the form of CADx (Computer Aided Diagnosis) systems, whose main objective is to help in the process of diagnosing the disease. In this context, the present work aims to present a methodology based on Deep Learning Features and Support Vector Machines for automatic classification of breast lesions in images of histopathological exams. When applied to the BreakHis image base, the methodology proved to be promising, with an accuracy of 97.3%.*

Resumo. *O câncer de mama é uma das formas mais comuns da doença no mundo. Estima-se que cerca de 8% da população feminina será afetada em algum ponto de suas vidas por esta patologia. Apesar de todos os avanços para o tratamento desta doença, o diagnóstico precoce continua sendo fundamental para a eficácia de qualquer técnica aplicada. Aliado a este propósito, diversas técnicas e algoritmos têm sido desenvolvidos na forma de sistemas CADx (Computer Aided Diagnosis), os quais têm como principal objetivo auxiliar no processo de diagnóstico da doença. Neste contexto, o presente trabalho visa apresentar uma metodologia baseada em Deep Learning Features e Support Vector Machines para classificação automática de lesões mamárias em imagens de exames histopatológicos. Quando aplicado à base de imagens BreakHis, a metodologia mostrou-se promissora, com acurácia de 97.3%.*

1. Introdução

O câncer de mama é uma das formas mais comuns da doença no mundo. Essa doença é caracterizada pela multiplicação desordenada de células da mama que, por sua vez, geram tumores que podem ser classificados quanto à sua malignidade. Estima-se que cerca de 8% da população feminina será afetada em algum ponto de suas vidas pela doença, o que a torna, assim, o segundo tipo mais frequente da patologia em mulheres, atrás apenas do câncer de pele [Amrane et al. 2018, Spanhol et al. 2016].

Em geral, o diagnóstico do câncer de mama é realizado a partir da análise feita por especialistas em imagens de exames histopatológicos. Durante este processo,

transformações como rotações e redimensionamentos podem ser aplicadas às imagens dos exames, provocando um aumento na complexidade da tarefa. Por esse motivo, mesmo quando feita por especialistas experientes, a análise histopatológica é suscetível a erros [Zhou et al. 2022].

Ao longo dos anos foram desenvolvidas técnicas e ferramentas para auxiliar especialistas na tarefa de detectar e diagnosticar doenças em exames de imagens. Dentre esse conjunto de ferramentas temos os sistemas CADe (*Computer Aided Detection*), aplicados à detecção, e CADx (*Computer Aided Diagnosis*), voltados ao diagnóstico de patologias. Estes sistemas podem empregar diferentes métodos e visam tornar mais automática a segmentação e classificação de regiões de interesse. Além disso, esses sistemas também são capazes de detectar patologias nos estágios mais iniciais, o que é de grande valor para os casos de câncer de mama [e Flávio Araújo 2022, Firmino et al. 2016].

Durante os últimos 40 anos diversas pesquisas vêm sendo realizadas com o intuito de detectar e diagnosticar o câncer e, apesar da complexidade da tarefa, resultados promissores foram obtidos [Spanhol et al. 2016]. É possível encontrar pesquisas para detecção e classificação de câncer em que são aplicadas técnicas de processamento digital de imagens, aprendizagem de máquina, entre outras. Além disso, bases de dados públicas com imagens de raios X, exames histopatológicos e ressonâncias magnéticas podem ser facilmente acessadas, o que é muito importante para o desenvolvimento e aprimoramento de metodologias.

Com o intuito de contribuir com as pesquisas desta área, neste trabalho é proposta uma metodologia que combina características extraídas por redes neurais convolucionais (CNNs), ou *Deep Learning Features*, e um classificador baseado em *Support Vector Machines* (SVM), que visa auxiliar na tarefa de classificar lesões mamárias em imagens de exames histopatológicos quanto à sua malignidade. Durante a etapa de extração de características foram testadas algumas arquiteturas de CNNs, com validação da performance realizada na base de exames BreakHis, inicialmente apresentada por [Spanhol et al. 2015].

O restante deste trabalho está organizado da seguinte forma: na Seção 2 temos um levantamento dos principais trabalhos encontrados na literatura nesta linha de pesquisa; na Seção 3 são descritas as etapas nas quais a metodologia desenvolvida neste trabalho foi dividida; na Seção 4 são discutidos os resultados alcançados pela metodologia proposta e, por fim, na Seção 5 são apresentadas as principais conclusões deste trabalho assim como melhorias a serem aplicadas em implementações futuras.

2. Trabalhos Relacionados

Diversos estudos vêm sendo realizados com o intuito de aperfeiçoar ferramentas CADx para auxiliar no diagnóstico automático do câncer de mama [Spanhol et al. 2016]. Esta seção contém um resumo dos principais trabalhos encontrados na literatura descrevendo, de forma sucinta, a metodologia aplicada em cada um deles bem como os resultados alcançados.

O trabalho de Ataky e Lameiras Koerich (2022) apresenta um novo descritor de textura baseado em medições de biodiversidade e distinção taxonômica, para extração de características de imagens de exames histopatológicos presentes na base BreakHis. Pos-

teriormente, as informações extraídas foram usadas para treinar um classificador SVM, atingindo acurácia de 97.3%.

O método proposto por Bisneto e Araújo (2022) combina *Deep Learning Features* e características de textura bioinspiradas para realizar a classificação de lesões mamárias. Na fase de extração das *Deep Learning Features*, foram testadas diversas arquiteturas de redes neurais convolucionais e as características extraídas foram usadas para treinar o classificador *Random Forest*. Além disso, foram coletados resultados de um subconjunto da base BreakHis, que correspondem às imagens com ampliação de $40\times$, nas quais os autores afirmam que obtiveram acurácia de 92.9%.

O trabalho apresentado por Spanhol (2016) faz uso da base de imagens BreakHis e propõe a utilização de uma CNN AlexNet, inicialmente apresentada por Krizhevsky (2013), para auxiliar a classificação de lesões na mama com acurácia de 90% no melhor cenário.

No trabalho de Alom (2019) é proposto um modelo que combina *Residual Networks* (Resnet) e *Inception Networks* para classificação de lesões mamárias de imagens de exames histopatológicos. O autor classifica o modelo como *Inception Recurrent Residual Convolutional Neural Network* (IRRCNN) e usa algumas bases de imagens para treiná-lo, incluindo a base BreakHis, atingindo acurácia de 97.0%.

O artigo de Zhou (2022) propõe uma metodologia que faz uso de *Resolution Adaptive Networks* (RANet), *Speed-up Robust Features* (SURF) e SVM para realizar a classificação de lesões mamárias em imagens histopatológicas. Neste trabalho foram usadas as bases de imagens BreakHis e BACH 2018, atingindo acurácias de 99.14% e 99.25% respectivamente, nos melhores casos.

O trabalho de Erfankhah (2019) propõe uma adaptação do descritor *Local Binary Patterns* (LBP) para extração de características de imagens de exames histopatológicos para classificação de lesões mamárias. Para coletar resultados, o autor utiliza um classificador SVM. Além disso, são usadas algumas bases de imagens para treinar o modelo, incluindo a BreakHis, na qual foi obtida acurácia de 88.32%.

Como visto, os estudos citados acima fazem uso de métodos que variam desde algoritmos mais tradicionais de processamento digital de imagens até técnicas mais avançadas, que combinam aprendizagem de máquina, algoritmos de visão computacional, adaptação de descritores de textura já existentes e até mesmo criação de novos descritores e algoritmos. Também é possível notar que os trabalhos com melhores resultados empregaram metodologias divididas em diversas etapas de pré-processamento, treino e classificação.

A metodologia aqui proposta, por sua vez, visa validar o uso de características extraídas unicamente por CNNs, sem qualquer treinamento e usando transferência de aprendizado a partir da base *Imagenet*, utilizando um classificador SVM, quando aplicada à base BreakHis.

3. Metodologia

A metodologia deste trabalho está dividida em três etapas: pré-processamento de imagens, extração de características e classificação. Na Figura 1 é possível visualizar o fluxograma da metodologia proposta.

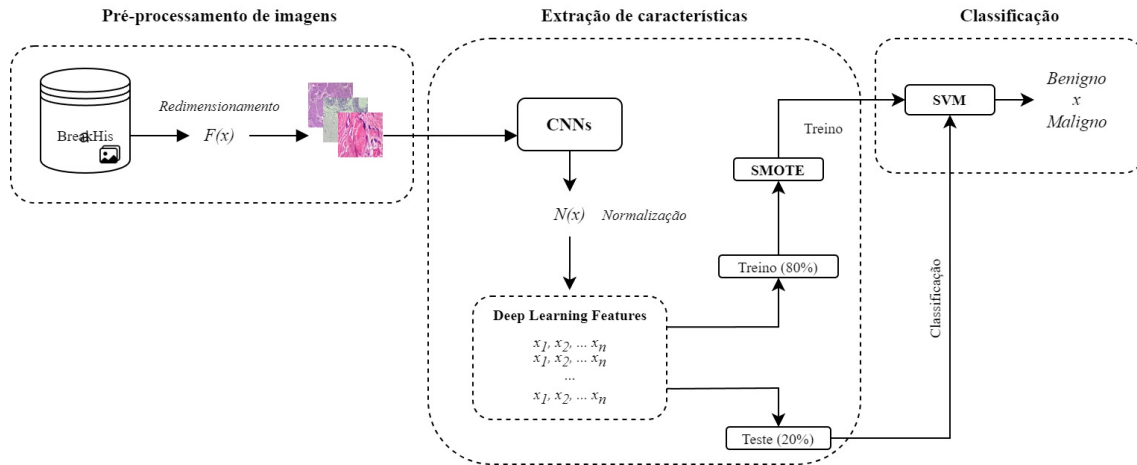


Figura 1. Diagrama da metodologia proposta.

3.1. Pré-Processamento de Imagens

Para realizar validação da metodologia proposta, foi utilizada a base pública de imagens BreakHis. Esta base contém um total de 7.909 imagens de exames histopatológicos que podem ser divididas em dois grandes subconjuntos considerando sua malignidade. Além disso, esta base também pode ser dividida de acordo com a magnificação de seus representantes, possuindo imagens com ampliações de $40\times$, $100\times$, $200\times$ e $400\times$ [Spanhol et al. 2015]. A Tabela 1 mostra o detalhamento da divisão de imagens na base BreakHis.

Originalmente as imagens da base utilizada possuem dimensão de 700×460 pixels, utilizando o modelo RGB com 8 bits em cada canal. No entanto, para este trabalho, as imagens foram redimensionadas como parte de uma etapa de pré-processamento, para que estas pudessem se adaptar as dimensões padrões das CNNs utilizadas para extração de características. Além disso, trabalhos como o apresentado por Bisneto e Araújo (2022), explora os resultados obtidos apenas para o subconjunto de imagens com ampliação de $40\times$. Este trabalho, por sua vez, utilizada toda a base para treinar o classificador, discutindo os resultados alcançados para cada ampliação.

Tabela 1. Detalhamento do número de imagens na base BreakHis.

	$40\times$	$100\times$	$200\times$	$400\times$
Benigno	652	644	623	588
Maligno	1370	1437	1390	1232

3.2. Extração de Características

Para a extração de características das imagens foram utilizadas redes neurais convolucionais. O vetor de informações extraído de cada imagem é conhecido como *Deep Learning Features* e, na metodologia aqui apresentada, foi obtido acessando a saída da penúltima camada da rede a ser usada, que está imediatamente anterior a camada de predição e possui um vetor com dimensão $1 \times n$, onde n é quantidade de características extraídas da imagem submetida. Esta técnica pode ser encontrada em trabalhos como o de Bisneto e Araújo (2022).

Neste trabalho foram testadas as seguintes arquiteturas: DenseNET-169 e DenseNET-201 [Huang et al. 2017], MobileNET e MobileNET V2 [Howard et al. 2017], ResNET-50, ResNET-101 e ResNET-152 [He et al. 2016], VGG16 e VGG19 [Simonyan e Zisserman 2014]. A Tabela 2 descreve as redes elencadas acima junto com o tamanho do vetor de características extraído da penúltima camada de cada uma delas e dimensão padrão de imagem. Vale ressaltar que todas as arquiteturas utilizadas neste trabalho estão implementadas na biblioteca Keras, disponível para a linguagem Python [Nikhil e Kekar 2017].

Tabela 2. Dimensão do vetor de características e dimensão padrão de imagem.

Arquitetura	Quantidade de características	Dimensão padrão de imagem
DenseNET-169	1664	224, 224, 3
DenseNET-201	1920	224, 224, 3
MobileNET	1000	224, 224, 3
MobileNET V2	1280	224, 224, 3
ResNET-50	2048	224, 224, 3
ResNET-101	2048	224, 224, 3
ResNET-152	2048	224, 224, 3
VGG16	4096	224, 224, 3
VGG19	4096	224, 224, 3

Durante esta etapa também foi empregada uma técnica conhecida como transferência de aprendizagem. Em suma, esta técnica consiste em usar um modelo para performar uma tarefa semelhante a qual ele foi originalmente treinado para executar [Weiss et al. 2016]. Neste cenário, todas as redes utilizadas para extração das *Deep Learning Features* foram inicializadas com os pesos gerados após o treinamento da arquitetura com a base *Imagenet* [Kornblith et al. 2019].

A Tabela 1 mostrou o detalhamento da divisão de imagens na base BreakHis e nesta foi possível perceber que há um desbalanceamento entre os representantes de cada classe. Esse desbalanceamento pode gerar um viés para uma das classes durante o treinamento do classificador a ser utilizado. Com o intuito de evitar tal viés, foi empregada uma técnica conhecida como SMOTE. A SMOTE, originalmente descrita por Chawla (2002), é uma técnica de *Data Augmentation* que consiste em criar dados sintéticos para a classe minoritária, balanceando, assim, o conjunto de características extraídas. Vale ressaltar que a SMOTE foi aplicada somente ao conjunto de treino, afim de não comprometer os resultados obtidos.

Por fim, após o balanceamento, a base foi normalizada subtraindo cada representante do conjunto de dados (x) pelo valor mínimo de todo o conjunto (min) e então dividindo esse valor pela diferença entre os valores máximo (max) e o mínimo (min) de todo o conjunto de características extraídas. Essa técnica de normalização é expressa pela Equação 1.

$$N(x) = \frac{x - \min}{\max - \min} \quad (1)$$

3.3. Treino e Classificação

Para classificação foram testados os classificadores *Random Forest*, *Decision Tree*, MLP, AdaBoost e SVM. No entanto, o classificador escolhido foi o SVM, uma vez que apresentou melhor acurácia nos testes realizados. Além disso, trabalhos como os de Ataky e Lameiras Koerich (2022) e Zhou (2022) também utilizam este classificador, o que contribuiu para a escolha deste modelo, visto que ambos os trabalhos referenciados apresentaram acurácia acima de 97% para a base BreakHis em suas metodologias. Todos os modelos para classificação citados estão implementados na biblioteca *Scikit-Learn* [Ekaba e Bisong 2019] para linguagem Python. Vale ressaltar que foram testadas, empiricamente, diversas configurações para os hiperparâmetros utilizados na pré-configuração do classificador SVM e a combinação com a qual foram obtidos os melhores resultados é mostrada na Tabela 3.

Tabela 3. Hiperparâmetros utilizados para pré-configurar o classificador SVM.

Hiperparâmetro	Inicialização
kernel	'rbf'
C	2.7
coef0	0.001
random_state	42

4. Resultados e Discussão

Para o treino e coleta de resultados do classificador SVM foi utilizada uma técnica de validação conhecida como *Hold-Out*. Esta técnica é um tipo de validação cruzada e consiste em embaralhar e particionar o conjunto de dados a ser utilizado em amostras de treino e teste. A partição de treino é utilizada para treinar o classificador, enquanto que a partição de testes é usada na etapa de coleta de resultados [Levman et al. 2021]. Na metodologia aqui apresentada foram utilizadas amostras de 80% e 20% para as partições de treino e teste, respectivamente.

Foram executados um total de 10 treinamentos individuais para cada arquitetura utilizada na extração de características, coletando 5 métricas principais: Acurácia, *Cohen's Kappa Score*, Precisão, *Recall* e *F1-Score*. Os valores presentes na Tabela 4 representam a média dos resultados obtidos em cada execução para a métrica em questão e a Figura 2 mostra as variações dos resultados obtidos para acurácia, divididos por subconjunto de ampliação.

Podemos perceber que a metodologia proposta neste trabalho alcança acurácias de até 97.3% quando utilizada a arquitetura ResNET-50 na etapa de extração das *Deep Learning Features*. Além disso, é possível perceber que, como dito em Bisneto e Araújo (2022), no caso geral, as melhores acurácias foram alcançadas no subconjunto de imagens com ampliação de $40\times$ da base BreakHis.

Ainda analisando os resultados apresentados na Tabela 4, é possível perceber que CNNs que geram um vetor de características muito diverso tendem a obter menor acurácia quando utilizados para treinar um classificador SVM na metodologia deste trabalho. Podemos destacar as redes VGG16 e VGG19, que geraram um vetor de 4096 posições e obtiveram acurácias de 93.5% e 92.4% respectivamente, nos melhores casos.

Tabela 4. Resultados alcançados para cada CNN e ampliações de imagem.

Arquitetura	Ampliação	Acurácia (%)	Kappa	Precisão	Recall	F1
DenseNET-169	40×	96.4%	0.91	0.96	0.98	0.97
	100×	95.1%	0.88	0.95	0.98	0.97
	200×	96.0%	0.90	0.96	0.97	0.97
	400×	94.6%	0.87	0.95	0.96	0.96
DenseNET-201	40×	96.8%	0.92	0.97	0.98	0.97
	100×	95.0%	0.88	0.95	0.97	0.96
	200×	95.6%	0.89	0.96	0.96	0.96
	400×	94.4%	0.87	0.95	0.96	0.95
MobileNET	40×	94.0%	0.88	0.95	0.97	0.96
	100×	91.2%	0.78	0.93	0.94	0.93
	200×	92.4%	0.81	0.94	0.94	0.94
	400×	91.4%	0.79	0.93	0.94	0.93
MobileNET V2	40×	92.4%	0.93	0.97	0.98	0.98
	100×	90.7%	0.77	0.91	0.95	0.93
	200×	91.3%	0.79	0.92	0.95	0.93
	400×	89.0%	0.74	0.95	0.95	0.95
ResNET-50	40×	97.3%	0.94	0.97	0.98	0.98
	100×	95.7%	0.90	0.96	0.97	0.97
	200×	96.2%	0.91	0.97	0.97	0.97
	400×	94.2%	0.86	0.96	0.95	0.96
ResNET-101	40×	96.5%	0.92	0.97	0.98	0.97
	100×	95.4%	0.89	0.96	0.97	0.97
	200×	96.4%	0.91	0.97	0.97	0.97
	400×	94.7%	0.88	0.96	0.95	0.96
ResNET-152	40×	97.1%	0.93	0.97	0.98	0.98
	100×	95.5%	0.89	0.96	0.97	0.97
	200×	96.2%	0.90	0.97	0.97	0.87
	400×	94.5%	0.87	0.96	0.95	0.96
VGG16	40×	93.1%	0.84	0.94	0.96	0.95
	100×	93.5%	0.84	0.95	0.96	0.95
	200×	93.2%	0.84	0.94	0.96	0.95
	400×	91.0%	0.80	0.92	0.95	0.93
VGG19	40×	92.2%	0.81	0.93	0.95	0.94
	100×	91.3%	0.80	0.92	0.95	0.93
	200×	92.4%	0.82	0.93	0.95	0.94
	400×	89.9%	0.77	0.91	0.94	0.92

5. Conclusões e Trabalhos Futuros

Partindo do que foi apresentado, é possível afirmar que a metodologia proposta alcança valores excelentes quando aplicada a base BreakHis, superando os resultados obtidos por Bisneto e Araújo (2022), Spanhol (2016) e Alom (2016), que chegaram a acurácias de 92.9%, 90.0% e 97.0% em seus melhores casos, respectivamente. Além disso, o modelo apresentado obteve acurácia semelhante a de Ataky e Lameiras Koerich (2022), que

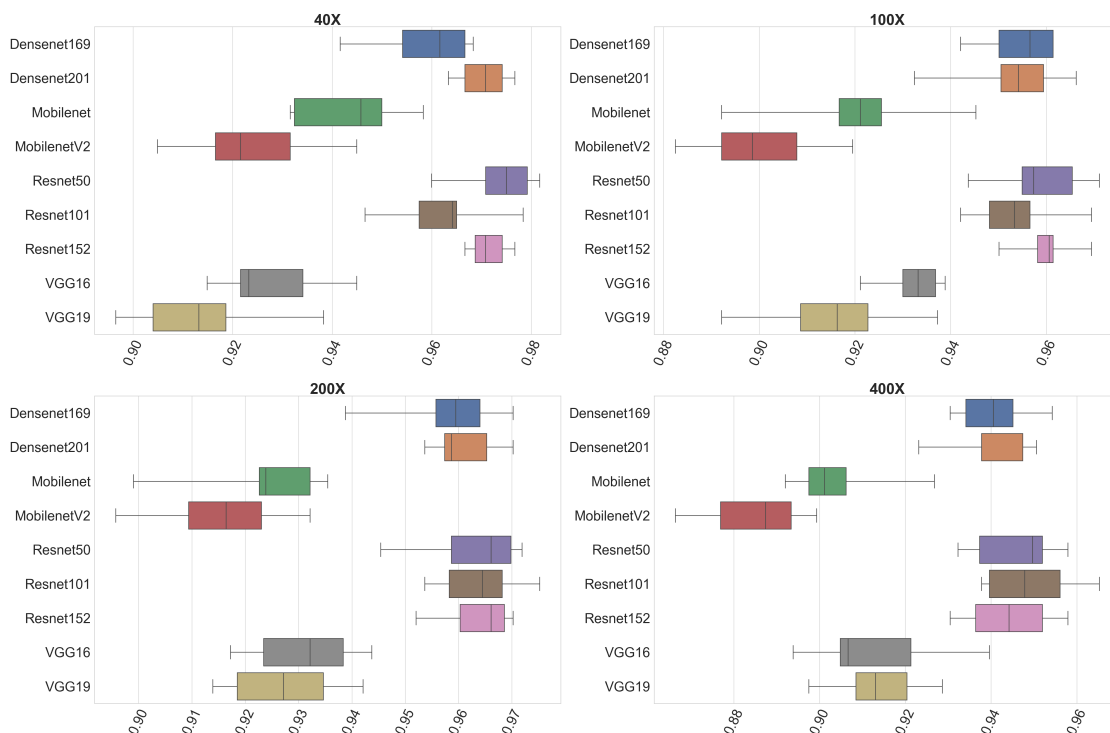


Figura 2. Variação de acurácia para 10 execuções.

atingiu valores de 97.3% no melhor cenário, não se distanciando do melhor valor encontrado na literatura, mostrado em Zhou (2022), que alcançou acurácias de até 99.14% para classificação binária da base BreakHis. Assim, acredita-se que a metodologia aqui apresentada pode, de fato, integrar sistemas CADx para auxiliar especialistas no diagnóstico de câncer de mama, partindo da análise de imagens de exames histopatológicos.

Como trabalhos futuros, pretende-se testar mais arquiteturas de redes neurais convolucionais para extração de características. Além disso, é possível aplicar técnicas de *Fine-Tuning* com o intuito de automatizar a escolha dos parâmetros para o classificador escolhido bem como aprimorar as características extraídas por CNNs. Por fim, deseja-se analisar os resultados da metodologia apresentada em outras bases de imagens.

Referências

- Alom, M. Z., Yakopcic, C., Nasrin, M. S., Taha, T. M., and Asari, V. K. (2019). Breast cancer classification from histopathological images with inception recurrent residual convolutional neural network. *Journal of digital imaging*, 32:605–617.
- Amrane, M., Oukid, S., Gagaoua, I., and Ensari, T. (2018). Breast cancer classification using machine learning. In *2018 Electric Electronics, Computer Science, Biomedical Engineerings' Meeting (EBBT)*, pages 1–4.
- Ataky, S. T. M. and Lameiras Koerich, A. (2022). A novel bio-inspired texture descriptor based on biodiversity and taxonomic measures. *Pattern Recognition*, 123:108382.
- Chawla, N. V., Bowyer, K. W., Hall, L. O., and Kegelmeyer, W. P. (2002). Smote: synthetic minority over-sampling technique. *Journal of artificial intelligence research*, 16:321–357.

- e Flávio Araújo, T. B. (2022). Diagnóstico automático de lesões malignas de mama em imagens histopatológicas usando mensuração mútua de características de textura bioinspiradas e deep learning. In *Anais do XXII Simpósio Brasileiro de Computação Aplicada à Saúde*, pages 246–255, Porto Alegre, RS, Brasil. SBC.
- Ekaba and Bisong (2019). Introduction to scikit-learn. *Building Machine Learning and Deep Learning Models on Google Cloud Platform: A Comprehensive Guide for Beginners*, pages 215–229.
- Erfankhah, H., Yazdi, M., Babaie, M., and Tizhoosh, H. R. (2019). Heterogeneity-aware local binary patterns for retrieval of histopathology images. *IEEE Access*, 7:18354–18367.
- Firmino, M., Angelo, G., Morais, H., Dantas, M. R., and Valentim, R. (2016). Computer-aided detection (cade) and diagnosis (cadx) system for lung cancer with likelihood of malignancy. *Biomedical engineering online*, 15(1):1–17.
- He, K., Zhang, X., Ren, S., and Sun, J. (2016). Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778.
- Howard, A. G., Zhu, M., Chen, B., Kalenichenko, D., Wang, W., Weyand, T., Andreetto, M., and Adam, H. (2017). Mobilenets: Efficient convolutional neural networks for mobile vision applications. *arXiv preprint arXiv:1704.04861*.
- Huang, G., Liu, Z., Van Der Maaten, L., and Weinberger, K. Q. (2017). Densely connected convolutional networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 4700–4708.
- Kornblith, S., Shlens, J., and Le, Q. V. (2019). Do better imagenet models transfer better? In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 2661–2671.
- Krizhevsky, A., Sutskever, I., and Hinton, G. (2013). Imagenet classification with deep convolutional networks. In *Proceedings of the 26th Annual Conference on Neural Information Processing Systems (NIPS)*, pages 1106–1114.
- Levman, J., Jennings, M., Kabaria, P., Rouse, E., Nangaku, M., Berger, D., Gondra, I., Takahashi, E., and Tyrrell, P. (2021). Hold-out validation for the assessment of stability and reliability of multivariable regression demonstrated with magnetic resonance imaging of patients with schizophrenia. *International Journal of Developmental Neuroscience*, 81(7):655–662.
- Nikhil and Ketkar (2017). Introduction to keras. *Deep learning with python: a hands-on introduction*, pages 97–111.
- Simonyan, K. and Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*.
- Spanhol, F. A., Oliveira, L. S., Petitjean, C., and Heutte, L. (2015). A dataset for breast cancer histopathological image classification. *IEEE Transactions on Biomedical Engineering*, 63(7):1455–1462.

- Spanhol, F. A., Oliveira, L. S., Petitjean, C., and Heutte, L. (2016). Breast cancer histopathological image classification using convolutional neural networks. In *2016 International Joint Conference on Neural Networks (IJCNN)*, pages 2560–2567.
- Weiss, K., Khoshgoftaar, T. M., and Wang, D. (2016). A survey of transfer learning. *Journal of Big data*, 3(1):1–40.
- Zhou, Y., Zhang, C., and Gao, S. (2022). Breast cancer classification from histopathological images using resolution adaptive network. *IEEE Access*, 10:35977–35991.