

Apoio ao Diagnóstico de Câncer de Mama por meio de Wavelet e Vision Transformers

Pedro Guagneli Furlan¹, Thiago Pirola Ribeiro¹

¹Faculdade de Computação – Universidade Federal de Uberlândia (UFU)
Uberlândia - MG – Brazil

pedro.furlan@ufu.br, tpribeiro@ufu.br

Abstract. *Breast cancer is one of the leading causes of death among women worldwide, and is one of the most studied diseases in the world. Neural networks have evolved considerably and are widely used in Computer-Aided Diagnosis (CAD) systems. However, reducing false negatives remains a critical challenge; as the most severe type of error, an incorrect negative assessment may discourage patients from seeking medical follow-up, posing serious health risks. This study proposes and evaluates the use of the wavelet transform technique in conjunction with a Vision Transformer (ViT) architecture for mammogram classification. The training process was structured into two main stages: the first uses the wavelet as a guide to help the network better identify edges; the second performs a final fine-tuning using the original preprocessed images in order to capture these features more effectively. The results demonstrated that the use of only the Wavelet improved the F1-score by 9.73% and the recall by 8.44%, representing a significant increase in the evaluation metrics.*

Resumo. *O câncer de mama é uma das principais causas de morte entre as mulheres em todo o mundo, sendo uma das doenças mais estudadas no mundo. As redes neurais têm evoluído consideravelmente, sendo amplamente utilizadas em Sistemas Computacionais de Apoio ao Diagnóstico (CAD). Contudo, a redução de falsos-negativos ainda permanece como um ponto de atenção, por ser o erro mais grave, uma avaliação negativa incorreta pode induzir a paciente a não procurar o médico, gerando um grave risco à saúde. Este trabalho propõe e avalia a utilização da técnica da Transformada Wavelet em uma rede com a arquitetura Vision Transformer (ViT) para classificar mamografias. O treinamento foi estruturado em duas partes principais: a primeira utiliza a técnica como um guia para a rede identificar melhor as bordas; a segunda realiza um ajuste final com as imagens originais pré-processadas, para tentar captar essas bordas. Os resultados demonstraram que o uso apenas da Wavelet melhorou o F1-score em 9,73% e o recall em 8,44%, representando um aumento relevante das métricas.*

1. Introdução

Segundo o Instituto Nacional do Câncer [INCA 2026], o câncer é a segunda causa mais comum de óbito no mundo, com taxa de incidência crescente, sendo o câncer de mama a neoplasia maligna mais comum entre mulheres. Estimativas do [IARC/WHO 2022] indicam que, em 2022, foram diagnosticados aproximadamente 2,3 milhões de novos

casos de câncer de mama no mundo, além de 666.103 óbitos associados à doença. No Brasil, esse tipo de câncer também ocupa o primeiro lugar em incidência entre mulheres, desconsiderando os tumores de pele não melanoma, com 78.610 novos casos estimados para cada ano do triênio 2026–2028 [INCA 2026].

No contexto brasileiro, a mamografia por raio X é amplamente utilizada no rastreamento e diagnóstico do câncer de mama, especialmente por seu baixo custo, empregando compressão e feixes com espectro energético ajustado para aumentar o contraste entre estruturas glandulares e possíveis lesões. Com o avanço da mamografia digital, houve melhoria significativa na qualidade das imagens e, conseqüentemente, no apoio ao diagnóstico. Paralelamente, métodos tradicionais de aprendizado de máquina e, sobretudo, abordagens de *deep learning* têm contribuído para aumentar a acurácia, melhorar a detecção de lesões e reduzir falsos-positivos [Ou et al. 2021, Gao et al. 2023]. Nesse cenário, os Sistemas Computacionais de Apoio ao Diagnóstico (CAD — *Computer-Aided Diagnosis*) atuam como uma segunda opinião ao profissional de saúde, sem substituir o julgamento clínico, e têm sido aprimorados por redes neurais capazes de aprender representações discriminativas diretamente das imagens [Yeasmin et al. 2024]. Assim, este trabalho tem como objetivo avaliar o uso da Transformada Wavelet em conjunto com *Vision Transformers* aplicados a imagens de mamografia, visando contribuir para a redução de falsos-positivos e falsos-negativos.

2. Materiais e Métodos

A metodologia proposta para este trabalho é dividida em quatro etapas principais: aquisição da base de dados; pré-processamento e divisão das imagens em conjuntos de treino e teste; realização de dois treinamentos distintos; por fim, a análise dos resultados obtidos.

2.1. Conjunto de Dados e Pré-Processamento

O dataset escolhido foi o Mini-DDSM [Lekamlage et al. 2020], que é uma versão compactada da DDSM (*Digital Database for Screening Mammography*) [Heath et al. 2001]. O dataset Mini-DDSM é composto por imagens em formato JPG de 1.938 pacientes, totalizando 7.751 mamografias. Os dados estão classificados como benigno, maligno ou normal, sendo que as classes são balanceadas. Para o desenvolvimento da pesquisa, foram separados 80% dos dados para treinamento e 20% para validação.

O dataset contém imagens com resoluções variadas e, com isso, foi aplicado um pré-processamento para redimensionar as imagens para 224×224 pixels, tamanho compatível com o utilizado na rede escolhida. Foi aplicada a técnica de espelhamento para o alinhamento das mamas nas imagens para a esquerda, pois as mamas apresentavam orientações distintas (esquerda e direita), para facilitar o treinamento. Outro tratamento aplicado foi a retirada do fundo da imagem que continha marcas de identificação no canto da mamografia.

2.2. Vision Transformer

A rede neural escolhida foi a arquitetura *Vision Transformer* (ViT) [Google 2021], pré-treinada com o banco de dados ImageNet-21k [Deng et al. 2009] (14 milhões de imagens e 21.843 classes) e submetida a um *fine-tuning* com o ImageNet-1k (1,3 milhão de imagens e 1.000 classes).

Esta arquitetura foi proposta em 2020 com o objetivo de revolucionar as redes neurais já existentes, principalmente as Redes Neurais Convolucionais (CNNs), que utilizam filtros locais para processar os pixels próximos, procurando por padrões. A principal diferença do ViT é que ele segmenta a imagem em blocos - chamados de *patches* - de 16×16 pixels [Dosovitskiy et al. 2020]. Cada bloco é transformado em um vetor linear e por meio do mecanismo de *Self-Attention*, a rede compara o *patch* com todos os demais, calculando a relevância de cada um em relação aos outros e classifica a imagem. Dessa forma, ao contrário das CNNs, que analisam apenas pontos próximos, o ViT compara e avalia toda a imagem, permitindo ao modelo identificar características com maior facilidade.

2.3. Transformada Wavelet

A Transformada Wavelet é uma ferramenta matemática que decompõe imagens em diferentes níveis de resolução, utilizando funções transladadas para registrar a localização das variações e funções escalonadas para identificar tanto estruturas amplas quanto detalhes finos [Mallat 1989]. Na decomposição, a imagem original é dividida em quatro sub-bandas: aproximação (LL), que preserva as características estruturais e de baixa frequência, e três sub-bandas de alta frequência, responsáveis por evidenciar bordas e detalhes nas direções horizontal (HL), vertical (LH) e diagonal (HH) (Figura 1). Essa separação permite destacar padrões que poderiam ser difíceis de perceber na imagem original e, para uso na rede, as sub-bandas foram concatenadas em um único bloco de imagem.

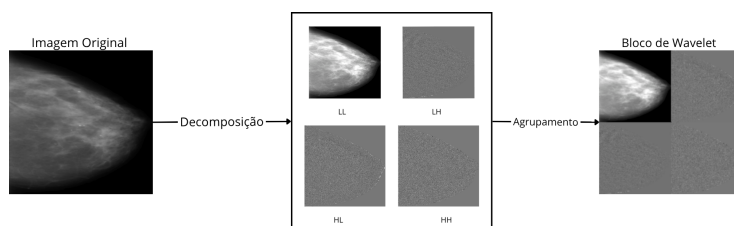


Figura 1. Decomposição da imagem original nas sub-bandas LL, LH, HL e HH, posteriormente concatenadas para formar o bloco que será utilizado.

2.4. Procedimento Experimental

O processo experimental foi dividido em duas partes principais: o treinamento de um modelo base e outro utilizando a técnica proposta. Em ambos os casos, utilizou-se a mesma arquitetura de rede, o *Vision Transformer*, com o mesmo pré-treinamento e o mesmo banco de imagens já pré-tratado. Como o modelo pré-treinado original tem mais de 1.000 classes, foi necessário redefinir a camada de classificação para apenas três categorias: maligno, benigno e normal.

No primeiro experimento, referente ao modelo base, não foram aplicadas técnicas adicionais de processamento, como a técnica wavelet, apenas ajustes de hiperparâmetros da rede. Após o treinamento, a melhor versão do modelo foi selecionada e suas métricas foram salvas para posterior comparação.

No segundo experimento, referente ao modelo com Wavelet, o treinamento foi realizado em duas etapas. Inicialmente, a rede foi treinada com a mesma base de dados, porém com a aplicação da Transformada Wavelet às imagens, com o objetivo de direcionar o aprendizado para bordas e transições de intensidade, auxiliando na localização e

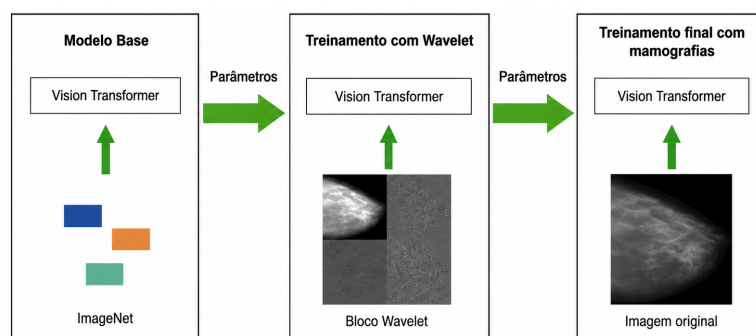


Figura 2. Sequência e dados de treinamento do modelo proposto com Wavelet.

identificação de possíveis tumores. Em seguida, o mesmo modelo foi treinado novamente com as imagens originais pré-processadas, sem a aplicação da Wavelet, buscando fazer com que a rede preservasse a capacidade de reconhecer essas estruturas sem o auxílio direto da técnica, contribuindo para uma identificação mais precisa da neoplasia.

3. Métricas de Avaliação

A matriz de confusão, que mostra os acertos e erros da rede, é essencial para avaliar e calcular as métricas utilizadas. Quando um modelo avalia uma imagem, ele pode gerar quatro resultados possíveis: verdadeiro-positivo (TP), falso-positivo (FP), verdadeiro-negativo (TN) e falso-negativo (FN). Esses resultados são utilizados por algumas métricas para calcular o desempenho da rede, neste trabalho foram utilizadas: Acurácia, Precisão, Recall e F1-Score.

4. Resultados e Discussões

Os experimentos realizados resultaram em duas versões diferentes: a primeira com treinamento direto, usando apenas as imagens originais pré-processadas; já o outro modelo aplicou a técnica de wavelet nas mesmas imagens e no segundo *fine-tuning*, utilizou apenas as imagens pré-processadas comuns. Na Tabela 1 está a comparação das métricas finais obtidas para ambos os modelos no conjunto de validação.

Tabela 1. Comparativo de desempenho entre o treinamento direto e o modelo com wavelet.

Modelo	Acurácia	Precisão	Recall	F1-Score
ViT (Direto)	0,680	0,677	0,687	0,678
ViT + Wavelet	0,742	0,753	0,745	0,744

Analisando os dados da Tabela 1, observa-se que o modelo treinado com a wavelet apresentou um desempenho superior ao treinado diretamente. O *F1-score* indica que a rede evoluiu e manteve um bom equilíbrio na precisão com uma evolução de 0,066. Além disso, o *recall*, que é uma das métricas mais importantes para esse projeto, teve uma melhoria considerável de 0,058, indicando uma redução na quantidade de falsos-negativos. Esses resultados evidenciam que a técnica sozinha não gera um aumento expressivo, porém quando combinada com outras técnicas pode ajudar a alcançar grandes resultados.

5. Conclusão

Neste trabalho foi proposta uma análise do uso da técnica da Transformada Wavelet em redes com arquitetura ViT, para classificar mamografias e tentar reduzir a ocorrência de falsos-negativos. O objetivo de utilizar essa técnica na primeira fase do *fine-tuning* foi ajudar o modelo a captar detalhes importantes com mais facilidade, graças às transformações causadas pela Wavelet, facilitando identificar as bordas dos tumores.

Os resultados demonstraram que utilizar esse método na primeira fase do treinamento do modelo, proporcionou uma melhora no desempenho geral da rede, servindo como um guia inicial para encontrar as bordas da lesão. Além disso, o aumento da métrica *recall*, indica que o wavelet ajudou a reduzir o número de falsos-negativos e identificar melhor se realmente não havia nenhum tumor. Pretende-se melhorar ainda mais o desempenho da rede, e testar novas técnicas que podem impulsionar a Wavelet a ser mais assertiva e reduzir os falsos-negativos.

Referências

- Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K., and Fei-Fei, L. (2009). Imagenet: A large-scale hierarchical image database. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 248–255, Miami, FL, USA. IEEE.
- Dosovitskiy, A. et al. (2020). An image is worth 16x16 words: Transformers for image recognition at scale. *arXiv preprint arXiv:2010.11929*.
- Gao, Y. et al. (2023). The application of traditional machine learning and deep learning techniques in mammography: a review. *Frontiers in Oncology*, 13:1213045.
- Google (2021). Vision transformer (base-sized model). <https://huggingface.co/google/vit-base-patch16-224>.
- Heath, M., Bowyer, K., Kopans, D., Moore, R., and Kegelmeyer, W. P. (2001). The digital database for screening mammography. In Yaffe, M. J., editor, *Proceedings of the Fifth International Workshop on Digital Mammography*, pages 212–218. Medical Physics Publishing. ISBN 1-930524-00-5.
- IARC/WHO (2022). Globocan 2022: Cancer today. Technical report, International Agency for Research on Cancer, Lyon.
- INCA (2026). *Estimativa 2026: incidência de câncer no Brasil*. INCA, Rio de Janeiro. Versão eletrônica. Acesso em: 13 maio 2026.
- Lekamlage, C. D., Afzal, F., Westerberg, E., and Cheddad, A. (2020). Mini-ddsm: Mammography-based automatic age estimation. In *Proceedings of the 3rd International Conference on Digital Medicine and Image Processing (DMIP 2020)*, pages 1–6, Kyoto, Japan. ACM.
- Mallat, S. G. (1989). A theory for multiresolution signal decomposition: the wavelet representation. *IEEE transactions on pattern analysis and machine intelligence*, 11(7):674–693.
- Ou, X., Chen, X., Xu, X., et al. (2021). Recent development in x-ray imaging technology: Future and challenges. *Research*, 2021:9892152.
- Yeasmin, M. N. et al. (2024). Advances of ai in image-based computer-aided diagnosis: A review. *Array*, page 100357.