

Impacto dos Marcadores de Posicionamento Anatômico na Detecção de Câncer de Mama por Termografia

João Augusto Pilato de Castro¹, Fabrício Araújo Filgueiras¹,
Luíza Machado Costa Nascimento¹, Heder Soares Bernardino¹, Saulo Moraes Villela¹

¹Universidade Federal de Juiz de Fora, Juiz de Fora, MG, Brasil

{pilato.joao, fabricio.filgueiras, luiza.machado}@estudante.ufjf.br,

{heder.bernardino, saulo.moraes}@ufjf.br

Abstract. *Breast cancer remains one of the main public health challenges, with early diagnosis crucial. In this context, thermography has been investigated as a complementary strategy because it is non-invasive, free of ionizing radiation, and has low operational costs. In the literature on classifying thermographic images of the breast, region of interest segmentation strategies are adopted to eliminate noise and highlight relevant physiological information. However, these strategies can also remove elements that play a functional role in the learning process. This study investigates the influence of a marker commonly present in thermographic images on the performance of deep learning models for breast cancer detection. The results indicate that preserving the marker improves performance, suggesting that it acts as a spatial reference and highlighting the need to critically evaluate segmentation and preprocessing strategies.*

Resumo. *O câncer de mama permanece como um dos principais desafios de saúde pública, sendo fundamental o diagnóstico precoce. Nesse contexto, a termografia tem sido investigada como estratégia complementar por não ser invasiva, livre de radiação ionizante e de baixo custo operacional. Na literatura sobre classificação de imagens termográficas da mama, estratégias de segmentação da região de interesse são adotadas para eliminar ruídos e destacar informações fisiológicas relevantes. No entanto, podem também remover elementos que desempenham papel funcional no processo de aprendizado. Este estudo investiga a influência de um marcador comumente presente em imagens termográficas no desempenho de modelos de aprendizado profundo para identificação do câncer de mama. Os resultados indicam que preservar o marcador melhora o desempenho, sugerindo que ele atua como uma referência espacial e evidenciando a necessidade de avaliar criticamente estratégias de segmentação e pré-processamento.*

1. Introdução

Dados recentes indicam que o câncer de mama permanece como um dos principais desafios de saúde pública. Nesse contexto, a termografia surge como uma modalidade complementar promissora por não ser invasiva, sem radiação ionizante e de baixo custo operacional [Saadizadeh 2021]. No entanto, a interpretação clínica de termogramas é desafiadora, pois padrões térmicos costumam ser sutis e influenciados por fatores fisiológicos que podem atuar como variáveis de confusão [Bandyopadhyay 2010].

Com o avanço de modelos de aprendizado profundo, as Redes Neurais Convolucionais (*Convolutional Neural Networks* – CNNs) têm sido cada vez mais utilizadas para detectar padrões complexos frequentemente imperceptíveis à análise visual humana [Krizhevsky et al. 2012, Gamal et al. 2023]. Embora avanços recentes tenham se concentrado em melhorias arquiteturais, muitos *pipelines* de termografia ainda adotam estratégias baseadas na segmentação da região de interesse (*region of interest* – ROI), com o objetivo de remover ruído de fundo e enfatizar a região anatômica relevante [Fabbrizzi et al. 2022]. Essa prática assume que elementos não fisiológicos são irrelevantes para o aprendizado do modelo. Como resultado, artefatos relacionados à aquisição e marcadores de referência são frequentemente removidos, descartando informações que podem contribuir para o desempenho [de Castro et al. 2026].

Nesse contexto, este trabalho investiga o impacto de um elemento não fisiológico recorrente em termogramas mamários: o marcador de posicionamento anatômico (TAG). Foi avaliado, de forma controlada, como a presença e as propriedades do TAG (posição, escala e assinatura térmica) influenciam o desempenho de um classificador baseado em CNN. Os resultados mostram que preservar o TAG melhora métricas como ROC_AUC e F1-score, enquanto removê-lo ou deslocá-lo degrada o desempenho, sugerindo que o marcador atua como uma referência espacial contextual. Assim, este trabalho contribui com a investigação sistemática do impacto desse marcador na classificação de termogramas mamários, demonstrando que sua preservação melhora o desempenho dos modelos e destaca a importância de avaliar criticamente estratégias de pré-processamento.

2. Trabalhos Relacionados

Revisões recentes indicam que CNNs e *Transfer Learning* são abordagens predominantes na análise automatizada de termogramas, com arquiteturas como VGG e ResNet atuando como *baselines* recorrentes [Simonyan and Zisserman 2015, He et al. 2015]. Há também uma forte dependência de um único conjunto de dados: mais de 77% dos estudos utilizam o Visual Lab (DMR-IR)¹ [Singh and Arora 2024], o que facilita comparações, mas pode amplificar vieses técnicos e de aquisição.

Apesar do uso frequente de pré-processamento e aumento de dados [Alshehri and AlSaeed 2023], a literatura raramente analisa sistematicamente o impacto de artefatos específicos de aquisição, como os marcadores, sobre desempenho e interpretabilidade [Nguyen Chi et al. 2025]. Este trabalho contribui diretamente nesse ponto ao investigar, de maneira controlada, a influência do TAG.

3. Base de Dados

Utilizou-se a base pública Visual Lab [Silva et al. 2014], amplamente empregada na literatura em termografia mamária. O conjunto completo é composto por 1463 imagens infravermelhas, distribuídas entre casos saudáveis e doentes, adquiridas sob protocolo padronizado em múltiplos ângulos (frontal, 45° e 90° à esquerda e à direita).

Para reduzir a variabilidade geométrica introduzida por diferentes perspectivas, foi considerada apenas a vista frontal, que originalmente contém 297 imagens (176 saudáveis e 121 com câncer). Após a remoção de imagens com anomalias ou inconsistências de aquisição e a separação de um subconjunto destinado ao treinamento dos

¹<https://visual.ic.uff.br/dmi/>

modelos de segmentação, foram consideradas 250 imagens válidas para os experimentos de classificação, sendo 153 saudáveis e 97 com diagnóstico de câncer.

Um artefato visual proeminente nessas imagens é o TAG, um adesivo quadrangular utilizado como referência geométrica durante a aquisição (Figura 1).

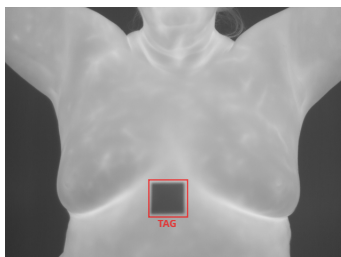


Figura 1. Exemplo de termograma frontal com TAG.

4. Metodologia

4.1. Configurações de Dados (Subconjuntos)

A partir do conjunto original (ORIG), foram construídos subconjuntos para isolar propriedades do TAG: (i) SEG: segmentação da mama (ROI), removendo tecido extra-mamário e, consequentemente, o TAG; (ii) SEG+TAG: segmentação preservando explicitamente o TAG; (iii) WT (*Without TAG*): remoção completa do TAG; (iv) FS (*Fixed Size*): TAG com tamanho normalizado; (v) MV (*Moved*): TAG deslocado para a posição fixa no canto da imagem; (vi) FT (*Fixed Temperature*): TAG com temperatura normalizada.

Essas configurações permitem avaliar separadamente o efeito de presença/ausência, escala, posição e assinatura térmica do marcador, mantendo o restante do pipeline inalterado.

4.2. Pré-processamento, Treinamento e Avaliação

Os dados foram particionados em treino (60%), validação (20%) e teste (20%), com separação por paciente para evitar *data leakage*. As imagens foram normalizadas por min-max usando mínimo e máximo calculados apenas no conjunto de treinamento e aplicados aos demais conjuntos. As imagens foram redimensionadas para 224×224 via *letterbox resizing*. O aumento de dados foi aplicado apenas no conjunto de treinamento (rotações pequenas, zoom, variação de brilho e espelhamento horizontal com probabilidade de 50%).

Como classificador, foi usada uma CNN baseada na VGG-16 [Simonyan and Zisserman 2015]. O treinamento foi realizado com *binary cross-entropy* e Adam (10^{-4}), por até 500 épocas, com *early stopping* (*patience*=50 e melhoria mínima=0,01). O melhor modelo no conjunto de validação foi avaliado no teste. O processo metodológico pode ser visto na Figura 2.

5. Experimentos Computacionais

Para cada configuração dos subconjuntos de dados descritos na seção anterior, obtidos a partir de diferentes tratativas de imagens realizadas no Photoshop, foram realizadas

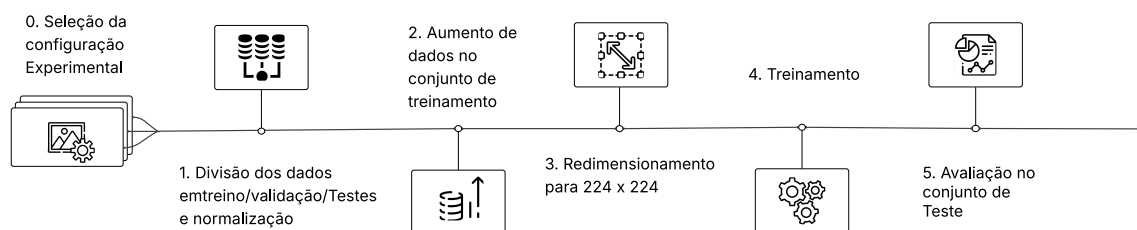


Figura 2. Processo metodológico.

30 execuções independentes. Como etapa preliminar à aplicação de testes paramétricos, avaliou-se a hipótese de normalidade das distribuições por meio do teste de Shapiro-Wilk.

Para as métricas ROC_AUC, F1-score, Recall e Precisão, a condição foi satisfeita, permitindo a aplicação de ANOVA unidirecional de Welch ($\alpha = 0,05$), seguida de pós-teste de Games-Howell para comparações par a par.

Os experimentos foram executados utilizando TensorFlow 2.14 com suporte a GPU (CUDA 12.2), em uma NVIDIA GeForce RTX 4090. O código está disponível publicamente².

5.1. Resultados e Discussões

A Tabela 1 apresenta as médias e desvios padrão das métricas de desempenho no conjunto de teste (ROC_AUC, F1-score, Recall, Precisão e Accuracy).

Tabela 1. Médias e desvios padrão das métricas de desempenho.

Configuração	Acurácia	Precisão	Recall	F1-Score	ROC_AUC
FT (TAG temp. fixa)	0,875 ± 0,039	0,839 ± 0,062	0,836 ± 0,097	0,833 ± 0,053	0,955 ± 0,023
ORIG (original)	0,871 ± 0,045	0,823 ± 0,081	0,847 ± 0,073	0,832 ± 0,063	0,944 ± 0,024
FS (tamanho fixo)	0,851 ± 0,043	0,797 ± 0,071	0,827 ± 0,087	0,807 ± 0,057	0,939 ± 0,024
SEG+TAG	0,853 ± 0,048	0,808 ± 0,077	0,806 ± 0,128	0,799 ± 0,075	0,924 ± 0,033
MV (TAG movido)	0,831 ± 0,039	0,780 ± 0,068	0,775 ± 0,093	0,773 ± 0,055	0,899 ± 0,039
SEG (somente ROI)	0,831 ± 0,053	0,791 ± 0,103	0,755 ± 0,107	0,766 ± 0,080	0,906 ± 0,040
WT (sem TAG)	0,825 ± 0,042	0,774 ± 0,082	0,764 ± 0,106	0,762 ± 0,060	0,905 ± 0,040

Observa-se uma hierarquia consistente: configurações que preservam o TAG (ORIG, FT, FS, SEG+TAG) apresentam desempenho superior às que removem ou invalidam o marcador (WT, SEG, MV). Em particular, deslocar o TAG (MV) degrada o desempenho para patamares semelhantes à remoção, sugerindo que a relação espacial do marcador com a anatomia é crítica.

A análise estatística global confirma que existe uma diferença significativa entre as configurações para todas as métricas avaliadas: ROC_AUC ($p = 4,25 \times 10^{-11}$), F1-score ($p = 3,04 \times 10^{-6}$), Recall ($p = 3,79 \times 10^{-4}$) e Precisão ($p = 7,33 \times 10^{-3}$).

Após a rejeição da hipótese nula global pela ANOVA, foram realizadas comparações par a par utilizando o teste de Games-Howell. A Figura 3 apresenta as matrizes triangulares inferiores dos p -valores para as métricas F1-score e ROC_AUC. Nas

²<https://github.com/joaoaugustopc/breast-thermography-project>

comparações par a par, a evidência mais consistente é que preservar o TAG é mais importante do que manter exatamente sua escala ou assinatura térmica, enquanto alterar sua posição é particularmente prejudicial. Isso sustenta a interpretação de que o TAG funciona como referência espacial contextual.

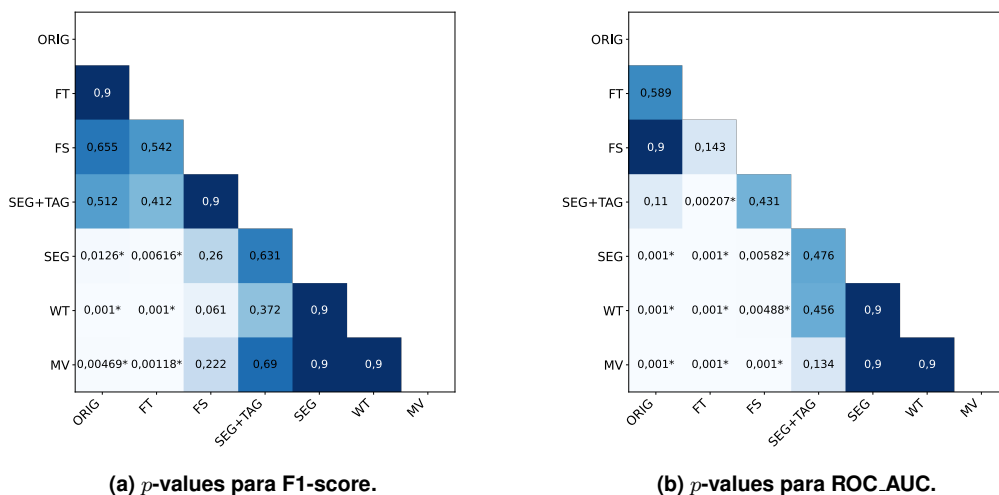


Figura 3. Comparações par a par (Games-Howell).

6. Conclusões

Diante da relevância do câncer de mama como um dos principais desafios de saúde pública, estudos voltados ao desenvolvimento de ferramentas computacionais para apoio ao diagnóstico tornam-se cada vez mais importantes. Nesse contexto, compreender como decisões técnicas, estratégias de pré-processamento e manipulação de dados, impactam o aprendizado dos modelos é fundamental para garantir resultados confiáveis e interpretáveis. Diante disso, foi investigado o impacto de um indício visual não fisiológico (TAG) no treinamento de CNNs para a classificação de termogramas mamários. Verificou-se que preservar o TAG melhora consistentemente o desempenho, enquanto removê-lo ou deslocá-lo degrada métricas como ROC_AUC e F1-score, com diferenças estatisticamente significativas.

Os resultados sugerem que o TAG não deve ser interpretado apenas como um artefato a ser eliminado; sua contribuição está ligada ao contexto geométrico e à organização espacial da imagem. Assim, o pré-processamento deve ser tratado como uma etapa analítica crítica, pois remover elementos contextuais pode reduzir a capacidade de generalização e alterar o que o modelo aprende.

Esses achados oferecem subsídios para pesquisas futuras na área, ao evidenciar que determinados artefatos de aquisição podem desempenhar um papel importante no processo experimental, afetando não apenas o desempenho reportado nos estudos, mas também decisões relacionadas à construção e padronização de novas bases de dados.

Como trabalho futuro, pretende-se investigar marcadores geométricos artificiais para verificar se referências estruturais simplificadas reproduzem os efeitos observados. Além disso, embora a base Visual Lab seja amplamente utilizada na literatura, este estudo foi conduzido em um único conjunto de dados e utilizando apenas uma única arquitetura,

a VGG-16, o que impõe limitações à generalização dos resultados. Assim, investigações futuras envolvendo outras bases e diferentes arquiteturas são importantes para avaliar a robustez da hipótese em diferentes contextos experimentais e cenários de aquisição.

Agradecimentos

Trabalho realizado com o apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) – Código de Financiamento 001, CNPq (313452/2025-3), FAPEMIG (APQ-03313-22 e APQ-01832-22), UFJF e RePesq..

Referências

- Alshehri, A. and AlSaeed, D. (2023). Breast cancer diagnosis in thermography using pre-trained vgg16 with deep attention mechanisms. *Symmetry*, 15(3):582.
- Bandyopadhyay, S. K. (2010). Survey on segmentation methods for locating masses in a mammogram image. *International Journal of Computer Applications*, 9(11):25–28.
- de Castro, J. A. P., Filgueiras, F. A., Nascimento, L. M. C., Bernardino, H. S., and Villela, S. M. (2026). Impact of anatomical positioning markers on breast cancer detection thermography. In *Proceedings of the 2026 International Conference on Computational Science and Its Applications (ICCSA 2026)*, LNCS. Springer.
- Fabrizzi, S., Papadopoulos, S., Ntoutsis, E., and Kompatsiaris, I. (2022). A survey on bias in visual datasets. *Computer Vision and Image Understanding*, 223:103552.
- Gamal, S., Atef, H., Youssef, D., Ismail, T., and El-Azab, J. (2023). Early breast cancer screening from thermography via deep pre-trained edge detection with extreme gradient boosting. In *Intell. Methods, Systems, and Applications (IMSA)*, pages 430–433.
- He, K., Zhang, X., Ren, S., and Sun, J. (2015). Deep residual learning for image recognition. *arXiv preprint arXiv:1512.03385*.
- Krizhevsky, A., Sutskever, I., and Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. In *Adv. Neural Inf. Processing Syst.*, volume 25, pages 1097–1105. Curran Associates, Inc.
- Nguyen Chi, T., Le Thi Thu, H., Doan Quang, T., and Taniar, D. (2025). A lightweight method for breast cancer detection using thermography images with optimized CNN feature and efficient classification. *Journal of Imaging Informatics in Medicine*, 38(3):1434–1451.
- Saadizadeh, S. (2021). Breast cancer detection in thermal images using glrlm algorithm. In *Intl. Conf. on Electrical, Communication, and Computer Engineering (ICECCE)*, pages 1–6, Kuala Lumpur, Malaysia.
- Silva, L. F., Saade, D. C. M., Sequeiros, G. O., Silva, A. C., Paiva, A. C., Bravo, R. S., and Conci, A. (2014). A new database for breast research with infrared image. *Journal of Medical Imaging and Health Informatics*, 4(1):92–100.
- Simonyan, K. and Zisserman, A. (2015). Very deep convolutional networks for large-scale image recognition. In *International Conference on Learning Representations (ICLR)*.
- Singh, J. and Arora, A. S. (2024). Breast cancer detection using infrared thermography: A survey of texture analysis and machine learning approaches. *Journal of Medical Imaging and AI Research*.