

Cross-Validation Deep Learning for Breast Cancer Detection Using DMR-IR Infrared Images

Validação Cruzada em Deep Learning para Detecção de Câncer de Mama Utilizando Imagens Infravermelhas do DMR-IR

Renata dos Santos Melo^{1*}, Henrique Coelho Fernandes¹, André Ricardo Backes²

Abstract: Breast cancer detection is a global health priority. While traditional methods have limitations, infrared thermography offers a promising, non-invasive alternative by detecting subtle thermal changes that can indicate tumors. This paper assessed five pre-trained Convolutional Neural Networks (CNNs) for breast cancer detection using DMR-IR thermal images, employing a 5-fold cross-validation. Among the tested models, ResNet50 achieved the best overall performance, with the highest average accuracy (92.79%), precision (95.00%), specificity (98.67%), sensitivity (72.00%), and F1-score (79.43%). The model was trained using raw thermal images from three anatomical views (frontal, lateral 90°, and lateral 45°), totaling five images per patient, an approach still uncommon in the literature. These results highlight ResNet50's strong potential for reliable and clinically applicable breast cancer detection using thermography.

Keywords: breast cancer — infrared imaging — deep learning — CNN — cross-validation

Resumo: A detecção do câncer de mama é uma prioridade global em saúde. Embora os métodos tradicionais apresentem limitações, a termografia infravermelha oferece uma alternativa promissora e não invasiva, ao detectar alterações térmicas sutis que podem indicar tumores. Este trabalho avaliou cinco Redes Neurais Convolucionais (CNNs) pré-treinadas para a detecção de câncer de mama utilizando imagens térmicas do banco DMR-IR, empregando validação cruzada com 5 *folds*. Entre os modelos testados, a ResNet50 obteve o melhor desempenho geral, com a maior acurácia média (92.79%), precisão (95.00%), especificidade (98.67%), sensibilidade (72.00%) e F1-score (79.43%). O modelo foi treinado utilizando imagens térmicas brutas de três vistas anatômicas (frontal, lateral 90° e lateral 45°), totalizando cinco imagens por paciente, uma abordagem ainda pouco comum na literatura. Esses resultados destacam o forte potencial da ResNet50 para uma detecção confiável e clinicamente aplicável do câncer de mama por meio da termografia.

Palavras-Chave: câncer de mama — imagem infravermelha — deep learning — CNN — validação cruzada

¹ Faculdade de Computação, Universidade Federal de Uberlândia (UFU), Brazil

² Departamento de Computação, Universidade Federal de São Carlos (UFSCar), Brazil

*Corresponding author: renatasm@ufu.br

DOI: <http://dx.doi.org/10.22456/2175-2745.150755> • Received: 07/10/2025 • Accepted: 27/10/2025

CC BY-NC-ND 4.0 - This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License.

1. Introduction

Breast cancer is the most frequently diagnosed cancer in women and a leading cause of cancer-related deaths globally. Early-stage detection greatly improves treatment success [1]. While breast self-examination is commonly encouraged, especially in resource-limited regions, its diagnostic reliability is low. Mammography remains the standard screening tool, though its sensitivity decreases in women with dense breast tissue, particularly younger individuals [2]. To improve detection, additional techniques such as ultrasound, MRI, and thermography are employed [3, 4]. Thermography, a non-invasive method that records skin temperature via infrared

imaging, can reveal abnormal thermal patterns caused by the increased metabolic activity and vascularization of malignant tumors [5].

The analysis of thermographic images for the identification of potential breast anomalies can be significantly augmented by computational techniques [6]. Specifically, machine learning algorithms can be trained to discern patterns associated with cancerous growths [4]. Contemporary research highlights the substantial potential of integrating thermography with computational methodologies for the early detection of breast cancer [7, 8, 9]. The combination of thermography and computational methods offers key benefits for breast can-

cer screening: it is non-invasive and free of ionizing radiation, allowing for early tumor detection and improved patient comfort. Additionally, it is cost-effective and more accessible than mammography, requiring only an infrared camera and a computer, making it suitable for use in smaller clinical settings [10].

In recent years, CNNs have gained widespread adoption for the analysis of thermographic imagery [11]. CNNs excel at recognizing intricate image features, rendering them highly valuable for medical image interpretation [12]. In breast cancer detection through thermography, CNNs can identify abnormal thermal patterns in breast tissue, helping to localize potential tumors [13]. Studies show that CNNs outperform traditional image analysis methods by automatically extracting relevant features and continuously improving accuracy as more data becomes available [10].

2. Related Work

Pramanik et al. [14] proposed a lightweight model that uses SqueezeNet 1.1 for deep feature extraction, combined with a hybrid feature selection algorithm. Similarly, Le Thi Thu et al. [15] employed several pre-trained CNNs (such as ResNet34, Xception, InceptionV3, MobileNetV2, and DenseNet121) for feature extraction from frontal thermal images, optimizing them with feature filters for efficient classification. Both works, while demonstrating high accuracy, predominantly focus on the analysis of frontal thermal images.

However, relying solely on frontal images can introduce bias into the data and hinder comprehensive detection capabilities. As pointed out by Tsietsos et al. [16], most CADx (Computer-Aided Diagnosis) systems based on thermography primarily focus on frontal breast thermograms, which can lead to missing lesions that develop on the sides of the breast. Breast anatomy, including the tail of Spence, can harbor lesions that are not visible in a single frontal view, making multi-view analysis essential for comprehensive diagnostic coverage.

To address this limitation, Madhavi and Thomas [17] highlighted the importance of fusing texture features from frontal and lateral thermograms, arguing that analyzing lateral views, in conjunction with the frontal view, is crucial for accurate diagnosis, overcoming the partial obscurity that can occur in single views. Sánchez-Cauce et al. [18] proposed a multi-input convolutional neural network that combines frontal and lateral thermal images (right 90° and left 90°), along with clinical data, demonstrating that the inclusion of lateral views improves the classification model's performance.

Tiwari et al. [19] introduced a multi-view breast cancer detection system that concatenates frontal, left, and right thermal images to create a more informative temperature map, using deep transfer learning models like VGG16. Mammoottill et al. [20] expanded this approach, using five different thermal views (frontal, right/left 45° lateral, and right/left 90° lateral) in their CNNs, and observed that the addition of clinical data significantly improved accuracy. This work underscores the

relevance of a more comprehensive breast perspective.

Tiwari et al. [21] developed Breast Cancer-Caps, a capsule network-based system that uses multi-view thermal images (concatenated frontal, left, and right), demonstrating the effectiveness of more advanced architectures for this purpose. Tsietsos et al. [16] also integrated multiple views (frontal, left 90°, and right 90°) with clinical data in a multi-input deep learning approach, emphasizing the ability to detect lesions in often-overlooked lateral areas. Mahoro and Akhloufi [22] explored thermogram classification using deep CNNs and Transformers, which are capable of capturing intricate details in image data, including frontal and lateral positions. Samantaraya et al. [23] focused specifically on the effectiveness of angular view thermal images (lateral and oblique) in isolation, using CNNs and radiomic features, filling a gap in understanding their independent efficacy.

The literature consistently demonstrates that incorporating lateral and oblique images, and fusing features from different views, can mitigate bias risks and improve diagnostic accuracy, offering a more complete and less biased representation of the breast.

3. Materials and Methods

3.1 Convolutional Neural Networks (CNNs)

Convolutional Neural Networks (CNNs) are specialized artificial neural networks adept at processing structured data like images, learning hierarchical visual features through convolutional, pooling, and fully connected layers. The core convolution operation extracts patterns such as edges and textures using filters, enabling CNNs to excel in diverse computer vision tasks, including medical image classification [24, 25]. Leveraging CNNs has significantly improved the accuracy of early breast cancer detection, leading to enhanced diagnostic capabilities and improved patient outcomes [26, 27, 22].

Various neural network architectures, including AlexNet, DenseNet, Inception, SqueezeNet, VGG16, VGG19, U-Net, MobileNet, and EfficientNet, are widely employed for thermographic image analysis in breast cancer detection due to their pattern extraction and generalization capabilities in medical imaging. Furthermore, classical machine learning methods like Support Vector Machines (SVMs) remain relevant, particularly when combined with feature extraction techniques such as Gray-Level Co-Occurrence Matrix (GLCM) and Histogram of Oriented Gradients (HOG) [28, 29, 21, 30, 22, 31, 32, 33, 14].

3.2 Thermogram dataset and pre-processing

We used the Database for Mastology Research with Infrared Image (DMR-IR) [34], a public repository of infrared thermograms from healthy individuals and breast cancer patients. This dataset, sourced from Antônio Pedro University Hospital, includes both thermal images (frontal and lateral views at 45° and 90°) and comprehensive clinical/personal data. The clinical data encompasses details such as age, race, personal and family medical history, reported symptoms, and lifestyle

habits. The acquisition and use of this data were ethically approved by the hospital and registered with the Brazilian Ministry of Health under CAEE: 01042812.0.0000.5243. Images were captured using a FLIR SC620 thermal camera, with dimensions of 640×480 pixels, and are publicly accessible online ¹.

Initially, DMR-IR thermograms are obtained in grayscale. However, for better visualization and analysis of temperature variations, these images were converted to colormap, as shown in Figure 1.

The DMR-IR dataset uses static and dynamic protocols to assess breast tissue under different thermal conditions. Static protocols involve image capture after a rest period (10–15 minutes) to ensure thermal stability. Dynamic protocols apply thermal stimulation, such as fan cooling or alcohol, followed by sequential image capture, typically generating six images per minute.

In this study, we used the static protocol, once it was the only one that had 3 different angles, totaling 5 images per patient. Within all the patients, 76 healthy and 21 with cancer were selected, 380 healthy images and 105 with cancer, totaling 485 images. Patients who had images with the breast region cut out, with other visible objects, out of focus, missing one of the 5 images, without changing the angle with objects and lowered arms, were discarded, as shown in the example in Figure 2.

3.3 Evaluation metrics

Various metrics are available to assess the performance of classifiers, with sensitivity, specificity, accuracy, and the F1-Score being the most widely reported in the literature and applied in this study. They are described as follows: True Positive (TP) refers to the number of samples that are actually diseased and correctly classified as such; True Negative (TN) denotes the number of samples that are actually healthy and correctly identified as healthy; False Positive (FP) represents the number of healthy samples that were incorrectly classified as diseased; and False Negative (FN) corresponds to the number of diseased samples that were incorrectly classified as healthy [35].

4. Experimental setting

To construct the training and test sets, we used thermograms from a total of 97 patients, 76 classified as healthy and 21 diagnosed with breast cancer. To preserve the independence between sets and avoid data leakage, all five images from a given patient were kept within the same set, either training or test. The model evaluation was carried out using 5-fold stratified cross-validation ($k = 5$), ensuring that images from the same patient did not appear simultaneously in both training and test sets. In each fold, approximately 80% of patients were used for training and 20% for testing, maintaining the class distribution.

¹<http://visual.ic.uff.br/dmi/>

To address class imbalance, data augmentation was applied exclusively to the training set. The goal was to artificially increase the number of samples from the minority class (*sick*) to match those of the majority class (*healthy*). Augmentation techniques included random rotations with a maximum amplitude of 45° , horizontal and vertical flips (each applied with a probability of 0.5), brightness, contrast, and saturation adjustments (all limited to a factor of 0.6, and hue to 0.3), as well as Gaussian blur with a kernel size randomly chosen between 5 and 9, and a sigma ranging from 0.1 to 5.

All images were resized to match the input dimensions required by each CNN: 224×224 for ResNet50 and DenseNet models, and 240×240 for EfficientNet-B0. Standard normalization using RGB channel statistics was applied, following ImageNet preprocessing conventions.

We implemented a custom multi-input architecture composed of five parallel networks, each processing one thermal image per patient (frontal, 90° , and 45° lateral views). Each network is based on a deep CNN pretrained on ImageNet, with its original classification layer replaced by an identity mapping to extract high-level feature vectors. The output feature vectors from the five subnetworks are concatenated into a single representation of size $5 \times (\text{feature size})$, which is then passed through a fully connected classification head.

This classification head is composed of a linear layer reducing the dimension to 128, followed by batch normalization, ReLU activation, dropout (rate = 0.4), and a final linear layer mapping to the two output classes (healthy vs. sick). This structure (feature concatenation $\rightarrow 128 \rightarrow 2$) was consistently applied across all evaluated models.

All networks were fine-tuned using the Adam optimizer (learning rate = $1e-3$, weight decay = $1e-4$), trained with cross-entropy loss for up to 150 epochs, batch size 32, and evaluated under a 5-fold stratified cross-validation scheme to ensure robust performance estimates.

The implementation was developed in PyTorch, with support from torchvision, scikit-learn, and Pillow, and training was performed on a GPU-equipped machine (Windows 10 Pro, Intel Core i7, 32GB RAM, NVIDIA GeForce GTX 1050 Ti – 4GB).

5. Results and Discussion

Table 1 presents the classification accuracy obtained across the five cross-validation folds for each evaluated CNN architecture. In addition to the per-fold results, the table reports the average accuracy and the standard deviation (SD), which helps quantify the variability in performance across different data splits. Among the models, ResNet50 achieved the highest mean accuracy (92.79%), with the lowest variability (SD = 2.80), indicating both strong and consistent performance.

EfficientNetB0 and DenseNet201 followed closely with identical average accuracies (91.74%), though EfficientNetB0 exhibited greater variability (SD = 4.20) compared to DenseNet201 (SD = 2.87). DenseNet121 showed a slightly lower mean (89.74%) and moderate variability. Interestingly, while

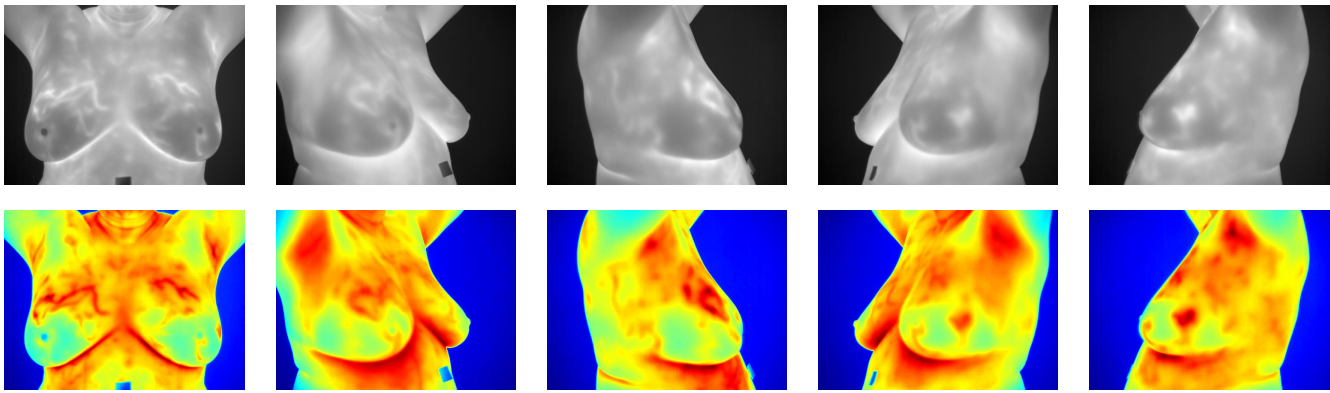


Figure 1. Set of images of a single patient in grayscale (top row) and colormap (bottom row).

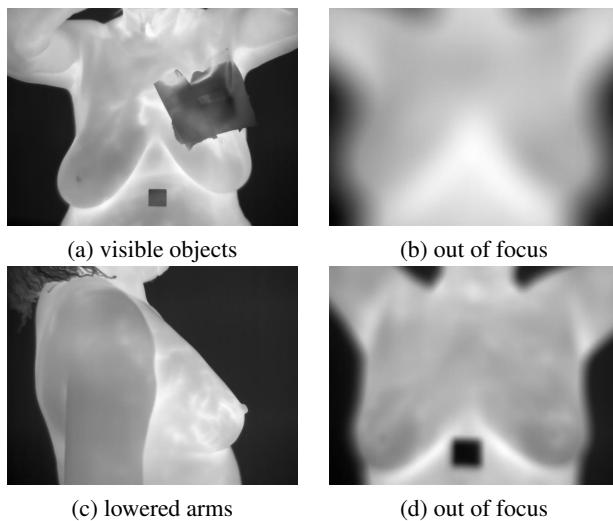


Figure 2. Removed images from the dataset due to artifacts or anomalies.

InceptionV3 had the lowest average accuracy (90.74%), it also displayed one of the highest standard deviations ($SD = 4.19$), highlighting fluctuations across folds. Overall, despite similar average accuracies across the architectures, the standard deviation values underscore the importance of evaluating model stability through cross-validation, especially when dealing with small or heterogeneous medical imaging datasets.

To further explore the learning behavior of individual models, Figure 3 and Figure 4 present the training and validation accuracy curves for fold 1 of ResNet50 and DenseNet121, respectively. These curves demonstrate the potential of both architectures, which achieved high accuracy levels early in training and sustained competitive performance over time. These insights reinforce the importance of fold-level analysis in understanding model robustness and generalization capacity.

Table 2 presents the average performance metrics of five pre-trained CNN models evaluated on the DMR-IR dataset using 5-fold cross-validation. Among the models, ResNet50 achieved the highest overall accuracy (92.79%), precision

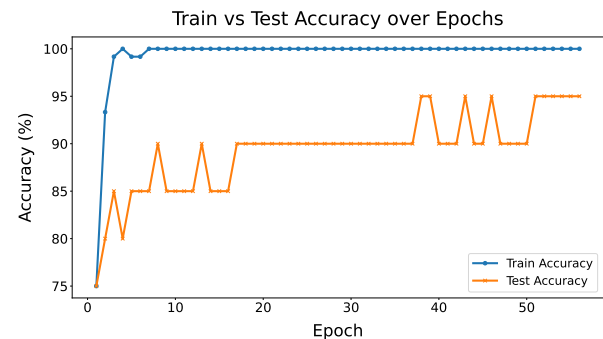


Figure 3. ResNet50 train and test accuracy - 1-fold

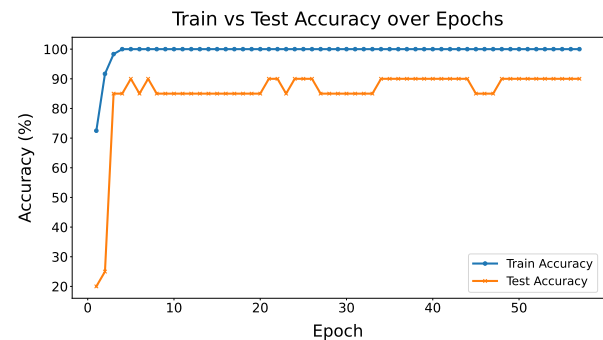


Figure 4. DenseNet121 train and test accuracy - 1-fold

(95.00%), specificity (98.67%), and F1-score (79.43%), indicating strong and balanced performance. It also reached the highest sensitivity (72.00%), a critical metric in medical diagnosis for minimizing false negatives. This makes ResNet50 particularly valuable for breast cancer detection, where correctly identifying diseased patients is essential.

While other models such as EfficientNetB0 and DenseNet201 also showed competitive performance—with accuracies of 91.74% and F1-scores close to 79%—they exhibited slightly lower sensitivity values (71.00%), which may lead to a higher number of missed positive cases. InceptionV3, despite achieving the same specificity (98.67%) and precision (95.00%) as ResNet50, had the lowest sensitivity (63.00%), indicating reduced ability to detect cancer cases.

Table 1. Accuracy (%) per fold, mean, and standard deviation (SD) for each CNN model

Method	1-fold	2-fold	3-fold	4-fold	5-fold	Mean $\pm$ SD
ResNet50	95.00	90.00	94.74	89.47	94.74	92.79 $\pm$ 2.80
EfficientB0	95.00	90.00	94.74	84.21	94.74	91.74 $\pm$ 4.20
DenseNet121	90.00	85.00	89.47	89.47	94.74	89.74 $\pm$ 3.45
DenseNet201	95.00	90.00	89.47	89.47	94.74	91.74 $\pm$ 2.87
InceptionV3	95.00	85.00	89.47	89.47	94.74	90.74 $\pm$ 4.19

Table 2. Average performance metrics

Method	Accuracy (%)	Precision (%)	Sensitivity (%)	Specificity (%)	F1-score (%)
ResNet50	92.79	95.00	72.00	98.67	79.43
EfficientB0	91.74	89.33	71.00	97.33	78.85
DenseNet121	89.74	90.00	67.00	97.42	71.90
DenseNet201	91.74	91.00	71.00	97.33	78.62
InceptionV3	90.74	95.00	63.00	98.67	74.05

We analyzed the average confusion matrix of ResNet50 across the five cross-validation folds (Figure 5). The results demonstrate the model's consistent ability to distinguish between healthy and sick cases, with only a single instance of a cancer patient incorrectly classified as healthy across all folds, highlighting its strong sensitivity and overall robustness.

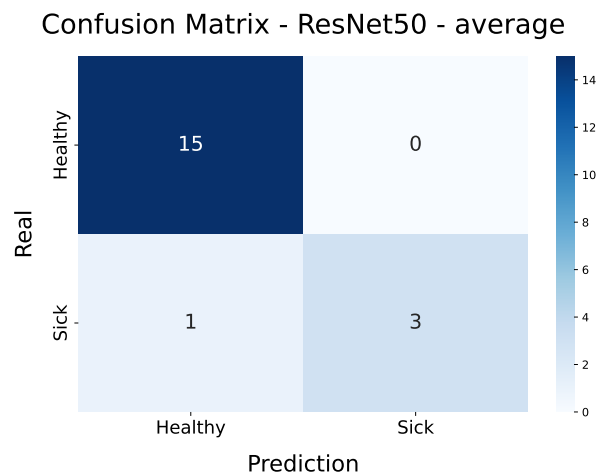
**Figure 5.** Confusion matrix ResNet50

Table 3 compares our method with previous studies using thermographic imaging for breast cancer detection, in the column of the table, the number of images refers to each patient. A key distinction of our approach is the use of all three anatomical views, frontal, lateral 90°, and lateral 45°, totaling five images per patient. This multi-angle configuration offers a more comprehensive thermal profile and remains uncommon in the literature.

Studies such as those by Venkatachalam et al. [32] and Nguyen et al. [15] rely solely on frontal images and still achieve very high accuracy values, surpassing 95% and even 99%. However, their performance may be influenced by controlled datasets or experimental designs that differ significantly in preprocessing or model complexity.

Some studies, like Madhavi et al. [17], Tsietso et al. [16],

Sánchez et al. [18], Tiwari et al. [19], and Mammoottil et al. [20], utilize three images per patient, typically combining frontal and lateral 90° views. These works demonstrate the potential benefit of adding anatomical diversity, with reported accuracies reaching up to 97%. However, many of them incorporate additional information beyond raw images: Madhavi et al. use manually defined regions of interest (ROIs) and texture features; Tsietso et al. also rely on ROIs; and Sánchez et al., Tiwari et al., and Mammoottil et al. include clinical features in their models. These additions can significantly influence results.

By contrast, our study does not use clinical data, ROIs, or handcrafted features. Our pipeline is fully image-based and end-to-end, relying only on the five thermal views of each patient processed through a deep learning model.

Focusing specifically on the work by Mammoottil et al. (2022), their reported accuracy of 85.4% and specificity of 84.4% are comparable to the results obtained in our study using ResNet50 (accuracy of 92.79% and specificity of 98.67%). Notably, the result from Mammoottil refers to the configuration without clinical data, which makes it a fairer baseline for comparison with our fully image-based method. However, our model presents a higher precision (95.00%), indicating that while it is better at avoiding false positives (as reflected in its higher specificity), it may still miss some true positive cases.

A notable strength of our approach lies in the exclusive use of raw thermographic images, without applying any preprocessing techniques, feature engineering, or manual enhancement steps. This design choice aims to simulate a realistic and clinically applicable diagnostic scenario, where automation is key and minimal human intervention is preferred. While omitting preprocessing enhances clinical applicability, it may reduce sensitivity, particularly in detecting true positives. Techniques such as image enhancement or ROI extraction could improve detection performance, but must be applied carefully to avoid excluding diagnostically relevant regions, such as the tail of Spence. Additionally, our evaluation employed a five-

Table 3. Comparison performance with literature

Reference	Number of images	Accuracy (%)	Precision (%)	Specificity (%)	ROC/AUC (%)
[32]	1	95.00	94.00	94.11	-
[15]	1	99.62	99.63	-	-
[17]	3	96.00	-	92.00	-
[16]	3	90.48	93.33	83.33	94.00
[18]	3	97.00	-	100	99.00
[20]	3	85.40	82.40	84.40	89.40
[19]	3	98.00	98.00	98.00	-
ResNet50 (Ours)	5	92.79	95.00	98.67	84.48

fold cross-validation strategy to improve the robustness and generalizability of the results by assessing the model across varied patient splits. Although this approach demands substantial computational resources, it provides a more reliable estimate of real-world performance than a single train-test split.

6. Conclusion

In conclusion, our evaluation of five pre-trained CNNs for breast cancer detection using thermal images from the DMR-IR dataset. Our robust 5-fold cross-validation approach provided a comprehensive assessment of each model's performance.

Among the tested models, ResNet50 achieved the best overall results, with the highest average accuracy (92.79%), precision (95.00%), specificity (98.67%), and F1-score (79.43%). It also obtained the highest sensitivity (72.00%), which is particularly critical in breast cancer screening, where minimizing false negatives is essential. These results demonstrate that ResNet50 offers a strong balance between detecting cancer cases and avoiding false alarms, making it the most effective model in our study.

Although other models like EfficientB0 and DenseNet201 also achieved high accuracy and specificity, their slightly lower sensitivity and F1-scores suggest reduced effectiveness in identifying true positive cases compared to ResNet50. Notably, InceptionV3, despite matching ResNet50 in precision and specificity, presented the lowest sensitivity (63.00%), limiting its reliability in clinical screening contexts.

A major strength of our approach lies in the use of raw thermograms from three anatomical views (frontal, lateral 90°, and lateral 45°), a five-image configuration that provides a richer thermal representation of the breast region. Unlike other studies that rely on clinical data, handcrafted features, or manual ROI selection, our pipeline is fully image-based and end-to-end, reflecting a more practical and scalable clinical application.

Future efforts will concentrate on enhancing both sensitivity and overall model robustness. We plan to explore advanced data augmentation techniques and investigate automated ROI extraction to further improve feature learning and diagnostic accuracy.

Acknowledgements

This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) – Finance Code 001. A.R.B. gratefully acknowledges the financial support of CNPq (Grant #302790/2024-1). H.F. gratefully acknowledges the financial support of CNPq (Grant #312530/2023-4). The funder had no role in the design of the study, data collection, analysis, publishing, or preparation of the manuscript.

Author Contributions

Renata dos Santos Melo contributed to experiments, conception, design, statistical analysis, and writing; Henrique Coelho Fernandes contributed to conception, statistical analysis, writing, and funding; André Ricardo Backes contributed to supervision, writing, statistical analysis, and funding.

References

- [1] INCA. *Estimativa 2014: Incidência de câncer no Brasil*. 2014. Disponível em: <https://www.inca.gov.br/bvscontrolecancer/publicacoes/Estimativa_2014.pdf>.
- [2] VIEIRA, S. C. et al. *Oncologia básica*. Teresina, MA: Fundação Quixote, 2012.
- [3] SMETHERMAN, D. H. Screening, imaging, and image-guided biopsy techniques for breast cancer. *Surgical Clinics*, Elsevier, v. 93, n. 2, p. 309–327, 2013.
- [4] ETEHADTAVAKOL, M.; NG, E. Y. Breast thermography as a potential non-contact method in the early detection of cancer: a review. *Journal of Mechanics in Medicine and Biology*, World Scientific, v. 13, n. 02, p. 1330001, 2013.
- [5] NG, E.-K. Improved sensitivity and specificity of breast cancer thermography. In: *Cancer Imaging*. [S.l.]: Elsevier, 2008. p. 435–444.
- [6] MORALES-CERVANTES, A. et al. An automated method for the evaluation of breast cancer using infrared thermography. *EXCLI journal*, Leibniz Research Centre for Working Environment and Human Factors, v. 17, p. 989, 2018.

- [7] CHEBBAH, N. K.; OUSLIM, M.; BENABID, S. New computer aided diagnostic system using deep neural network and svm to detect breast cancer in thermography. *Quantitative InfraRed Thermography Journal*, Taylor & Francis, v. 20, n. 2, p. 62–77, 2023.
- [8] FAUST, O. et al. Application of infrared thermography in computer aided diagnosis. *Infrared Physics & Technology*, Elsevier, v. 66, p. 160–175, 2014.
- [9] SANTOS, L. C. et al. A computing platform to analyze breast abnormalities using infrared images. *Medical & Biological Engineering & Computing*, Springer, v. 61, n. 2, p. 305–315, 2023.
- [10] KHAN, A. A.; ARORA, A. S. Thermography as an economical alternative modality to mammography for early detection of breast cancer. *Journal of Healthcare Engineering*, Hindawi, v. 2021, 2021.
- [11] HUSAINI, M. A. S. A. et al. A systematic review of breast cancer detection using thermography and neural networks. *IEEE Access*, IEEE, v. 8, p. 208922–208937, 2020.
- [12] KAYALIBAY, B.; JENSEN, G.; SMAGT, P. van der. Cnn-based segmentation of medical imaging data. *arXiv preprint arXiv:1701.03056*, 2017.
- [13] CIVILIBAL, S.; CEVIK, K. K.; BOZKURT, A. A deep learning approach for automatic detection, segmentation and classification of breast lesions from thermal images. *Expert Systems with Applications*, Elsevier, v. 212, p. 118774, 2023.
- [14] PRAMANIK, R.; PRAMANIK, P.; SARKAR, R. Breast cancer detection in thermograms using a hybrid of ga and gwo based deep feature selection method. *Expert Systems with Applications*, Elsevier, v. 219, p. 119643, 2023.
- [15] CHI, T. N. et al. A lightweight method for breast cancer detection using thermography images with optimized cnn feature and efficient classification. *Journal of Imaging Informatics in Medicine*, Springer, p. 1–18, 2024.
- [16] TSIETSO, D. et al. Multi-input deep learning approach for breast cancer screening using thermal infrared imaging and clinical data. *IEEE Access*, IEEE, v. 11, p. 52101–52116, 2023.
- [17] MADHAVI, V.; THOMAS, C. B. Multi-view breast thermogram analysis by fusing texture features. *Quantitative InfraRed Thermography Journal*, Taylor & Francis, v. 16, n. 1, p. 111–128, 2019.
- [18] SÁNCHEZ-CAUCE, R.; PÉREZ-MARTÍN, J.; LUQUE, M. Multi-input convolutional neural network for breast cancer detection using thermal images and clinical data. *Computer Methods and Programs in Biomedicine*, Elsevier, v. 204, p. 106045, 2021.
- [19] TIWARI, D.; DIXIT, M.; GUPTA, K. Deep multi-view breast cancer detection: A multi-view concatenated infrared thermal images based breast cancer detection system using deep transfer learning. *Traitement du Signal*, v. 38, n. 6, p. 1699–1711, 2021.
- [20] MAMMOOTIL, M. J. et al. Detection of breast cancer from five-view thermal images using convolutional neural networks. *Journal of Healthcare Engineering*, Wiley Online Library, v. 2022, n. 1, p. 4295221, 2022.
- [21] TIWARI, D.; DIXIT, M.; GUPTA, K. Breast cancer-caps: a breast cancer screening system based on capsule network utilizing the multiview breast thermal infrared images. *Turkish Journal of Electrical Engineering and Computer Sciences*, v. 30, n. 5, p. 1804–1820, 2022.
- [22] MAHORO, E.; AKHLOUFI, M. A. Breast cancer classification on thermograms using deep cnn and transformers. *Quantitative InfraRed Thermography Journal*, Taylor & Francis, v. 21, n. 1, p. 30–49, 2024.
- [23] SAMANTARAY, S. et al. Breast cancer diagnostics using angular view infrared images. In: *SPIE. Medical Imaging 2024: Imaging Informatics for Healthcare, Research, and Applications*. [S.l.], 2024. v. 12931, p. 181–189.
- [24] PATTERSON, J.; GIBSON, A. *Deep learning: A practitioner's approach*. [S.l.]: "O'Reilly Media, Inc.", 2017.
- [25] O'SHEA, K.; NASH, R. An introduction to convolutional neural networks. *arXiv preprint arXiv:1511.08458*, 2015.
- [26] ALSHEHRI, A.; ALSAEED, D. Breast cancer detection in thermography using convolutional neural networks (cnns) with deep attention mechanisms. *Applied Sciences*, MDPI, v. 12, n. 24, p. 12922, 2022.
- [27] MACAŞOI, I.-R.; NEAGOE, V.-E. Breast cancer detection using thermal infrared image analysis based on dempster-shafer decision fusion of cnn classifiers. In: *IEEE. 2023 15th International Conference on Electronics, Computers and Artificial Intelligence (ECAI)*. [S.l.], 2023. p. 01–06.
- [28] ABDEL-NASSER, M.; MORENO, A.; PUIG, D. Breast cancer detection in thermal infrared images using representation learning and texture analysis methods. *Electronics*, MDPI, v. 8, n. 1, p. 100, 2019.
- [29] MAHMOUD, A. Automatic breast thermography images classification based on deep neural networks. *Annals of Optimization Theory and Practice*, Firouzabad Institute of Higher Education, v. 3, n. 2, p. 71–79, 2020.
- [30] TIWARI, D.; DIXIT, M.; GUPTA, K. Performance comparison of the resnet50 and inceptionv3 deep transfer learning models over the breast cancer thermogram dataset. *BSSS Journal of Computer*, v. 13, n. 1, p. 1–10, 2022. ISSN 0975-7228.
- [31] HUSAINI, M. A. S. A. et al. Thermal-based early breast cancer detection using inception v3, inception v4 and modified inception mv4. *Neural Computing and Applications*, Springer, v. 34, n. 1, p. 333–348, 2022.

- [32] VENKATACHALAM, N. et al. Automated breast boundary segmentation to improve the accuracy of identifying abnormalities in breast thermograms. *IETE Journal of Research*, Taylor & Francis, v. 70, n. 2, p. 1462–1471, 2024.
- [33] MOHAMED, E. A. et al. A novel cnn pooling layer for breast cancer segmentation and classification from thermograms. *Plos one*, Public Library of Science San Francisco, CA USA, v. 17, n. 10, p. e0276523, 2022.
- [34] SILVA, L. et al. A new database for breast research with infrared image. *Journal of Medical Imaging and Health Informatics*, American Scientific Publishers, v. 4, n. 1, p. 92–100, 2014.
- [35] BARATLOO, A. et al. Part 1: Simple definition and calculation of accuracy, sensitivity and specificity. *Emerg (Tehran)*, v. 3, n. 2, p. 48–49, 2015. PMID: 26495380; PMCID: PMC4614595.