ESANN 2026 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium) and online event, 22-24 April 2026, i6doc.com publ., ISBN 9782875870964. Available from http://www.i6doc.com/en/.

XAI-Enabled Custom CNN for Cross-Modal Generalization in Breast Cancer Detection

Maram[4], Amal Jlassi[4], Abir Baâzaoui[1,2] and Walid Barhoumi[1,3]

1- Higher Institute of Computer Science, University of Tunis ElManar, LR16ES06 Laboratoire de recherche en Informatique, Modelisation et Traitement de l'Information et de la Connaissance, Ariana, Tunisia

2- Higher Institute of Computer Science and Management, Univ-K, Tunisia

3- National Engineering School of Carthage, University of Carthage

4- University of Ibn Khaldoun (UIK), La Soukra, Tunisia

Abstract. This paper presents a unified deep learning framework for breast cancer detection that generalizes effectively across mammography and histopathology. Using fine-tuned CNN architectures evaluated under a consistent cross-modal protocol, the method achieves stable, high accuracy on both imaging types, demonstrating robustness to domain shifts and heterogeneous clinical conditions. Another key contribution is the integration of model-agnostic (LIME, SHAP) and model-specific (Grad-CAM) explainability techniques, enabling a balanced trade-off between performance and interpretability. This hybrid XAI strategy provides clinically meaningful visual and feature-level insights, supporting transparent, reliable, and multi-modal diagnostic decision-making.

1 INTRODUCTION

Breast cancer remains a major public health concern, posing a significant challenge in the fight against cancer among women worldwide. The key to successful management of this disease lies in early detection, an essential backbone for improving survival rates and reducing the devastating impact of breast cancer [1]. For example, breast cancer remains the leading cancer among Tunisian women. In a recent cohort of 1,262 patients, the 5-year overall survival reached 72%. Current incidence and mortality rates are 50.17 and 14.04 per 100,000 women annually [2]. Breast cancer diagnosis relies on heterogeneous imaging modalities, including mammography, magnetic resonance imaging (MRI), ultrasound, and histopathology. Mammography is the primary imaging modality for breast cancer screening because of its non-invasive nature, affordability [3], and low radiation exposure [4], enabling the detection of microcalcifications and early-stage abnormalities. When further assessment is required, MRI is used due to its higher sensitivity in identifying suspicious lesions [5]. Despite the value of imaging for early detection, definitive breast cancer diagnosis depends on histopathology, which analyzes biopsied tissue at the cellular level to differentiate benign from malignant tumors and evaluate tumor aggressiveness [6]. This microscopic assessment provides essential information for diagnostic confirmation, treatment planning, and accurate pre-surgical evaluation that cannot be obtained through imaging alone.

ESANN 2026 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium) and online event, 22-24 April 2026, i6doc.com publ., ISBN 9782875870964. Available from http://www.i6doc.com/en/.

The integration of artificial intelligence, particularly deep learning, has greatly improved the accuracy and efficiency of breast cancer diagnosis [7, 8]. However, most existing approaches are modality-specific and show limited generalization across domains. To address this limitation, this work proposes a robust method capable of consistent performance across two distinct imaging modalities, mammography and histopathology, whose differences in scale, texture, and diagnostic features make cross-modal robustness a key indicator of real-world clinical applicability.

Furthermore, The black-box nature of deep learning models often hinders their clinical adoption, highlighting the need for Explainable Artificial Intelligence (XAI) to elucidate model decisions [9]. Model-agnostic methods such as SHAP and LIME, which analyze prediction changes under input perturbations, have proven effective in identifying clinically relevant regions in mammography [10, 11] and histopathology [12, 13]. In parallel, model-specific techniques like Grad-CAM and saliency maps leverage CNN internals to generate detailed visual explanations. Combining high-performing deep learning models with both model-agnostic and model-specific XAI methods is therefore crucial to balance diagnostic accuracy and interpretability in medical imaging.

To this end, this paper introduces a unified and generalizable deep learning framework for breast cancer detection that operates reliably across two distinct imaging modalities. By employing fine-tuned CNN model, the study demonstrates stable and high accuracy on both image types, highlighting the robustness of the approach to domain shifts and heterogeneous clinical environments. Yet, another key contribution lies in integrating both model-agnostic (LIME) and model-specific (Grad-CAM) XAI techniques, enabling a balanced trade-off between predictive performance and interpretability, and providing clinically meaningful visual and feature-level insights.

2 Proposed Method

This paper introduces a transparent CNN-based framework for cross-modal breast cancer detection using mammography and histopathology images, aiming to enhance both diagnostic accuracy and interpretability. The approach combines cross-modal image preprocessing with a fine-tuned, task- and modality-aware CNN that adapts existing learned representations without architectural novelty. Chosen over deeper predefined models such as ResNet or EfficientNet, this design prioritizes a balanced trade-off between performance, transparency, and clinical interpretability. Finally, explainability methods are applied to the model outputs to support reliable and clinically meaningful decision-making.

2.1 Cross-Modal Image Preprocessing

In this section, we ensure the quality and consistency of the dual-modal images by correcting missing values, normalizing variables, creating new features, and adjusting column names and data types. After verifying data integrity, images are extracted to construct a substantially reduced, curated dataset. To

ESANN 2026 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium) and online event, 22-24 April 2026, i6doc.com publ., ISBN 9782875870964. Available from http://www.i6doc.com/en/.

maintain dataset consistency and computational efficiency, all images from both mammography and histopathology modalities were resized to a fixed resolution of 50×50 pixels using linear interpolation to reduce computational cost, mitigate overfitting, and maintain sufficient visual information for binary classification.

2.2 Model Processing

The proposed fine-tuned CNN is a deep convolutional model for binary classification of mammography and histopathology images, built from successive 3×3 convolutional blocks with batch normalization and max-pooling. The resulting feature maps are flattened into fixed-length vectors to preserve fine-grained spatial information important for cross-modal lesion characterization, which may be partially lost when using global pooling. The resulting features are processed through fully connected layers with ReLU activation and dropout to mitigate overfitting, followed by a two-unit softmax output. The dataset was split into 80% training and 20% testing with a fixed random state (42), and data augmentation was applied to improve generalization. Training was conducted for 25 epochs with batch size gradually increased from 25 to 75, adaptive learning-rate adjustment, and early stopping based on validation loss with learning-rate reduction on plateau to ensure stable convergence and optimal generalization.

2.3 Custom CNN Explainability

To enhance the interpretability of the proposed fine-tuned CNN, three XAI techniques were applied: Integrated Gradients, Saliency Maps, and LIME. Integrated Gradients highlights the regions of each image that contribute most to the model's prediction by accumulating gradients along interpolated inputs. Saliency Maps provide a gradient-based visualization, identifying the pixels that have the strongest influence on the predicted class. LIME (Local Interpretable Model-agnostic Explanations) takes a different approach by generating superpixel-based explanations that approximate the model's behavior locally, highlighting the areas that support the predicted label. By combining these complementary techniques, a clearer understanding of the CNN's internal reasoning is achieved, ensuring transparency, trustworthiness, and visual validation of the learned representations.

3 Experimental Results

3.1 Dataset and Performance Metrics

This study utilizes two complementary datasets, leveraging both microscopic and macroscopic perspectives of breast cancer: a large-scale histopathology and the CBIS-DDSM mammography datasets. The first one comprises 277,524 RGB image patches extracted from breast tissue biopsies. The class distribution includes 78,786 cancerous patches (28.4%) and 198,738 non-cancerous patches (71.6%). The CBIS-DDSM (Curated Breast Imaging Subset of the Digital Database for

ESANN 2026 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium) and online event, 22-24 April 2026, i6doc.com publ., ISBN 9782875870964. Available from http://www.i6doc.com/en/.

Screening Mammography) dataset contains curated and standardized full-field digital mammograms, each accompanied by expert radiological annotations describing lesion type, pathology outcome, and region-of-interest segmentation.

To measure the effectiveness and suitability of the proposed custom model, several metrics are used, including accuracy (ACC), $F_1 - Score$, and area under the curve (AUC).

3.2 Fine-tuned CNN Results

To qualitatively demonstrate the effectiveness of the proposed custom CNN, Fig. 1 presents representative test samples from histopathology and mammography datasets, processed independently by the model. For each modality, the predicted diagnosis (cancerous/non-cancerous for histopathology and benign/malignant for mammography) is compared with the corresponding ground-truth annotation provided in the dataset. This qualitative evaluation illustrates the model's ability to learn modality-specific features and accurately classify breast tissue.

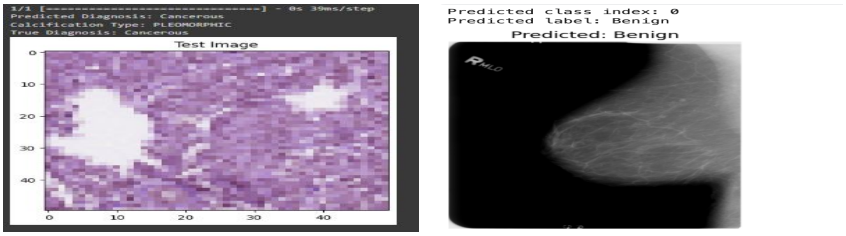

Fig. 1: Detection results on histopathology (left image) and mammogram (right image), showing predicted diagnosis.

In comparison with other models trained on the same histopathology dataset using transfer learning approaches (Table 1), our custom CNN demonstrates exceptionally high accuracy, highlighting its effectiveness and ability to generalize well to the specific characteristics of the mammogram and histopathology datasets.

3.3 Explainability Experiments

Fig. 2 presents the explainability results for a sample image from the histopathology dataset. In the first column, we show the original images along with their true and predicted labels. The following columns illustrate the outputs of three different explainability techniques applied to our custom CNN: Integrated Gradients, Saliency Maps, and LIME. Together, these visualizations help us understand how the CNN makes its decisions, showing which tissue patterns are most influential for classification and confirming that the model focuses on the relevant parts of the images.

ESANN 2026 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium) and online event, 22-24 April 2026, i6doc.com publ., ISBN 9782875870964. Available from http://www.i6doc.com/en/.

Table 1: Performance evaluation of the state-of-the-art models, against the proposed custom CNN model under histopathology (CNN-Histo), and mammogram (CNN-Mammo) datasets.

Model	AUC	$F_1 - Score$	ACC
Resnet-50-TL	0.861	0.858	0.861
Mobilenet-TL	0.854	0.855	0.854
Resnet-50v2-TL	0.843	0.854	0.843
Resnet-101v2-TL	0.832	0.833	0.832
Densenet-201-TL	0.821	0.823	0.821
Xception-TL	0.816	0.82	0.816
Densenet-169-TL	0.812	0.816	0.812
Resnext-101-TL	0.806	0.82	0.806
Nasnet-1-TL	0.804	0.802	0.804
Inception-v3-TL	0.778	0.781	0.779
Inception-Resnet-v2-TL	0.744	0.746	0.744
VGG16-TL	0.644	0.654	0.644
Proposed CNN-Histo	**0.96**	**94,24**	**97.38**
Proposed CNN-Mammo	**0.98**	**97.94**	**98.89**

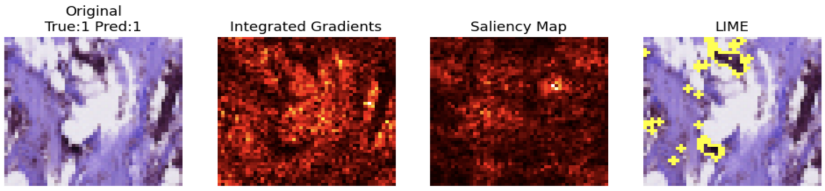

Fig. 2: Explainable AI results: Integrated Gradients, Saliency Maps, and LIME.

To further examine the model's internal reasoning on CBIS-DDSM, Gradient-weighted Class Activation Mapping (Grad-CAM) was applied to representative test image (Fig. 3). The visualization consists of three components: original mammogram (first column), Grad-CAM activation heatmaps indicating regions of highest model influence (second column), and overlaid heatmaps where color intensity represents attention strength (final columns). The model tends to focus on diffuse, poorly localized regions instead of specific lesions, indicating insufficient learning of diagnostically relevant features. These visualizations reinforce the quantitative findings and underscore the need for longer training, better class balancing, and greater malignant sample diversity to achieve clinically meaningful representations.

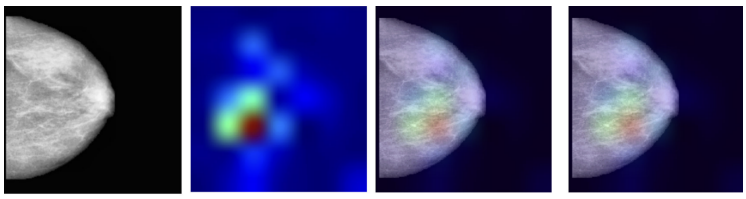

Fig. 3: Grad-CAM explainability results on mammography image.

ESANN 2026 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium) and online event, 22-24 April 2026, i6doc.com publ., ISBN 9782875870964. Available from http://www.i6doc.com/en/.

4 Conclusion

This study evaluated fine-tuned CNN model for breast cancer detection using two complementary imaging modalities. This model showed strong discriminatory capability, effectively capturing cellular-level morphological patterns associated with malignancy. Explainability analyses further clarified these differences: Grad-CAM produced well-focused, clinically coherent activation maps. Future research will prioritize enhancing model robustness through hierarchical feature extraction mechanisms.

References

[1] M. Abderrahim, A. BaÃ¢zaoui, and W. Barhoumi, Parallel hierarchical multi-view feature fusion based on canonical correlation analysis for mammogram re- trieval. *Vietnam Journal of Computer Science*, 2024.

[2] N. Mejri, H. Rachdi, L. Kochbati, and H. Boussen, General Oncology Care in Tunisia, In: H.O. Al-Shamsi, I.H. Abu-Gheida, F. Iqbal, A. Al-Awadhi (eds) *Cancer in the Arab World*, Springer, Singapore.

[3] M. Lawson, S. Partridge, D.S. Hippe, and et al., Comparative performance of contrast-enhanced mammography, abbreviated breast mri, and standard breast mri for breast cancer screening, *Radiology*, 308, 2023.

[4] N. Tamam, H. Salah, M. Rabbaa, M. Abuljoud, A. Sulieman, M. Alkhorayef, and D. Bradley, Evaluation of patients radiation dose during mammography imaging procedure. *Radiat. Phys. Chem.*, 188, 2021.

[5] YF. Lian, DQ. Liu, H. Wang, AF. Lian, Y.F. Liu, DQ. Wang, and TI. Hui, Meta-analysis of the synergistic effect of magnetic resonance imaging and mammography in breast cancer detection in women with dense breasts, *American Journal of Translational Research*, 17, 2025.

[6] HH. Luong, PP. Hong, DV. Minh, TNL Quang, AD. The, N. Thai-Nghe, HT. Nguyen, Principal component analysis and fine-tuned vision transformation integrating model explainability for breast cancer prediction, *Visual Computing for Industry Biomedicine and Art*, 8, 2025.

[7] F. Abdullakutty, Y. Akbari, S. Al-Maadeed, A. Bouridane, RR. Hamoudi, Enhancing the Prediction of Breast Cancer Progression Through Multi-modal Data Transformation, *Cognitive computation*, 17, 2025.

[8] M. Abderrahim, A. BaÃ¢zaoui, and W. Barhoumi, Survey on multi-source medical imaging fusion for classification and retrieval: Current status and available datasets *SN Computer Science*, 2025.

[9] O. Loukil, A. Baâzaoui, and W. Barhoumi, Mammography lexicon-based explainable artificial intelligence for diagnosis and visual interpretation of breast cancer. *In Advanced Concepts for Intelligent Vision Systems*, Springer, 2025.

[10] T. Murugan, P. Karthikeyan, and P. Sekar, Efficient breast cancer detection using neural networks and explainable artificial intelligence. *Neural Comput Applic*, 37, 2025.

[11] S. Barseghyan, A. Babajanyan, Z. Balyan and et al., Explainable ai for breast cancer risk prediction: evaluating the accuracy-explainability trade-off, *Mach. Learn. Comput. Sci. Eng*, 2025.

[12] T.K. Murugan, P. Karthikeyan, and P. Sekar, Efficient breast cancer detection using neural networks and explainable artificial intelligence, *Neural Computing and Applications* 37, 3759–3776, 2025.

[13] K.S. Manojee, and A.R. Kannan, Patho-Net: enhancing breast cancer classification using deep learning and explainable artificial intelligence, *American journal of cancer research*, 15(2), 754–768, 2025.