Explainability in Breast Cancer Detection

Ijaz Ahmad[1], Alessia Amelio[2], Daniela Cardone[2], Eliezer Zahid Gill[2] and Francesca Scozzari[3,*]

[1]Department of Human, Legal and Economic Sciences, Leonardo da Vinci University, Chieti, Italy
[2]Department of Engineering and Geology, University of Chieti–Pescara, Pescara, Italy
[3]Laboratory of Computational Logic and AI, Department of Economic Studies, University of Chieti–Pescara, Pescara, Italy

Abstract

Breast cancer is one of the most prevalent and lethal conditions among women across the globe, requiring timely and accurate diagnosis to contribute to better patient outcomes. Recent studies explored the risk factors connected to breast cancer. In premenopausal women and those with BRCA genetic susceptibility, air pollution predisposes to breast cancer because environmental toxins are more capable of inducing harmful results in these vulnerable groups, particularly those residing in densely populated urban areas with elevated pollution concentrations, and in the neighborhood of construction sites. Recent decades have seen deep learning emerging as a general-purpose piece of computer-assisted diagnosis software, enabling classification and segmentation tasks in the domain of medical imaging. These models are particularly effective at detecting weak patterns within imaging data imperceptible to the human eye, drastically enhancing diagnostic efficiency. This article focuses on the task of breast cancer classification using ultrasound images. Our results pinpoint ResNet50 as the best-performing model, which has a remarkable 98.72% accuracy rate. We further interpret the model's outcome using the XAI tool Grad-CAM by examining its capability to provide interpretable explanations. The XAI method provides clinically relevant and interpretable explanations, as supported by analysis using both the original images and their corresponding segmented masks.

Keywords

Breast Cancer, Air Pollution, Ultrasound Images, Explainable AI, Grad-CAM

1. Introduction

Breast cancer is a global health issue that strains healthcare systems, with early diagnosis and treatment improving outcomes and reducing mortality. Awareness and screening are key to controlling the disease. Besides genetics, environmental factors like air pollution contribute significantly to breast cancer risk [1]. Air pollution disrupts hormones, causes oxidative damage, and alters gene expression, increasing tumor risk. Several studies [2, 3, 4] link long-term exposure to fine particulate matter (PM2.5, PM10) and other pollutants to breast cancer, with building demolition particles influencing cancer cell behavior.

Artificial Intelligence (AI) is advancing breast cancer diagnostics, particularly through deep learning models applied to ultrasound images for early detection. More in general, deep learning plays a key role in medicine by enabling a wide range of applications that support both clinical practice and research. It is used for segmentation and classification of medical images [5] and for diagnosis and risk prediction [6, 7]. It also contributed to drug discovery [8, 9, 10], to enhance clinical decision support [11, 12] and to promote personalized medicine [13, 14].

However, ensuring transparency through Explainable AI (XAI) techniques is essential for clinical adoption [1, 15], since it improves personalized screening and counseling, especially in high-pollution areas. In fact, despite its benefits, the "black box" nature of deep learning models still limits clinical adoption. Transparent AI models build trust and support personalized strategies [16]. Interpretability is

Ital-IA 2025: 5th National Conference on Artificial Intelligence, organized by CINI, June 23-24, 2025, Trieste, Italy
*Corresponding author.
ijaz.ahmad@unidav.it (I. Ahmad); alessia.amelio@unich.it (A. Amelio); daniela.cardone@unich.it (D. Cardone); eliezerzahid.gill@phd.unich.it (E. Z. Gill); francesca.scozzari@unich.it (F. Scozzari)
0000-0001-6804-8025 (I. Ahmad); 0000-0002-3568-636X (A. Amelio); 0000-0002-1506-1995 (D. Cardone); 0000-0001-6274-9867 (E. Z. Gill); 0000-0002-2105-4855 (F. Scozzari)

Cases	Images	Dataset distribution	Images
Benign	437	Training	546
Malignant	210	Validation	78
Normal	133	Testing	156
Total	**780**	**Total**	**780**

Table 1
Details of the BUSI images used in the experiments

particularly important in breast cancer diagnosis, where understanding tumor characteristics guides treatment.

This research focuses on classifying breast cancer using ultrasound imaging, which offers radiation-free, real-time results, using XAI techniques such as Gradient-weighted Class Activation Mapping (Grad-CAM) [17] to ensure that deep learning models are accurate and transparent. Part of these results appeared in [18].

2. The proposed approach

This study combines XAI with the deep learning model ResNet50 to detect breast cancer using ultrasound images. Grad-CAM is used to generate visual explanations by highlighting areas of interest in a way that aligns with clinical insights to improve the model's transparency. The research is based on 780 labeled breast ultrasound images from the Breast Ultrasound Images (BUSI) dataset, which are divided into three categories: benign, malignant, and normal. The model's performance is evaluated using accuracy, precision, recall, and F1 score. The explainable visual outputs of the Grad-CAM heatmaps help clinicians interpret and validate the predictions from the model. The Grad-CAM method was employed to calculate the gradient of the target class with respect to the activations of the last convolutional layer for the ResNet50 model. The resulting gradients were averaged globally to calculate the importance weights for the feature map. The heatmap was created by performing a weighted sum of the feature maps followed by the application of the ReLU activation. The heatmap is used to identify the most impactful regions for the model's decision.

2.1. Dataset and preprocessing details

In this study, we utilize the BUSI dataset, which contains 780 preprocessed ultrasound images in PNG format, each image labeled as normal, benign, or malignant, along with corresponding segmentation masks. The dataset is divided into training (70%, 546), validation (10%, 78), and testing (20%, 156) subsets, each with a well-balanced class distribution across all partitions; these details are summarized in Table 1. This balanced distribution facilitates strong training as well as reliable testing of deep learning models for breast cancer detection. Mask overlay enhances visibility and marks affected regions clearly, increasing the accuracy in diagnostics over color mapping or zooming. Mask usage is essential since the identification of images without them can decrease accuracy by rendering delicate disease patterns more challenging to identify. The overlaid masks used in the current research were obtained from clinically verified segmentation, consistently marking clinically relevant regions.

2.2. Experimental setup and model training

The experiments were conducted in Python 3.10.13 using a NVIDIA A40 GPU, leveraging Tensor Cores, and CUDA Cores for accelerated visual computing. The pre-trained model with ImageNet weights was fine-tuned by removing top fully connected layers, freezing convolutional layers, and adding two dense layers of 128 units with ReLU activation, followed by a softmax output layer for three-class classification. Grad-CAM was applied by computing gradients of the predicted class score with respect to the last convolutional layer activations, averaging these gradients, and generating heatmaps that highlight image regions most relevant to the model's predictions. These heatmaps were overlaid on original images using a color map for visualization.

In this investigation, two types of experiments were performed: one on the classification using full ultrasound images without mask overlay and the other with segmentation overlay of the masks onto the original images. Given the primary focus is interpretability in the model, placing overlay over ultrasound images offers a valuable mechanism for cross-validation of affected areas. With this, we can measure whether the regions of interest (ROIs) of Grad-CAM match the clinician-validated regions, which helps build trust in the explanations given by the model and allows clinical relevance. Furthermore, we assess how the addition of segmentation masks affects both interpretability and classification accuracy. With this methodology, the model decision process is made transparent and clinically relevant.

3. Results and discussion

The proposed model with mask overlay achieves a remarkable accuracy of 98.72% and F1 score of 0.99, showcasing its robustness and effectiveness in classifying ultrasound images (see Table 2). These results emphasize the model's strong potential for accurate detection and diagnosis. The evaluation metrics collectively offer a comprehensive assessment of the model's predictive abilities, highlighting the superior performance of the approach. When combined with deep learning, ultrasound emerges as a cost-effective, efficient, and highly reliable method for breast cancer detection, especially in patients with dense breast tissue, where mammography tends to be less accurate. It is especially beneficial for susceptible populations like premenopausal women and those with BRCA germline susceptibility, who are further affected by environmental toxins and are at increased risk. Routine ultrasound screenings with the help of deep learning can enhance early detection and treatment in these individuals. While there are still limitations like ultrasound variability in imagery and aberrant pattern occurrence hindering conventional DL algorithms, innovations like Denoising Diffusion Probabilistic Models (DDPM) have tremendous potential for detecting such complex patterns, overcoming these limitations and advancing ultrasound-based diagnostics. Full ultrasound images include background tissue, noise, and artifacts like shadow and speckles, which can mislead the model and decrease accuracy. Masked ultrasound images only include essential regions such as tumors or lesions, which help the model disregard unnecessary artifacts and learn the meaningful aspects more efficiently. Targeted training typically results in enhanced classification performance.

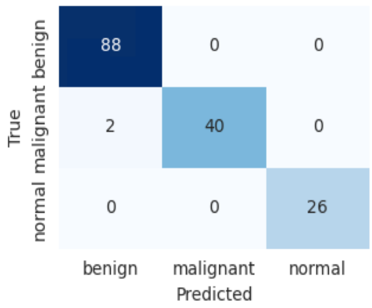
For Figures 1a and 1c, we have computed precision, recall and F1 score for each category separately. The results of Figure 1a are shown in Table 3a. With mask overlay, the model effectively distinguishes between the classes. It demonstrates perfect recall for benign and normal, perfect precision for malignant and normal, and near-perfect precision for benign. The high F1 scores across all classes indicate a strong balance between precision and recall. The model's performance is particularly robust for the normal class, with perfect scores in all metrics.

The results of Figure 1c are shown in Table 3b. Without mask overlay, the model performs worse in distinguishing between malignant, benign, and normal classes. For malignant cases, it achieves good precision (0.80) and poor recall (0.69), with a balanced F1 score of 0.74. For benign cases, the model has a good recall (0.89) and good precision (0.80), resulting in an F1 score of 0.84. For normal cases, it has good precision (0.82) but lower recall (0.72), yielding an F1 score of 0.77.

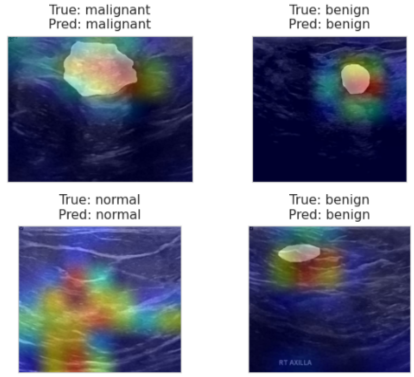
Figures 1b and 1d illustrate the effectiveness of the proposed model through Grad-CAM heatmaps generated from both masked and original ultrasound images. The visualizations in both figures evidently

Model Name	Image Type	Accuracy (%)	Precision	Recall	F1 Score
ResNet50	Masked Images	98.72	0.99	0.98	0.99
ResNet50	Original Images	80.13	0.81	0.77	0.78

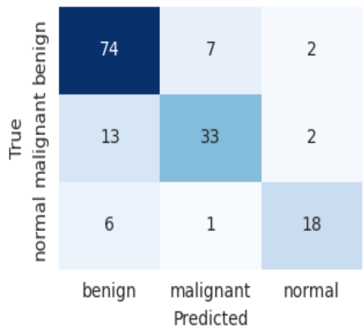
Table 2
Result of ResNet-50 model for masked images and original images



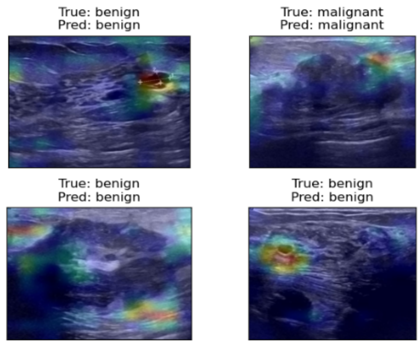
(a) Confusion matrix of ResNet50 masked images



(b) Grad-CAM heatmaps on masked images



(c) Confusion matrix of ResNet50 original images



(d) Grad-CAM heatmaps on original images

Figure 1: Confusion matrices and Grad-CAM heatmaps for ResNet50 on masked and original images.

identify infected areas, which indicates that the model can localize the relevant regions. Classification accuracy is relatively lower in the case of original ultrasound images than in the case of masked images. This is mainly because of ultrasound artifacts like speckle noise, acoustic shadows, and irrelevant anatomical structures, which can mislead the deep learning model and lower its accuracy. Masked images can eliminate these artifacts significantly by segmenting lesions and silencing background noises, thereby allowing the model to learn more discriminative and accurate features. Hence, the use of preprocessing or advanced segmentation techniques is critical to enhance model accuracy in the original ultrasound images. Overall, the results clearly demonstrate that deep learning combined with explainability methods like Grad-CAM provides valuable insights for breast cancer detection of the infected regions in both figures. However, the accuracy is lower in the case of original images, which means that original ultrasound images require novel preprocessing techniques to remove the artifact and related anomalies from ultrasound images. From the results, one thing is very clear: these approaches have the ability to classify the images correctly and help in breast cancer detection. As future work, we plan to apply abstract interpretation-based techniques [19, 20, 21] to verify the correctness of the code.

Class	Precision	Recall	F1 Score
Benign	0.98	1.00	0.99
Malignant	1.00	0.95	0.98
Normal	1.00	1.00	1.00

(a) Masked images.

Class	Precision	Recall	F1 Score
Benign	0.80	0.89	0.84
Malignant	0.80	0.69	0.74
Normal	0.82	0.72	0.77

(b) Original images.

Table 3
Precision, recall, and F1 score metrics for the Benign, Malignant, and Normal classes

4. Related works

AI has demonstrated strong performance in breast cancer classification [22, 23]. In particular, transfer learning approaches outperformed traditional methods in image classification [24]. In this context, the link between air pollution and breast cancer risk has gained significant attention, particularly in urban areas where traffic-related pollutants may elevate the risk [25]. In Tehran, pollutants such as ethylbenzene, xylene, and NO_2 were associated with advanced-stage breast cancer, especially in younger women and those from lower socio-economic backgrounds [26]. A systematic review confirmed a connection between long-term exposure to pollutants like PM2.5, benzo[a]pyrene, and NO_x, increasing breast cancer risk [27]. Similarly, a meta-analysis showed a slight increase in risk with NO_2 exposure, though the evidence remains inconclusive [28]. Many studies have highlighted the importance of XAI for augmenting trust in AI-based tools [29, 30]. ResNet50, combined with XAI methods like Integrated Gradient and Occlusion, provided interpretable cancer classifications with a high accuracy [4]. Other contributions include ensemble classifiers combining Xception, InceptionV3, and ResNet101, enhanced by Grad-CAM and attention mechanisms [31] and TransXAI for improving result interpretability [32].

Conflict of interest

We wish to confirm that there are no known conflicts of interest associated with this publication and there has been no significant financial support for this work that could have influenced its outcome.

Acknowledgments

This work was partially supported by the PNRR project FAIR — Future AI Research (PE00000013), Spoke 9 – Green-aware AI, under the NRRP MUR program funded by the NextGenerationEU. This study has been partially performed in the Laboratory of Computational Logic and AI and in the HPC Laboratory of the University of Chieti-Pescara, and in the Research Laboratory “Hugo Gernsback" of Unidav.

Declaration on Generative AI

The authors have not employed any Generative AI tools.

References

[1] Y. Sun, X. Meng, D. Sun, et al., Long-term exposure to fine particulate matter interacting with individual conditions increase breast cancer incidence: a large-scale chinese cohort, Environmental Sciences Europe 37 (2025) 1–11.

[2] J. Santos, C. Pereira, M. Fiore, et al., Air pollution and breast cancer risk: An umbrella review, Environments 12 (2025) 153.

[3] I. Godet, D. Gilkes, Brca1 and brca2 mutations and treatment strategies for breast cancer, Integrative Cancer Science and Therapeutics 4 (2017) 10.15761/ICST.1000228.

[4] C.-W. Cheng, G.-T. Sheu, J.-S. Chou, et al., Fine particulate matter pm2.5 generated by building demolition increases the malignancy of breast cancer mda-mb-231 cells, Chemosphere 265 (2021) 129028.

[5] G. Thakur, A. Thakur, S. Kulkarni, et al., Deep learning approaches for medical image analysis and diagnosis, Cureus 16 (2024) e59507.

[6] M. Singh, A. Kumar, N. Khanna, et al., Artificial intelligence for cardiovascular disease risk assessment in personalised framework: a scoping review, eClinicalMedicine 73 (2024) 102660.

[7] M. Khalifa, M. Albadawy, Artificial intelligence for clinical prediction: Exploring key domains and essential functions, Computer Methods and Programs in Biomedicine Update 5 (2024) 100148.

[8] R. Alizadehsani, S. Oyelere, S. Hussain, et al., Explainable artificial intelligence for drug discovery and development: A comprehensive survey, IEEE Access (2024). doi:10.1109/ACCESS.2024.3373195.
[9] J. Jiménez-Luna, F. Grisoni, G. Schneider, Drug discovery with explainable artificial intelligence, Nature Machine Intelligence 2 (2020) 573–584.
[10] F. Wong, S. Omori, A. Li, et al., An explainable deep learning platform for molecular discovery, Nature Protocols (2024). doi:10.1038/S41596-024-01084-X, epub ahead of print 1 April 2024.
[11] J. Amann, D. Vetter, S. Blomberg, et al., To explain or not to explain?—artificial intelligence explainability in clinical decision support systems, PLOS Digital Health 1 (2022) e0000016.
[12] S. Kim, D. Kim, M. Kim, et al., Xai-based clinical decision support systems: A systematic review, Applied Sciences 14 (2024) 6638.
[13] Z. Zhou, M. Hu, M. Salcedo, et al., Xai meets biology: A comprehensive review of explainable ai in bioinformatics applications (2023). Https://arxiv.org/pdf/2312.06082v1 (accessed 27 May 2025).
[14] M. Gimeno, E. San José-Enériz, S. Villar, et al., Explainable artificial intelligence for precision medicine in acute myeloid leukemia, Frontiers in Immunology 13 (2022) 977358.
[15] O. Nicolis, D. De Los Angeles, C. Taramasco, A contemporary review of breast cancer risk factors and the role of artificial intelligence, Frontiers in Oncology 14 (2024) 1356014.
[16] J. Ahn, S. Shin, S. Yang, et al., Artificial intelligence in breast cancer diagnosis and personalised medicine, Journal of Breast Cancer 26 (2023) 405.
[17] R. Selvaraju, M. Cogswell, A. Das, et al., Grad-cam: Visual explanations from deep networks via gradient-based localization, Proceedings of the IEEE International Conference on Computer Vision 2017-October (2017) 618–626.
[18] I. Ahmad, A. Amelio, F. Ali, A. Merla, F. Scozzari, N. Ahmad, A comparative analysis of artificial intelligence methods for breast cancer interpretation, in: Proceedings of the International Joint Conference on Neural Networks (IJCNN), 2025. To appear.
[19] G. Amato, M. Rubino, F. Scozzari, Inferring linear invariants with parallelotopes, Science of Computer Programming 148 (2017) 161–188. doi:10.1016/j.scico.2017.05.011.
[20] G. Amato, M. C. Meo, F. Scozzari, On collecting semantics for program analysis, Theoretical Computer Science 823 (2020) 1–25. doi:10.1016/j.tcs.2020.02.021.
[21] G. Amato, F. Scozzari, The ScalaFix equation solver, in: M. Chechik, J.-P. Katoen, M. Leucker (Eds.), Formal Methods, 25th International Symposium, FM 2023, Lübeck, Germany, Proceedings, volume 14000 of *LNCS*, 2023, pp. 142–159. doi:10.1007/978-3-031-27481-7_10.
[22] N. Niehoff, M. Terry, D. Bookwalter, et al., Air pollution and breast cancer: An examination of modification by underlying familial breast cancer risk, Cancer Epidemiology, Biomarkers & Prevention 31 (2022) 422–429.
[23] R. Wang, P. Wang, Y. Zhou, et al., Association between long-term ambient air pollution exposure and the incidence of breast cancer: A meta-analysis based on updated evidence, Ecotoxicology and Environmental Safety 289 (2025) 117472.
[24] J. Madrigal, C. Pruitt, J. Fisher, et al., Carcinogenic industrial air pollution and postmenopausal breast cancer risk in the national institutes of health aarp diet and health study, Environment International 191 (2024) 108985.
[25] A. White, P. Bradshaw, G. Hamra, Air pollution and breast cancer: A review, Current Epidemiology Reports 5 (2018) 92–100.
[26] Z. Khorrami, M. Pourkhosravani, A. Karamoozian, et al., Ambient air pollutants and breast cancer stage in tehran, iran, Scientific Reports 14 (2024) 3873.
[27] J. Tippila, N. Wah, K. Akbar, et al., Ambient air pollution exposure and breast cancer risk worldwide: A systematic review of longitudinal studies, International Journal of Environmental Research and Public Health 21 (2024) 1713.
[28] D. Praud, F. Deygas, A. Amadou, et al., Traffic-related air pollution and breast cancer risk: A systematic review and meta-analysis of observational studies, Cancers 15 (2023) 927.
[29] C. Smotherman, B. Sprague, S. Datta, et al., Association of air pollution with postmenopausal breast cancer risk in uk biobank, Breast Cancer Research 25 (2023) 83.

[30] M. Fiore, M. Palella, E. Ferroni, et al., Air pollution and breast cancer risk: An umbrella review, Environments 12 (2025) 153.
[31] M. Duboeuf, A. Amadou, T. Coudon, L. Grassot, M. Ramel-Delobel, E. Faure, P. Salizzoni, J. Gulliver, G. Severi, F. R. Mancini, B. Fervers, D. Praud, Long-term exposure to air pollution at residential and workplace addresses and breast cancer risk: A case-control study nested in the french e3n-générations cohort from 1990 to 2011, European Journal of Cancer 210 (2024) 114293.
[32] R. Zeineldin, M. Karar, Z. Elshaer, et al., Explainable hybrid vision transformers and convolutional network for multimodal glioma segmentation in brain mri, Scientific Reports 14 (2024) 3713.