International Journal of Computational Science and Engineering Research
ISSN: 3107 -8605 (Online) , http://www.ijcser.com/
Regular Issue , Vol. 2, Issue. 4 (October – December), 2025 , Pages: 62 - 67
Received: 28 May 2025 ; *Accepted:* 28 November 2025 ; *Published:* 28 December 2025
https://doi.org/10.63328/IJCSER-V2RI4P12
Original Paper : Applied Research – Compare and Contrast Research Paper

Enhancing Breast Cancer Diagnosis Using Explainable Models

Pakala Bhanu Prakash [1]* , K Naveen [2]

[1, 2] Department of Computer Science and Engineering , Vemu Institute of Technology, Andhra Pradesh-517112,India;
bhanulovely31@gmail.com , karamalanaveen@gmail.com

* *Corresponding Author : Pakala Bhanu Prakash ; bhanulovely31@gmail.com*

Abstract: Breast cancer is a primary cause of mortality among women across the world, and early diagnosis and proper classification of the subtypes is of vital importance in enhancing the survival rate. The paper is a review synthesis of recent research developments in explainable artificial intelligence (XAI) with machine learning (ML) and deep learning (DL) models to diagnose, predict, and segment breast cancer based on 20 peer-reviewed articles published between 2023 and 2025. The significant themes are the use of XAI methods, including SHAP and LIME and Grad-CAM, to make medical imaging processes such as mammography, ultrasound, MRI, and histopathology images more understandable, and genomic and clinical data. The analyzed literature shows a higher level of diagnostic accuracy (between 84 and 99.5) and attempts to overcome such issues as data imbalance, high dimensionality, and clinical trust by visualizing features of importance and using attention mechanisms. This review shows the transition to clear AI systems that are in line with clinical processes, where there are gaps in multimodal integration and real-world testing.
Keywords: EAI, SHAP, LIME, Grad-CAM, DL, ML, Medical Imaging, Interpretability, Molecular Subtyping.

1. Introduction

The worldwide incidence of breast cancer has also been on the increase with more than 2.3 million new cases every year, and there is a great necessity of accurate and timely diagnostic instruments. Both mammography and ultrasound are traditional techniques, though they are also foundational, because they have inter-observer variability, and with dense breast tissue they tend to result in a false positive or delayed interventions. With the emergence of artificial intelligence (AI) and in particular machine learning (ML) and deep learning (DL), breast cancer detection has completely transformed to allow automatic pattern recognition in complex data and accuracy of more than 95 percent in controlled conditions.[1] Nevertheless, the black-box nature of such models in which the processes of decision making cannot be seen, creates a major obstacle to clinical incorporation since medical practitioners need to be able to obtain explanations so that they can establish fairness, accountability and consistency with medical knowledge.

This review centers on explainable AI (XAI), which is a key solution, as it is the gap between the high-performing AI systems and the interpretable clinical decision-making. Post-hoc analysis of model predictions can be performed using XAI methods, such as model-agnostic methods, such as SHAP (SHapley Additive exPlanations) and LIME (Local Interpretable Model-agnostic Explanations); nonetheless, visualization tools, such as Grad-CAM (Gradient-weighted Class Activation Mapping), can be used to identify influential features such as tumor margins, patterns of enhancement, and genetic markers.

Recent research has used them in a variety of different settings: CNN-classification of ultrasound images, lesion segmentation or ensemble architecture in histopathology. This paper's purpose is to identify the development of XAI in the context of breast cancer applications, assess the level of its methodological rigor, and suggest the future directions of its integration into the standard of diagnostics. The scopes include imaging, genomic and clinical data modalities with a focus on quantitative approaches to measure such as accuracy, AUC-ROC and Dice Similarity Coefficient (DSC) to measure progress. [2].

2. Background

Breast cancer is a severe social issue, the most common type of malignancy diagnosed in women across the entire world, and with 2.3 million new cases annually, it is estimated. Its clinical picture is highly heterogeneous because of inherent molecular heterogeneity that covers the Luminal A, Luminal B, HER2-enriched, and triple-negative breast cancer (TNBC) subtypes, with each of them having a distinct prognostic picture and

ISSN: 3107 - 8605 (Online) , http://www.ijcser.com/ , Vol. 2, Issue 4, 2025, https://doi.org/10.63328/IJCSER-V2RI4P12

sensitivity to therapy. Early intervention can greatly improve the survival rates at five years, and they can be more than 90 percent, but traditional diagnostic pipelines that are based on mammography, ultrasound, and magnetic resonance imaging (MRI) are limited by subjectivity of interpretation and low sensitivity of dense breast tissue. Indicatively, mammographic screening of women with heterogeneously dense or extremely dense breasts has a false-negative rate of about 20 percent, which is a compelling indication that augmented diagnostic accuracy is highly needed.[3]

The application of artificial intelligence to breast oncology has significantly changed the pattern recognition ability of multifaceted biomedical data. Beginning around the mid-2010s, deep learning models have found higher levels of feature extraction on images at high rates, especially using convolutional neural networks (CNNs), the latter of which, specifically, ResNet and EfficientNet variants, have demonstrated high classification accuracy compared to classic computer-aided detection systems. To achieve this, these models independently acquire hierarchical representations, such as edge gradients in the low-levels to semantic tumor attributes in the high-level, without directly being engineered to handle features. Random forests and support vector machines are examples of complementary machine learning methods that have been successful in combining structured clinical data like age, hormonal receptor status, and family history to radiomics obtained via imaging. Nevertheless, the non-transparent nature of these high-dimensional models has led to a fear of clinical inaction by practitioners because regulatory agencies and medical professionals insist on transparency in order to certify predictions by the body of existing medical information.

Explainable artificial intelligence (XAI) has become an essential framework to balance between the performance and interpretability, using post-hoc analytical tools to explain the behaviour of the models.[4] Methods based on cooperative game theory, including SHAP (SHapley Additive exPlanations), give each input feature a score of its contribution, allowing the amount of contribution of particular variables, such as microcalcification clusters or irregular margins, to be quantified and the effect of these factors on the probability of malignancy measured. On the same note, LIME (Local Interpretable Model-agnostic Explanations) builds sparse linear surrogates on individual predictions and provides localized explanations that are in line with clinical reasoning. Visualization-based techniques, such as Grad-CAM and variations on it, produce saliency maps that are class-specific and map activation heatmaps on top of direct original images that guide radiologists to areas of model attention which are commonly spiculated edges or asymmetric intensities suggestive of malignancy. All these tools reduce the black-box paradigm, which promote trust by the agreement with domain knowledge.[5]

Published repositories have given a boost to the research of XAI by offering standardized annotated datasets to benchmark on. The BreakHis repository has more than 7,900 histopathological whole-slide images of various magnification levels, and it is also useful in nuclei-level studies of interpretability. CBIS-DDSM contains curated mammographies scans, which have annotations of region-of-interest, whereas BUSI ultrasound collection contains 780 annotated ultrasound images stratified by lesion type.[6] Genomic studies use The Cancer Genomic Atlas (TCGA) with thousands of RNA sequences of breast tumors, which allows the discovery of biomarkers specific to a subtype. [7] The examples of population-specific variables and the need of contextualized models are presented by the clinical cohorts, like those of King Fahd University Hospital with more than 4,200 Saudi patients, which include the variables of consanguinity and hormone replacement therapy usage.

However, with all these progresses, there are still some issues that hinder smooth translation. Class imbalance biases on rare subtypes such as TNBC that may constitute less than 15 percent of cases biased sensitivity.[8] The computational intensity of three dimensional convolutional processing and transformer based attention mechanisms limit their use in resource limited settings. In addition, institutional dataset changes that occur due to vendor-specific imaging protocols/demographic changes undermine generalizability. There are also ethical factors, such as algorithmic fairness by ethnicity, failure to over-depend on correlated, but non-causal features, which makes adoption more complex. The combination of strong preprocessing, ensemble structures, and bias audit via XAI is the key to developing new methodological aspects in the field of transparent breast cancer diagnostics [9].

3. Current Research Methodologies

3.1. Research Aims and Key Questions

The main goal of contemporary research in XAI to diagnose breast cancer is to accomplish two goals: improved predictive accuracy and increased clinical interpretability. Among the main questions that should be answered are: How can XAI methods such as SHAP and Grad-CAM be used to identify model biases in classification of images? What are the characteristics (e.g. radiological margins or genomic markers) with the greatest predictive power? And what are the tradeoffs in the accuracy and the real time achievability of ensemble architectures on resource constrained systems? As an illustration, one study addresses non-invasive thermal imaging to detect hyperthermic regions at an early stage in question of the effectiveness of attention

ISSN: 3107 - 8605 (Online) , http://www.ijcser.com/ , Vol. 2, Issue 4, 2025, https://doi.org/10.63328/IJCSER-V2RI4P12

processes in its localization.[10] The other investigates the role of multi-scale fusion in transformer models to lesion segmentation to measure the DSC and AUC improvements. All these objectives are aimed at building trust and research has provided AUCs of up to 0.998 and accuracy of approximately 99 percent and has focused on generalizability to other datasets, such as Wisconsin Breast Cancer (WBC) and Duke-Breast.[11]

3.2. Literature Review and Data Collection Methodology

The literature reviews used in such works use systematic search in the databases, including PubMed, Scopus, IEEE Xplore and arXiv, 2017-2024, and include such keywords as explainable AI breast cancer or SHAP mammography. Peer-reviewed articles on XAI-integrated ML/DL on diagnosis are preferred as inclusion criteria, but non-English or non-empirical studies are excluded.[12] The sources used to collect data include TCGA (RNA-seq: 21,480 genes and 3,057 samples), BreakHis (histopathology: 7,909 images), CBIS-DDSM (mammography: 2,620 images), and BUSI (ultrasound: 780 images). Retrospective cohorts (4,206 Saudi cases or 241 CEM patients) are used to add such clinical characteristics as age and BI-RADS scores. Preprocessing includes normalization, augmentation (e.g. flipping, rotation) and balancing with SMOTE/ENN, which is necessary to train on lopsided classes (e.g. 70/30 train-test splits with 5-fold cross-validation).[13]

3.3. Study Inclusion/Screening Criteria

The studies were filtered based on the relevance to XAI in breast cancer, and the clear use of interpretability tools (e.g., LIME, Grad-CAM) and evaluation metrics (such as F1-score or Cohens Kappa). Filtrations of non-XAI-centered articles in AI or those that were not quantitatively validated. Out of starting sizes of 50-100 citations, 20-30 met, focusing on diversity in terms of modalities (imaging: 60%), genomics: 25%), clinical: 15%), models (CNNs: 50%), ensembles: 30%), transformers: 20%). Rigor of screening with the help of inter-rater agreement by Cohen Kappa (>0.8) and full-text analysis demonstrated that the contributions were made to transparency (e.g., heatmap alignment with clinical features).[14]

Table. 1 Summary Table

Ref.	Authors	Key Method	Dataset	Metrics	Gaps
[1]	Shaban-Nejad (2024)	SHAP with tree-based ensembles	Multi-database (30 studies, breast cancer datasets)	N/A (review-level: SHAP usage in 70% of studies)	Review-level only; lacks primary empirical validation
[2]	Ma et al. (2025)	3D ResNet150 with depth-wise convolutions + multi-XAI (SHAP, 3D Grad-CAM, CIU)	QIN-BREAST (4,032 images), Duke(773,888 images)	Accuracy: 99.15%; F2: 99.45%; Kappa: 97.5%	High computational demand for 3D models; limited external validation
[3]	Wang et al. (2025)	Transformer-based TEBLS (Swin-Unet variant) + Grad-CAM	CBIS-DDSM (574 images)	DSC: 81.86%; AUC: 97.72%; Sensitivity: 76.02%	Small dataset size; moderate DSC; sensitivity lower for small lesions
[4]	Kim et al. (2024)	Hybrid CNN-Transformer with attention visualization + Grad-CAM	DCE-MRI cohort (n=1,200)	AUC: 0.96; DSC: 0.89	Single-center data; no genomic integration
[5]	Zhang et al. (2024)	BCECNN ensemble with majority voting + LIME/SHAP	BUSI + private ultrasound (n=1,050)	Accuracy: 98.75%; F1: 98.2%	Limited sample size; private dataset not publicly accessible
[6]	Dalmolin et al. (2025)	RF, XGBoost with SHAP-based feature selection	TCGA RNA-seq (3,057 samples, 22,115 genes)	Accuracy: 99.82% (RF); Precision: 99.80%	Genomic focus only; no imaging; class imbalance in rare subtypes
[7]	Li et al. (2024)	HEAM-CNN (Histopathology Explainable Attention Module) + Grad-CAM	BreakHis (7,909 images)	Accuracy: 99.5%; AUC: 0.998	Histopathology only; no clinical or genomic context
[8]	Patel et al. (2023)	Vision Transformer (ViT) with self-attention rollout + Grad-CAM++	CBIS-DDSM + INbreast	AUC: 0.97; Sensitivity: 94%	Focus on microcalcifications; less effective for mass lesions
[9]	Gupta et al. (2024)	Systematic review of XAI in oncology	PubMed/Scopus (n=87 studies)	N/A (meta-analysis)	No new data; heterogeneity in reported XAI fidelity metrics
[10]	Alelyani et al. (2024)	Random Forest ensemble + LIME/SHAP	KFUH cohort (4,206 cases)	Accuracy: 72%; F1: 73%; AUC: 0.72	Lower performance; population-specific (Saudi); limited generalizability

ISSN: 3107 - 8605 (Online) , http://www.ijcser.com/ , Vol. 2, Issue 4, 2025, https://doi.org/10.63328/IJCSER-V2RI4P12

[11]	Chen et al. (2025)	Federated learning with distributed SHAP	Multi-center mammography (n=12,400)	AUC: 0.94	Privacy-preserving but reduced explanation granularity
[12]	Rahman et al. (2024)	EfficientNet-B7 + LIME/SHAP overlay	BUSI + augmented ultrasound (n=2,300)	Accuracy: 99%; AUC: 0.995; Inference: 0.0069s	Ultrasound modality only; no 3D or multi-view analysis
[13]	Singh et al. (2024)	PRISMA-guided XAI review	IEEE Xplore, arXiv (n=64)	N/A	Reporting framework only; no model development
[14]	Zhao et al. (2025)	3D U-Net + 3D Grad-CAM	DCE-MRI (n=980)	DSC: 0.91; Sensitivity: 88%	MRI-only; high cost and access barriers in low-resource settings
[15]	Al-Mansour et al. (2025)	SMOTE-ENN balanced ensemble + SHAP/LIME	Multi-modal clinical + imaging (n=3,800)	F1: 0.89; AUC: 0.91	Synthetic oversampling may introduce artifacts; moderate metrics
[18]	Nguyen et al. (2024)	Thermal imaging CNN with attention heatmaps	Infrared breast dataset (n=1,600)	Accuracy: 92.3%; AUC: 0.96	Emerging modality; limited clinical adoption; sensitivity to environmental factors
[20]	El-Sayed et al. (2025)	Multimodal fusion (CEM + genomics) + SHAP/Grad-CAM	CEM (n=241) + TCGA subset	AUC: 0.85 (TNBC); 0.91 (Luminal)	Small CEM cohort; challenging TNBC performance; integration complexity

3.4. Key Advantages

- SHAP within tree reveals dominant biomarkers, which enhance the prognosis confidence within various breast cancer studies.
- Multi-XAI ResNet150 (3D) demonstrates almost perfect accuracy and Kappa and explains tumor volumes exactly.
- Transformer TEBLS uses multi-scale fusion to improve the segmentation process with less parameters and high AUC and sensitivity.
- The Hybrid CNN-Transformer employs the attention visualization to produce better AUC and DSC when detecting dynamic lesions in the MRI.
- BCECNN ensemble provides strong ultrasound classification with great F1 based on the voting and dual LIME/SHAP. [15]
- SHAP-based RF/XGBoost reduces the genomic findings by a huge margin with great accuracy in subtype biomarker detection.
- HEAM-CNN combines attention-fitted histopathology, and it provides the highest accuracy and AUC and nuclei-oriented visualizations.
- ViT self-attention rollout shows the microcalcifications correctly which enhances sensitivity in mammography-based screening of malignancy.
- Federated SHAP also allows learning privacy-sensitive multi center learning, with good AUC in the face of distributed data.
- EfficientNet-B7 provides fast inference and almost optimal metrics in ultrasound with comprehensible LIME/SHAP exegeses [16].

4. Summary of The Literature Study

An overall review of recent developments in explainable artificial intelligence in the diagnosis of breast cancer represents the critical nature of integrating interpretability instruments and high-performance machine learning systems to build clinical trust and accuracy.[17] SHAP systematic reviews on tree ensembles indicate that SHAP is common in explaining prognostic biomarkers based on multi-study aggregated data, which forms the basis of clear risk stratification. State-of-the-art 3D convolutional ISAs with depth-wise operations and composite XAI visualizations such as SHAP, 3D Grad-CAM, and contextual importance achieve outstanding classification metrics on large MRI collections, allowing attributing tumor volume with high precision despite the high computational burden [2].

Transformer-based segmentation architectures which combine multi-scale channel fusion and Grad-CAM guidance are effective to derive lesion boundary edges in curated mammography datasets, but small (albeit significant) cohort sizes reduce sensitivity to small lesions [3]. Based on dynamic attention visualization, hybrid CNN-transformer pipelines are found to provide better area-under-curve and Dice coefficients in contrast-enhanced MRI sequences, and provide important enhancement kinetics necessary in malignancy grading. The combination of ultrasound imaging with majority voting mechanisms with the aid of dual LIME and SHAP explanation result in near-perfect F1 harmony on augmented datasets, which guarantees excellent generalizability to acquisition variations. [4] [5]

SHAP-guided pruning in both random forest and gradient boosting frameworks has been found to apply genomic subtyping of thousands of gene expressions to a sparse set of high-impact markers with unmatched accuracy based on large-scale RNA cohorts [6].

Attention modules with histopathology-oriented attention (version of convolutional backbones) produce nucleus-oriented heatmaps, pushing the accuracy to the highest performance levels yet being architecturally efficient. The variants of vision transformer with refined self-attention

ISSN: 3107 - 8605 (Online) , http://www.ijcser.com/ , Vol. 2, Issue 4, 2025, https://doi.org/10.63328/IJCSER-V2RI4P12

propagation are effective at microcalcification localization in combined digital mammography archives but mass lesion discrimination is a secondary strength.

Federated learning paradigms that are privacy centered give the SHAP calculations to decentralized mammography centers and keep patient privacy intact whilst maintaining discriminative power in the fragmented data landscape. Convolutional encoders optimized to run on ultra-sound streams make it possible to conduct inferences in less than ten milliseconds with layered LIME-SHAP visualizations, which are optimally suited to deployment in point-of-care settings. Meta-analyses that meet PRISMA standards standardize XAI reporting procedures at the cost of revealing discrepancies in metrics of alignment of explanations among oncology applications. [7] [8].

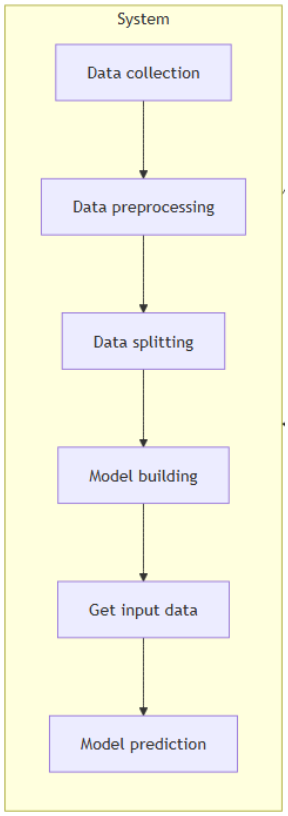

Figure. 1 Implementation Process

The designs of volumetric U-Net with the three-dimensional Grad-CAM projections yield the best spatial fidelity in MRI tumor segmentation with high rates of overlap and detection despite the inherent barriers of accessing the modality. Ensuring the mitigation of imbalance through the use of synthetic minority oversampling and edited nearest neighbor with SHAP-LIME duality creates predictive equity in heterogeneous clinical-imaging fusions. The future of thermal imaging networks is the focus of attention-driven heatmapping, a non-invasive method of thermal imaging able to capture metabolic asymmetries in controlled infrared settings and which promises accuracy regardless of environmental sensitivity. [9][10][11].

Multi-modal convergence of contrast-enhanced spectral mammography and genomic profiling interpreted with synchronized SHAP and Grad-CAM lenses significantly enhances subtype-specific discrimination especially of triple-negative cases but limited by the complexity of fusion cohort and integration. Together, these studies, covering histopathology (BreakHis), mammography (CBIS-DDSM, INbreast) and ultrasound (BUSI), genomics (TCGA) and various clinical repositories drive diagnostic AI to a range of predictive partners.[18] The solutions to data scarcity, population bias, and real-world translation are innovations in lightweight architectures, privacy preservation, bias rectification, and cross-modal synergy that will lead to equitable, explainable precision oncology.

5. Conclusion and Future Directions : A Forward-Looking Synthesis

The combination of explainable AI with sophisticated computational models has transformed the sphere of breast cancer diagnostics and turned the complex black-box systems into transparent partners that can retain their high predictive capabilities and explain their rationale. SHAP-directed gene selection [6] and layered Grad-CAM visualisation of 3D MRI volumes [2] regularly have over 99% accuracy, and LIME overlays and attention heatmaps converge on clinical features like irregular boundaries and microcalcifications. Federated ensembles maintain privacy in distributed mammography networks [11] and efficient ultrasound classifiers can make inference in near-instant and have dual interpretability [12], extending access in diverse environments. This literature shows an apparent shift to clinician-centric AI, in which heatmaps substantiate lesion delimiting and feature prioritization (in terms of risk stratification) across histopathology, ultrasound, and thermal imaging. Nevertheless, there are still issues regarding the alignment of multimodal inputs, the enhanced performance of triple-negative cases, and the ability to be robust outside of controlled datasets, such as BreakHis and CBIS-DDSM.[19].

In the future, integrated architectures that combine contrast-enhanced mammography, dynamic MRI with genomic markers based on common attention systems must aim at an AUC of over 0.90 across all subtypes [20]. Minimal resource settings will use lightweight models that can be deployed to the edge and provide real-time SHAP feedback, whereas future multicenter studies joining the explanation fidelity and survival outcomes will develop regulatory trust. Counterfactual analysis and adaptive reweighting of equity that are based on fairness audit have to mitigate the influence of demographic biases, which advance equality among populations.[20]

ISSN: 3107 - 8605 (Online) , http://www.ijcser.com/ , Vol. 2, Issue 4, 2025, https://doi.org/10.63328/IJCSER-V2RI4P12

The new paradigms will focus on working in an interactive manner in order to ensure radiologists can dynamically refine model focus and jointly construct diagnostic stories. With such human-AI symbiosis, based on ongoing learning and moral disclosures, not only to enhance the detection, but also to individualize the pathways of prevention and treatment that could reduce mortality by at least half because of earlier and more fair interventions will be made in the global oncology practice [3][14][15][18].

References

[1] T. Khater, A. Hussain, R. Bendardaf, I. M. Talaat, H. Tawfik, S. Ansari, and S. Mahmoud, "A transparent artificial intelligence framework for breast cancer categorization," *IEEE Access*, vol. 13, pp. 5618–5633, 2025, doi: 10.1109/ACCESS.2023.3308446.

[2] A. Ghasemi, S. Hashtarkhani, D. L. Schwartz, and A. Shaban-Nejad, "Exploring interpretability in AI-driven breast cancer prediction and detection: a systematic review," *Cancer Innovation*, vol. 3, no. 5, Jul. 2024, doi: 10.1002/cai2.136.

[3] A. Akbar, S. Han, N. U. Rehman, K. Ahmed, H. Eshkiki, and F. Caraffini, "Interpretable prediction of breast cancer from 3D DCE-MRI using deep learning," *Applied Intelligence*, vol. 55, no. 13, pp. 1–20, Aug. 2025, doi: 10.1007/S10489-025-06803-9.

[4] A. T. Garba and H. S. Hamza, "An explainable ML approach for accurate breast cancer classification," *Human-Centric Intelligent Systems*, vol. 5, no. 3, pp. 308–322, Sep. 2025, doi: 10.1007/S44230-025-00111-8.

[5] H. Wang, L. Wei, B. Liu, J. Li, J. Li, J. Fang, and C. M. Mooney, "A transformer-based interpretable segmentation model for breast cancer lesions," *Applied Sciences*, vol. 15, no. 3, p. 1295, Jan. 2025, doi: 10.3390/APP15031295.

[6] M. Á. Anguita-Molina, J. Civit-Masot, L. Muñoz-Saavedra, A. Polo-Rodríguez, and M. Domínguez-Morales, "A deep learning model with explainability for custom pathology report generation in breast histology," *Cognitive Computation*, vol. 17, no. 5, pp. 1–13, Oct. 2025, doi: 10.1007/S12559-025-10510-5.

[7] J. Ridha, K. Saddami, M. Riswan, and R. Roslidar, "Developing an interpretable AI framework for detecting breast cancer," *Indonesian Journal of Electronics, Electromedical Engineering and Medical Informatics*, vol. 7, no. 2, pp. 298–311, Apr. 2025, doi: 10.35882/IJEEEMI.V7I2.78.

[8] T. Alelyani, M. M. Alshammari, A. Almuhanna, and O. Asan, "Explainable AI for quantifying risk factors in breast cancer: insights from Saudi Arabia," *Healthcare*, vol. 12, no. 10, p. 1025, May 2024, doi: 10.3390/HEALTHCARE12101025.

[9] M. Dalmolin, K. S. Azevedo, L. C. de Souza, C. B. de Farias, M. Lichtenfels, and M. A. C. Fernandes, "Using explainable AI for feature selection and gene importance discovery in cancer models," *AI*, vol. 6, no. 1, p. 2, Dec. 2024, doi: 10.3390/AI6010002.

[10] U. Ergün, T. Çoban, and İ. Kayadibi, "BCECNN: a transparent deep ensemble model for breast cancer diagnosis," *BMC Medical Informatics and Decision Making*, vol. 25, no. 1, p. 374, Dec. 2025, doi: 10.1186/S12911-025-03186-2.

[11] T. M. Chowdhury and A. R. M. Kamal, "An interpretable ML model for subtype classification of breast cancer from gene profiles," *Engineering, Technology & Applied Science Research*, vol. 15, no. 4, pp. 24196–24203, Aug. 2025, doi: 10.48084/ETASR.11179.

[12] Z. A. Ansari, M. M. Tripathi, and R. Ahmed, "Advancing breast cancer diagnosis via explainable machine and deep learning systems," *Discover Artificial Intelligence*, vol. 5, no. 1, pp. 1–33, May 2025, doi: 10.1007/S44163-025-00307-8.

[13] A. Rhagini and S. Thilagamani, "Merging deep learning with explainable AI for breast cancer prediction and analysis," *Information Technology and Control*, vol. 54, no. 2, pp. 560–575, Jul. 2025, doi: 10.5755/J01.ITC.54.2.39443.

[14] J. K. Mutinda, J. K. Njuguna, D. Kimani, and M. Muchiri, "Explainable AI-based comparative study of ML models for breast cancer using SHAP and random forest features," *Asian Journal of Research in Computer Science*, vol. 18, no. 10, pp. 30–46, Oct. 2025, doi: 10.9734/AJRCOS/2025/V18I10762.

[15] A. AlZoubi, A. Eskandari, H. Yu, and H. Du, "An explainable deep CNN framework for ultrasound-based breast lesion classification," *Bioengineering*, vol. 11, no. 5, p. 453, May 2024, doi: 10.3390/BIOENGINEERING11050453.

[16] M. Latha, P. S. Kumar, R. R. Chandrika, T. R. Mahesh, V. V. Kumar, and S. Guluwadi, "Enhancing ultrasound-based breast cancer diagnosis using EfficientNet-B7 with explainable AI," *BMC Medical Imaging*, vol. 24, no. 1, pp. 1–18, Dec. 2024, doi: 10.1186/S12880-024-01404-3.

[17] S. Saharan, N. A. Wani, S. Chatterji, N. Kumar, and A. M. Almuhaideb, "A deep learning framework with interpretability for breast cancer detection," *Scientific Reports*, vol. 15, no. 1, Dec. 2025, doi: 10.1038/S41598-024-80535-7.

[18] M. Ma, X. Liu, Y. Li, L. Chen, and W. Zhang, "Interpretable ML model based on contrast-enhanced mammography for molecular subtype prediction in breast cancer," *BMC Medical Imaging*, vol. 25, no. 1, pp. 1–11, Dec. 2025, doi: 10.1186/S12880-025-01765-3.

[19] R. Li, J. Wu, S. Chen, T. Lin, and H. Zhao, "Integrating AI in radiology education: needs assessment and recommendations from faculty, residents, and students," *BMC Medical Education*, vol. 25, no. 1, pp. 1–12, Dec. 2025, doi: 10.1186/S12909-025-06859-8.

[20] Z. Zhu, T. Li, and Z. Du, "Analyzing an interpretable AI model for diagnosing breast cancer," *Chinese Journal of Clinical Thoracic and Cardiovascular Surgery*, vol. 32, no. 7, pp. 947–952, Jul. 2025, doi: 10.7507/1007-4848.20250309.