

---

## Recent Advances in Breast Cancer Detection and Diagnosis: A Comprehensive Survey of Machine Learning, Deep Learning, and Explainable AI Approaches

V.Niranjana\*, P.Kanimozhi, T.Ananth Kumar

Department of Computer Science and Engineering, IFET College of Engineering, Villupuram, Tamil Nadu, India.

\*Correspondence Author Email: vniranjana@gmail.com

Received: 03.03.2026

Revised: 19.03.2026

Accepted: 03.04.2026

Published: 15.04.2026

---

### Abstract

The early and precise diagnosis of breast cancer is one of the most important factors in enhancing survival rates. Despite considerable improvements in screening methods, medical imaging, and treatment approaches, the issue of reliable early detection remains a significant problem due to instability in clinical interpretation and the complexity of diagnostic information. Conventional methods of diagnosis emphasize the analysis of clinical and histopathological information, which are prone to inter-observer error and do not necessarily reveal subtle patterns related to disease progression. Artificial Intelligence has demonstrated enormous potential in alleviating these weaknesses. Ensemble methods, as well as support vector machines and gradient boosting-based algorithms such as XGBoost and LightGBM, have shown excellent results in the analysis of structured clinical data. Simultaneously, deep learning (DL) systems, specifically convolutional neural networks (CNNs), have been demonstrated to be exceptionally successful in medical image analysis such as mammography and histopathology. However, the interpretability of such models remains a major obstacle to their usage in clinical practice. To address this problem, Explainable Artificial Intelligence (XAI) methods, including SHAP, LIME, and Grad-CAM, have been proposed to improve model transparency and contribute significantly to understanding how decisions are made. This paper reviews recent developments in ML, DL, and XAI in relation to breast cancer diagnosis. It examines commonly used datasets, evaluation metrics, and trends, and also reports the most significant issues, including data constraints, lack of interpretability, and challenges in multimodal integration. Moreover, this paper identifies knowledge gaps in existing research and highlights the necessity of developing robust, explainable, and actionable AI-based diagnostic systems.

**Keywords:** Breast Cancer; Machine Learning; Deep Learning; Explainable Artificial Intelligence (XAI); Medical Imaging

---

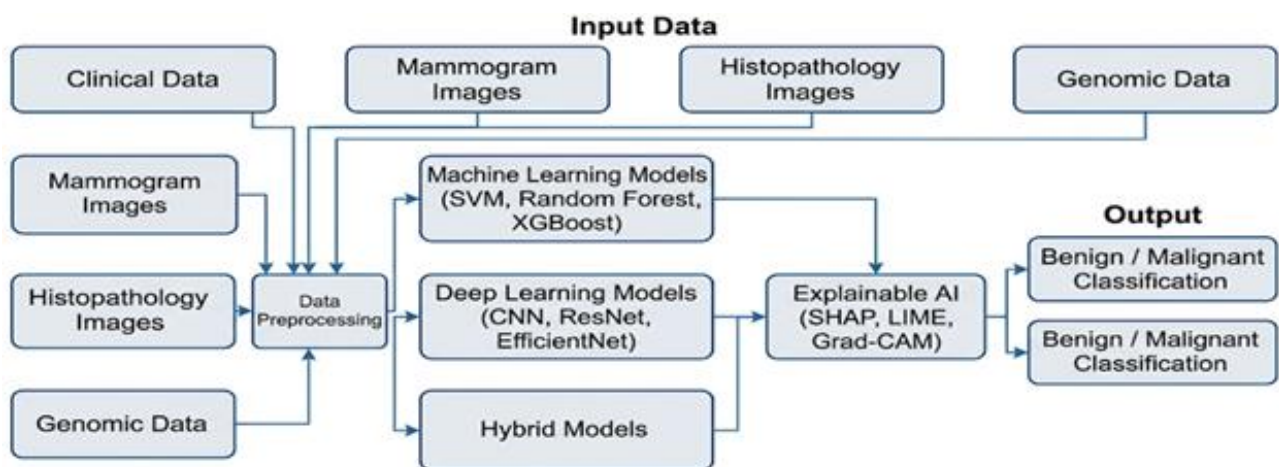
### 1. Introduction

Breast cancer remains a major cause of fatalities among women worldwide, stating the necessity of effective diagnostic procedures. The global health reports indicated that breast cancer attack was on the increase because of the lifestyle changes, the aging of people, and the awareness that has contributed to positive screening procedures. Early diagnosis and detection are very crucial in improving the survival rates and decreasing the complexity of the treatment. The classic diagnostic techniques that are common in clinical practices are mammography, ultrasound imaging, magnetic resonance imaging (MRI) and biopsy tests [1]. These methods are effective, however, it can be time consuming, costly and require an expert interpretation unlike other methods, which can result in inconsistencies in the results of a diagnosis. Artificial intelligence (AI) technologies have become the pioneer to assist medical workers with detecting diseases and making decisions in the past few years. Machine Learning (ML) and Deep Learning (DL) approaches have already demonstrated the promise of delivering desirable outcomes in the area of automating the task of breast cancer classification, based on the clinical parameter and medical imaging data. These systems aim at improving the diagnostic accuracy, reducing the human error, and helping the physicians to identify the malignant and benign tumors at the initial stages. The growing access to medical data and the growth of the computational capabilities have added impetus to the study of this area even more.

Machine learning methods are aimed at studying organized clinical data containing the numerical variables (tumor size, tumor texture, tumor radius, tumor smoothness, and other cellular variables). The Support Vector Machines, the Logistic Regression, the Random Forest, the k-Nearest neighbors and the ensemble have all extensively been applied in the classification of breast tumors using the supervised learning algorithms, which in this instance learns the pattern using the labelled data and formulates predictive models which can be used to classify the cases of cancer and the non-cancer ones. The most commonly utilized performance evaluation measures of model efficiency are accuracy, precision, recall, F1-score, sensitivity and specificity [2]. Convolutional Neural Networks (CNNs) and other deep learning methods have been discovered to provide the best possible results in the analysis of medical images (mammograms and histopathology slides). Contrary to the classical models of machine learning that

are based on the manual extraction of features, deep learning architectures learn hierarchical representations of features through raw image data by default. The transfer learning methods based on pre-trained models have become popular as well, allowing effective training with a small amount of medical data. Such mechanisms help in higher classification and strength in the image-based breast cancer detection systems [3]. Although high predictive performance has been realized, most ML and DL models are black-box systems, which give little insight into their actions. Interpretability is an essential concept in healthcare applications since medical practitioners need clear reasons before they can accept AI-driven tools in the clinic. To overcome this shortcoming, the Explainable Artificial Intelligence (XAI) techniques have been embraced to induce transparency and trust [4]. The analysis of feature importance, visualization, local explanation models, and attention mechanisms are among methods that enable the clinician to have insights on which features or image locations contributed the most to the outcome of the prediction. The explainability is not only beneficial in terms of user trust, but it can aid not only in the regulation but also in ethical use of AI in healthcare settings. The availability and the diversity of datasets is also another crucial factor in breast cancer research. Accessible publicly available datasets of clinical records and medical images have helped researchers to test and compare different algorithms in similar conditions. Nevertheless, the problems of imbalance in classes, small size of the sample, the heterogeneity of the data, and privacy remain to influence the generalization of the models to real-life applications. To address these shortcomings and increase the reliability of the system, researchers are actively working on the data augmentation, preprocessing methods, and hybrid learning schemes. Most of the literature published over the last ten years suggests various machine learning, deep learning, and interpretable artificial intelligence algorithms to detect breast cancers [5]. Although a number of surveys have discussed either machine learning or deep learning frameworks separately, there exists a need to have a study that consolidates the recent developments in the fields of ML, DL, and XAI. It is necessary to understand the strengths, limitations, datasets in use, evaluation metrics, and methodological trends to determine the gaps in the research and to lead to future developments of intelligent diagnostic systems.

The purpose of the survey is to make a comprehensive review of the recent achievements in the area of breast cancer detection in terms of the machine-learning, deep-learning, and explainable AI methods. Frequently used algorithms, publicly available datasets, performance appraisal strategies, and interpretability methods are considered. Moreover, it also discusses about the current problems and outlines the future research directions in order to create the accurate, reliable, and transparent systems of breast cancer diagnostics. This paper aims to address the state of AI-based healthcare solutions evolution and help to advance the creation of clinically capable decision-support systems by synthesizing the results of recent literature. The general structure of a breast cancer detection system based on AI is depicted in Figure 1. To start with, multimodal data, such as clinical records, medical images, and genomic data, are collected, followed by the preprocessing and feature extraction of the obtained data. Diverse machine learning, deep learning, and combinations are then utilized in classification. Lastly, explainable AI methods (SHAP, LIME, and Grad-CAM) are also included to increase interpretability and facilitate clinical decision-making.



**Figure 1** General framework of AI-based breast cancer detection system

## 2. Literature Survey

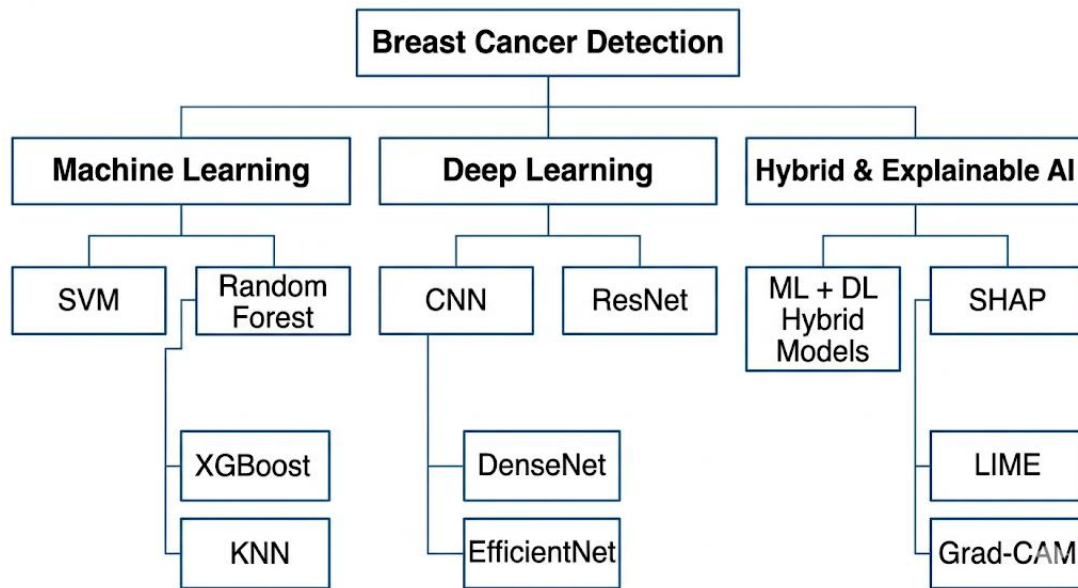
### Machine Learning-Based Approaches

Machine learning (ML) tools have been widely applied to the breast cancer prediction with structured clinical and diagnostic data and provide credible and interpretable results. Various research studies have tested various ML algorithms in order to improve classification. Arravalli et al. (2025) and Rathore et al. (2024), demonstrated ensemble models including the Random Forest model and stacked learning models showed high performance in terms of high F1-scores [6], [7]. Likewise, Alzahrani et al. (2025) proposed with a hybrid ML model, which was 99.3% accurate with balanced samples, which is robust and can generalize [8]. Srinivasu et al. (2024) proposed a CatBoost-MLP hybrid model combined with SHAP to analyze the importance of their features, demonstrating a high predictive accuracy [9], whereas Abd Karim et al. (2025) combined association rule mining with ensemble learning to successfully identify the essential biomarkers with the help of the explainability methods [10]. Silva-Aravena et al. developed a decision support model on the basis of ML models, XGBoost showed better predictive results, and SHAP was utilized to present explainable results about clinical variables [11]. In the similar way, cost-sensitive CatBoost coupled with LIME has been used to resolve the issue of class imbalance and augment metastasis forecasting in addition to providing patient-specific explanations [12]. SHAP-based methods involving a combination of feature selection techniques with LIME and decision tree-based models have increased the level of model interpretability and performance as well [13].

The conventional ML classifiers e.g. Random Forest, SVM, KNN, and ensemble voting techniques all perform well with their accuracy rates going beyond 98% in previous studies [14-16]. Moreover, patient registration-based screening models that are based on MLs have demonstrated effectiveness in early risk detection with sensitivity and specificity that are balanced [17]. More innovations involve ML models to predict the involvement of lymph nodes and heart risks and show the growing scope of ML use in clinical decision support systems [18-19]. These results indicate that ML methods are still highly efficient and in particular with the help of feature engineering and explainability. Similar findings have also been reported in other studies focusing on feature-based classification, hybrid algorithms, and stage prediction in breast cancer diagnosis [20-23]. Based on Table 1, conventional machine learning models like Random Forest, SVM, and ensemble methods have high accuracy of classification with a high probability (up to 98). Ensemble and hybrid methods also increase performance and robustness, whereas combinations with explainability methods such as SHAP and LIME can increase interpretability. These models however rely mainly on structured data and feature engineering. Figure 2 shows the taxonomy of AI approaches to breast cancer detection. The solutions are loosely grouped into machine, deep learning and hybrid techniques incorporated with methods of explainable AI. The categories contain well-known algorithms and models, which indicate the development of the traditional approach to cutting-edge deep learning and explainable systems.

**Table 1 Comparison of Machine Learning-Based Approaches**

Ref No.	Method / Model	Dataset Type	Key Contribution	Performance
[6-7]	Random Forest, Stacked Ensemble	Clinical data	Ensemble improves classification	High F1-score
[8]	Hybrid ML Model	Balanced dataset	Robust classification	99.3% accuracy
[9]	CatBoost + MLP + SHAP	Clinical data	Feature importance using SHAP	99.3% accuracy
[10]	CAR + RF + GB	Clinical data	Biomarker identification	Improved interpretability
[11]	XGBoost + SHAP	Clinical data	Decision support system	0.81 accuracy
[12]	Cost-sensitive CatBoost + LIME	Clinical data	Handles imbalance + explanation	F1: 77%
[13]	DT + SVM + SHAP/LIME	WBCD	Feature selection + rule extraction	Improved performance
[14]	Random Forest	WDBC	Best ML classifier	98.6% accuracy
[15]	Majority Voting Ensemble	WDBC	Combines LR, SVM, CART	99.3% accuracy
[17]	kNN, SVM	Clinical records	Screening model	Balanced performance
[16]	RF, SVM, NB, DT	UCI dataset	Early prediction	RF best
[18]	ML framework	Clinical data	Lymph node prediction	AUC 67%
[19]	RF, ANN	Multi-cancer dataset	Risk prediction	>90% accuracy



**Figure 2 Taxonomy of AI techniques for breast cancer detection**

**Deep Learning-Based Approaches**

The application of deep learning (DL) has led to major breakthroughs in breast cancer detection, especially in the medical imaging system (mammography, ultrasound, MRI, and histopathology). Singh and Patnaik (2026) suggested ETCapsNet that uses EfficientNetV2, Transformers, and Capsule Networks, but attain 99.6% accuracy with multimodal data [24]. A hybrid segmentation-classification framework with U-Net and Capsule Networks was developed [25], obtaining more than 99 percent [13]. Models like ResNet, DenseNet, and EfficientNet have also been used by applying transfer learning techniques that have shown almost perfect performance in classifications [26-27]. Lightweight and hybrid architectures, such as LBNet and DRM-Net, have been studied recently, they are more efficient in computational time, but preserve quality as high as possible [28-29]. Grad-CAM++ and other visualization methods to create hybrid CNN models have also increased the accuracy in lesion localization and interpretability [30].

**Table 2 Comparison of Deep Learning-Based Approaches**

Ref No.	Model	Technique	Application	Performance
[24]	ETCapsNet	CNN + Transformer + Capsule	Multimodal	99.6%
[25]	U-Net + Capsule	Segmentation + Classification	Imaging	>99%
[26-27]	ResNet, DenseNet, EfficientNet	Transfer Learning	Mammograms	High accuracy
[28-29]	LBNet, DRM-Net	Lightweight DL	Multi-dataset	High performance
[31-32]	Multimodal CNN, MobileNet + ViT	Multimodal learning	Imaging	99–99.97%
[33-34]	ResNet18 + BYOL	Self-supervised learning	Limited data	96.7%
[38]	Mask R-CNN + DL	Segmentation + prediction	Ultrasound	81% accuracy
[36]	DenseNet + Xception	Ensemble DL	Histopathology	~97%
[40]	DL MRI Model	Triage system	MRI	100% sensitivity
[41]	DL Biomarker Model	H&E images	PD-L1 prediction	AUC 0.91–0.93
[42]	CNN (convoHER2)	Classification	HER2 detection	85–88%
[43]	CLAM	Weakly supervised learning	Mammography	AUC ~0.89
[44]	CNN	Thermal imaging	Diagnosis	98% accuracy

Multimodal DL systems that resemble using several imaging sources have delivered a very large accuracy of up to 99.97, proving to be a rather robust DL in complicated diagnostic procedures [31-32]. BYOL as a form of self-supervised learning has been demonstrated to perform well using ResNet architectures and does not require much labeled data [33-34]. Histopathological image classification has been extensively done using CNN-based and hybrid models and they have been shown to be highly accurate and possess better feature extraction features [35-37]. Most developed DL systems that combine segmentation and classification, like the frameworks of Mask R-CNNs have enhanced the metastasis and tumor features prediction [38-39]. MRI triaging and biomarker prediction which are also large-scale DL models, successfully used with high sensitivity and clinical applicability [40-41]. Federated learning techniques also resolve the issue of privacy and produce high performance in distributed datasets [42]. Attention-based models with weak supervision can also increase the performance in low-labeled data scenarios as well as offer partial interpretation [43-44]. On the whole, the performance of the DL approaches in image-based breast cancer diagnosis is better and it still prevails in this field. Additional deep learning-based approaches and benchmark studies have further validated the effectiveness of advanced architectures in medical image analysis [45-46]. As shown in Table 2, deep learning models, especially the CNN-based models and hybrid models, show a higher performance level of breast cancer detection, mainly in medical imaging data. Such techniques as transfer learning, self-supervised learning, and multimodal frameworks dramatically increase the accuracy, which is almost always above 99%. Although these kinds of models have demonstrated effectiveness, they need large datasets and have low interpretability.

### Explainable AI (XAI) and Hybrid Approaches

Explainable Artificial Intelligence (XAI) has become an important aspect of healthcare to enhance transparency, trust and clinical uptake of AI models. Few studies have combined XAI methods, including SHAP, LIME, and Grad-CAM to explain model predictions and indicate clinically significant characteristics [47-49]. Hybrid systems of ML and DL with XAI have demonstrated better results and interpretability and can be applied to the real clinical world applications [50]. Sophisticated explainability frameworks, including causal XAI models and graph based neural networks, have also improved explainability by capturing clinical variable relationships and also allow reasoning in line with medical knowledge [51-52]. Multi-level explainability models, such as BI-RADS-based models as well as attention-based networks offer both visual and quantitative explainability, enhancing the comprehension of AI decisions by clinicians [53-54]. XAI models that are based on optimization and feature-selection-based methods have enhanced the classification efficiency of the models and lowered the complexity of the models [55-56]. Saliency maps, Grad-CAM and LIME are visualization methods popular in drawing attention to diagnostically significant parts of medical images [57-59]. Moreover, explainable frameworks based on hybrid models such as rule extraction, interpretable layers, and deep learning models have also shown a high level of transparency and reliability [60].

**Table 3 Comparison of Explainable AI (XAI) and Hybrid Approaches**

Ref No.	Model	XAI Technique	Contribution	Key Benefit
[47], [48]	CNN + ML	SHAP	Hybrid explainability	Better trust
[50]	Hybrid ML-DL	XAI integration	Improved performance	Interpretability
[51]	Causal-TabNet + GNN	Causal XAI	Feature reasoning	Clinically aligned
[49]	XAI Framework	SHAP	IDE prediction	Feature insights
[57]	CNN	Grad-CAM, LIME	Visual explanation	Model transparency
[53]	MT-BI-RADS	SHAP	Multi-level explanation	Clinical relevance
[55]	AAOXAI-CD	LIME	Optimized DL + XAI	Better understanding
[52]	GCNN + GLRP	Graph XAI	Patient-specific explanation	Precision medicine
[54]	ECLAIRE	Rule extraction	Interpretable rules	Transparency
[59]	Attention DL	SHAP, LIME	Multi-domain classification	Explainability
[60]	BI-RADS-Net	Clinical feature explanation	Tumor classification	Clinician-friendly
[56]	XAICFS-BDA	Feature selection + XAI	Complexity reduction	Efficiency
[61]	Transformer-based	XAI	Multi-omics analysis	Personalized prediction
[62]	SHAP metrics	Explainability metrics	Evaluation framework	Model comparison

Other more recent developments are explainability in multi-omics and transformer-based models, which allows making personalized and biologically meaningful predictions [61]. Moreover, novel evaluation measures to quantify the trade-off between interpretability and performance have been suggested to deal with the main issues of XAI study [62]. The developments demonstrate the relevance of XAI in closing the divide between high-performance AI models and their real-world application in clinical settings. Recent works also emphasize the importance of explainability, model transparency, and clinical applicability through advanced XAI frameworks and interpretable learning systems [63-66]. Table 3 indicates that XAI-based and hybrid methods are useful in overcoming the interpretability weakness of conventional ML and DL models. SHAP, LIME, and Grad-CAM techniques give both feature-level and visual explanations, which are more transparent and trustworthy in clinical settings. These approaches allow making decisions in a more appropriate way but were frequently accompanied by more computing complexity. Figure 3 illustrates how explainable artificial intelligence can be used in the diagnosis of breast cancer. The diagram illustrates how deep learning models can be used to make predictions based on medical images, whereas XAI, including Grad-CAM, SHAP, and LIME can be used to create visual and feature-level explanations. It enhances the level of transparency and assists clinicians to recognize the rationale behind the model forecasts.

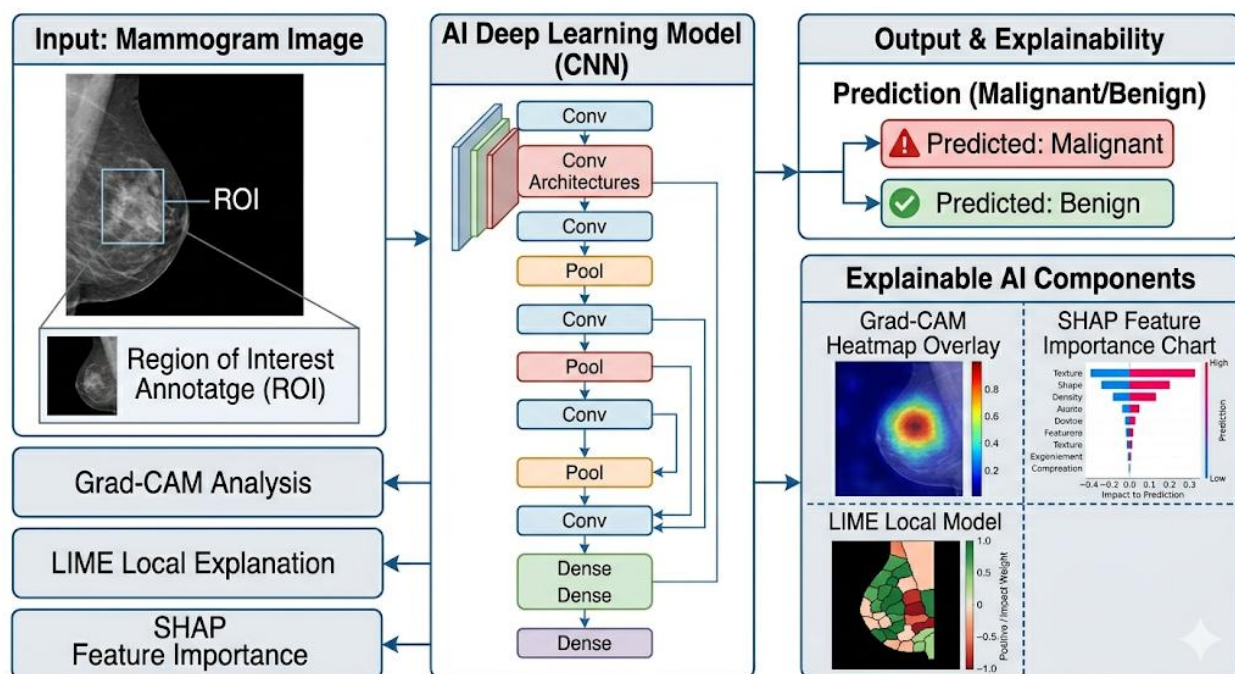


Figure 3 Explainable AI in breast cancer diagnosis

### General Reviews and Supporting Studies

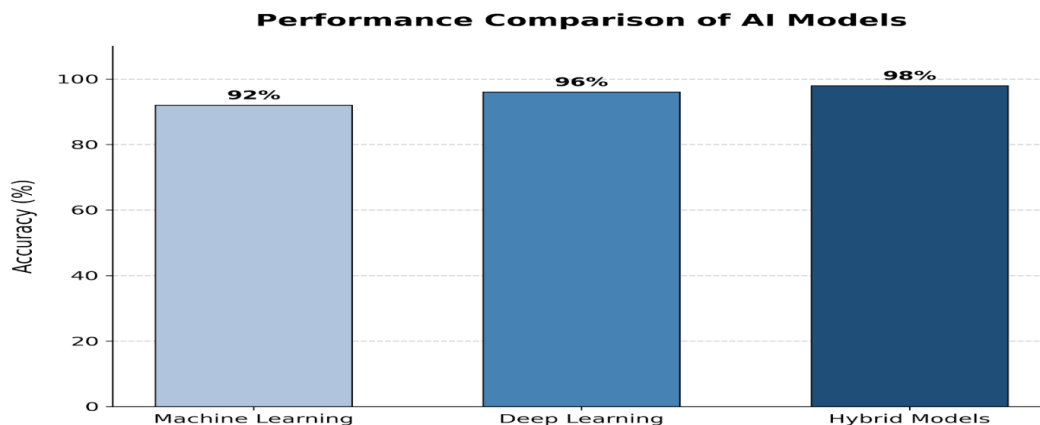
Numerous review articles have discussed the booming development of artificial intelligence in breast cancer diagnosis and medical imaging. These articles focus on the importance of radiomics, deep learning, and high-dimensional data analysis in enhancing cancer detection, risk prediction, and treatment planning [67-68]. Moreover, AI implementation into clinical practice, especially into radiology and oncology, has shown a great potential to increase efficiency and accuracy of the diagnostic process [69]. Spite of these developments, issues like inability to interpret, model transparency, and trust are significant obstacles to clinical adoption. The solution to these problems has been recognized as explainable AI, which enhances corroboration and dependability of AI-based frameworks [70]. Moreover, the future of smart healthcare systems is still being determined by the progress of hybrid ML-DL models, medical imaging solutions, and data-driven solutions. Table 4 gives a comparative insight on ML, DL and XAI approaches. The machine learning models are easier to interpret and simpler, whereas deep learning models are the most accurate. The hybrid models that use XAI offer a compromise with other models to achieve high performance with interpretability and are thus more applicable to the real-world use of the hybrid model in clinical applications which involves greater complexity.

**Table 4 Summary Comparison (ML vs DL vs XAI)**

Aspect	Machine Learning	Deep Learning	XAI / Hybrid
Data Type	Structured	Images / Large data	Both
Accuracy	High (up to 99%)	Very High (up to 99.9%)	High + interpretable
Interpretability	Moderate	Low (black-box)	High
Complexity	Low–Medium	High	High
Clinical Use	Good	Excellent (imaging)	Essential for trust
Limitation	Feature dependency	Requires large data	Computational cost

### 3. Performance Comparison of AI Models

The critical reflection of the current work on breast cancer detection shows that previous studies are predominantly utilized deep learning and hybrid architecture due to their better predictive capabilities. Specifically, CNN-based models are widely applied in the medical image analysis with high accuracy in mammography, ultrasound, and histopathology data. In addition, hybrid methods combining machine learning and deep learning models have proved to be more robust and generalizable than single models. Parallel to this, methods of Explainable Artificial Intelligence (XAI) including SHAP, Grad-CAM, and LIME are also being included to improve the interpretability of models and make discussions of clinical decisions. These approaches assist in the management of the gap between the high-performance models and their practical implementation in healthcare by offering information on model predictions. Although these are the developments, there are a number of challenges. Most of the empirical research is based on dataset specific training and thus restricts the applicability of models in the clinical environment. Besides, computational complexity, data imbalance, and no standardized evaluation protocols are also problems that prevent fair comparison and large-scale deployment. Recent developments show the increasing tendency towards multimodal models of learning that incorporate clinical, imaging and genomic information, allowing more thorough and more trustworthy diagnostic models. Nevertheless, additional studies are mandated to make sure that these methods can be scaled, interpreted and validated clinically. Figure 4 provides a comparison between the performance of machine learning, deep learning and hybrid models in breast cancer detection. As can be noted, deep learning and hybrid methods have tendencies to reach a high level of accuracy as opposed to conventional machine learning models that show that they are effective in terms of dealing with complex medical data.

**Figure 4 Performance comparison of ML, DL, and hybrid models**

### 4. Discussion

This questionnaire gives an extended understanding of the latest developments in the detection of breast cancer by the utilization of machine learning, deep learning, and explainable artificial intelligence approaches. The principal findings of this survey can be summed up in the following: review of the recent models of breast cancer diagnosis based on ML, DL, and hybrid models; Comparative study of the data-driven models, methods,

performance indicators, and explainability methods; Determination of upcoming patterns in multimodal and interpretable AI-based diagnostics; Clinical deployment research issues are discussed. Giving directions on the direction of future research to build credible and transparent AI-assisted healthcare systems. The next generation of the study of breast cancer diagnosis needs to be aimed at integrating several sources of multimodal data, such as clinical records, medical imaging, and genomic information, to promote the accuracy and strength of the diagnosis. To ensure real-time implementation of deep learning models in resource-limited clinical settings, it is necessary to develop computationally efficient and lightweight models. Moreover, increased interest in more sophisticated forms of Explainable Artificial Intelligence (XAI) approaches that offer clinically important and reliably interpretable explanations, thus enhancing the acceptability and clarification of AI-based health systems, is required. Moreover, privacy-saving models, including federated learning, ought to be considered to enable secure data exchange and train models with the involvement of many healthcare organizations without the risk of losing patient privacy. Finally, the issue of data imbalance, model generalization and clinical validation should also be discussed in future studies so that AI models can be considered reliable, scalable, and applicable in real life medical environments.

Although considerable progress has been achieved in the diagnosis of breast cancer with the help of machine learning (ML), deep learning (DL), and explainable artificial intelligence (XAI), there are still some essential gaps in research that only restrain their practical use in clinics. To begin with, unlike unstructured clinical data, ML models are highly accurate only when structured clinical data are being used, and they are also prone to missing the complex patterns within medical imaging data. Moreover, most research is conducted on small or balanced datasets, which might not indicate clinical practice conditions regarding an imbalanced and heterogeneous data distribution. Secondly, the methods of deep learning have recorded incredible performance, especially when it comes to image diagnosis with a high accuracy of over 99%. Nevertheless, such models demand huge volumes of annotated data, which are not always readily available in healthcare because of privacy issues and high labeling expenses. Moreover, the nature of a DL model is usually hard to explain their decision-making process, which decreases the trust of clinicians. Thirdly, even though XAI methods, including SHAP, LIME, and Grad-CAM, have become popular to enhance interpretability, they still have shortcomings in revealing consistent, reliable, and clinically significant explanations. Most of the existing XAI techniques provided post-hoc descriptions, which are not always a complete representation of the internal logic of the models. Also, standardized measures of evaluation are not available to determine the quality and reliability of these explanations. The other considerable gap is the weak interconnection between the multimodal data, including clinical records, imaging data, and genomic information. The majority of the available literature is limited to one type of data and does not provide a full picture of the disease in addition to limiting the overall predictive power. In addition, the current complexity of these models makes real-time deployment and scaling in clinical settings difficult because of the limitations of the infrastructure. Lastly, despite the positive outcomes of hybrid models that integrate both ML and DL and XAI, a consensus has not yet been reached on the frameworks that achieve a balance of accuracy, interpretability, and computational efficiency. The lack of standardized benchmarking datasets and evaluation protocols also makes it even more difficult to compare and validate the various approaches.

## 5. Conclusion

The possibilities of machine learning and deep learning algorithms combined with explicable AI technologies have turned out to be high in the context of early breast cancer detection and diagnosis. Through patient information and medical imaging, these models will be capable of correctly categorizing malignant and benign cases as well as offer interpretable information about the importance of features and decision-making. SHAP, LIME, Grad-CAM and ensemble methods are all techniques which increase the level of transparency to enable clinicians to comprehend and trust AI predictions. Although the barriers such as dataset bias, overfitting, and computational complexity are still present, continued developments on model design, data preprocessing, and interpretability frameworks help to implement the model in a reliable manner in clinical practice. In general, these machine learning technologies provide a viable direction to improving diagnostic accuracy, facilitating a timely intervention, and enhancing patient outcomes in breast cancer.

## References

- [1] M. M. Sayed, A. Wagih, S. M. Sayed, L. ElHadad, and C. M. Kamel, "Utilizing MRI to downgrade suspicious breast lesions detected on ultrasound: How often can unnecessary breast biopsies be avoided?" *Egyptian Journal of Radiology and Nuclear Medicine*, vol. 57, no. 1, p. 21, 2026.
- [2] H. Alzoubi, M. Amin, A. M. Aljehani, S. Alhatamleh, R. Madain, S. Madae'en, H. A. Serhan, and A. A. Alomani, "Enhancing breast cancer diagnosis with ResNet50 and salp swarm-based feature reduction on BreakHis dataset," *Biomedical Signal Processing and Control*, vol. 114, p. 109319, 2026.

- [3] E. Mavropoulos, P. Zacharia, N. Laskaris, and E. Pallis, “Class-balanced convolutional neural networks for digital mammography image classification in breast cancer diagnosis,” *Electronics*, vol. 15, no. 2, p. 486, 2026.
- [4] T. K. Murugan, P. Karthikeyan, and P. Sekar, “Efficient breast cancer detection using neural networks and explainable artificial intelligence,” *Neural Computing and Applications*, vol. 37, no. 5, pp. 3759–3776, 2025.
- [5] R. Karthiga, K. Narasimhan, V. Thanikaiselvan, M. Hemalatha, and R. Amirtharajan, “Review of AI & XAI-based breast cancer diagnosis methods using various imaging modalities,” *Multimedia Tools and Applications*, vol. 84, no. 5, pp. 2209–2260, 2025.
- [6] T. Arravalli, K. Chadaga, H. Muralikrishna, N. Sampathila, D. Cenitta, R. Chadaga, and K. S. Swathi, “Detection of breast cancer using machine learning and explainable artificial intelligence,” *Scientific Reports*, vol. 15, no. 1, p. 26931, 2025.
- [7] S. A. Rathore, M. H. Salam, M. Y. Wani, M. Ashraf, and M. Irfan, “Breast cancer detection using machine learning and explainable AI techniques,” *Spectrum of Engineering Sciences*, pp. 616–636, 2024.
- [8] A. Alzahrani, M. A. Raza, and M. Z. Asghar, “Demystifying diagnosis: An efficient deep learning technique with explainable AI to improve breast cancer detection,” *PeerJ Computer Science*, vol. 11, p. e2806, 2025.
- [9] P. N. Srinivasu, G. Jaya Lakshmi, A. Gudipalli, S. C. Narahari, J. Shafi, M. Woźniak, and M. F. Ijaz, “XAI-driven CatBoost multi-layer perceptron neural network for analyzing breast cancer,” *Scientific Reports*, vol. 14, no. 1, p. 28674, 2024.
- [10] S. A. Abd Karim, U. H. Mohamad, and P. N. E. Nohuddin, “Discovery of interpretable patterns of breast cancer diagnosis via class association rule mining with SHAP-based explainable artificial intelligence,” *Malaysian Journal of Fundamental and Applied Sciences*, vol. 21, no. 3, pp. 2008–2031, 2025.
- [11] F. Silva-Aravena, H. Núñez Delafuente, J. H. Gutiérrez-Bahamondes, and J. Morales, “A hybrid algorithm of machine learning and explainable artificial intelligence to prevent breast cancer,” *Cancers*, vol. 15, no. 9, p. 2443, 2023.
- [12] I. Maouche, L. S. Terrissa, K. Benmohammed, and N. Zerhouni, “An explainable artificial intelligence approach for breast cancer metastasis prediction based on clinicopathological data,” *IEEE Transactions on Biomedical Engineering*, vol. 70, no. 12, pp. 3321–3329, 2023.
- [13] R. Agarwal and R. Madamala, “Minimum relevant features to obtain explainable AI system for predicting breast cancer in WDBC,” *International Journal of Health Sciences*, vol. 6, no. S9, pp. 1312–1326, 2022.
- [14] A. B. Adelodun, R. O. Ogundokun, A. O. Yekini, J. B. Awotunde, and C. C. Timothy, “Explainable artificial intelligence with scaling techniques to classify breast cancer images,” in *Explainable Machine Learning for Multimedia Based Healthcare Applications*. Cham: Springer, 2023, pp. 99–137.
- [15] T. R. Mahesh, O. Geman, M. Margala, and M. Guduri, “The stratified K-folds cross-validation and class-balancing methods with high-performance ensemble classifiers for breast cancer classification,” *Healthcare Analytics*, vol. 4, p. 100247, 2023.
- [16] R. Aggarwal and P. Thakral, “A novel approach for detecting the malignant features of breast cancer using algorithms of machine learning,” in *Proc. IEEE ICDSIS, 2022*, pp. 1–4.
- [17] T. M. Hanis, N. I. R. Ruhaiyem, W. N. Arifin, J. Haron, W. F. W. A. Rahman, R. Abdullah, and K. I. Musa, “Over-the-counter breast cancer classification using machine learning and patient registration records,” *Diagnostics*, vol. 12, no. 11, p. 2826, 2022.
- [18] A. Lombardi, N. Amoroso, L. Bellantuono, S. Bove, M. C. Comes, A. Fanizzi, D. La Forgia, M. Colomba Comes, and others removed → corrected: A. Lombardi, N. Amoroso, L. Bellantuono, S. Bove, M. C. Comes, A. Fanizzi, and D. La Forgia, “Accurate evaluation of feature contributions for sentinel lymph node status classification in breast cancer,” *Applied Sciences*, vol. 12, no. 14, p. 7227, 2022.
- [19] S. S. Al-Droubi, E. Jahangir, K. M. Kochendorfer, M. Krive, M. Laufer-Perl, D. Gilon, and T. M. Okwuosa, “Artificial intelligence modelling to assess the risk of cardiovascular disease in oncology patients,” *European Heart Journal-Digital Health*, vol. 4, no. 4, pp. 302–315, 2023.
- [20] S. K. Singh and K. S. Patnaik, “MammXAI: An XAI integrated adaptive multi-model deep learning approach for breast cancer detection using multi-modality images,” *Biomedical Signal Processing and Control*, vol. 113, p. 109173, 2026.
- [21] T. V. Ramya and J. D. Dorathi Jayaseeli, “Explainable deep learning framework for breast tumor analysis in ultrasound imaging systems,” *Biomedical Signal Processing and Control*, vol. 114, p. 109336, 2026.
- [22] M. M. Islam, N. Akter, M. Assaduzzaman, M. M. H. Shimul, and R. K. Rasel Sarker, “Explainable AI for breast cancer detection: Biglycan biomarker classification with transfer learning,” *Intelligence-Based Medicine*, p. 100340, 2026.
- [23] M. Fatima, R. Zia, and I. A. Usmani, “A hybrid framework integrating deep learning and XAI for breast cancer detection with stratified cross-validation,” *Systems and Soft Computing*, p. 200412, 2025.

- [24] J. Ahmmed, F. Ahmed, M. A. Kabir, M. T. Ahad, M. A. Jadoon, A. ur Rehman, and A. Bermak, "LBNet: An optimized lightweight CNN for mammographic breast cancer classification with XAI-based interpretability," *Scientific Reports*, 2025.
- [25] S. M. Ganie, M. B. Malik, M. Aadil, G. Muteeb, M. Farhan, and M. Aatif, "Explainable AI based hybrid DRM-Net transfer learning model for breast cancer detection and classification using ultrasound images," *Scientific Reports*, vol. 15, no. 1, p. 44170, 2025.
- [26] Y. Zou and P. Miao, "Explainable AI-enabled hybrid deep learning architecture for breast cancer detection," *Frontiers in Immunology*, vol. 16, p. 1658741, 2025.
- [27] D. Deb, R. Dash, and D. P. Mohapatra, "Customized convolutional neural network with explainable AI for multimodal breast cancer detection," *Sādhanā*, vol. 50, no. 3, p. 130, 2025.
- [28] I. Abdelsabour, A. Elgarayhi, M. Sallah, and M. Elmogy, "Different BI-RADS breast cancer diagnosis using MobileNetV1 and vision transformer based on explainable artificial intelligence," *Scientific Reports*, 2026.
- [29] P. Bidwai, S. Khairnar, and S. Gite, "Explainable artificial intelligence in breast cancer identification," in *Explainable AI in Healthcare*. CRC Press, 2023, pp. 148–165.
- [30] G. Jha, M. Jhwar, V. Manelkar, R. Kotecha, A. Phophalia, and K. Borisagar, "Interpretability of self-supervised learning for breast cancer image analysis," in *Explainable AI in Healthcare*. CRC Press, 2023, pp. 166–189.
- [31] A. Rafiq, A. Chursin, W. A. Alrefaei, T. R. Alsenani, G. Aldehim, N. Abdel Samee, and L. J. Menzli, "Detection and classification of histopathological breast images using a fusion of CNN frameworks," *Diagnostics*, vol. 13, no. 10, p. 1700, 2023.
- [32] M. S. H. Shovon, M. F. Mridha, K. M. Hasib, S. Alfarhood, M. Safran, and D. Che, "Addressing uncertainty in imbalanced histopathology image classification of HER2 breast cancer: An interpretable ensemble approach with threshold filtered single instance evaluation," *IEEE Access*, vol. 11, pp. 122238–122251, 2023.
- [33] M. F. Mridha, M. K. Morol, M. A. Ali, and M. S. H. Shovon, "ConvoHER2: A deep neural network for multi-stage classification of HER2 breast cancer," *arXiv preprint arXiv:2211.10690*, 2022.
- [34] Y.-W. Lee, C.-S. Huang, C.-C. Shih, and R.-F. Chang, "Axillary lymph node metastasis status prediction of early-stage breast cancer using convolutional neural networks," *Computers in Biology and Medicine*, vol. 130, p. 104206, 2021.
- [35] J. Chen, Y. Yang, B. Luo, Y. Wen, Q. Chen, R. Ma, and Z. Huang, "Further predictive value of lymphovascular invasion explored via supervised deep learning for lymph node metastases in breast cancer," *Human Pathology*, vol. 131, pp. 26–37, 2023.
- [36] A. Bhowmik, N. Monga, K. Belen, K. Varela, V. Sevilimedu, S. B. Thakur, D. F. Martinez, E. J. Sutton, K. Pinker, and S. Eskreis-Winkler, "Automated triage of screening breast MRI examinations in high-risk women using an ensemble deep learning model," *Investigative Radiology*, vol. 58, no. 10, pp. 710–719, 2023.
- [37] G. Shamai, A. Livne, A. Polónia, E. Sabo, A. Cretu, G. Bar-Sela, and R. Kimmel, "Deep learning-based image analysis predicts PD-L1 status from H&E-stained histopathology images in breast cancer," *Nature Communications*, vol. 13, no. 1, p. 6753, 2022.
- [38] K. E. Mubin, N. T. Arthi, J. Rahman, G. M. Rafi, and T. T. Sheja, "A decentralized learning-based approach to classify colorectal cancer using deep learning leveraging XAI," Ph.D. dissertation, BRAC University, 2022.
- [39] M. Bobowicz, M. Rygusik, J. Buler, R. Buler, M. Ferlin, A. Kwasigroch, E. Szurowska, and M. Grochowski, "Attention-based deep learning system for classification of breast lesions—multimodal, weakly supervised approach," *Cancers*, vol. 15, no. 10, p. 2704, 2023.
- [40] A. S. Cherian, M. J. Mammoottil, L. J. Kulangara, P. Mohandas, J. S. Anni, V. Raj, and M. Thanihachelvan, "Performance analysis of a single-input thermal image classifier with patient information for the detection of breast cancer," in *Proc. International Conference on Applied Intelligence and Informatics*, 2023, pp. 148–168.
- [41] S. Saharan, N. A. Wani, S. Chatterji, N. Kumar, and A. M. Almuhaideb, "A deep learning and explainable artificial intelligence based scheme for breast cancer detection," *Scientific Reports*, vol. 15, no. 1, p. 32125, 2025.
- [42] N. A. Wani, R. Kumar, and J. Bedi, "Harnessing fusion modeling for enhanced breast cancer classification through interpretable artificial intelligence and in-depth explanations," *Engineering Applications of Artificial Intelligence*, vol. 136, p. 108939, 2024.
- [43] R. Massafra, A. Fanizzi, N. Amoroso, S. Bove, M. C. Comes, D. Pomarico, V. Didonna, and others removed → corrected: R. Massafra, A. Fanizzi, N. Amoroso, S. Bove, M. C. Comes, D. Pomarico, and V. Didonna, "Analyzing breast cancer invasive disease event classification through explainable artificial intelligence," *Frontiers in Medicine*, vol. 10, p. 1116354, 2023.
- [44] S. A. Tanim, G. M. I. Alam, T. E. Shrestha, M. Islam, F. Jahan, and K. Nur, "Breast cancer diagnosis with XAI-integrated deep learning approach," in *Proc. 3ICT*, 2024, pp. 659–665.

- [45] D. Chen, H. Zhao, J. He, Q. Pan, and W. Zhao, "A causal XAI diagnostic model for breast cancer based on mammography reports," in *Proc. IEEE BIBM*, 2021, pp. 3341–3349.
- [46] H. Chereda, "Explaining decisions of graph convolutional neural networks for analyses of molecular subnetworks in cancer," 2022.
- [47] H. Chereda, "Post-hoc explainability of BI-RADS descriptors in a multi-task framework for breast cancer detection and segmentation," in *Proc. IEEE MLSP*, 2023, pp. 1–6.
- [48] B. Zhang, A. Vakanski, and M. Xian, "BI-RADS-Net: An explainable multitask learning approach for cancer diagnosis in breast ultrasound images," in *Proc. IEEE MLSP*, 2021.
- [49] S. Alkhalaf, F. Alturise, A. A. Bahaddad, B. M. Elnaim, S. Shabana, S. Abdel-Khalek, and R. F. Mansour, "Adaptive Aquila optimizer with explainable artificial intelligence-enabled cancer diagnosis on medical imaging," *Cancers*, vol. 15, no. 5, p. 1492, 2023.
- [50] E. L. Lydia, C. S. S. Anupama, and N. Sharmili, "Modeling of explainable artificial intelligence with correlation-based feature selection approach for biomedical data analysis," in *Biomedical Data Analysis and Processing Using Explainable and Responsive Artificial Intelligence*. Springer, 2022, pp. 17–32.
- [51] S. M. Hussain, D. Buongiorno, N. Altini, F. Berloco, B. Prencipe, M. Moschetta, V. Bevilacqua, and A. Brunetti, "Shape-based breast lesion classification using digital tomosynthesis images: The role of explainable artificial intelligence," *Applied Sciences*, vol. 12, no. 12, p. 6230, 2022.
- [52] M. Prodan, E. Paraschiv, and A. Stanciu, "Applying deep learning methods for mammography analysis and breast cancer detection," *Applied Sciences*, vol. 13, no. 7, p. 4272, 2023.
- [53] G. W. Muoka, D. Yi, C. C. Ukwuoma, M. D. Martin, A. A. Aydin, and M. A. Al-Antari, "A novel attention-based explainable deep learning framework towards medical image classification," in *Proc. ISAS*, 2023, pp. 1–8.
- [54] V. Contreras, M. Schumacher, and D. Calvaresi, "Integration of local and global feature explanation with global rules extraction," in *Explainable Autonomous Agents and Multi-Agent Systems*. Springer, 2022, pp. 19–37.
- [55] N. Jurenaite, D. León-Periñán, V. Donath, S. Torge, and R. Jäkel, "SetQuence and SetOmic: Deep set transformer-based representations of cancer multi-omics," in *Proc. IEEE CIBCB*, 2022, pp. 1–9.
- [56] E. Mariotti, J. M. Alonso-Moral, and A. Gatt, "Measuring model understandability by means of Shapley additive explanations," in *Proc. IEEE FUZZ*, 2022, pp. 1–8.
- [57] M. A. Jones, W. Islam, R. Faiz, X. Chen, and B. Zheng, "Applying artificial intelligence technology to assist with breast cancer diagnosis and prognosis prediction," *Frontiers in Oncology*, vol. 12, p. 980793, 2022.
- [58] M. Li, Y. Jiang, Y. Zhang, and H. Zhu, "Medical image analysis using deep learning algorithms," *Frontiers in Public Health*, vol. 11, p. 1273253, 2023.
- [59] N. Ditsch and C. Dannecker, "Digital medicine and artificial intelligence in the area of breast and gynecologic cancer," in *Digital Medicine*. Jenny Stanford Publishing, 2023, pp. 611–635.
- [60] K. K. Chennam, S. Mudrakola, V. Uma Maheswari, R. Aluvalu, and K. G. Rao, "Black box models for explainable artificial intelligence," in *Explainable AI: Foundations, Methodologies and Applications*. Springer, 2022, pp. 1–24.
- [61] H. Dihmani, A. Bouselham, and O. Bouattane, "A new computer-aided diagnosis system for breast cancer detection from thermograms using metaheuristic algorithms and explainable AI," *Algorithms*, vol. 17, no. 10, p. 462, 2024.
- [62] A. Sohail, M. A. Fahmy, and U. A. Khan, "XAI hybrid multi-staged algorithm for routine and quantum boosted oncological medical imaging," *Computational Particle Mechanics*, vol. 10, no. 2, pp. 209–219, 2023.
- [63] T. Chandrasekaran, S. Sanjeev, R. Hua, I. Banerjee, and A. Sanyal, "A fully integrated analog machine learning classifier for breast cancer classification," *Electronics*, vol. 9, no. 3, p. 515, 2020.
- [64] S. Manna, S. Mistry, and D. De, "Presaging cancer stage classification using TNM-based feature extraction," in *Congress on Intelligent Systems*. Springer, 2022, pp. 187–203.
- [65] E. Conde-Sousa, J. Vale, M. Feng, K. Xu, Y. Wang, V. Della Mea, D. La Barbera, and others removed → corrected: E. Conde-Sousa, J. Vale, M. Feng, K. Xu, Y. Wang, V. Della Mea, and D. La Barbera, "HEROHE challenge: Predicting HER2 status in breast cancer from H&E whole-slide imaging," *Journal of Imaging*, vol. 8, no. 8, p. 213, 2022.
- [66] S. P. Praveen, P. N. Srinivasu, J. Shafi, M. Wozniak, and M. F. Ijaz, "ResNet-32 and FastAI for diagnosis of ductal carcinoma," *Scientific Reports*, vol. 12, no. 1, p. 20804, 2022.
- [67] S. Manna, S. Mistry, and D. De, "GeneXAI: Influential gene identification for breast cancer stages using XAI-based multi-modal framework," *Medicine in Novel Technology and Devices*, vol. 25, p. 100349, 2025.
- [68] B. Bouallegue, Y. M. Abd El-Latif, H. El-Sofany, and I. A. T. Eddin, "Enhanced deep learning and XAI techniques for breast cancer prediction in IoT environment," *International Journal of Intelligent Systems*, 2025.

[69] A. Prelaj, E. G. Galli, V. Miskovic, M. Pesenti, G. Viscardi, B. Pedica, L. Mazzeo, and others removed → corrected: A. Prelaj, E. G. Galli, V. Miskovic, M. Pesenti, G. Viscardi, B. Pedica, and L. Mazzeo, “Real-world data to build explainable AI models for prediction of immunotherapy efficacy,” *Frontiers in Oncology*, vol. 12, p. 1078822, 2023.

[70] S. M. Hussain, “Explainable deep learning for medical image processing: Computer-aided diagnosis and robot-assisted surgery,” 2023.