Intelligent Oncology 2 (2026) 100037

Contents lists available at ScienceDirect

Intelligent Oncology

journal homepage: www.keaipublishing.com/en/intonc/

Review

Multimodal medical imaging AI for breast cancer diagnosis: A comprehensive review

Ting-Ruen Wei, Yuling Yan*

Santa Clara University, Santa Clara CA 95053, USA

ARTICLE INFO

Keywords:
Breast cancer
Artificial intelligence
Machine learning
Deep learning
Multimodal

ABSTRACT

Traditional artificial intelligence (AI)-based methods for breast cancer diagnosis often rely on a single modality, such as ultrasound images. With the rise of multimodal approaches, multiple data sources, including imaging from diverse medical modalities, structured clinical information, and unstructured medical reports, are increasingly integrated to provide richer and more informative signals for model training. This survey reviews the data modalities employed in AI-based breast cancer research, examines common multimodal combinations and fusion strategies, and discusses their applications across clinical tasks such as diagnosis, treatment planning, and outcome prediction. By consolidating current literature and identifying critical gaps, this survey aims to guide future research toward the development of reliable, clinically relevant multimodal AI systems for use in breast cancer management.

1. Introduction

Breast cancer remains one of the leading causes of cancer-related mortality among women worldwide.[1] Early detection and accurate diagnosis are essential for improving treatment outcomes and survival rates. Traditional screening methods such as clinical breast exams, mammography, ultrasound, and magnetic resonance imaging (MRI) are widely used to detect abnormalities. However, each imaging modality introduces its own challenges: mammography suffers reduced sensitivity in women with dense breast tissue, ultrasound is highly operator-dependent, and MRI has variable specificity and remains expensive and less accessible. These modality-specific limitations often necessitate the use of multiple imaging techniques in combination to obtain a comprehensive assessment. Radiologists found that coronal ultrasound images enhanced radiologist performance compared to transverse images, illustrating how even within the same imaging technique, different views can offer varying diagnostic value.[2]

While early artificial intelligence (AI) systems often relied on a single data source, such as mammograms or ultrasound images, there is a growing recognition that incorporating multiple data modalities can enhance model performance. This trend mirrors developments in the broader AI landscape, where the transition from large language models[3] to large vision-language models[4] has led to more powerful and generalizable systems. In the medical domain, recent multimodal foundation models such as BioGPT,[5] MedCLIP,[6] LLaVa-Med,[7] and PathCLIP[8] have demonstrated the capacity to integrate imaging and textual data for tasks including report generation, visual question answering, and image-text alignment. In breast cancer, combining imaging modalities, clinical data, genomic profiles, and clinical reports can provide a more comprehensive understanding of the patient's condition, supporting improved clinical decision-making and patient care.

Despite a growing body of research in this area, existing reviews often focus on specific imaging techniques or AI applications, lacking a comprehensive examination of how multimodal data are used in AI for breast cancer. To address this gap, we conducted an extensive survey of 88 papers that utilize multiple data modalities or image perspectives in training AI models for breast cancer classification, including lesion classification, treatment response prediction, and outcome estimation. By systematically examining the types of modalities used, this paper aims to provide a consolidated resource and practical guidance for researchers and practitioners working at the intersection of AI and breast cancer diagnosis.

Numerous survey papers have explored the role of AI in breast cancer, each emphasizing different aspects such as clinical applications,[9] patient perspectives,[10] data fusion strategies,[11] and imaging techniques. Shah et al.[12] reviewed imaging modalities including ultrasound, mammography, MRI, and histopathology, with a particular focus on mammography-based methods due to the wide availability of datasets. Chugh et al.[13] expanded on this by incorporating digital breast tomosynthesis (DBT), and Saleh et al.[14] highlighted advancements

* Corresponding author.
E-mail address: yyan1@scu.edu (Y. Yan).

https://doi.org/10.1016/j.intonc.2025.12.005
Received 23 September 2025; Received in revised form 27 December 2025; Accepted 27 December 2025
Available online 31 December 2025
2950-2616/© 2026 The Authors. Publishing services by Elsevier B.V. on behalf of KeAi Communications Co. Ltd. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

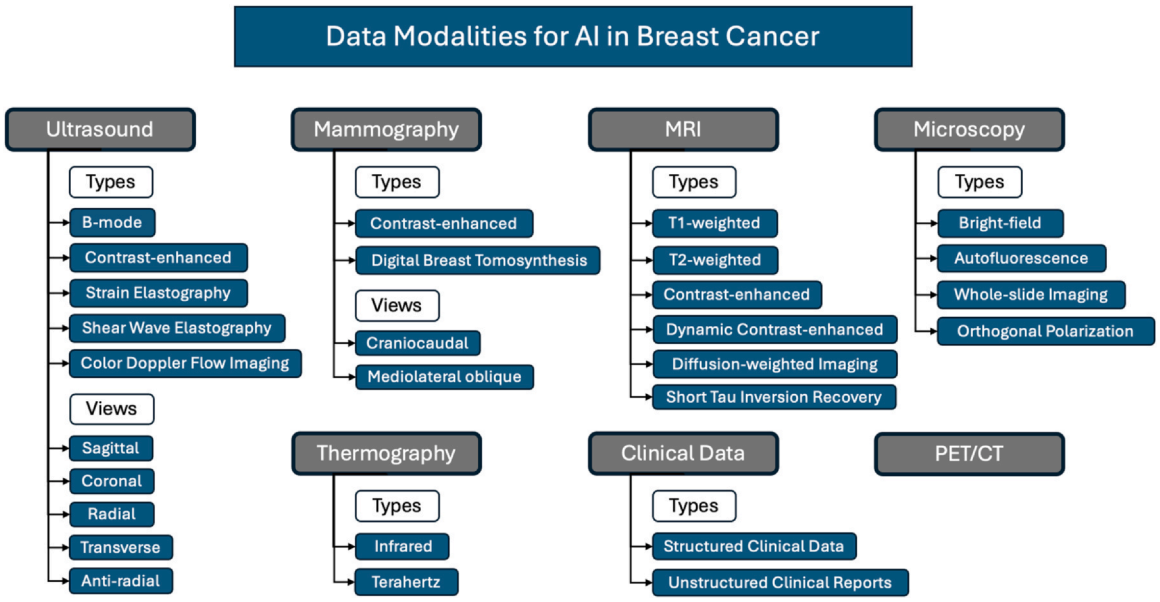


Figure 1. Overview of data modalities and imaging perspectives for AI in breast cancer. Abbreviations: 1. MRI: magnetic resonance imaging 2. PET/CT: positron emission tomography/computed tomography.

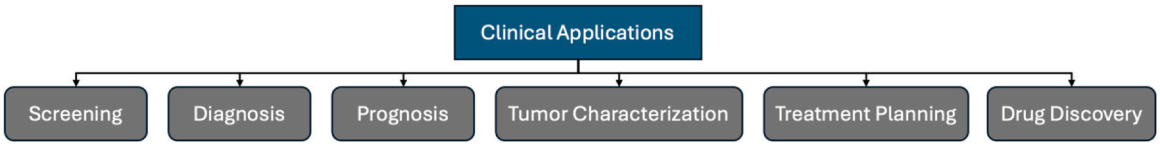


Figure 2. Clinical applications of AI for breast cancer.

within these mainstream imaging techniques. Meanwhile, other survey papers concentrate on mammography and DBT,[15–17] genetic sequencing and histopathological imaging,[18] radiomics-based approaches,[19] and the integration of molecular and imaging data.[20] As multimodality becomes a growing trend in current research,[21,22] this survey focuses on the various data modalities used in AI for breast cancer. Unlike previous work, we provide separate overviews of each data modality, including medical images and textual reports, as well as distinct imaging perspectives such as transverse and sagittal views. The full scope is shown in Figure 1, covering various clinical applications including screening, diagnosis, prognosis, tumor characterization, treatment planning, and drug discovery (Figure 2). We examine how each modality or perspective is integrated with others in the literature for various clinical applications, offering practical guidance for both researchers and practitioners.

2. Data fusion strategies

Multimodal AI systems combine heterogeneous data sources to create more comprehensive models. As shown in Figure 3, integration strategies are typically categorized into early, intermediate, and late fusion, each offering different advantages depending on data alignment, clinical complexity, and model design.

Early fusion integrates modalities at the raw data or low-level feature stage, such as combining spatially aligned inputs like stacked MRI sequences or different imaging perspectives. This approach allows the model to learn cross-modal patterns directly from low-level signals in a single forward pass, but it requires precise alignment and is sensitive to missing or noisy inputs.

Intermediate fusion is the dominant approach in the medical domain. Each modality is independently processed via feature extractors or engineered features, and the resulting representations are merged. This approach supports heterogeneous data types and allows cross-modal interactions through attention or learned projection layers. Importantly, combining radiomics features from different imaging modalities is considered intermediate fusion, because the fusion occurs after modality-specific feature extraction.

Late fusion combines decisions rather than features, integrating outputs from separate models using methods such as averaging or weighted voting. This strategy increases the interpretability and robustness, and it allows independent models to contribute without requiring the complete set of input data.

3. Ultrasound

Ultrasound is a crucial imaging modality in the diagnosis and management of breast cancer, particularly due to its real-time imaging capabilities, lack of ionizing radiation, and effectiveness in dense breast tissue where mammography may be limited.[23] Breast ultrasound relies on a range of imaging modalities that provide complementary anatomical, vascular, and mechanical information, greatly enhancing diagnostic accuracy. B-mode,[24] or brightness mode, is the fundamental technique that produces real-time grayscale images by capturing echoes reflected from internal breast structures. It allows clear visualization of tissue architecture and is especially effective at distinguishing between cystic lesions and solid masses. Color Doppler flow imaging (CDFI)[25] adds a functional layer by displaying blood flow within and around lesions using color overlays, which helps identify increased or disorganized vascularity commonly associated with malignant tumors. Contrast-enhanced ultrasound (CEUS)[26] provides even more detailed vascular assessment through the injection of micro-bubble contrast agents, enabling real-time visualization of microcirculation and perfusion patterns. This is useful in evaluating tumor aggressiveness and vascular architecture that may not be visible on conventional Doppler. Elastography techniques[27] offer mechanical insights by assessing tissue stiffness, a property often elevated in malignant lesions. Shear wave

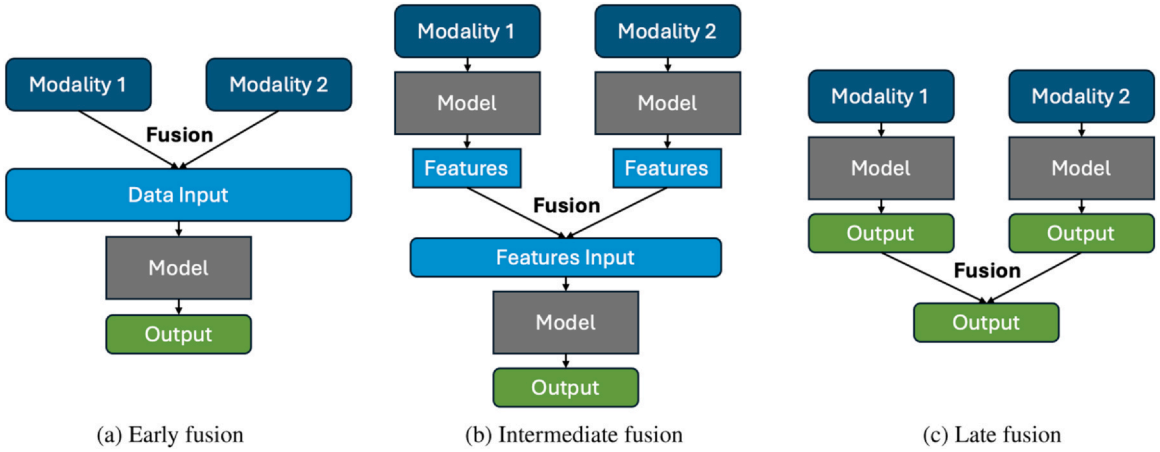


Figure 3. Fusion strategies. Early fusion takes place at the input level, intermediate fusion at the feature level, and late fusion at the decision level.

elastography (SWE)[28] quantitatively measures the speed at which shear waves travel through tissue, generating color-coded stiffness maps where stiffer areas often indicate malignancy. An advanced form of this, known as shear wave absolute vibro elastography (SWAVE),[29] offers more precise, real-time elasticity measurements, making it valuable for detecting subtle differences in tissue consistency. Strain elastography (SE),[30] on the other hand, is a qualitative method that assesses how much tissue deforms in response to compression. Malignant lesions generally exhibit minimal deformation due to their firm structure, while benign lesions tend to compress more easily.

Ultrasound imaging perspectives are essential for comprehensive breast assessment, as they provide spatial orientation and improve lesion characterization.[31] The radial plane involves positioning the transducer so that the ultrasound beam radiates outward from the center of the breast, following the natural orientation of the ductal structures. This view is especially useful for evaluating ductal abnormalities, such as intraductal masses or ductal dilatation. The anti-radial plane is oriented perpendicular to the radial plane and provides a cross-sectional view of the ducts and surrounding tissues. This helps in visualizing lesions that may not be fully appreciated in the radial view and in assessing their full extent. The sagittal view, also known as the longitudinal plane, divides the breast into right and left sections. It provides a vertical image that is valuable for measuring the height and depth of a lesion. The transverse view, sometimes referred to as the axial plane, divides the breast into superior and inferior portions and offers a horizontal cross-section of tissue. It is helpful for evaluating the width and lateral spread of lesions. The coronal view, which divides the breast into anterior and posterior sections, is less commonly used in standard 2D ultrasound but can be reconstructed from 3D imaging or elastography. It is particularly beneficial for assessing lesion depth and spatial relationships within the full volume of the breast. By fusing complementary modalities and imaging perspectives, multimodal ultrasound provides a more robust foundation for accurate prediction.

3.1. Fusion of ultrasound modalities

A substantial number of studies have focused on combining multiple ultrasound modalities, most commonly B-mode, CDFI, and elastography, to enhance classification performance by capturing complementary aspects of tumor characteristics.[32] Wang et al.[33] further extended this by fusing B-mode, CDFI, SWE, and SE using a shared ResNet backbone,[34] while employing reinforcement learning to learn optimal combining weights for the final classification decision. Building on this, the AW3M framework[35] enhances performance under real-world constraints by addressing the issue of missing modalities, introducing a mechanism to maintain effectiveness even when only a subset of the expected input data are available.

While some studies incorporate the full set of modalities, others selectively utilize subsets based on clinical context or data availability, such as B-mode and SE,[36] B-mode and SWAVE,[37] and B-mode and CDFI.[38] Finally, modality fusion has been extended to include dynamic imaging. Guo et al.[39] combined static B-mode images with CEUS videos to incorporate perfusion dynamics into the classification task. This points to a growing interest in combining both spatial and temporal modalities, and in developing architectures that can adapt to variability in imaging protocols, view availability, and modality completeness.

More recently, contrastive learning has been employed to ensure semantic alignment across modalities. Yan et al.[40] introduced a training objective that encouraged representations from different ultrasound modalities, B-mode, CDFI, and elastography, to converge in a shared embedding space, improving robustness and cross-modality consistency in cases where input modalities are missing or noisy. Similarly, Qian et al.[21] leveraged this trimodal ultrasound information, combining it with mammography and clinical metadata for diagnosis, demonstrating the benefits of multimodal fusion across imaging types and data sources.

3.2. Multi-view fusion in ultrasound

Fusing specific ultrasound imaging perspectives can also promote the classification performance. Wei et al.[41,42] introduced radial and anti-radial views at the data level to facilitate transfer learning with DenseNet[43] for diagnosis. Other studies have adopted combinations such as transverse and coronal views,[44] and longitudinal, transverse, and coronal views.[45,46] Qian et al.[47] utilized transverse and longitudinal views across B-mode, CDFI, and elastography for Breast Imaging Reporting and Data System classification, and Chen et al.[48] enhanced the flexibility to operate effectively on only the longitudinal or transverse view. On the other hand, Luo et al.[49] adopted a different approach by simulating spatial diversity through multi-resolution cropping of tumor regions, treating each resolution-scale pair as a unique perspective to stregthen representational richness.

For comparison, Table 1 provides a comprehensive summary of multimodal fusion approaches in selected ultrasound studies, including details on the modalities used (e.g., B-mode, CDFI); fusion strategies (early fusion at the input level, intermediate fusion at the feature level, and late fusion at the decision level); AI model architectures; patient dataset sizes; clinical objectives (e.g., diagnosis, screening, or tumor subtype classification); and model performance as measured by the area under the curve (AUC) or accuracy. Overall, studies employing intermediate or early-intermediate fusion strategies combined with conventional convolutional neural network (CNN) or ResNet architectures achieved high AUC values (greater than 95%). In contrast, one of the two studies using transformer-based models with similar fusion

Table 1
Summary of ultrasound studies.

Study	Modalities	Fusion Strategy	AI Method	Clinical Purpose	Dataset Size	Performance
Multi-modal Fusion						
Zhou et al.[32]	B-mode/CDFI/SWE	Intermediate	CNN	Subtype	807 patients	AUC: 96%
Wang et al.[33]	B-mode/SE/SWE/CDFI	Intermediate	ResNet	Diagnosis	835 patients	Acc: 95%
Yan et al.[40]	B-mode/CDFI/Elastography	Intermediate	CNN	Diagnosis	248 patients	AUC: 98%
Multi-view Fusion						
Wei et al.[41]	Radial/Anti-radial	Early	DenseNet	Diagnosis	578 patients	AUC: 98%
Ding et al.[45]	Long/Trans/Coronal	Intermediate	Transformer	Diagnosis	214 patients	Acc: 95%
Wang et al.[44]	Trans/Coronal	Early	CNN	Diagnosis	316 lesions	AUC: 95%
Other Fusion						
Wei et al.[42]	US/Report	Intermediate	Transformer	Diagnosis	290 patients	AUC: 85%
Qian et al.[21]	US/MG/Clinical	Intermediate	ResNet	Diagnosis	5,025 patients	Acc: 90%
Chen et al.[22]	US/MG	Intermediate	CNN	Screening	790 patients	AUC: 97%

Abbreviations: 1. B-mode: brightness mode 2. CDFI: color Doppler flow imaging 3. SWE: shear wave elastography 4. SE: strain elastography 5. MG: mammography 6. US: ultrasound 7. CNN: convolutional neural network 8. Acc: accuracy 9. AUC: area under the curve. Long, Trans, Clinical, and Report denote longitudinal view, transverse view, structured clinical data, and unstructured clinical report, respectively.

strategies reported lower AUC (85%). Notably, this study was the first to fuse inputs from two distinct modalities, ultrasound images and radiology text reports, and its limited dataset size likely contributed to the reduced performance.

3.3. Integration with radiomic features

Radiomic analysis has gained attention in breast ultrasound research for its ability to quantitatively characterize lesion morphology, texture, and heterogeneity beyond what is perceptible to the human eye. By extracting high-dimensional, handcrafted features from medical images, radiomics enables the capture of subtle image patterns associated with tumor biology. Several studies have applied this approach in combination with CEUS, which provides dynamic vascular information complementary to structural imaging. Li et al.[50] extracted radiomic features from B-mode, SWE, and CEUS images, using a support vector machine (SVM) optimized with a genetic algorithm for binary classification. Xu et al.[51] computed radiomic features from five anatomically defined regions of interest in B-mode and CEUS images. These features are refined through a neural network before classification with a SVM. Zhu et al.[52] examined the association between radiomic features derived from B-mode, SE, CDFI, and CEUS with breast cancer molecular subtypes. Their use of logistic regression highlights the broader applicability of radiomic features, which can be effectively leveraged by traditional machine learning algorithms without the need for deep neural networks to process image representations.

3.4. Fusion of ultrasound with other modalities

Ultrasound can also integrate with other data sources to enhance breast cancer classification such as DBT,[53] textual reports,[42] and mammography[22,54–56] to leverage complementary anatomical information. Zhang et al.[57] incorporated mammography for molecular subtype prediction, and Al-Tam et al.[58] emphasized interpretability in the combined framework. Beyond imaging, other studies have integrated ultrasound with clinical data.[59,60] Ye et al.[61] used such fusion to guide medical decisions, and He et al.[62] incorporated both mammographic and ultrasound images along with demographic data to build a comprehensive model for biopsy recommendation.

4. Mammography

Mammography remains a fundamental tool in breast cancer screening and diagnosis, widely used due to its ability to detect early-stage tumors through high-resolution X-ray imaging. It is often the first-line modality in population-wide screening programs and plays a critical role in risk assessment, lesion characterization, and biopsy planning. In recent years, AI algorithms have been increasingly applied to mammographic data to support automated detection, classification, and triaging. Deep learning and radiomic approaches have demonstrated promise in identifying subtle features that may not be visually discernible, improving diagnostic accuracy and workflow efficiency. As a result, mammographic analysis has become a central focus in the development of AI-assisted breast imaging tools.

To overcome the limitations of standard 2D imaging, advanced mammographic techniques have been developed to provide greater contrast and structural detail. DBT[63,64] reconstructs volumetric images from multiple low-dose X-ray projections, offering improved lesion localization and reducing the impact of tissue overlap, which is a common source of false positives and negatives in traditional mammography. Another innovation, contrast-enhanced spectral mammography (CESM),[65] uses iodine-based contrast agents and dual-energy acquisitions to highlight areas of abnormal vascularity, enhancing the visibility of malignant lesions.

Standard mammographic examinations typically include two primary imaging views: the craniocaudal (CC) and the mediolateral oblique (MLO). The CC view captures a top-down image of the breast, while the MLO view is angled to include both breast tissue and the upper axillary region. These orthogonal perspectives are designed to provide comprehensive coverage of the breast and improve lesion detectability by offering different anatomical contexts. AI models often leverage both views during training and inference to learn spatial correlations and improve classification performance. When used in conjunction with advanced modalities such as DBT or CESM, these views form the basis for more detailed multi-angle analysis, further enhancing the potential of computer-aided diagnosis systems.

4.1. Multi-view fusion in mammography

A wide range of studies utilize the standard CC and MLO mammographic views together to provide complementary perspectives of the breast. Several studies[66–70] have employed straightforward fusion strategies to combine CC and MLO images for binary classification. Song et al.[71] leveraged CESM by using both low-energy and dual-energy subtracted images in CC and MLO views. More advanced modeling techniques have also been developed to exploit inter-view relationships. A multi-task Siamese network[72] can jointly match the masses across both views,[73] and the Transformer model can fuse both views effectively.[74,75] Truong et al.[76] and Bobowicz et al.[77] further extended this by integrating both views using cross-attention and multiple instance learning[78] to train on unannotated data. Similarly, Liu et al.[79] represented the two views as a bipartite graph, enabling the use of graph convolutional networks[80] to model view-level interactions. These views can also incorporate with radiologist gaze information to enhance

interpretability and model guidance,[81] and clinical metadata and ultrasound modalities to enhance subtype prediction.[21]

4.2. Fusion of mammography with other modalities

A growing number of studies enhance mammographic analysis by integrating it with diverse data sources to improve diagnostic performance. Several works have concentrated on combining mammography with ultrasound, either through handcrafted features[53,54] or deep learning.[22,53,55–58] Others have combined mammography with clinical or demographic data for various tasks including subtype classification,[82] cancer prediction, and biopsy recommendation.[62,82–86] Radiomics also plays a role in cross-modal learning. Radiomic scores serve as a proxy for case difficulty in the curriculum learning framework.[87] In more specialized combinations, mammography can combine with histopathology[88] and dynamic contrast-enhanced MRI (DCE-MRI)[89] to leverage multimodality. More recent approaches leverage advanced data representations to improve multimodal breast cancer analysis. Vo et al.[90] applied frozen vision-language models to combine mammography images with electronic health records (EHR). Specifically, they utilized EVA-CLIP,[91] a refined variant of the original CLIP framework,[92] to encode mammography images and EHR data separately through dedicated visual and textual encoders, enabling effective data representation. These multimodal strategies illustrate the adaptability of mammographic analysis and its potential when enriched with heterogeneous clinical and imaging data.

Table 2 presents a comprehensive summary of multi-view mammography studies, including the modalities used, fusion strategies, AI models, patient dataset sizes, clinical objectives, and model performance measured by AUC or accuracy. The studies employing intermediate or late fusion strategies in combination with conventional CNN or ResNet architectures, or an SVM classifier, achieved consistently high AUC values for diagnostic tasks (ranging from 85% to 97%). In contrast, the two studies focused on screening used intermediate fusion and CNN architectures and reported markedly different AUCs (73% vs. 97%). This discrepancy appears to be closely linked to the substantial difference in dataset size, suggesting a potential issue with model generalization in the smaller study. The research on tumor subtype classification, which also adopted an intermediate fusion strategy and a CNN model, achieved an AUC of 93%.

5. MRI

MRI[93] is a cornerstone in breast cancer imaging due to its excellent soft tissue contrast and ability to provide both anatomical and functional information. In particular, it is invaluable in detecting tumors in dense breast tissue, assessing lesion extent, and monitoring treatment response. A variety of specialized MRI modalities are employed to capture different aspects of breast tumor biology. T1-weighted imaging (T1WI) and T2-weighted imaging (T2WI) sequences are standard anatomical imaging protocols. T1WI offers good contrast between fat and other tissues and is particularly effective after the administration of contrast agents. T2WI, on the other hand, is sensitive to differences in water content and is helpful in identifying edema, cysts, or necrosis within tumors. Contrast-enhanced MRI (CE-MRI)[94] leverages gadolinium-based contrast agents to improve the visibility of tumors by enhancing regions with abnormal vascularity. Malignant breast lesions often exhibit increased blood supply and leaky vessels due to angiogenesis, causing them to take up contrast more rapidly and intensely than surrounding normal tissue. Building on CE-MRI, DCE-MRI[95] captures a time series of T1WI before, during, and after contrast administration, allowing analysis of enhancement kinetics over time. DCE-MRI provides functional information about tumor perfusion and permeability, enabling characterization of lesions based on their contrast uptake and washout patterns. Diffusion-weighted imaging (DWI)[96] provides insight into the mobility of water molecules within tissue. Cancerous tissues tend to restrict water diffusion due to their high cellularity and dense extracellular matrix, leading to high signal intensity on DWI images. The corresponding apparent diffusion coefficient (ADC) maps quantify this diffusion and can help differentiate between benign and malignant lesions, with malignant tumors typically showing lower ADC values. Short tau inversion recovery (STIR)[97] is a fat-suppression technique often used in conjunction with other sequences. By nulling fat signals, STIR improves the visibility of lesions in fatty breast tissue, particularly in patients with dense breasts. Each of these MRI modalities provides complementary information. When integrated, they offer a powerful multimodal approach to breast cancer detection, classification, and treatment response assessment.

5.1. Multimodal fusion with MRI

Integrating multiple MRI modalities can markedly enhance model performance in breast cancer imaging by providing complementary anatomical and functional information. Morais et al.[98] employed a 3D CNN using DCE-MRI, T1WI, and T2WI to detect breast cancer. Similarly, Zhang et al.[99] leveraged a combination of T1WI, T2WI, DWI, DCE-MRI, and ADC maps for binary classification, showcasing the utility of integrating diverse imaging sequences to capture multiple biological characteristics of tumors. These MRI modalities are frequently combined with additional data sources, such as mammography[89] and clinical information, to improve both diagnostic and prognostic modeling. Several studies have demonstrated the benefit of integrating clinical variables to predict pathologic complete response (pCR) with various types of MRI, such as DCE-MRI[100,101]; T1WI and T2WI[102]; and T1WI, T2WI, DWI, and STIR.[103] Others aim to predict survival outcomes[104] and cancer recurrence[105] with multimodal MRI frameworks.

In addition to clinical data, more comprehensive approaches include the integration of genomics and whole slide imaging (WSI). Ye et al.[106]

Table 2
Summary of mammography studies.

Study	Modalities	Fusion Strategy	AI Method	Clinical Purpose	Dataset Size	Performance
Multi-view Fusion						
Song et al.[71]	LE/DES/CC/MLO	Intermediate	CNN	Diagnosis	95 patients	AUC: 97%
Tan et al.[66]	CC/MLO	Late	EfficientNet	Diagnosis	12,433 patients	AUC: 95%
Petrini et al.[67]	CC/MLO	Intermediate	EfficientNet	Diagnosis	3,103 images	AUC: 85%
Geras et al.[70]	CC/MLO	Intermediate	CNN	Screening	129,208 patients	AUC: 73%
Other Fusion						
Tian et al.[53]	DBT/CC/US	Intermediate/Late	SVM	Diagnosis	487 patients	AUC: 96%
Qian et al.[21]	CC/MLO/US/Clinical	Intermediate	ResNet	Diagnosis	5,025 patients	Acc: 90%
Chen et al.[22]	MG/US	Intermediate	CNN	Screening	790 patients	AUC: 97%
Tan et al.[56]	CC/MLO/US	Intermediate	CNN	Subtype	8,324 images	AUC: 93%

Abbreviations: 1. LE: low-energy 2. DES: dual-energy subtracted 3. CC: craniocaudal 4. MLO: mediolateral oblique 5. DBT: digital breast tomosynthesis 6. MG: mammography 7. US: ultrasound 8. CNN: convolutional neural network 9. Acc: accuracy 10. AUC: area under the curve 11. SVM: support vector machine. Clinical denotes structured clinical data.

Table 3
Summary of MRI studies.

Study	Modalities	Fusion Strategy	AI Method	Clinical Purpose	Dataset Size	Performance
Joo et al.[102]	T1WI/T2WI/Clinical	Intermediate	3D ResNet	pCR	536 patients	AUC: 89%
Yu et al.[104]	T1WI/T2WI/Clinical	Intermediate	3D DenseNet	Survival	1,199 patients	AUC: 90%
Gamal et al.[103]	T1WI/T2WI/STIR/DWI/Clinical	Intermediate/Late	LightGBM	pCR	109 patients	Acc: 88%
Verma et al.[101]	DCE-MRI/Clinical	Intermediate	3D CNN	pCR	207 patients	AUC: 88%
Ye et al.[106]	MRI/Geno/Clinical	Intermediate	Ridge	pCR	719 patients	AUC: 92%
Mao et al.[107]	T1WI/T2WI/DWI/WSI/Clinical	Intermediate	CNN	pCR	1,004 patients	AUC: 91%

Abbreviations: 1. T1: T1-weighted imaging 2. T2: T2-weighted imaging 3. STIR: short tau inversion recovery 4. MRI: magnetic resonance imaging 5. DCE-MRI: dynamic contrast-enhanced MRI 6. WSI: whole slide imaging 7. CNN: convolutional neural network 8. Acc: accuracy 9. AUC: area under the curve 10. DWI: diffusion-weighted imaging 11. pCR: pathologic complete response. Clinical and Geno denote structured clinical data and genomics, respectively.

combined MRI with proteomic, genomic, and clinical information using machine learning techniques to predict pCR in patients undergoing neoadjuvant therapy. Mao et al.[107] incorporated MRI, WSI, and clinical risk factors to assess treatment response, highlighting the potential of combining radiology and pathology data. Together, these studies reflect a growing trend in breast cancer research toward multimodal integration, where MRI serves as a central component alongside clinical, histological, and molecular data. Such approaches aim to improve the accuracy, generalizability, and clinical relevance of diagnostic and prognostic models, supporting more personalized and informed treatment strategies. Table 3 provides a comprehensive summary of MRI studies, including modality information (e.g., T1WI, T2WI, DWI), fusion strategies, AI models (e.g. 3D CNN, 3D ResNet, 3D DenseNet), patient dataset sizes, clinical objectives, and model performance as measured by AUC or accuracy. Overall, the reviewed studies achieved consistently high AUC or accuracy values, ranging from 88% to 92%, for both pCR and survival outcome prediction.

6. Other imaging techniques

6.1. Thermography

Thermographic imaging,[108] encompassing both infrared thermography (IRT) and terahertz (THz) imaging, is a non-invasive, radiation-free technique that detects physiological abnormalities by measuring heat distribution and electromagnetic responses from the body. As a functional imaging approach, thermography has gained attention in breast cancer research due to its ability to capture early physiological changes, such as increased vascularization, localized inflammation, and elevated metabolic activity, which often precede visible structural abnormalities. IRT[109] detects naturally emitted thermal radiation from the skin surface to produce temperature maps. In breast cancer screening, these maps can highlight asymmetries or localized hot spots linked to tumor growth. Despite its advantages, such as low cost, safety, and portability, IRT is limited to superficial measurements and is highly sensitive to external conditions, which has historically restricted its diagnostic reliability. On the other hand, THz imaging[110] offers shallow subsurface imaging by probing the tissue's dielectric properties. Malignant tissues tend to exhibit higher water content and distinct electromagnetic signatures, enabling differentiation from healthy tissue. While THz imaging provides molecular and compositional contrast, it faces practical challenges including limited penetration depth, slower imaging speed, and higher equipment cost. Together, IRT and THz represent complementary components of thermographic imaging, and Gezimati and Singh[111] combined both modalities with CNN for breast cancer classification.

6.2. Microscopy

Microscopic imaging[112] remains central to breast cancer diagnostics, providing detailed views of tissue architecture and cellular morphology. Clinical practice relies on bright-field microscopy (BFM) with hematoxylin and eosin staining to assess nuclear atypia, mitotic activity, ductal structures, and stromal invasion,[113] forming the basis of tumor classification, grading, and staging. Digital pathology and AI now augment this workflow through WSI, automated malignancy detection, and molecular marker prediction. Label-free techniques such as autofluorescence imaging (AFI) complement bright-field analysis by revealing metabolic and biochemical changes without stains. Endogenous fluorophores, such as nicotinamide adenine dinucleotide, flavin adenine dinucleotide, and collagen, vary during malignant transformation,[114] enabling real-time differentiation of cancerous and normal tissue, particularly useful for intraoperative margin assessment. Orthogonal polarization imaging (OPI) adds structural insight by suppressing surface reflections and highlighting subsurface features such as remodeled stromal collagen.[115] Together, these modalities integrate morphological, metabolic, and microstructural information to improve diagnostic accuracy and support surgical decision-making.

Recent studies have increasingly explored multimodal microscopic imaging with machine learning to enhance breast cancer classification. Wu et al.[116] combined BFM, AFI, and OPI with deep learning models to extract and fuse complementary information for diagnostic tasks. Other researchers combine BFM with other modalities to improve classification and prediction, such as mammography,[88] Ribonucleic acid sequences,[117,118] and gene expression data.[119,120] Similarly, Liu et al.[121] combined pathological images with gene expression data and copy number variants to predict molecular subtypes. To integrate with more data modalities, Mao et al.[107] incorporated WSI with MRI and clinical risk factors to forecast pCR. On the other hand, several studies have incorporated EHR or clinical data alongside imaging for diagnosis[122,123] and recurrence prediction.[124,125]

Table 4 summarizes representative thermography and microscopy studies, including the imaging modalities used, fusion strategies, AI models, patient dataset sizes, clinical objectives (diagnosis, recurrence, or tumor subtype classification), and model performance as measured by AUC or accuracy. In particular, in two comparative studies conducted for the same diagnostic purpose, one using a CNN and the other a feed-forward deep learning network, both achieved high performance, with reported accuracy of 98% and an AUC of 94%, respectively. However, a similar CNN used in another study for recurrence prediction achieved less favorable results (AUC of 72%) compared with the study that used a Transformer model for the same task (AUC of 89%), despite using the same fusion strategy. This discrepancy is likely due to the substantial difference in patient sample sizes (127 vs. 6,172).

6.3. PET and CT

Positron emission tomography (PET) and computed tomography (CT) are valuable imaging tools used together to evaluate breast cancer by combining metabolic and anatomical information. PET, often using fluorodeoxyglucose as a tracer, highlights areas of increased metabolic activity that may indicate cancer involvement in lymph nodes or distant organs. CT adds precise structural detail, helping clinicians assess tumor size, location, and surrounding tissue changes. When integrated

Table 4
Summary of thermography and microscopy studies.

Study	Modalities	Fusion Strategy	AI Method	Clinical Purpose	Dataset Size	Performance
Thermography						
Gezimati and Singh[111]	IRT/THz	Intermediate	CNN	Diagnosis	385 patients	AUC: 99%
Microscopy						
Wu et al.[116]	BFM/AFI/OPI	Early/Late	DL	Diagnosis	16 patients	AUC: 94%
Oyelade et al.[88]	HP/MG	Intermediate	CNN	Diagnosis	10,545 samples	Acc: 98%
Liu et al.[121]	HP/Gene/CNV	Intermediate	CNN	Subtype	1,098 samples	AUC: 94%
Boehm et al.[125]	HP/HP Report	Intermediate	Transformer	Recurrence	6,172 patients	AUC: 89%
Yang et al.[124]	HP/Clinical	Intermediate	CNN	Recurrence	127 patients	AUC: 72%

Abbreviations: 1. IRT: infrared thermography 2. THz: terahertz imaging 3. BFM: bright-field microscopy 4. AFI: autofluorescence imaging 5. OPI: orthogonal polarization imaging 6. HP: histopathological imaging 7. CNN: convolutional neural network 8. Acc: accuracy 9. AUC: area under the curve 10. MG: mammography 11. DL: deep learning 12. CNV: copy number variants. Clinica and Gene denote structured clinical data and gene expression data, respectively.

together, the modalities improve accuracy in detecting metastases, monitoring disease progression, and evaluating responses to therapy, making it an important component of comprehensive breast cancer management. Recent studies have explored the integration of PET and CT features with deep learning and machine learning approaches, including the use of DenseNet[43] and Transformer for subtype classification.[126,127] Additional work has used both modalities with XGBoost[128] to predict human epidermal growth factor receptor 2 (HER2) status.[129,130] PET and CT radiomics have also been integrated with machine learning methods to support prognostic assessment in breast cancer.[131] PET can also combine with other data modalities. Other research has combined PET with MRI using a SVM to enhance diagnostic accuracy[132] and fused PET with histopathological, genomic, and clinical information for treatment response prediction using a SVM.[133]

Table 5 summarizes representative PET and CT studies, including the imaging modalities, fusion strategies, AI models, patient dataset sizes, clinical objectives (diagnosis, pCR, tumor subtype classification, or HER2 status), and model performance as evaluated by AUC or accuracy. The two studies on tumor subtype classification using combined PET/CT modalities, both employing an intermediate fusion strategy but different AI models, a combined DenseNet/SVM model versus a Transformer, reported comparable results, with an accuracy of 94% and 95%, respectively. On the other hand, the remaining three studies presented in the Table employed traditional machine learning algorithms, including SVM and XGBoost, the latter being an ensemble learning method based on decision trees, for the tasks of pCR prediction, diagnosis, and HER2 status prediction. This choice is understandable given the small dataset sizes of 57, 102, and 217 patients, respectively. The corresponding AUCs were 82%, 98%, and 76%, respectively.

7. Clinical data

Clinical data are essential for breast cancer diagnosis, prognosis, and treatment planning. It can be divided into structured data, such as coded fields, checkboxes, and numerical entries, which are consistent and easily integrated into computational models, and unstructured data, including free-text clinical notes or radiology reports, which require natural language processing to extract actionable information. Both types provide complementary insights critical for patient care.

7.1. Structured clinical data

Demographic data are a key subset of structured clinical data, including patient age, sex, and ethnicity. Age is one of the most significant risk factors in breast cancer, with older individuals facing a higher likelihood of developing the disease. Including demographic information in computational models helps capture population-level variation and can contribute to more equitable and personalized decision-making in clinical settings. In many studies, structured clinical data are incorporated with thermography,[111] histopathology,[107,123,124] ultrasound,[59–61] MRI,[100–104] and mammography[21,62,82–87] for a wide range of clinical applications such as screening, diagnosis, pCR prediction, and recurrence estimation.

Another important source of clinical information is EHR, which offers structured data. Unlike unstructured clinical reports or narrative physician notes, structured components of EHR can be directly incorporated into machine learning models without the need for additional processing. These records support longitudinal analysis, allowing researchers and clinicians to track disease progression, monitor treatment response, and evaluate long-term outcomes. EHR has been incorporated with histopathological images,[122,123] MRI sequences,[101] and mammography.[90]

In parallel, many studies investigate the integration of clinical data with molecular-level information, including gene expression and copy number alterations,[121,134–137] genomics,[138,139] and protein expression data.[106,118] For instance, Karampuri et al.[140] explored similar integrations to predict the drug response in breast cancer. These works highlight the importance of combining structured clinical context with the underlying molecular landscape to support personalized breast cancer treatment.

7.2. Unstructured clinical reports

Unstructured clinical reports are narrative, free-text entries created by healthcare professionals as part of routine medical documentation.

Table 5
Summary of PET/CT studies.

Study	Modalities	Fusion Strategy	AI Method	Clinical Purpose	Dataset Size	Performance
Jia et al.[126]	PET/CT	Intermediate	DenseNet/SVM	Subtype	8,432 images	Acc: 94%
Ge et al.[127]	PET/CT	Intermediate	Transformer	Subtype	7,594 images	Acc: 95%
Chen et al.[130]	PET/CT	Intermediate	XGBoost	HER2	217 patients	AUC: 76%
Romeo et al.[132]	PET/MRI	Intermediate	SVM	Diagnosis	102 patients	AUC: 98%
Groheux et al.[133]	PET/HP/Geno	Intermediate	SVM	pCR	57 patients	AUC: 82%

Abbreviations: 1. PET: positron emission tomography 2. CT: computed tomography 3. MRI: magnetic resonance imaging 4. HP: histopathological imaging 5. SVM: support vector machine 6. HER2: human epidermal growth factor receptor 2 7. pCR: pathologic complete response 8. Acc: accuracy 9. AUC: area under the curve. Geno denotes Genomics.

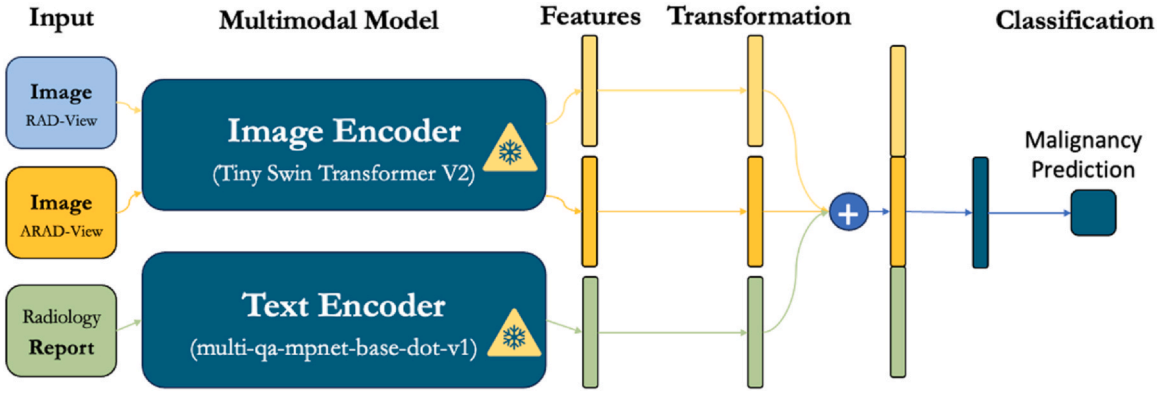


Figure 4. Model Architecture. Image and text encoders are used for independent feature extractions from the dual-view image pair and the report, which are subsequently aligned in a common space and merged through a single hidden layer for classification.[42]

Table 6
Summary of clinical data studies.

Study	Modalities	Fusion Strategy	AI Method	Clinical Purpose	Dataset Size	Performance
Structured Clinical Data						
Verma et al.[101]	Clinical/DCE-MRI	Intermediate	3D CNN	pCR	207 patients	AUC: 88%
Karampuri et al.[140]	Gene/Protein	Intermediate	Autoencoder	Treatment	14,963 samples	R^2: 92%
Liu et al.[121]	Gene/CNV/HP	Intermediate	CNN	Subtype	1,098 samples	AUC: 94%
Yu et al.[104]	Clinical/T1WI/T2WI	Intermediate	3D DenseNet	Survival	1,199 patients	AUC: 90%
Mao et al.[107]	Clinical/T1WI/T2WI/DWI/WSI	Intermediate	CNN	pCR	1,004 patients	AUC: 91%
Qian et al.[21]	Clinical/US/MG	Intermediate	ResNet	Diagnosis	5,025 patients	Acc: 90%
Unstructured Clinical Reports						
Wei et al.[42]	US/US Report	Intermediate	Transformer	Diagnosis	290 patients	AUC: 85%
Boehm et al.[125]	HP/HP Report	Intermediate	Transformer	Recurrence	6,172 patients	AUC: 89%

Abbreviations: 1. DCE-MRI: dynamic contrast-enhanced magnetic resonance imaging 2. CNV: copy number variants 3. T1: T1-weighted imaging 4. T2: T2-weighted imaging 5. DWI: diffusion-weighted imaging 6. HP: histopathological imaging 7. WSI: whole slide imaging 8. CNN: convolutional neural network 9. Acc: accuracy 10. AUC: area under the curve 11. MG: Mammography 12. US: ultrasound 13. pCR: pathologic complete response. 14. R^2: R-squared. Clinical denotes structured clinical data.

These reports encompass a wide range of clinical content, including diagnostic impressions, operative summaries, pathology interpretations, imaging findings, and progress notes. Unlike structured documentation formats that follow predefined templates or data fields, unstructured reports are composed in natural language, allowing clinicians to capture observations and clinical reasoning in a flexible and detailed manner.

The main advantage of unstructured clinical reports lies in their richness of clinical context and their capacity to reflect complex decision-making processes. However, this flexibility also presents significant challenges for data standardization, retrieval, and computational analysis. Variability in language, terminology, abbreviations, and report organization can make it difficult to extract consistent information across patient records or clinical settings.

Recent developments in natural language processing have considerably improved the ability to analyze unstructured clinical text. These techniques allow for the automated extraction of key clinical concepts such as diagnoses, treatment plans, responses to therapy, and disease progression.[141] Building on these foundations, current advances in large language models have further enhanced the accuracy, adaptability, and contextual understanding of clinical narratives. Models such as transformer-based architectures are capable of interpreting complex sentence structures, resolving ambiguities, and recognizing subtle clinical cues that earlier rule-based or statistical approaches often missed.

Large language models trained on biomedical corpora can be fine-tuned to extract highly specific clinical entities, map them to standardized terminologies, and even infer implicit relationships within the text. This has enabled more reliable identification of nuanced information such as temporal changes in disease status, eligibility criteria for clinical trials, and patient-specific risk factors. As a result, these models are increasingly being integrated into clinical decision support systems, EHR mining, and real-time information retrieval, supporting both research and point-of-care applications with unprecedented scale and precision.[142]

Building on the capabilities of language models, recent studies have explored the integration of textual and imaging data to improve clinical decision-making. Wei et al.[42] extracted features from unstructured clinical reports and combine with radial and anti-radial ultrasound views for diagnosis. As shown in Figure 4, these features are derived using pre-trained text encoders to capture language patterns and semantics. Similarly, Boehm et al.[125] trained a Transformer model to capture the patterns in the textual reports for histopathological images to predict recurrence scores. These approaches exemplify how medical contexts can be integrated with imaging modalities to enhance diagnostic accuracy.

Table 6 lists selected multimodal studies on structured or unstructured clinical data, including the modalities, fusion strategies, AI models, patient dataset sizes, clinical objectives, and model performance as evaluated by AUC, accuracy, or R-squared value. Notably, all of these studies employed an intermediate fusion strategy combined with autoencoders, CNNs, ResNets, or their 3D variants for tasks such as pCR prediction, treatment response, tumor subtype classification, and diagnosis. These studies consistently reported high AUC values, ranging from 88% to 94%.

8. Conclusions

This survey reviewed multimodal AI approaches for breast cancer management, encompassing a diverse range of data sources, including

imaging modalities such as ultrasound, mammography, MRI, CT, PET, and microscopy, alongside structured clinical data and unstructured clinical reports. By organizing the literature around individual data modalities, imaging perspectives, and their integration strategies, this review provides a modality-centric synthesis of how multimodal AI has been applied across the clinical workflow, highlighting its role in improving diagnosis, tumor characterization, treatment planning, and outcome prediction. Importantly, this review highlights multimodal data integration and multimodal modeling, rather than individual algorithmic choices on a single modality, as the key driver of recent progress in AI-assisted breast cancer management.

Across studies, intermediate, early, or late fusion strategies combined with deep learning architectures such as CNNs, ResNets, and their 3D variants consistently achieved high predictive performance, with AUC values typically ranging from 85% to 98% depending on the clinical task. Transformer-based models demonstrated promise in integrating heterogeneous modalities, although performance in some cases was limited by small dataset sizes. Traditional machine learning algorithms, including SVM and XGBoost, remain suitable for small datasets, achieving moderate to high AUCs depending on the task. Notably, discrepancies in performance were often linked to differences in dataset size, illustrating the critical influence of data availability and composition on model generalizability.

Overall, these findings demonstrate the effectiveness of multimodal fusion in advancing AI-driven breast cancer care. More detailed analyses reveal that multimodal fusion offers consistent performance gains when appropriately matched to data characteristics and clinical objectives. In particular, comprehensive studies of multimodal fusion approaches in selected ultrasound investigations that employed early or intermediate feature-level fusion combined with AI model architectures such as CNNs or ResNets, have demonstrated enhanced performance in clinical tasks, including diagnosis, screening, and tumor subtype classification, achieving AUC values exceeding 95%. Additionally, multi-view mammography studies using intermediate or late fusion strategies, in combination with CNN or ResNet architectures, have consistently achieved high AUC values across various diagnostic tasks, ranging from 85% to 97%. Similarly, representative MRI studies employing similar fusion strategies with 3D CNN models also showed notable performance improvements, with AUC or accuracy values ranging from 88% to 92% for predicting both pCR and survival outcomes. Importantly, the choice of AI architectures and fusion strategies should be guided by dataset size and specific task requirements.

Despite these encouraging results, several critical challenges remain. Many models are trained on relatively small or modality-imbalanced datasets, with limited external validation and underrepresentation of longitudinal data and diverse patient populations. Fusion strategies across studies often lack standardized, clinically guided protocols, thus limiting reproducibility and cross-study comparisons. Furthermore, the interpretability of multimodal models and their integration into routine clinical workflows remain insufficiently addressed, potentially constraining their adoption in clinical practice.

Moving forward, several directions are likely to advance the field. Developing large-scale, well-curated multimodal datasets that integrate imaging, clinical, and textual information is critical for training reliable models and ensuring rigorous, unbiased evaluation. Clinically relevant fusion strategies and multimodal foundation models have the potential to capture complex disease patterns while improving generalizability. Incorporating multi-view and longitudinal data, such as sequential mammograms, ultrasound images, or MRI scans collected before, during, and after chemotherapy, or tracking changes in tumor size, shape, or radiomic features over time, can enhance modeling of disease progression, treatment response, and recurrence risk. Equally important is embedding interpretability, uncertainty quantification, and human-AI collaboration into model design to ensure AI functions as an effective clinical decision-support system rather than an isolated predictor. Finally, rigorous external validation, prospective studies, and careful attention to regulatory and deployment considerations will be essential for translating multimodal AI from research into routine clinical practice for breast cancer management.

CRediT authorship contribution statement

Ting-Ruen Wei: Conceptualization, Methodology, Validation, Formal analysis, Investigation, Resources, Data Curation, Writing, Visualization. **Yuling Yan:** Conceptualization, Methodology, Formal analysis, Investigation, Writing, Supervision, Project administration, Funding acquisition.

Declaration of Competing Interest

Yuling Yan is an editorial board member for Intelligent Oncology and was not involved in the editorial review or the decision to publish this article. The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

1. Giaquinto AN, Sung H, Newman LA, et al. Breast cancer statistics 2024. *CA Cancer J Clin.* 2024;74(6):477–495.
2. Suzuki M, Nakayama R, Namba K, et al. Diagnostic performance of coronal view in comparison with transverse view of three-dimensional automated breast ultrasound. *Acta Radiol.* 2021;62(1):27–33.
3. Chang Y, Wang X, Wang J, et al. A survey on evaluation of large language models. *ACM Trans Intell Syst Technol.* 2024;15(3):1–45.
4. Zhang J, Huang J, Jin S, Lu S. Vision-language models for vision tasks: a survey. *IEEE Trans Pattern Anal Mach Intell.* 2024.
5. Luo R, Sun L, Xia Y, et al. BioGPT: generative pre-trained transformer for biomedical text generation and mining. *Brief Bioinforma.* 2022;23(6):bbac409.
6. Wang Z, Wu Z, Agarwal D, Sun J. Medclip: Contrastive learning from unpaired medical images and text. In: *Proceedings of the Conference on Empirical Methods in Natural Language Processing. Conference on Empirical Methods in Natural Language Processing.* Vol 2022. 2022:3876.
7. Li C, Wong C, Zhang S, et al. Llava-med: training a large language-and-vision assistant for biomedicine in one day. *Adv Neural Inf Process Syst.* 2023;36:28541–28564.
8. Sun Y, Zhu C, Zheng S, et al. Pathasst: A generative foundation ai assistant towards artificial general intelligence of pathology. In: *Proceedings of the AAAI Conference on Artificial Intelligence.* Vol 38. 2024:5034-5042.
9. Mendes J, Domingues J, Aidos H, Garcia N, Matela N. AI in breast cancer imaging: a survey of different applications. *J Imaging.* 2022;8(9):228.
10. Pesapane F, Giambersio E, Capetti B, et al. Patients' perceptions and attitudes to the use of artificial intelligence in breast cancer diagnosis: a narrative review. *Life.* 2024;14(4):454.
11. Nakach F, Idri A, Goceri E. A comprehensive investigation of multimodal deep learning fusion strategies for breast cancer classification. *Artif Intell Rev.* 2024;57(12):327.
12. Shah SM, Khan RA, Arif S, Sajid U. Artificial intelligence for breast cancer analysis: trends & directions. *Comput Biol Med.* 2022;142:105221.
13. Chugh G, Kumar S, Singh N. Survey on machine learning and deep learning applications in breast cancer diagnosis. *Cogn Comput.* 2021;13(6):1451–1470.
14. Saleh GA, Batouty NM, Gamal A, et al. Impact of imaging biomarkers and AI on breast cancer management: a brief review. *Cancers.* 2023;15(21):5216.
15. Yoon JH, Strand F, Baltzer PA, et al. Standalone AI for breast cancer detection at screening digital mammography and digital breast tomosynthesis: a systematic review and meta-analysis. *Radiology.* 2023;307(5):e222639.
16. Díaz O, Rodríguez-Ruíz A, Sechopoulos I. Artificial Intelligence for breast cancer detection: technology, challenges, and prospects. *Eur J Radiol.* 2024:111457.
17. Sechopoulos I, Teuwen J, Mann R. Artificial intelligence for breast cancer detection in mammography and digital breast tomosynthesis: state of the art. *Semin Cancer Biol.* 2021;72:214–225.
18. Nassif AB, Talib MA, Nasir Q, Afadar Y, Elgendy O. Breast cancer detection using artificial intelligence techniques: a systematic literature review. *Artif Intell Med.* 2022;127:102276.
19. Zhang T, Tan T, Samperna R, et al. Radiomics and artificial intelligence in breast imaging: a survey. *Artif Intell Rev.* 2023;56(1):857–892.
20. Llinas-Bertran A, Butjosa-Espín M, Barberi V, Seoane JA. Multimodal data integration in early-stage breast cancer. *Breast.* 2025:103892.
21. Qian X, Pei J, Han C, et al. A multimodal machine learning model for the stratification of breast cancer risk. *Nat Biomed Eng.* 2025;9(3):356–370.
22. Chen J, Pan T, Zhu Z, et al. A deep learning-based multimodal medical imaging model for breast cancer screening. *Sci Rep.* 2025;15(1):14696.
23. Guo R, Lu G, Qin B, Fei B. Ultrasound imaging technologies for breast cancer detection and management: a review. *Ultrasound Med Biol.* 2018;44(1):37–70.
24. Hangiandreou NJ. AAPM/RSNA physics tutorial for residents: topics in US: B-mode US: basic concepts and new technology. *Radiographics.* 2003;23(4):1019–1033.
25. Adler DD, Carson PL, Rubin JM, Quinn-Reid D. Doppler ultrasound color flow imaging in the study of breast cancer: preliminary findings. *Ultrasound Med Biol.* 1990;16(6):553–559.

26. Balleyguier C, Opolon P, Mathieu MC, et al. New potential and applications of contrast-enhanced ultrasound of the breast: own investigations and review of the literature. *Eur J Radiol.* 2009;69(1):14–23.
27. Goddi A, Bonardi M, Alessi S. Breast elastography: a literature review. *J Ultrasound.* 2012;15(3):192–198.
28. Liu B, Zheng Y, Huang G, et al. Breast lesions: quantitative diagnosis using ultrasound shear wave elastography—a systematic review and meta-analysis. *Ultrasound Med Biol.* 2016;42(4):835–847.
29. Eskandari H, Salcudean SE, Rohling R, et al. Identifying malignant and benign breast lesions using vibroelastography. In: *2013 IEEE International Ultrasonics Symposium (IUS)*. IEEE; 2013:25–28.
30. Carlsen JF, Ewertsen C, Lönn L, Nielsen MB. Strain elastography ultrasound: an overview with emphasis on breast cancer diagnosis. *Diagnostics.* 2013;3(1):117–125.
31. Hooley RJ, Scoutt LM, Philpotts LE. Breast ultrasonography: state of the art. *Radiology.* 2013;268(3):642–659.
32. Zhou BY, Wang LF, Yin HH, et al. Decoding the molecular subtypes of breast cancer seen on multimodal ultrasound images using an assembled convolutional neural network model: A prospective and multicentre study. *EBioMedicine.* 2021;74:103684.
33. Wang J, Miao J, Yang X, et al. Auto-weighting for breast cancer classification in multimodal ultrasound. In: *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer; 2020:190–199.
34. He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2016:770-778.
35. Huang R, Lin Z, Dou H, et al. AW3M: an auto-weighting and recovery framework for breast cancer diagnosis using multi-modal ultrasound. *Med Image Anal.* 2021;72:102137.
36. Misra S, Yoon C, Kim K, et al. Deep learning-based multimodal fusion network for segmentation and classification of breast cancers using B-mode and elastography ultrasound images. *Bioeng Transl Med.* 2023;8(6):e10480.
37. Shao Y, Hashemi HS, Gordon P, et al. Breast cancer detection using multimodal time series features from ultrasound shear wave absolute vibro-elastography. *IEEE J Biomed Health Inform.* 2021;26(2):704–714.
38. Sultan LR, Schultz SM, Cary TW, Sehgal CM. Machine learning to improve breast cancer diagnosis by multimodal ultrasound. In: *2018 IEEE International Ultrasonics Symposium (IUS)*. IEEE; 2018:1–4.
39. Guo D, Lu C, Chen D, et al. A multimodal breast cancer diagnosis method based on Knowledge-Augmented Deep Learning. *Biomed Signal Process Control.* 2024;90:105843.
40. Yan P, Gong W, Li M, et al. TDF-Net: Trusted Dynamic Feature Fusion Network for breast cancer diagnosis using incomplete multimodal ultrasound. *Inf Fusion.* 2024;112:102592.
41. Wei T, Hell M, Vierra A, et al. Breast Cancer Detection on Dual-View Sonography via Data-Centric Deep Learning. *IEEE Open J Eng Med Biol.* 2024.
42. Wei TR, Chang A, Kang Y, Patel M, Fang Y, Yan Y. Multimodal deep learning for enhanced breast cancer diagnosis on sonography. *Comput Biol Med.* 2025;194:110466.
43. Huang G, Liu Z, Van Der Maaten L, Weinberger KQ. Densely connected convolutional networks. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2017:4700-4708.
44. Wang Y, Choi EJ, Choi Y, Zhang H, Jin GY, Ko SB. Breast Cancer Classification in Automated Breast Ultrasound Using Multiview Convolutional Neural Network with Transfer Learning. *Ultrasound Med Biol.* 2020;46(5):1119–1132.
45. Ding W, Zhang H, Zhuang S, Zhuang Z, Gao Z. Multi-view stereoscopic attention network for 3D tumor classification in automated breast ultrasound. *Expert Syst Appl.* 2023;234:120969.
46. Zhou Y, Chen H, Li Y, Wang S, Cheng L, Li J. 3D multi-view tumor detection in automated whole breast ultrasound using deep convolutional neural network. *Expert Syst Appl.* 2021;168:114410.
47. Qian X, Pei J, Zheng H, et al. Prospective assessment of breast cancer risk from multimodal multiview ultrasound images via clinically applicable deep learning. *Nat Biomed Eng.* 2021;5(6):522–532.
48. Chen H, Li Y, Zhang J, et al. IRUM: An Image Representation and Unified Learning Method for Breast Cancer Diagnosis from Multi-View Ultrasound Images. *Int Workshop Mach Learn Med Imaging.* 2024:22–30.
49. Luo Y, Huang Q, Liu L. Classification of tumor in one single ultrasound image via a novel multi-view learning strategy. *Pattern Recognit.* 2023;143:109776.
50. Li Y, Liu Y, Zhang M, Zhang G, Wang Z, Luo J. Radiomics with attribute bagging for breast tumor classification using multimodal ultrasound images. *J Ultrasound Med.* 2020;39(2):361–371.
51. Xu Z, Wang Y, Chen M, Zhang Q. Multi-region radiomics for artificially intelligent diagnosis of breast cancer using multimodal ultrasound. *Comput Biol Med.* 2022;149:105920.
52. Zhu J, He H, Jiang X, Bao H, Chen F. Multimodal ultrasound features of breast cancers: correlation with molecular subtypes. *BMC Med Imaging.* 2023;23(1):57.
53. Tian R, Lu G, Zhao N, Qian W, Ma H, Yang W. Constructing the Optimal Classification Model for Benign and Malignant Breast Tumors Based on Multifeature Analysis from Multimodal Images. *J Imaging Inform Med.* 2024:1–15.
54. Atrey K, Singh BK, Bodhey NK. Multimodal classification of breast cancer using feature level fusion of mammogram and ultrasound images in machine learning paradigm. *Multimed Tools Appl.* 2023:1–22.
55. Muduli D, Dash R, Majhi B. Automated diagnosis of breast cancer using multi-modal datasets: A deep convolution neural network based approach. *Biomed Signal Process Control.* 2022;71:102825.
56. Tan T, Rodriguez-Ruiz A, Zhang T, et al. Multi-modal artificial intelligence for the combination of automated 3D breast ultrasound and mammograms in a population of women with predominantly dense breasts. *Insights into Imaging.* 2023;14(1):10.
57. Zhang T, Tan T, Han L, et al. Predicting breast cancer types on and beyond molecular level in a multi-modal fashion. *NPJ Breast Cancer.* 2023;9(1):16.
58. Al-Tam RM, Al-Hejri AM, Alshamrani SS, Al-antari MA, Narangale SM. Multimodal breast cancer hybrid explainable computer-aided diagnosis using medical mammograms and ultrasound Images. *Biocybern Biomed Eng.* 2024;44(3):731–758.
59. Yu Z, Hong Y, Chou C. Enhancing Breast Cancer Diagnosis: A Nomogram Model Integrating AI Ultrasound and Clinical Factors. *Ultrasound Med Biol.* 2024;50(9):1372–1380.
60. Pfob A, Sidey-Gibbons C, Barr RG, et al. The importance of multi-modal imaging and clinical information for humans and AI-based algorithms to classify breast masses (INSPiRED 003): an international, multicenter analysis. *Eur Radiol.* 2022;32(6):4101–4115.
61. Ye J, Liu Y, Ren S, et al. A multimodal information-interconnected network for medication guidance in HR + /HER2-breast cancer treatment. *Inf Fusion.* 2025:103326.
62. He T, Puppala M, Ezeana CF, et al. A deep learning–based decision support tool for precision risk assessment of breast cancer. *JCO Clin Cancer Inform.* 2019;3:1–12.
63. Vedantham S, Karellas A, Vijayaraghavan GR, Kopans DB. Digital breast tomosynthesis: state of the art. *Radiology.* 2015;277(3):663–684.
64. Chong A, Weinstein SP, McDonald ES, Conant EF. Digital breast tomosynthesis: concepts and clinical practice. *Radiology.* 2019;292(1):1–14.
65. Tagliafico AS, Bignotti B, Rossi F, et al. Diagnostic performance of contrast-enhanced spectral mammography: systematic review and meta-analysis. *Breast.* 2016;28:13–19.
66. Tan H, Wu Q, Wu Y, et al. Mammography-based artificial intelligence for breast cancer detection, diagnosis, and BI-RADS categorization using multi-view and multi-level convolutional neural networks. *Insights into Imaging.* 2025;16(1):1–12.
67. Petrini DG, Shimizu C, Roela RA, Valente GV, Folgueira MAAK, Kim HY. Breast cancer diagnosis in two-view mammography using end-to-end trained efficientnet-based convolutional network. *IEEE Access.* 2022;10:77723–77731.
68. Wu Q, Tan H, Qiao Z, et al. Cross-view contrastive mutual learning across masked autoencoders for mammography diagnosis. In: *International Workshop on Machine Learning in Medical Imaging*. Springer; 2023:74–83.
69. Chen Y, Wang H, Wang C, et al. Multi-view local co-occurrence and global consistency learning improve mammogram classification generalisation. In: *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer; 2022:3–13.
70. Geras KJ, Wolfson S, Shen Y, et al. High-resolution breast cancer screening with multi-view deep convolutional neural networks. *arXiv preprint arXiv:1703.07047*. Published online 2017.
71. Song J, Zheng Y, Zakir Ullah M, et al. Multiview multimodal network for breast cancer diagnosis in contrast-enhanced spectral mammography images. *Int J Comput Assist Radiol Surg.* 2021;16(6):979–988.
72. Li Y, Chen CP, Zhang T. A survey on siamese network: Methodologies, applications, and opportunities. *IEEE Trans Artif Intell.* 2022;3(6):994–1014.
73. Yan Y, Conze P, Lamard M, Quellec G, Cochener B, Coatrieux G. Towards improved breast mass detection using dual-view mammogram matching. *Med Image Anal.* 2021;71:102083.
74. Chen X, Zhang K, Abdoli N, et al. Transformers improve breast cancer diagnosis from unregistered multi-view mammograms. *Diagnostics.* 2022;12(7):1549.
75. Sun Z, Jiang H, Ma L, Yu Z, Xu H. Transformer based multi-view network for mammographic image classification. In: *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer; 2022:46–54.
76. Truong Vu YN, Guo D, Taha A, Su J, Matthews TP. M&m: Tackling false positives in mammography with a multi-view and multi-instance learning sparse detector. In: *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer; 2023:778-788.
77. Bobowicz M, Rygusik M, Buler J, et al. Attention-based deep learning system for classification of breast lesions—multimodal, weakly supervised approach. *Cancers.* 2023;15(10):2704.
78. Carbonneau M, Cheplygina V, Granger E, Gagnon G. Multiple instance learning: A survey of problem characteristics and applications. *Pattern Recognit.* 2018;77:329–353.
79. Liu Y, Zhang F, Zhang Q, Wang S, Wang Y, Yu Y. Cross-view correspondence reasoning based on bipartite graph convolutional network for mammogram mass detection. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2020:3812-3822.
80. Zhang S, Tong H, Xu J, Maciejewski R. Graph convolutional networks: a comprehensive review. *Comput Soc Netw.* 2019;6(1):1–23.
81. Ji C, Du C, Zhang Q, Mammo-net, et al. Integrating gaze supervision and interactive information in multi-view mammogram classification. In: *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer; 2023:68–78.
82. Ben Rabah C, Sattar A, Ibrahim A, Serag A. A Multimodal Deep Learning Model for the Classification of Breast Cancer Subtypes. *Diagnostics.* 2025;15(8):995.
83. Zhu X, Wolfgruber TK, Leong L, et al. Deep learning predicts interval and screening-detected cancer from screening mammograms: a case-case-control study in 6369 women. *Radiology.* 2021;301(3):550–558.
84. Akselrod-Ballin A, Chorev M, Shoshan Y, et al. Predicting breast cancer by applying deep learning to linked health records and mammograms. *Radiology.* 2019;292(2):331–342.
85. Barros V, Tlusty T, Barkan E, et al. Virtual biopsy by using artificial intelligence–based multimodal modeling of binational mammography data. *Radiology.* 2022;306(3):e220027.

86. Vo HQ, Yuan P, He T, Wong ST, Nguyen HV. Multimodal breast lesion classification using cross-attention deep networks. In: *2021 IEEE EMBS International Conference on Biomedical and Health Informatics (BHI)*. IEEE; 2021:1–4.
87. Nebbia G, Dadsetan S, Arefan D, et al. Radiomics-informed deep curriculum learning for breast cancer diagnosis. In: *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer; 2021:634–643.
88. Oyelade ON, Irunokhai EA, Wang H. A twin convolutional neural network with hybrid binary optimizer for multimodal breast cancer digital image classification. *Sci Rep*. 2024;14(1):692.
89. Mokni R, Gargouri N, Damak A, Sellami D, Feki W, Mnif Z. An automatic Computer-Aided Diagnosis system based on the Multimodal fusion of Breast Cancer (MF-CAD). *Biomed Signal Process Control*. 2021;69:102914.
90. Vo HQ, Wang L, Wong KK, et al. Frozen Large-scale Pretrained Vision-Language Models are the Effective Foundational Backbone for Multimodal Breast Cancer Prediction. *IEEE J Biomed Health Inform*. 2024.
91. Sun Q, Fang Y, Wu L, Wang X, Cao Y. Eva-clip: Improved training techniques for clip at scale. *arXiv preprint arXiv:2303.15389*. Published online 2023.
92. Radford A, Kim JW, Hallacy C, et al. Learning transferable visual models from natural language supervision. In: *International Conference on Machine Learning*. PmLR; 2021:8748–8763.
93. Mann RM, Cho N, Moy L. Breast MRI: State of the Art. *Radiology*. 2019;292(3):520-536.
94. Chung M, Calabrese E, Mongan J, et al. Deep learning to simulate contrast-enhanced breast MRI of invasive breast cancer. *Radiology*. 2022;306(3):e213199.
95. Jiang Y, Edwards AV, Newstead GM. Artificial intelligence applied to breast MRI for improved diagnosis. *Radiology*. 2021;298(1):38–46.
96. Amornsiripanitch N, Bickelhaupt S, Shin HJ, et al. Diffusion-weighted MRI for unenhanced breast cancer screening. *Radiology*. 2019;293(3):504–520.
97. LeBihan D, Iima M, Partridge SC. Fat-signal suppression in breast diffusion-weighted imaging: the Good, the Bad, and the Ugly. *Eur Radiol*. 2025;35(2):733–741.
98. Morais M, Calisto FM, Santiago C, Aleluia C, Nascimento JC. Classification of breast cancer in Mri with multimodal fusion. In: *2023 IEEE 20th International Symposium on Biomedical Imaging (ISBI)*. IEEE; 2023:1–4.
99. Zhang Z, Lan H, Zhao S. Analysis of the value of quantitative features in multimodal MRI images to construct a radio-omics model for breast cancer diagnosis. *Breast Cancer Targets Therapy*. 2024:305–318.
100. Holste G, Partridge SC, Rahbar H, Biswas D, Lee CI, Alessio AM. End-to-end learning of fused image and non-image features for improved breast cancer classification from mri. In: *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 2021:3294-3303.
101. Verma M, Abdelrahman L, Collado-Mesa F, Abdel-Mottaleb M. Multimodal spatio-temporal deep learning framework to predict response of breast cancer to neoadjuvant systemic therapy. *Diagnostics*. 2023;13(13):2251.
102. Joo S, Ko ES, Kwon S, et al. Multimodal deep learning models for the prediction of pathologic response to neoadjuvant chemotherapy in breast cancer. *Sci Rep*. 2021;11(1):18800.
103. Gamal A, Sharafeldeen A, Alnaghy E, et al. A novel machine learning approach for predicting neoadjuvant chemotherapy response in breast cancer: integration of multi-modal radiomics with clinical and molecular subtype markers. *IEEE Access*. 2024.
104. Yu Y, Ren W, Mao L, et al. MRI-based multimodal AI model enables prediction of recurrence risk and adjuvant therapy in breast cancer. *Pharmacol Res*. 2025:107765.
105. Rabinovici-Cohen S, Fernández XM, Grandal Rejo B, et al. Multimodal prediction of five-year breast Cancer recurrence in women who receive Neoadjuvant chemotherapy. *Cancers*. 2022;14(16):3848.
106. Ye Z, Yuan J, Hong D, Xu P, Liu W. Multimodal diagnostic models and subtype analysis for neoadjuvant therapy in breast cancer. *Front Immunol*. 2025;16:1559200.
107. Mao N, Dai Y, Zhou H, et al. A multimodal and fully automated system for prediction of pathological complete response to neoadjuvant chemotherapy in breast cancer. *Sci Adv*. 2025;11(18):eadr1576.
108. BaniAhmad AY, Alzubi JA, Vasanthan M, Kondaveeti SB, Shreyas J, Priyanka TP. Efficient hybrid heuristic adopted deep learning framework for diagnosing breast cancer using thermography images. *Sci Rep*. 2025;15(1):13605.
109. Alzahrani RM, Sikkandar MY, Begum SS, et al. Early breast cancer detection via infrared thermography using a CNN enhanced with particle swarm optimization. *Sci Rep*. 2025;15(1):25290.
110. Chavez T, Vohra N, Bailey K, El-Shenawee M, Wu J. Supervised Bayesian learning for breast cancer detection in terahertz imaging. *Biomed Signal Process Control*. 2021;70:102949.
111. Gezimati M, Singh G. Deep Learning for Multimodal Breast Cancer Characterization with Emergence of Terahertz and Infrared Imaging. *IEEE Trans Instrum Meas*. Published online 2025.
112. Kowal M, Filipczuk P, Obuchowicz A, Korbicz J, Monczak R. Computer-aided diagnosis of breast cancer based on fine needle biopsy microscopic images. *Comput Biol Med*. 2013;43(10):1563–1572.
113. Veta M, Pluim JP, Van Diest PJ, Viergever MA. Breast cancer histopathology image analysis: A review. *IEEE Trans Biomed Eng*. 2014;61(5):1400–1411.
114. Ostrander JH, McMahon CM, Lem S, et al. Optical redox ratio differentiates breast cancer cell lines based on estrogen receptor status. *Cancer Res*. 2010;70(11):4759–4766.
115. Groner W, Winkelman JW, Harris AG, et al. Orthogonal polarization spectral imaging: a new method for study of the microcirculation. *Nat Med*. 1999;5(10):1209–1212.
116. Wu J, Xu Z, Shang L, et al. Multimodal microscopic imaging with deep learning for highly effective diagnosis of breast cancer. *Opt Lasers Eng*. 2023;168:107667.
117. Yu Y, Cai G, Lin R, et al. Multimodal data fusion AI model uncovers tumor microenvironment immunotyping heterogeneity and enhanced risk stratification of breast cancer. *MedComm*. 2024;5(12):e70023.
118. Palmal S, Arya N, Saha S, Tripathy S. Integrative prognostic modeling for breast cancer: Unveiling optimal multimodal combinations using graph convolutional networks and calibrated random forest. *Appl Soft Comput*. 2024;154:111379.
119. Rani S, Ahmad T, Masood S, Saxena C. Diagnosis of breast cancer molecular subtypes using machine learning models on unimodal and multimodal datasets. *Neural Comput Appl*. 2023;35(34):24109–24121.
120. Subramanian V, Syeda-Mahmood T, Do MN. Multimodal fusion using sparse CCA for breast cancer survival prediction. In: *2021 IEEE 18th International Symposium on Biomedical Imaging (ISBI)*. IEEE; 2021:1429–1432.
121. Liu T, Huang J, Liao T, Pu R, Liu S, Peng Y. A hybrid deep learning model for predicting molecular subtypes of human breast cancer using multimodal data. *Irbm*. 2022;43(1):62–74.
122. Yan R, Zhang F, Rao X, et al. Richer fusion network for breast cancer classification based on multimodal data. *BMC Med Inform Decis Mak*. 2021;21(Suppl 1):134.
123. Yan R, Ren F, Rao X, et al. Integration of multimodal data for breast cancer classification using a hybrid deep learning method. In: *International Conference on Intelligent Computing*. Springer; 2019:460–469.
124. Yang J, Ju J, Guo L, et al. Prediction of HER2-positive breast cancer recurrence and metastasis risk from histopathological images and clinical information via multimodal deep learning. *Comput Struct Biotechnol J*. 2022;20:333–342.
125. Boehm KM, El Nahhas OS, Marra A, et al. Multimodal histopathologic models stratify hormone receptor-positive early breast cancer. *Nat Commun*. 2025;16(1):2106.
126. Jia L, Chen Z, Shao D, Abba AA. Breast cancer subtype classification based on PET/CT bimodal imaging feature fusion. In: *2023 IEEE 6th International Conference on Pattern Recognition and Artificial Intelligence (PRAI)*. IEEE; 2023:25–29.
127. Ge C, Chen Z. A PET/CT dual-modal breast cancer classification method based on vision transformer. In: *2025 8th International Conference on Computer Information Science and Application Technology (CISAT)*. IEEE; 2025:608–613.
128. Chen T. XGBoost: A Scalable Tree Boosting System. *Cornell University*. Published online 2016.
129. Qin L, Chen W, Ye Y, et al. Prediction of HER2 expression in gastric adenocarcinoma based on preoperative noninvasive multimodal 18F-FDG PET/CT imaging. *Acad Radiol*. 2024;31(8):3200–3211.
130. Chen Y, Wang Z, Yin G, et al. Prediction of HER2 expression in breast cancer by combining PET/CT radiomic analysis and machine learning. *Ann Nucl Med*. 2022;36(2):172–182.
131. Zhong H, Huang D, Wu J, Chen X, Chen Y, Huang C. 18F-FDG PET/CT based radiomics features improve prediction of prognosis: multiple machine learning algorithms and multimodality applications for multiple myeloma. *BMC Med Imaging*. 2023;23(1):87.
132. Romeo V, Clauser P, Rasul S, et al. AI-enhanced simultaneous multiparametric 18F-FDG PET/MRI for accurate breast cancer diagnosis. *Eur J Nucl Med Mol Imaging*. 2022;49(2):596–608.
133. Groheux D, Ferrer L, Vargas J, et al. FDG-PET/CT and multimodal machine learning model prediction of pathological complete response to neoadjuvant chemotherapy in triple-negative breast cancer. *Cancers*. 2025;17(7):1249.
134. Sun D, Wang M, Li A. A multimodal deep neural network for human breast cancer prognosis prediction by integrating multi-dimensional data. *IEEE/ACM Trans Comput Biol Bioinforma*. 2018;16(3):841–850.
135. Kayikci S, Khoshgoftaar TM. Breast cancer prediction using gated attentive multimodal deep learning. *J Big Data*. 2023;10(1):62.
136. Kayikci S, Khoshgoftaar T. A stack based multimodal machine learning model for breast cancer diagnosis. In: *2022 International Congress on Human-Computer Interaction, Optimization and Robotic Applications (HORA)*. IEEE; 2022:1–5.
137. Du X, Zhao Y. Multimodal adversarial representation learning for breast cancer prognosis prediction. *Comput Biol Med*. 2023;157:106765.
138. Guo W, Liang W, Deng Q, Zou X. A multimodal affinity fusion network for predicting the survival of breast cancer patients. *Front Genet*. 2021;12:709027.
139. Karim MR, Wicaksono G, Costa IG, Decker S, Beyan O. Prognostically relevant subtypes and survival prediction for breast cancer based on multimodal genomics data. *IEEE Access*. 2019;7:133850–133864.
140. Karampuri A, Kundur S, Perugu S. Exploratory drug discovery in breast cancer patients: a multimodal deep learning approach to identify novel drug candidates targeting RTK signaling. *Comput Biol Med*. 2024;174:108433.
141. Diab KM, Deng J, Wu Y, Yesha Y, Collado-Mesa F, Nguyen P. Natural language processing for breast imaging: a systematic review. *Diagnostics*. 2023;13(8):1420.
142. Liu F, Zhou H, Gu B, et al. Application of large language models in medicine. *Nat Rev Bioeng*. 2025:1–20.