Narrative Review Open Access

Archives Of Breast Cancer

DOI: 10.32768/abc.4829103847-562

Artificial Intelligence-Driven Multimodal Integration of Genomic, Pathological, and Clinical Data in Breast Cancer: A Narrative Review of Clinical Applications and Translational Challenges

Jamshad Taslimi[1], Alireza Minagar*[2]

[1] Tehran University of Medical Sciences (TUMS), Tehran, Iran

[2] Adjunct Faculty, School of Cybersecurity and Information Technology, University of Maryland Global Campus, Adelphi, MD, USA

* **Corresponding Author:** Alireza Minagar, MD, MBA, MS. Email: aminagar@gmail.com

ARTICLE INFO

Received:
19 May 2026

Revised:
13 August 2026

Accepted:
17 August 2026

Keywords:
breast neoplasms, deep learning, multimodal imaging, computer-assisted image processing, natural language processing, precision medicine, clinical decision support systems

ABSTRACT

Background: Breast cancer remains the most frequently diagnosed malignancy in women worldwide, and its clinical heterogeneity imposes formidable challenges on treatment standardization and outcome prediction. The emergence of artificial intelligence (AI) and multimodal data integration strategies is fundamentally reshaping how genomic, histopathological, radiological, and electronic health record (EHR) data are synthesized into actionable clinical intelligence.

Methods: We conducted a narrative review of peer-reviewed literature published between 2015 and 2026, focusing on AI-driven approaches to genomic profiling, digital pathology, radiomics, natural language processing (NLP) of clinical text, and multimodal data fusion in the context of breast cancer diagnosis, prognosis, and treatment planning.

Results: AI-based frameworks have demonstrated superior performance compared with conventional single-modality approaches in subtype classification, treatment response prediction, and biomarker discovery. Deep learning architectures applied to whole-slide imaging, convolutional neural networks in radiomics, graph-based genomic models, and transformer-based NLP pipelines each contribute distinct and complementary dimensions to precision breast oncology. Multimodal fusion models integrating 2 or more data streams consistently outperform unimodal counterparts in prognostic accuracy.

Conclusions: Multimodal AI represents a convergent approach capable of operationalizing the biological complexity of breast cancer at the point of clinical care. Realizing this potential requires rigorous prospective validation, harmonized data standards, explainability frameworks, and equitable implementation strategies. Future models must be adaptive, continuously learning, and grounded in real-world clinical evidence.

© 2026 The Author(s). This open-access article is distributed under the Creative Commons Attribution-NonCommercial 4.0 International License, which permits sharing and adapting the material for any purpose except commercial purposes, provided appropriate credit is given, a link to the license is provided, and any changes are indicated.

INTRODUCTION

Breast cancer is the most commonly diagnosed cancer among women globally, accounting for an estimated 2.3 million new cases and approximately 666 000 deaths in 2022.[1] Its molecular and phenotypic heterogeneity, spanning luminal A and B, HER2-enriched, triple-negative, and claudin-low subtypes, necessitates individualized therapeutic strategies that extend far beyond anatomic staging.[2] Despite advances in endocrine therapy, targeted agents, immunotherapy, and poly(ADP-ribose) polymerase (PARP) inhibitors, selecting optimal regimens, anticipating resistance, and identifying patients at risk for late recurrence remain difficult.[3,4]

Artificial intelligence (AI), broadly encompassing machine learning (ML), deep learning (DL), and natural language processing (NLP), has emerged as a substantial force across the oncology continuum.[5] In breast cancer, AI has shown strong performance in mammographic screening, histopathological grading, gene expression profiling, treatment response prediction, and survivorship monitoring.[6,7] However, these applications have largely developed in disciplinary silos, imaging AI divorced from genomics, pathology models disconnected from clinical context, and genomic classifiers unaware of radiological phenotypes.[8] The persistence of these silos is partly

technical: jointly modeling high-dimensional genomic arrays with the far larger pixel space of whole-slide images yields a feature space that dwarfs the number of available patients, so naive integration risks fitting noise rather than biology. This is the curse of dimensionality, and it motivates the fusion strategies discussed below. Common mitigations include regularized or embedded feature selection such as LASSO, and metaheuristic optimizers such as particle swarm optimization, which reduce the effective number of input features before or during fusion, though they add tuning burden and can overfit if the selection step is not kept strictly within the training partition.

The convergence of high-dimensional biological data with scalable AI architectures now permits a more integrated approach, in which genomic, pathological, radiological, and EHR-derived data are jointly modeled to produce prognostic and therapeutic predictions.[9] This multimodal approach mirrors the biological reality of breast cancer, wherein molecular alterations manifest as pathological phenotypes, imaging signatures, and clinical trajectories that are mutually informative.[10]

This narrative review synthesizes current evidence across 4 AI-driven data modalities: genomics, digital pathology, radiomics, and clinical text mining, examines frameworks for their integration, and evaluates the translational pathway toward adaptive precision breast oncology. We also address the persistent challenges of model explainability, validation heterogeneity, data governance, and health equity that must be confronted before multimodal AI can be responsibly operationalized in clinical practice. We emphasize that the contribution of this review is integrative and critical rather than architectural: we do not propose a new fusion model, but instead organize a fragmented and changing literature, make explicit the methodological and validation limitations that temper its headline results, and define the conditions under which multimodal integration could progress from retrospective demonstration to genuine clinical utility. The recurring finding that multimodal models outperform unimodal ones is therefore treated not as a conclusion in itself but as a starting point for examining how, and how reliably, that advantage has been established.

METHODS

This study is a narrative review. We searched PubMed, Scopus, Web of Science, IEEE Xplore, and Google Scholar for peer-reviewed articles published between January 2015 and April 2026. Search terms combined breast cancer keywords ("breast cancer," "breast neoplasm," "breast carcinoma") with terms covering each of the data modalities under review: "artificial intelligence," "machine learning," "deep learning," "convolutional neural network," "transformer," "radiomics," "digital pathology," "whole-slide imaging," "genomics," "transcriptomics," "spatial transcriptomics," "natural language processing," "large language model," "multimodal," and "data fusion."

The primary search strategy executed in PubMed/MEDLINE was as follows:

("Breast Neoplasms"[Mesh] OR "breast cancer"[tiab] OR "breast carcinoma"[tiab]) AND ("Artificial Intelligence"[Mesh] OR "Machine Learning"[Mesh] OR "Deep Learning"[Mesh] OR "Neural Networks, Computer"[Mesh] OR "convolutional neural network"[tiab] OR "transformer"[tiab] OR "radiomics"[tiab] OR "natural language processing"[tiab] OR "multimodal"[tiab] OR "data fusion"[tiab]) AND ("2015/01/01"[Date - Publication] : "2026/04/30"[Date - Publication])

This strategy was adapted to the native syntax of Scopus, Web of Science, and IEEE Xplore; Google Scholar was searched with equivalent free-text terms.

We screened titles and abstracts for relevance to one or more of the 4 target modalities (genomics and pharmacogenomics, digital pathology, radiomics and breast imaging, and clinical text mining) and to multimodal integration, and hand-searched reference lists of key articles. From several hundred records identified across the 5 databases, approximately 70 sources met these criteria. Because this is a narrative rather than a systematic review, we emphasized studies reporting external or multi-institutional validation, larger cohorts, and transparent methodology, together with regulatory-cleared systems and guideline documents, rather than weighting all retrieved studies equally; proof-of-concept and single-center reports are cited where they illustrate emerging directions and are identified as such.

Inclusion criteria were peer-reviewed original research, systematic reviews, and authoritative regulatory or guideline documents, published in English between January 2015 and April 2026, addressing AI applied to one or more target modalities in breast cancer. We excluded non-English publications, preprints, editorials and commentaries, conference abstracts lacking full peer review, and studies without a substantive AI or multimodal component. A small number of foundational references predating 2015 were retained where essential for conceptual context.

Because this is a narrative review, we did not apply a formal risk-of-bias tool and did not pool results quantitatively. Selection was guided by topical relevance, methodological strength, clinical applicability, and recency. Both authors reviewed the final reference list to confirm scope coverage and balance across modalities. The review was structured to follow the SANRA quality criteria for narrative reviews. Specifically, in line with the SANRA items, we have justified the article's importance to the readership, stated specific aims, described the literature search and selection process, supported statements with current references, presented scientific reasoning rather than mere enumeration, and reported representative data with appropriate caveats. We nonetheless acknowledge the limitations inherent to the narrative format: the absence of a pre-registered protocol and formal quantitative synthesis means the selection of studies, although guided by the explicit criteria above, remains susceptible to author judgment

and selection bias. We have sought to mitigate this through the documented search strategy, inclusion and exclusion criteria, dual-author review, and the explicit critical-appraisal and risk-of-bias discussion provided in later sections.

RESULTS

AI-driven genomic profiling and pharmacogenomics

Gene expression classifiers and transcriptomic AI

Multigene expression assays, including Oncotype DX, MammaPrint, PAM50/Prosigna, and EndoPredict, have markedly refined prognostic stratification and chemotherapy decision-making in hormone receptor-positive, HER2-negative early-stage breast cancer.[11] These platforms, however, rely on predetermined gene panels and linear modeling assumptions that may not fully capture the nonlinear interdependencies inherent in transcriptomic data.

Recent work has applied graph neural networks (GNNs) and transformer-based architectures to bulk and single-cell RNA sequencing (scRNA-seq) data in breast cancer, enabling the identification of recurrence-associated gene modules with superior discriminative accuracy.[12] Graph-based models, which encode gene-gene interaction networks derived from STRING or KEGG databases as edges, capture co-regulatory relationships that are biologically interpretable and clinically actionable.[13] In a recent study, a graph neural network (GNN) model applied to whole-slide images from The Cancer Genome Atlas (TCGA) breast cancer cohort ranked patients based on their disease-specific survival times, demonstrating a significant improvement in survival prediction over existing methods.[14] Such single-cohort TCGA results warrant cautious interpretation, however: in the absence of external validation and strict separation of training and test patients, reported gains may partly reflect overfitting or information leakage rather than generalizable predictive value, a recurring concern in the machine-learning reproducibility literature that applies to many of the survival-prediction studies surveyed here. A related and frequently underreported risk is hyperparameter selection: when architectural and training hyperparameters are tuned on the same data used to report performance, rather than on a held-out validation set or through nested cross-validation, the resulting accuracy estimates are optimistically biased. Transparent reporting of the tuning protocol, alongside external validation, is therefore essential before such performance figures can be taken at face value.

Single-cell transcriptomic profiling, now integrable with spatial transcriptomics platforms such as 10x Visium and Slide-seq, adds a cellular compositional layer to tumor characterization. AI algorithms applied to spatially resolved gene expression data have mapped immunosuppressive tumor microenvironment (TME) architectures associated with immunotherapy resistance, providing a mechanistic rationale for combination regimens in triple-negative breast cancer (TNBC).[15]

Somatic mutation profiles and mutational signature analysis

Whole-exome sequencing (WES) and targeted next-generation sequencing (NGS) panels have catalogued recurrent somatic mutations in breast cancer, most prominently in PIK3CA, TP53, CDH1, GATA3, MAP3K1, and the homologous recombination (HR) repair pathway genes BRCA1 and BRCA2.[16]

Mutational signature analysis, originally formalized in the COSMIC Mutational Signatures v3.4 framework, decomposes somatic mutation patterns into biologically interpretable signatures reflecting etiologic processes such as APOBEC deamination (Signatures 2 and 13), HR deficiency (Signature 3), and aging-associated deamination (Signature 1).[17] Deep learning approaches based on autoencoder architectures, trained on large-scale somatic mutation catalogs, have demonstrated improved signature extraction performance compared with conventional non-negative matrix factorization (NMF).[18] A central limitation of these unsupervised approaches is their limited interpretability: whereas COSMIC signatures map to known mutational processes, the latent factors learned by autoencoders do not necessarily correspond to any defined mechanism, and post hoc analysis is required before they can be assigned biological meaning or used clinically. These findings have direct biomarker implications. Signature 3 positivity reflects homologous recombination deficiency and correlates with platinum and PARP inhibitor sensitivity in research settings.[19] It does not, however, drive clinical PARP inhibitor selection. In breast cancer, the approved indications for olaparib and talazoparib are restricted to patients with germline BRCA1/2 mutations identified through companion diagnostic testing.[4] The situation with immune checkpoint inhibitors is analogous. Pembrolizumab combined with chemotherapy is approved for unresectable locally recurrent or metastatic TNBC expressing PD-L1 at a combined positive score (CPS) ≥10, measured with an FDA-approved assay.[20] MSI-H/dMMR and TMB-high status are tumor-agnostic indications that may occasionally apply in breast cancer, but they do not replace PD-L1 testing in standard practice. For these reasons, AI-based estimation of TMB, MSI, or related genomic features remains investigational and should not be used as a stand-alone criterion for therapy selection unless analytically and clinically validated in the intended population.

Pharmacogenomics and AI-guided treatment personalization

Pharmacogenomics, the study of how germline genetic variation modulates drug response, represents a critical but underutilized dimension of breast cancer precision medicine.[21] CYP2D6 polymorphisms, which determine tamoxifen activation to its active metabolite endoxifen, have long been recognized as determinants of adjuvant endocrine therapy efficacy in HR-positive breast cancer. Yet pharmacogenomic testing remains inconsistently integrated into routine

clinical workflows, partly due to the complexity of interpreting multi-gene pharmacogenomic panels in the context of polypharmacy and co-medication interactions.[22] It should be noted that the routine clinical utility of CYP2D6-guided tamoxifen dosing remains contested, and major guideline bodies differ on whether genotyping should inform standard practice, owing to inconsistent outcome associations across studies.

AI-driven pharmacogenomic decision support systems (PGx-DSS) are emerging as scalable solutions to this gap.[23] These systems integrate patient germline variant data with drug-gene interaction knowledge bases (CPIC, PharmGKB, and DPWG) and real-world outcome registries to generate patient-specific dosing recommendations and alternative agent suggestions.[24] Precision endoxifen prediction frameworks have been proposed to integrate CYP2D6 with additional genetic variants (such as CYP3A4/5 and CYP2C19) and clinical covariates, addressing the well-documented imprecision of using CYP2D6 genotype alone as a surrogate for active metabolite exposure during tamoxifen therapy.[25]

Beyond endocrine therapy, AI-guided pharmacogenomics is being applied to anthracycline cardiotoxicity risk (RARG, TNNT2 variants), taxane-induced neuropathy (TUBB2A, CYP1B1 variants), and CDK4/6 inhibitor neutropenia risk stratification. Integration of these germline pharmacogenomic features into multimodal AI models, alongside tumor somatic profiles and clinical variables, offers the prospect of pre-treatment computational simulation of individualized pharmacokinetic-pharmacodynamic trajectories.[26]

More broadly, AI-driven pharmacology is emerging as a bridge between computational modeling and therapeutic decision-making, spanning drug-target identification, toxicity prediction, and dose optimization.[27] For breast cancer, this positions pharmacogenomic AI as a route to translate computational findings into actionable therapeutic interventions, but only where models are built on transparent, shared data and confirmed through prospective validation rather than retrospective benchmarks alone.

Artificial intelligence in digital pathology

Computational histopathology and whole-slide image analysis

Digital pathology, enabled by whole-slide imaging (WSI) scanners producing gigapixel-resolution tissue images, provides a rich morphological record of tumor architecture, cellular atypia, stromal interactions, and immune infiltration. The application of convolutional neural networks (CNNs) and, more recently, vision transformers (ViTs) to WSI has produced models capable of tasks ranging from automated grading to molecular feature prediction directly from hematoxylin and eosin (H&E) stains.[28]

The CLAM (Clustering-constrained Attention Multiple Instance Learning) framework, developed at Harvard University and Brigham and Women's Hospital, exemplifies weakly supervised DL approaches that train slide-level classifiers without requiring exhaustive pixel-level annotations.[29] Applied to breast cancer histology, weakly supervised deep learning models can recover hormone receptor status, HER2 status, and intrinsic molecular subtype (the standard luminal, HER2-enriched, and basal-like categories) from H&E morphology alone, indicating that gene-expression phenotypes are partly readable from tissue features.[30]

Tumor-infiltrating lymphocyte (TIL) quantification from H&E slides, which has emerged as an independent prognostic biomarker in TNBC and HER2-positive breast cancer, is now automatable at scale through validated DL-based segmentation pipelines. AI-assisted TIL scoring is being investigated to improve reproducibility, while International TIL Working Group recommendations provide the manual scoring framework. Prospective integration of automated TIL scores into immune checkpoint inhibitor eligibility algorithms is under active investigation.[31]

Spatial transcriptomics and AI-powered microenvironment mapping

The convergence of spatial transcriptomics with AI represents one of the most substantial frontiers in breast cancer biology. Spatial gene expression data, overlaid onto tissue morphology, enables the mapping of tumor-stromal-immune communication axes at cellular and subcellular resolution.[32] Computational deconvolution of spatial transcriptomics in breast cancer has identified immunosuppressive cellular niches in which extracellular matrix-producing myofibroblastic cancer-associated fibroblasts (CAFs) co-localize with regulatory T cells and immunosuppressive macrophage populations. These niches define microenvironment ecotypes that contribute to immune evasion and may influence response to immune checkpoint blockade.[33] These analyses are, at present, largely exploratory and descriptive: most are reported on small, single-institution cohorts without standardized outcome benchmarks (such as AUC-ROC or concordance-index against an external test set), so their clinical predictive value, as opposed to biological plausibility, remains to be established prospectively.

Multimodal architectures integrating histopathology with bulk transcriptomic profiles have been applied across cancer types, including breast cancer, to extract complementary morphological and molecular features for outcome prediction.[34] Such models can resolve prognostic signals that single-modality approaches miss, supporting more refined patient stratification.

Despite these advances, the performance of digital-pathology models is sensitive to factors that single-center studies can obscure. Scanners from different vendors produce images with differing color and resolution, and staining varies between laboratories, so models trained at one site can degrade at another unless stain normalization is applied. Manual annotations introduce inter-observer variability that becomes label noise, and this label noise can degrade model performance and contribute

to failed external validation. Adoption is further constrained by infrastructure: many centers outside large academic systems lack the slide scanners, storage, and computational resources required for routine whole-slide analysis. Digital-pathology AI is therefore best regarded as promising but contingent on stain normalization, external validation across scanners and sites, and broader access to digitization.

Radiomics and AI in breast imaging

Mammography, digital breast tomosynthesis, and MRI

Radiomics, the high-throughput extraction of quantitative imaging features from medical images, has transformed breast cancer screening, diagnosis, and treatment monitoring.[35] AI models applied to full-field digital mammography (FFDM) and digital breast tomosynthesis (DBT) have demonstrated sensitivity and specificity for malignancy detection comparable to or exceeding that of radiologists in reader studies, particularly for dense breast tissue where conventional visual interpretation is challenged.[36]

The FDA-cleared Transpara (Screenpoint Medical) and iCAD ProFound AI (iCAD, Inc.) systems exemplify commercially deployed DL-based mammographic CADe (detection) and CADx (diagnosis) tools. A large retrospective analysis from the Karolinska Institute demonstrated that AI-assisted reading of screening mammograms could reduce radiologist workload by more than half while maintaining non-inferior cancer detection rates, supporting AI-first reading protocols in high-volume screening programs.[37]

Dynamic contrast-enhanced magnetic resonance imaging (DCE-MRI) provides superior sensitivity for invasive lobular carcinoma detection, high-risk screening surveillance, and pre-surgical local staging. Radiomic features extracted from DCE-MRI time-intensity curves, including peak enhancement ratio, wash-out fraction, and intratumoral heterogeneity metrics, have been integrated into ML models that predict pCR after NACT, enabling adaptive treatment intensification or de-escalation.[38] Beyond pCR prediction, convolutional neural networks applied to breast MRI have been used to classify intrinsic molecular subtype directly from imaging features, supporting non-invasive characterization of tumor biology and complementing approaches based on hand-crafted radiomic features.[39] Emerging adjuncts are also drawing AI attention: contrast-enhanced mammography and abbreviated breast MRI offer higher sensitivity at lower cost and shorter acquisition time than full DCE-MRI, and early AI models applied to both are being investigated for lesion detection and characterization, although these applications remain at an investigational stage.

AI-guided interventional guidance and surgical planning

Beyond diagnostic imaging, AI is increasingly embedded in interventional and surgical workflows. AI-based decision support systems are emerging to assist radiologists during ultrasound-guided breast biopsies, with the potential to increase specificity and reduce unnecessary invasive procedures.[40] In the surgical domain, AI-powered 3D reconstruction of pre-operative MRI data facilitates oncoplastic resection planning by generating patient-specific virtual surgical models that optimize lumpectomy cavity geometry while minimizing healthy tissue sacrifice.[41]

Beyond preoperative planning, a critical challenge in partial mastectomy is achieving clear surgical margins. To address the low accuracy of detecting microscopic margin positivity, one computer vision model based on the InceptionV3 architecture predicts pathologic margin status directly from intraoperative specimen mammograms. When internally validated on a dataset of 821 images of resected breast tumors, this deep learning model predicted margin positivity with a sensitivity of 84%, a specificity of 42%, and an area under the receiver operating characteristic curve (AUROC) of 0.71.[42]

Deep learning has also been combined with digital breast tomosynthesis to map the cutting surface of resected specimens. In a preliminary assessment of specimens from 46 patients undergoing breast-conserving treatment, an image segmentation model was trained to identify the tumor area and its distance to the surgical margin. This deep learning approach effectively localized the tumor area, achieving a Mean Intersection over Union (MIoU) of 0.91, a global accuracy of 99%, and a Dice coefficient of 93% on the training dataset.[43] Together, these advances demonstrate that applying deep learning to intraoperative imaging can provide highly accurate, timely surgical margin assessments, potentially reducing the need for costly and burdensome re-excision surgeries.[42,43]

Natural language processing and clinical text mining in breast cancer

EHR data extraction and phenotyping

Electronic health records (EHRs) contain longitudinal, richly contextualized clinical narratives, pathology reports, radiology impressions, clinical notes, and treatment summaries, which encode critical breast cancer phenotypic information in unstructured free text. NLP, particularly transformer-based language models fine-tuned on biomedical corpora (BioBERT, PubMedBERT, GatorTron), has dramatically improved the automated extraction of clinically actionable entities from these records.[44]

Named entity recognition (NER) models trained on annotated oncology corpora can extract breast cancer receptor status, tumor grade, lymph node involvement, margin status, and systemic therapy regimens from pathology and operative reports with precision exceeding 0.92 and recall exceeding 0.89.[45] These NLP-derived structured variables can populate clinical data warehouses at scale, enabling real-world evidence generation without manual chart abstraction, a capability with direct implications for comparative effectiveness research and post-market surveillance of novel therapeutics.[46]

Large language models and clinical decision support

Large language models (LLMs) have extended the role of NLP in oncology beyond entity extraction into reasoning over clinical text.[47] These include general-purpose systems such as GPT-4, Claude, and Gemini, as well as domain-tuned models such as Med-PaLM 2. Early evidence suggests that these models can synthesize multi-step clinical narratives, including tumor board summaries and adverse-event attribution. However, performance varies substantially across tasks and models, and current evidence does not support their unsupervised use for guideline-grade clinical decisions.

Practical deployment in breast cancer care still requires expert review at the point of recommendation. Retrieval-augmented generation (RAG) reduces hallucination, a central limitation of LLMs, by grounding model output in vetted source material such as NCCN guidelines, ClinicalTrials.gov, and institutional outcome databases.[48] Applied to oncology, RAG approaches improve concordance with reference guidelines and reduce the rate of unsupported recommendations. More recent work has combined LLM-based interpretation of pathology and clinical text with downstream deep learning classifiers to recommend treatment regimens. In a 2026 study, Meta-Llama-3-8B-Instruct extracted TNM stage and tumor size from real-world pathology reports of more than 5000 breast cancer patients, and the extracted features were passed to ensemble and transformer models for multi-label treatment prediction.[49]

Multimodal AI fusion frameworks

Architectures for cross-modal integration

Combining genomic, pathological, radiological, and clinical text data in a single predictive model is technically demanding. Each modality has its own data type, scale, and pattern of missingness, and the architecture must absorb this without collapsing under the imbalance.[50] Three fusion strategies dominate this literature. Figure 1 summarizes the overall architecture of multimodal integration, from modality-specific encoders through fusion to downstream clinical applications.

Early fusion concatenates feature vectors from each modality before model training, allowing the network to learn cross-modal interactions from the outset. It is computationally simple but struggles when modalities differ greatly in dimensionality.

Intermediate fusion, typically implemented through cross-modal attention mechanisms in transformer models, weights each modality dynamically based on what the prediction task requires.[51] It suits settings where modalities carry complementary rather than redundant information, such as histology paired with bulk transcriptomics.

Late fusion trains each modality separately and combines the outputs through ensemble methods, weighted averaging, stacking, or Bayesian model combination. It is the most clinically practical of the 3 strategies: each modality-specific model can be validated, audited, and updated independently.[52]

Ensemble strategies are best judged across malignancies. Bagging, boosting, and stacking have been applied to other cancers, including cervical cancer, to combine complementary classifiers.[53] Such cross-domain work also exposes a recurring pitfall: reported performance frequently reflects training-set or internal metrics that degrade sharply on held-out data, so an ensemble can appear highly accurate while generalizing poorly. Adapting these architectures to breast cancer therefore depends less on the ensemble mechanism itself than on disciplined train-test separation, external validation, and transparent reporting to avoid overfitting to a single cohort.

These architectural differences carry directly onto the ward. The strategy that performs best in benchmark studies is not the strategy that is easiest to deploy: intermediate fusion reports the strongest discrimination but is the most computationally demanding, the least interpretable at the point of care, and the least tolerant of the incomplete data that characterizes routine oncology practice, where a patient may have imaging and pathology but no sequencing, or sequencing performed years apart from imaging. Late fusion sacrifices some cross-modal signal but degrades gracefully when a modality is absent, keeps per-modality predictions auditable, and maps onto a regulatory pathway in which components are cleared individually. For near-term clinical implementation, the modular architectures are therefore the more realistic candidates, with intermediate fusion reserved for settings where complete paired data and computational infrastructure are both assured. Table 1 compares the principal fusion strategies across these dimensions.

Representative multimodal models in breast cancer

PORPOISE (Pathology-Omic Research Platform for the Identification of Survival-associated Embeddings), developed by Chen and colleagues, co-trains histopathology image encoders and bulk RNA-seq expression profiles in a multi-task survival prediction framework[34], building on earlier co-attention transformer architectures for cross-modal histology-genomics interaction.[54] When applied to 14 cancer types including breast cancer, PORPOISE demonstrated concordance indices (a measure of how well a model ranks patients by outcome) superior to both unimodal pathology and genomic models, with heatmap-based attention visualization identifying morphological and transcriptomic features co-contributing to survival risk stratification.

Pan-cancer multimodal models trained on TCGA have further demonstrated that WSI–genomic fusion captures biologically complementary variance: histomorphological features explain stromal and immune variation not captured by bulk transcriptomics, while genomic features resolve molecular subtypes not visible at the tissue scale.[55] These findings support the biological rationale for multimodal integration and suggest that neither imaging nor genomics alone is sufficient for optimal breast cancer risk stratification.

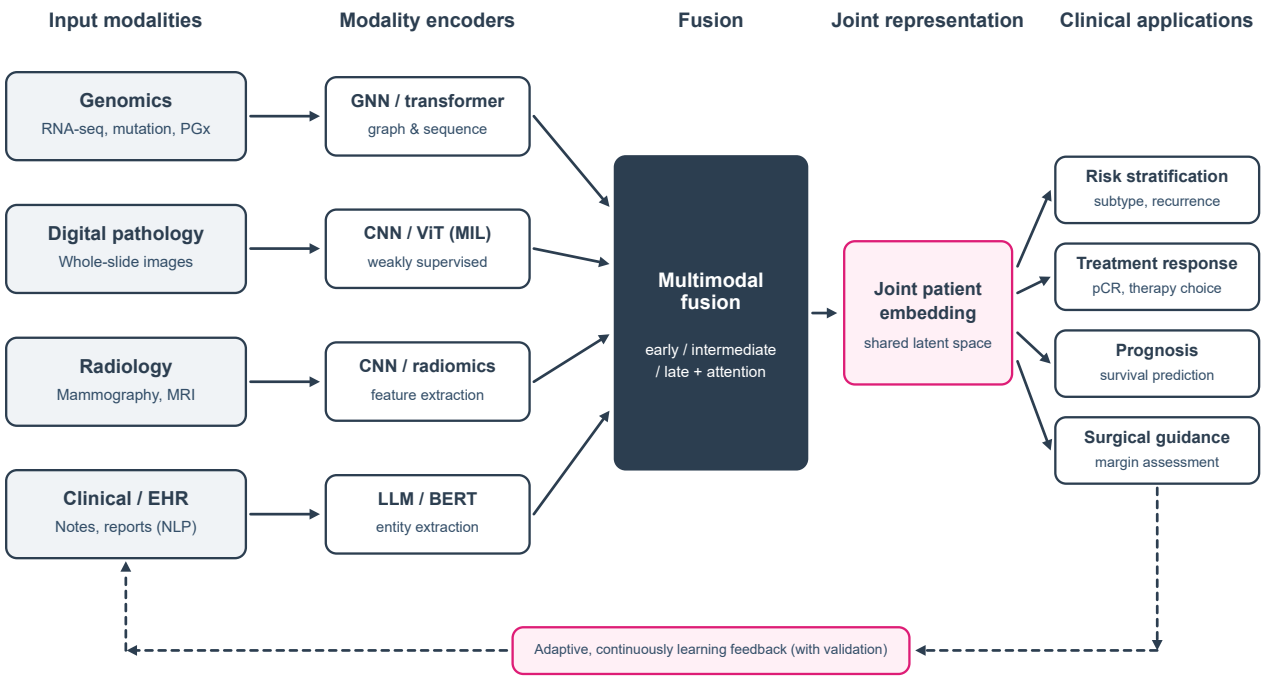


Figure 1. Conceptual Schematic of Multimodal AI Integration in Breast Cancer
Genomic, digital-pathology, radiological, and clinical/EHR data are processed by modality-specific encoders, combined through early, intermediate, or late fusion into a joint patient representation, and mapped to clinical applications. The dashed loop denotes adaptive, continuously learning models updated under validation.

Federated learning (FL) frameworks are increasingly proposed as privacy-preserving mechanisms for training multimodal breast cancer models across institutional boundaries without raw data sharing.[56] FL enables model parameters to be aggregated across hospital nodes using protocols such as FedAvg and FedProx, preserving patient data locality while enabling large-scale model training. Early FL pilots in breast cancer pathology report performance approaching that of centrally trained models.[57] Table 2 summarizes representative multimodal models and systems in breast cancer, together with their input modalities, fusion strategy, architecture, and validation status.

DISCUSSION

Regulatory requirements and validation standards

Translating AI models from research to regulated clinical use requires compliance with an increasingly structured regulatory framework. In the United States, the FDA has authorized a cumulative total of 1451 AI-enabled medical devices as of the end of 2025, with radiology and imaging applications representing roughly 76% of these authorizations.[58] For software as a medical device (SaMD), the De Novo and 510(k) pathways require analytical and clinical validity, with particular scrutiny of generalizability across demographic subgroups and equipment vendors.

The Total Product Lifecycle framework for AI-enabled SaMD acknowledges that models may require post-market updates. Predetermined change control plans (PCCPs) let sponsors specify in advance the nature, scope, and validation requirements for updates, reducing regulatory burden for continuous learning systems.[59] For multimodal breast cancer AI, PCCP design must address the modality-specific update cadence, e.g., genomic classifier recalibration following NGS panel revisions, without compromising cross-modal integration integrity.

In the European Union, AI-enabled devices fall under the Medical Device Regulation and the EU AI Act, both requiring evidence of safety, performance, and post-market monitoring. Across jurisdictions, the central challenge is the same: the distance between retrospective in silico performance, often derived from curated cohorts such as TCGA, and prospective, multi-site clinical utility. Closing that gap requires prospective validation with prespecified endpoints and evidence of clinical benefit rather than discrimination metrics alone.

Predictive-model studies vary widely in quality. We did not undertake a formal quantitative appraisal, but interpreted the cited models with reference to PROBAST and the reporting expectations of TRIPOD and TRIPOD-AI. Recurrent concerns include small or single-institution cohorts, performance reported from internal validation only, possible data leakage between training and test partitions, inconsistent handling of missing modalities, and limited reporting of calibration. These considerations temper the headline performance figures reported in many primary studies.

Table 1. Comparison of multimodal fusion strategies in breast cancer AI with respect to handling of missing modalities, interpretability, computational requirements, scalability, and suitability for clinical implementation.

Fusion strategy	Handling of missing modalities	Interpretability	Computational requirements	Scalability	Suitability for clinical implementation
Early (feature-level)	Poor. Concatenation requires all modalities per patient; absent inputs demand imputation or case exclusion, and imputation of high-dimensional omics is unreliable.	Low. Cross-modal interactions are learned in a single joint space, so attribution to a specific modality is difficult.	Moderate at training, low at inference. Dimensionality imbalance between omics and imaging features often necessitates aggressive reduction.	Limited. Adding a modality requires retraining the whole model.	Best suited to research cohorts with complete paired data rather than routine care, where missingness is the norm.
Intermediate (attention-based)	Moderate. Attention weights can downweight absent or low-quality modalities, but most implementations still assume a fixed input set.	Moderate. Attention weights offer modality-level and sometimes region-level attribution, though they are not causal explanations.	High. Transformer-based cross-modal attention is the most demanding to train and requires GPU infrastructure many centers lack.	Moderate. New modalities can be added as additional attention streams without full redesign.	Strongest predictive performance reported, but infrastructure and interpretability demands currently limit deployment outside academic centers.
Late (decision-level)	Good. Each modality is modeled independently, so a missing input removes one contributing model rather than invalidating the case.	High. Per-modality predictions remain inspectable and auditable before combination.	Low to moderate. Unimodal models can be trained and updated separately, and inference is parallelizable.	High. Modalities, and individual models, can be added, replaced, or retired independently.	Most compatible with existing clinical workflows and with regulatory pathways that clear single-modality devices individually.
Federated (distributed training)	Variable. Site-level modality availability differs, and non-independent, non-identically distributed data across sites compounds missingness.	Depends on the underlying fusion architecture; federation adds no interpretability of its own.	High, with substantial communication overhead and coordination cost across participating sites.	High in principle for data volume and site count; governance rather than computation is the binding constraint.	Addresses privacy and data-sharing barriers directly, but harmonization, governance, and validation across sites remain unresolved.

Deployability also depends on inference efficiency, a constraint often omitted from accuracy-focused reports. This is illustrated outside breast cancer by knowledge distillation, in which a vision-transformer-CNN teacher is compressed into a lightweight convolutional student; one cardiac-MRI study reported accuracy comparable to that of its teacher with a roughly 3-fold gain in inference efficiency, while noting single-center evaluation and no external validation.[60] Such efficiency-fidelity trade-offs are directly relevant to multimodal breast-cancer models, where gigapixel whole-slide inputs and high-dimensional genomic features make inference cost a genuine barrier to point-of-care use, although cross-domain transfer of these techniques must itself be confirmed in breast-specific cohorts. Scalability is therefore an evaluation axis distinct from accuracy: a model that performs well on a single curated cohort may still fail to scale to high-volume, multi-site screening workflows if its computational, storage, and latency demands exceed what community-based and resource-limited centers can sustain. Reporting that omits these costs gives an incomplete picture of clinical readiness.

Explainability and clinician trust

The 'black box' perception of DL models remains a significant barrier to clinical adoption, particularly in high-stakes oncological decision-making. Post-hoc explainability methods, including Grad-CAM for image models, SHAP (SHapley Additive exPlanations) for tabular and genomic models, and attention weight visualization for transformer architectures, provide approximations of feature importance that support model auditing and clinician interpretation.[61] LIME (Local Interpretable Model-agnostic Explanations) complements these methods by approximating a complex model's behavior around a single prediction with a simple, locally faithful surrogate, yielding case-level explanations that clinicians can inspect for an individual patient. Its explanations, like those of SHAP, are nonetheless approximations: they can be unstable across runs and are sensitive to how input features are perturbed and sampled, so in the studies surveyed here, such attributions are best treated as hypotheses about model behavior rather than definitive accounts of why a prediction was made. In concrete terms, a black-box model might flag a mammogram as high-risk or predict that a tumor will not respond to a given regimen while offering the

Table 2. Representative AI models and systems for breast cancer, summarizing input data modalities, fusion strategy, AI architecture, validation status with key reported performance, and potential clinical application. Superscripts refer to the manuscript reference list; performance values are given only where reported in the cited source.

Model / system	Data modalities	Fusion strategy	AI architecture	Validation status & key result	Clinical application
PORPOISE [34]	WSI + bulk RNA-seq	Intermediate (co-attention)	Image + genomic encoders; multi-task survival	Retrospective; 14 TCGA cohorts; higher C-index than unimodal models	Prognostic survival-risk stratification
Co-attention transformer fusion [54]	WSI + genomic profiles	Intermediate (co-attention)	Transformer co-attention	Retrospective benchmark cohorts	Cross-modal prognostic modeling
CLAM (weakly supervised MIL) [29]	WSI (H&E)	Single-modality	Attention-based MIL (CNN)	Weakly supervised, multi-institutional; recovers HR/HER2/subtype	Receptor / HER2 / subtype prediction; grading
GNN survival model [14]	WSI + gene-interaction graph	Single-modality (graph)	Graph neural network	Retrospective TCGA; improved survival ranking (single cohort)	Disease-specific survival ranking
Transpara ; iCAD ProFound AI [37]	Mammography (FFDM / DBT)	Single-modality	Deep CNN (commercial)	FDA-cleared; commercially deployed	Screening detection (CADe) and diagnosis (CADx)
Karolinska AI reading study [37]	Screening mammography	Single-modality	Deep CNN (AI reading)	Large retrospective; workload reduced >50%, non-inferior detection	Workload reduction / AI-first triage
InceptionV3 margin model [42]	Specimen mammography	Single-modality	InceptionV3 CNN	Internal only (821 images); sensitivity 84%, specificity 42%, AUROC 0.71	Intraoperative surgical-margin assessment
LLM + ensemble pipeline (Meta-Llama-3) [49]	Pathology text + structured features	Late / ensemble	LLM (Meta-Llama-3) + ensembl e/transformer	Retrospective; >5000 patients	Treatment-regimen recommendation
Federated learning pilots [57]	Multimodal pathology, multi-site	Privacy-preserving training	FedAvg / FedProx over CNNs	Pilot; approaching centralized performance	Privacy-preserving multi-site training

clinician no inspectable basis for that judgment; the difficulty is practical, because a recommendation that cannot be interrogated is hard to weigh and hard to override when it is wrong for an individual patient.

Multimodal explainability presents additional complexity, as attributing a prognostic prediction jointly to pathological, genomic, and clinical features requires cross-modal attribution methods that remain an active research area. Cross-modal attention maps, which highlight the WSI regions that most influenced the incorporation of a specific genomic feature, represent a promising approach.[62] Prospective evaluation of explainability interfaces through clinician usability studies is essential before widespread deployment; evidence suggests that poorly designed explanation interfaces can paradoxically reduce clinician trust and decision quality compared with no explanation at all.

Health equity and algorithmic bias

AI models trained predominantly on data from European-ancestry patients enrolled in academic medical center trials may systematically underperform in Black, Hispanic, Asian, and Indigenous patient populations, groups that bear disproportionate breast cancer mortality burdens and are underrepresented in benchmark genomic and imaging datasets.[63] Subgroup performance auditing, stratified by self-reported race/ethnicity, age, breast density category, and socioeconomic status, should be a mandatory component of any breast cancer AI validation study.

Algorithmic mitigation strategies, including adversarial debiasing during model training, fairness-constrained optimization objectives, and diverse federated node representation, can reduce demographic performance disparities but do not eliminate them.[64] Structural solutions, including equitable data collection infrastructure and community-engaged research partnerships at Federally Qualified Health Centers and safety-net hospitals, are necessary complements to technical debiasing approaches. Concrete initiatives illustrate how this can be operationalized: the NIH All of Us Research Program deliberately oversamples populations historically underrepresented in biomedical datasets; the Human Heredity and

Health in Africa (H3Africa) initiative builds genomic research capacity and African reference data to counter the European-ancestry skew of resources such as TCGA and gnomAD; and regional breast-cancer registries and biobanks in low- and middle-income countries are beginning to expand the diversity of imaging and genomic training data. Linking multimodal breast-cancer AI development to such resources is a practical prerequisite for equitable performance, rather than an afterthought.

Limitations of the current evidence base

Drawing these threads together, several limitations qualify the evidence reviewed here. First, methodological rigor varies widely: many of the strongest reported results derive from retrospective, single-institution cohorts with internal validation only, and headline metrics such as the concordance indices and AUROC values cited above should be read against the risk-of-bias considerations outlined in the Regulatory Requirements and Validation Standards subsection. Second, dataset heterogeneity is pervasive: scanner and staining variability in digital pathology, acquisition differences across imaging platforms, and ancestry skew in genomic reference resources all constrain how far a model trained at one site transfers to another, as noted in the digital-pathology and health-equity discussions. Third, reproducibility remains limited by incomplete reporting of preprocessing, hyperparameters, and training and test partitions, and by the scarcity of shared code and external test sets, which together make independent replication difficult. Fourth, barriers to clinical implementation extend beyond accuracy to inference cost and scalability, regulatory clearance, interpretability at the point of care, and the scanning, storage, and computing infrastructure required for routine whole-slide and multimodal analysis. Taken together, these constraints indicate that the field's most promising results are best regarded as proof of capability rather than evidence of clinical readiness, and they define the prospective, externally validated, and equity-aware studies needed to close the gap.

A balanced appraisal also requires attention to the studies in which external validation has failed rather than succeeded. Howard and colleagues demonstrated that whole-slide images in The Cancer Genome Atlas carry site-specific digital signatures that deep learning models detect readily, that these signatures survive standard color normalization and augmentation, and that these signatures produce biased accuracy for prediction of survival, genomic mutations, and tumor stage across more than 3000 patients and 6 cancer subtypes; ethnicity could itself be inferred from the same signatures.[72] Because several of the pan-cancer whole-slide-image and genomic fusion models discussed above were trained on this resource, a proportion of their reported performance may reflect institutional provenance rather than tumor biology, and preserved-site cross-validation is now the appropriate minimum standard. Hsu and colleagues externally validated the Digital Mammography DREAM Challenge ensemble in a demographically diverse Los Angeles screening population of 37 317 examinations. The ensemble had previously been validated in Swedish and Washington State cohorts. In this new external validation, discrimination fell to an AUROC of 0.85 from 0.90 in the development cohort and 0.92 in prior external cohorts; sensitivity in women with a personal history of breast cancer dropped to 0.596, compared with 0.850 for the interpreting radiologist.[73] Finally, an independent head-to-head evaluation of 3 commercial mammography algorithms found AUROCs ranging from 0.920 to 0.956, with only the best matching the average radiologist.[36] Taken together, these findings indicate that performance reported in a development cohort—and even in a single external cohort—is only a weak predictor of performance in a new population, and that generalizability must be demonstrated in the intended target population rather than assumed from it.

Future directions

The limitations set out above do not merely qualify the findings; they define the research agenda. Each priority below is stated as a response to a specific constraint identified in the previous section, and each is framed in terms of what would have to be demonstrated prospectively, implemented in routine practice, and evaluated for clinical impact before the corresponding technology could be considered ready for care delivery:

Large-scale self-supervised vision-language models pretrained on millions of paired pathology image-report datasets (e.g., CONCH, PLIP, UNI) are establishing universal feature representations that can be efficiently fine-tuned for downstream breast cancer tasks with minimal labeled data. These foundation models offer the prospect of democratizing high-performance AI diagnostics at institutions lacking the computational infrastructure for de novo model training.[65] This priority responds directly to the implementation barriers identified above: the test of success is not benchmark accuracy but whether a fine-tuned foundation model, deployed on commodity hardware at a non-academic center, reproduces its reported performance prospectively on that center's own consecutive cases.

The development of breast cancer AI that performs equitably across diverse populations will require concerted investment in genomic and imaging infrastructure for currently underrepresented regions. Breast cancer in the Middle East and North Africa (MENA), for example, often presents at a younger age and with distinct hereditary susceptibility patterns compared with Western cohorts, yet MENA populations remain markedly underrepresented in TCGA, gnomAD, and other foundational reference datasets.[66] Region-specific genomic sequencing initiatives, federated learning collaborations spanning low- and middle-income countries, and locally validated polygenic risk score recalibration are emerging priorities that will determine whether next-generation breast cancer AI narrows or widens global outcome disparities.[67] This

responds to the dataset heterogeneity and ancestry skew described above, and to the documented failure of screening models to generalize to demographically diverse cohorts. The requirement is prospective validation stratified by ancestry, breast density, and prior treatment history, reported as subgroup performance rather than pooled discrimination.

Cell-free DNA (cfDNA) and circulating tumor DNA (ctDNA) analysis, combined with ML-based copy number alteration (CNA) detection and methylation-based tissue-of-origin classifiers, enables minimally invasive tumor monitoring that is complementary to imaging and tissue-based AI.[68] Serial ctDNA analysis may complement imaging by providing minimally invasive information on tumor burden, molecular evolution, and treatment response in breast cancer. ctDNA detection has been shown in some studies to precede clinical or radiographic relapse, but routine ctDNA-guided treatment adaptation remains investigational because evidence of clinical utility is still evolving.[69] Because this evidence base is almost entirely observational, the priority here is interventional: trials in which ctDNA-guided treatment adaptation is randomized against standard surveillance, with disease-free survival and quality of life as endpoints, rather than further studies of detection lead time.

Static AI models trained on historical cohorts are ill-suited to the dynamically evolving therapeutic biology of breast cancer, with new agents, combination strategies, and molecular targets emerging on an accelerating timeline. Continuously learning AI architectures, incorporating online learning, transfer learning from registry data updates, and Bayesian model updating, represent the next-generation infrastructure for precision oncology clinical decision support.[70] This addresses both the temporal obsolescence of statically trained models and the regulatory constraint discussed above. Continuous learning is only deployable within a predetermined change control plan that specifies performance floors, drift monitoring, and automatic rollback, and its evaluation must include whether updating improves outcomes rather than only metrics.

The integration of AI into breast cancer care must extend beyond biological and therapeutic modeling to encompass patient-reported outcomes (PROs), quality of life (QoL) trajectory prediction, and psychosocial risk stratification. NLP analysis of patient narrative descriptions of symptom burden and treatment tolerability, combined with wearable-derived physiological data streams, can generate longitudinal models of survivorship risk that trigger timely supportive care interventions.[71] This is the strand where clinical impact, as distinct from predictive accuracy, is actually measured. The relevant endpoints are unplanned admissions, symptom burden, treatment completion, and patient-reported quality of life, assessed in implementation studies embedded in routine care rather than in retrospective cohorts.

Cutting across every priority above is the gap between retrospective performance and demonstrated clinical benefit, which is the single limitation most consistently identified in this review. Closing this gap requires a shift in study design rather than in model architecture: prospective, multi-site evaluation in the intended target population with preserved-site partitioning; prespecified subgroup analyses; reporting against PROBAST and TRIPOD; and, for systems intended to alter management, comparative studies with patient outcomes as the primary endpoint. Implementation research on workflow integration, clinician override behavior, and downstream resource use is equally necessary, since a model that is accurate but ignored, or accurate but acted upon inconsistently, delivers no benefit.

CONCLUSIONS

Breast cancer's biological complexity demands clinical intelligence tools that are equally multidimensional. AI-driven multimodal integration of genomic, pathological, radiological, and clinical data offers a scientifically compelling and clinically pragmatic framework for advancing precision oncology beyond the limitations of single-modality approaches. The convergence of genomics and pharmacogenomics, computational histopathology, radiomic phenotyping, and clinical NLP within unified predictive models represents a shift from retrospective classification to prospective, adaptive, patient-specific decision intelligence.

Realizing this potential requires a commensurate commitment to rigorous prospective validation across demographically diverse populations, standardized interoperability frameworks for multimodal data exchange, transparent and clinician-interpretable model architectures, and regulatory pathways that accommodate the adaptive nature of continuously learning clinical AI. Equally essential is the embedding of health equity as a foundational design principle, not a post-hoc audit, in every phase of multimodal AI development and deployment.

The next decade will likely determine whether AI in breast cancer fulfills its promise as a population-level precision medicine equalizer or remains a high-performance tool accessible only to patients treated at well-resourced academic centers. The answer will depend not only on algorithmic advances but on the deliberate choices made by oncologists, data scientists, regulators, health systems, and patient advocates in designing, validating, and deploying these systems with equity, transparency, and clinical rigor as non-negotiable standards. Finally, these tools augment rather than replace the clinician: sound breast-cancer care continues to depend on empathy, clear communication, and shared decision-making, and the equitable benefit of multimodal AI will rest as much on diverse training data and rigorous external validation as on algorithmic performance.

ETHICAL CONSIDERATIONS

Not applicable. This was a narrative review of previously published literature and did not involve

human participants, animal subjects, or identifiable patient data.

DATA AVAILABILITY

Data sharing is not applicable to this article as no new data were created or analyzed in this narrative review. All information cited is available in the referenced publications.

CONFLICT OF INTERESTS

The authors declare no conflicts of interest relevant to the content of this manuscript.

FUNDING

Not applicable.

ACKNOWLEDGMENTS

The authors thank the editors and reviewers of Archives of Breast Cancer for the opportunity to contribute to the Special Thematic Collection on Transforming Breast Cancer Care: AI and Advanced Technologies. We also acknowledge the broader research community whose work in artificial intelligence, computational pathology, radiomics, and clinical informatics forms the foundation of this review.

AI DISCLOSURE

During the preparation of this work, the authors used large language models (Claude and ChatGPT) for copyediting, language refinement, structuring, and reference formatting. AI tools were not used to generate original scientific content, analyze data, or formulate conclusions. All scientific reasoning and clinical framing were developed by the human authors, who reviewed all AI-assisted output and take full responsibility for the integrity of the publication.

AUTHOR CONTRIBUTIONS

JT: Methodology, Investigation, Data curation, Writing – original draft, Writing – review & editing. AM: Conceptualization, Methodology, Investigation, Visualization, Supervision, Project administration, Writing – review & editing.

REFERENCES

1. Bray F, Laversanne M, Sung H, Ferlay J, Siegel RL, Soerjomataram I, et al. Global cancer statistics 2022: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA A Cancer J Clinicians*. 2024;74(3):229-63. doi: 10.3322/caac.21834
2. Perou CM, Sørlie T, Eisen MB, van de Rijn M, Jeffrey SS, Rees CA, et al. Molecular portraits of human breast tumours. *Nature*. 2000;406(6797):747-52. doi: 10.1038/35021093
3. Cardoso F, Kyriakides S, Ohno S, Penault-Llorca F, Poortmans P, Rubio IT, et al. Early breast cancer: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up. *Annals of Oncology*. 2019;30(8):1194-220. doi: 10.1093/annonc/mdz173
4. Litton JK, Rugo HS, Ettl J, Hurvitz SA, Gonçalves A, Lee KH, et al. Talazoparib in Patients with Advanced Breast Cancer and a Germline BRCA Mutation. *N Engl J Med*. 2018;379(8):753-63. doi: 10.1056/nejmoa1802905
5. Topol EJ. High-performance medicine: the convergence of human and artificial intelligence. *Nat Med*. 2019;25(1):44-56. doi: 10.1038/s41591-018-0300-7
6. McKinney SM, Sieniek M, Godbole V, Godwin J, Antropova N, Ashrafian H, et al. International evaluation of an AI system for breast cancer screening. *Nature*. 2020;577(7788):89-94. doi: 10.1038/s41586-019-1799-6
7. Abu Abeelh E, Abuabeileh Z. Screening Mammography and Artificial Intelligence: A Comprehensive Systematic Review. *Cureus*. 2025. doi: 10.7759/cureus.79353
8. Boehm KM, Khosravi P, Vanguri R, Gao J, Shah SP. Harnessing multimodal data integration to advance precision oncology. *Nat Rev Cancer*. 2022;22(2):114-26. doi: 10.1038/s41568-021-00408-3
9. Acosta JN, Falcone GJ, Rajpurkar P, Topol EJ. Multimodal biomedical AI. *Nat Med*. 2022;28(9):1773-84. doi: 10.1038/s41591-022-01981-2
10. Cheerla A, Gevaert O. Deep learning with multimodal representation for pancancer prognosis prediction. *Bioinformatics*. 2019;35(14):i446-i454. doi: 10.1093/bioinformatics/btz342
11. Sparano JA, Gray RJ, Makower DF, Pritchard KI, Albain KS, Hayes DF, et al. Adjuvant Chemotherapy Guided by a 21-Gene Expression Assay in Breast Cancer. *N Engl J Med*. 2018;379(2):111-21. doi: 10.1056/nejmoa1804710
12. Rhee S, Seo S, Kim S. Hybrid Approach of Relation Network and Localized Graph Convolutional Filtering for Breast Cancer Subtype Classification. *Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence*. 2018;3527-34. doi: 10.24963/ijcai.2018/490
13. Szklarczyk D, Gable AL, Lyon D, Junge A, Wyder S, Huerta-Cepas J, et al. STRING v11: protein–protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic Acids Res*. 2019;47(D1):D607-D613. doi: 10.1093/nar/gky1131
14. Mackenzie CC, Dawood M, Graham S, Eastwood M, ul Amir Afsar Minhas F. Neural graph modelling of whole slide images for survival ranking. *Proceedings of the First Learning on Graphs Conference* [Internet]. PMLR; 2022 Oct 8;198:48:1-48:10. [cited 8 10 2026]. Available from: https://proceedings.mlr.press/v198/mackenzie22a.html
15. Wu SZ, Al-Eryani G, Roden DL, Junankar S, Harvey K, Andersson A, et al. A single-cell and spatially resolved atlas of human breast cancers. *Nat Genet*. 2021;53(9):1334-47. doi: 10.1038/s41588-021-00911-1
16. Ciriello G, Gatza ML, Beck AH, Wilkerson MD, Rhie SK, Pastore A, et al. Comprehensive Molecular Portraits of Invasive Lobular Breast Cancer. *Cell*. 2015;163(2):506-19. doi: 10.1016/j.cell.2015.09.033
17. Alexandrov LB, Kim J, Haradhvala NJ, Huang MN, Tian Ng AW, Wu Y, et al. The repertoire of mutational signatures in human cancer. *Nature*. 2020;578(7793):94-101. doi: 10.1038/s41586-020-1943-3
18. Pancotti C, Rollo C, Codicè F, Birolo G, Fariselli P, Sanavia T. MUSE-XAE: MUtational Signature Extraction with eXplainable AutoEncoder enhances tumour types classification. *Bioinformatics*. 2024;40(5):btae320. doi: 10.1093/bioinformatics/btae320
19. Telli ML, Timms KM, Reid J, Hennessy B, Mills GB, Jensen KC, et al. Homologous Recombination Deficiency (HRD) Score Predicts Response to Platinum-Containing Neoadjuvant Chemotherapy in Patients with Triple-Negative Breast Cancer. *Clinical Cancer Research*. 2016;22(15):3764-73. doi: 10.1158/1078-0432.ccr-15-2477

20. Cortes J, Cescon DW, Rugo HS, Nowecki Z, Im SA, Yusof MM, et al. Pembrolizumab plus chemotherapy versus placebo plus chemotherapy for previously untreated locally recurrent inoperable or metastatic triple-negative breast cancer (KEYNOTE-355): a randomised, placebo-controlled, double-blind, phase 3 clinical trial. *The Lancet*. 2020;396(10265):1817-28. doi: 10.1016/s0140-6736(20)32531-9
21. Relling MV, Evans WE. Pharmacogenomics in the clinic. *Nature*. 2015;526(7573):343-50. doi: 10.1038/nature15817
22. Bray J, Sludden J, Griffin MJ, Cole M, Verrill M, Jamieson D, et al. Influence of pharmacogenetics on response and toxicity in breast cancer patients treated with doxorubicin and cyclophosphamide. *Br J Cancer*. 2010;102(6):1003-9. doi: 10.1038/sj.bjc.6605587
23. Relling MV, Klein TE. CPIC: Clinical Pharmacogenetics Implementation Consortium of the Pharmacogenomics Research Network. *Clin Pharmacol Ther*. 2011;89(3):464-7. doi: 10.1038/clpt.2010.279
24. Whirl-Carrillo M, Huddart R, Gong L, Sangkuhl K, Thorn CF, Whaley R, et al. An Evidence-Based Framework for Evaluating Pharmacogenomics Knowledge for Personalized Medicine. *Clin Pharmacol Ther*. 2021;110(3):563-72. doi: 10.1002/cpt.2350
25. Helland T, Alsomairy S, Lin C, Søiland H, Mellgren G, Hertz DL. Generating a Precision Endoxifen Prediction Algorithm to Advance Personalized Tamoxifen Treatment in Patients with Breast Cancer. *JPM*. 2021;11(3):201. doi: 10.3390/jpm11030201
26. Hertz DL, Freedman RA. Pharmacogenomics in breast cancer. *Pharmacogenomics*. 2018;19(2):77-80. doi: 10.2217/pgs-2017-0167
27. Abdollahi J, Mehrpour O. Opportunities and challenges: artificial intelligence in pharmacology. *2024 11th International Symposium on Telecommunications (IST)*. Tehran, Iran: IEEE; 2024;28-37. doi: 10.1109/IST64061.2024.10843462
28. Kather JN, Heij LR, Grabsch HI, Loeffler C, Echle A, Muti HS, et al. Pan-cancer image-based detection of clinically actionable genetic alterations. *Nat Cancer*. 2020;1(8):789-99. doi: 10.1038/s43018-020-0087-6
29. Lu MY, Williamson DFK, Chen TY, Chen RJ, Barbieri M, Mahmood F. Data-efficient and weakly supervised computational pathology on whole-slide images. *Nat Biomed Eng*. 2021;5(6):555-70. doi: 10.1038/s41551-020-00682-w
30. Schmauch B, Romagnoni A, Pronier E, Saillard C, Maillé P, Calderaro J, et al. A deep learning model to predict RNA-Seq expression of tumours from whole slide images. *Nat Commun*. 2020;11(1):3877. doi: 10.1038/s41467-020-17678-4
31. Salgado R, Denkert C, Demaria S, Sirtaine N, Klauschen F, Pruneri G, et al. The evaluation of tumor-infiltrating lymphocytes (TILs) in breast cancer: recommendations by an International TILs Working Group 2014. *Annals of Oncology*. 2015;26(2):259-71. doi: 10.1093/annonc/mdu450
32. Marx V. Method of the Year: spatially resolved transcriptomics. *Nat Methods*. 2021;18(1):9-14. doi: 10.1038/s41592-020-01033-y
33. Croizer H, Mhaidly R, Kieffer Y, Gentric G, Djerroudi L, Leclere R, et al. Deciphering the spatial landscape and plasticity of immunosuppressive fibroblasts in breast cancer. *Nat Commun*. 2024;15(1):2806. doi: 10.1038/s41467-024-47068-z
34. Chen RJ, Lu MY, Williamson DFK, Chen TY, Lipkova J, Noor Z, et al. Pan-cancer integrative histology-genomic analysis via multimodal deep learning. *Cancer Cell*. 2022;40(8):865-878.e6. doi: 10.1016/j.ccell.2022.07.004
35. Lambin P, Leijenaar RTH, Deist TM, Peerlings J, de Jong EEC, van Timmeren J, et al. Radiomics: the bridge between medical imaging and personalized medicine. *Nat Rev Clin Oncol*. 2017;14(12):749-62. doi: 10.1038/nrclinonc.2017.141
36. Salim M, Wåhlin E, Dembrower K, Azavedo E, Foukakis T, Liu Y, et al. External Evaluation of 3 Commercial Artificial Intelligence Algorithms for Independent Assessment of Screening Mammograms. *JAMA Oncol*. 2020;6(10):1581-8. doi: 10.1001/jamaoncol.2020.3321
37. Dembrower K, Wåhlin E, Liu Y, Salim M, Smith K, Lindholm P, et al. Effect of artificial intelligence-based triaging of breast cancer screening mammograms on cancer detection and radiologist workload: a retrospective simulation study. *The Lancet Digital Health*. 2020;2(9):e468-e474. doi: 10.1016/s2589-7500(20)30185-0
38. Li W, Newitt DC, Gibbs J, Wilmes LJ, Jones EF, Arasu VA, et al. Predicting breast cancer response to neoadjuvant treatment using multi-feature MRI: results from the I-SPY 2 TRIAL. *npj Breast Cancer*. 2020;6(1):63. doi: 10.1038/s41523-020-00203-7
39. Ha R, Mutasa S, Karcich J, Gupta N, Pascual Van Sant E, Nemer J, et al. Predicting Breast Cancer Molecular Subtype with MRI Dataset Utilizing Convolutional Neural Network Algorithm. *J Digit Imaging*. 2019;32(2):276-82. doi: 10.1007/s10278-019-00179-2
40. Akin M, Erdemli S, Dulgeroglu O, Taskin F, Tokat F. Potential role of AI decision support in reducing unnecessary ultrasound-guided breast biopsies. *Egypt J Radiol Nucl Med*. 2026;57(1):22. doi: 10.1186/s43055-026-01684-5
41. Weitz M, Pfeiffer JR, Patel S, Biancalana M, Pekis A, Kannan V, et al. Performance of an AI-powered visualization software platform for precision surgery in breast cancer patients. *npj Breast Cancer*. 2024;10(1):98. doi: 10.1038/s41523-024-00696-6
42. Chen KA, Kirchoff KE, Butler LR, Holloway AD, Kapadia MR, Kuzmiak CM, et al. Analysis of Specimen Mammography with Artificial Intelligence to Predict Margin Status. *Ann Surg Oncol*. 2023;30(12):7107-15. doi: 10.1245/s10434-023-14083-1
43. Shia WC, Kuo YH, Hsu FR, Lin J, Wu WP, Wu HK, et al. Evaluating the Margins of Breast Cancer Tumors by Using Digital Breast Tomosynthesis with Deep Learning: A Preliminary Assessment. *Diagnostics*. 2024;14(10):1032. doi: 10.3390/diagnostics14101032
44. Alsentzer E, Murphy J, Boag W, Weng WH, Jindi D, Naumann T, McDermott M. Publicly Available Clinical BERT Embeddings. *Proceedings of the 2nd Clinical Natural Language Processing Workshop*. 2019;72-8. doi: 10.18653/v1/W19-1909
45. Kehl KL, Elmarakeby H, Nishino M, Van Allen EM, Lepisto EM, Hassett MJ, et al. Assessment of Deep Natural Language Processing in Ascertaining Oncologic Outcomes From Radiology Reports. *JAMA Oncol*. 2019;5(10):1421-9. doi: 10.1001/jamaoncol.2019.1800
46. Di Maio M, Perrone F, Conte P. Real-world evidence in oncology: opportunities and limitations. *Oncologist*. 2020;25(5):e746-e752. doi: 10.1634/theoncologist.2019-0647
47. Singhal K, Azizi S, Tu T, Mahdavi SS, Wei J, Chung HW, et al. Large language models encode clinical knowledge. *Nature*. 2023;620(7972):172-80. doi: 10.1038/s41586-023-06291-2
48. Lewis P, Perez E, Piktus A, Petroni F, Karpukhin V, Goyal N, Küttler H, Lewis M, Yih WT, Rocktäschel T, Riedel S, Kiela D. Retrieval-augmented generation for knowledge-intensive NLP tasks. *Adv Neural Inf Process Syst* [Internet]. NeurIPS; 2020 Oct 8;33:9459-74. [cited 8 10 2026]. Available from: https://papers.neurips.cc/paper_files/paper/2020/hash/6b493230205f780e1bc26945df7481e5-Abstract.html
49. Park H, Ok S, Kang T, Park M. Integrating Large Language Models with Deep Learning for Breast Cancer Treatment Decision Support. *Diagnostics*. 2026;16(3):394. doi: 10.3390/diagnostics16030394
50. Lipkova J, Chen RJ, Chen B, Lu MY, Barbieri M, Shao D, et al. Artificial intelligence for multimodal data integration in oncology. *Cancer Cell*. 2022;40(10):1095-110. doi: 10.1016/j.ccell.2022.09.012
51. Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN, Kaiser Ł, Polosukhin I. Attention is all you need. *Adv Neural Inf Process Syst* [Internet]. NeurIPS; 2017 Oct 8;30:5998-6008. [cited 8 10 2026]. Available from: https://papers.nips.cc/paper_files/paper/2017/hash/3f5ee243547dee91fbd053c1c4a845aa-Abstract.html

52. Polikar R. Ensemble based systems in decision making. *IEEE Circuits Syst. Mag.* 2006;6(3):21-45. doi: 10.1109/mcas.2006.1688199
53. Akinlade O, Abdollahi J, Paul Etchie M, Ajagbe SA, Alabi OO, Ayo Adeyanju B. Ensuring effective cervical cancer diseases diagnosis system using ensemble machine learning model. *Int J Artif Intell*. 2026;15(3):2410-22. doi: 10.11591/ijai.v15.i3.pp2410-2422
54. Chen RJ, Lu MY, Weng WH, Chen TY, Williamson DFK, Manz T, Shady M, Mahmood F. Multimodal Co-Attention Transformer for Survival Prediction in Gigapixel Whole Slide Images. *2021 IEEE/CVF International Conference on Computer Vision (ICCV)*. 2021;4015-25. doi: 10.1109/ICCV48922.2021.00398
55. Steyaert S, Pizurica M, Nagaraj D, Khandelwal P, Hernandez-Boussard T, Gentles AJ, et al. Multimodal data fusion for cancer biomarker discovery with deep learning. *Nat Mach Intell*. 2023;5(4):351-62. doi: 10.1038/s42256-023-00633-5
56. Rieke N, Hancox J, Li W, Milletarì F, Roth HR, Albarqouni S, et al. The future of digital health with federated learning. *npj Digit. Med*. 2020;3(1):119. doi: 10.1038/s41746-020-00323-1
57. Ogier du Terrail J, Leopold A, Joly C, Béguier C, Andreux M, Maussion C, et al. Federated learning for predicting histological response to neoadjuvant chemotherapy in triple-negative breast cancer. *Nat Med*. 2023;29(1):135-46. doi: 10.1038/s41591-022-02155-w
58. US Food and Drug Administration. Artificial Intelligence-Enabled Medical Devices [Internet]. Silver Spring, MD: U.S. Food and Drug Administration; 2026 Mar 8. [cited 8 10 2026]. Available from: https://www.fda.gov/medical-devices/digital-health-center-excellence/artificial-intelligence-enabled-medical-devices
59. US Food and Drug Administration. Predetermined Change Control Plans for Machine Learning-Enabled Medical Devices: Guiding Principles [Internet]. Silver Spring, MD: U.S. Food and Drug Administration; 2023 Oct 8. [cited 8 10 2026]. Available from: https://www.fda.gov/medical-devices/artificial-intelligence-enabled-medical-devices/predetermined-change-control-plans-machine-learning-enabled-medical-devices-guiding-principles
60. Abdollahi J, Nouri-Moghaddam B, Mohajer A, Mirzaei A, Derakhshanfard N. Bridging global semantic representation and inference efficiency via hybrid knowledge distillation for clinically deployable cardiac MRI classification. *Intelligence-Based Medicine*. 2026;14:100366. doi: 10.1016/j.ibmed.2026.100366
61. Lundberg SM, Lee SI. A unified approach to interpreting model predictions. *Adv Neural Inf Process Syst* [Internet]. NeurIPS; 2017 Oct 8;30:4765-74. [cited 8 10 2026]. Available from: https://papers.nips.cc/paper/2017/hash/8a20a8621978632d76c43dfd28b67767-Abstract.html
62. Klauschen F, Dippel J, Keyl P, Jurmeister M, Bockmayr M, Mock A, et al. Toward explainable artificial intelligence for precision pathology. *Annu Rev Pathol*. 2024;19:541-70. doi: 10.1146/annurev-pathmechdis-051222-113147
63. Obermeyer Z, Powers B, Vogeli C, Mullainathan S. Dissecting racial bias in an algorithm used to manage the health of populations. *Science*. 2019;366(6464):447-53. doi: 10.1126/science.aax2342
64. Seyyed-Kalantari L, Zhang H, McDermott MBA, Chen IY, Ghassemi M. Underdiagnosis bias of artificial intelligence algorithms applied to chest radiographs in under-served patient populations. *Nat Med*. 2021;27(12):2176-82. doi: 10.1038/s41591-021-01595-0
65. Chen RJ, Ding T, Lu MY, Williamson DFK, Jaume G, Song AH, et al. Towards a general-purpose foundation model for computational pathology. *Nat Med*. 2024;30(3):850-62. doi: 10.1038/s41591-024-02857-3
66. Al Amri WS, Al Jabri M, Al Abri A, Hughes TA. Cancer Genetics in the Arab World. *Technol Cancer Res Treat*. 2025;24:15330338251336829. doi: 10.1177/15330338251336829
67. Al-Jumaan M, Chu H, Alsulaiman A, Camp SY, Han S, Gillani R, et al. Interplay of Mendelian and polygenic risk factors in Arab breast cancer patients. *Genome Med*. 2023;15(1):65. doi: 10.1186/s13073-023-01220-4
68. Wan JCM, Massie C, Garcia-Corbacho J, Mouliere F, Brenton JD, Caldas C, et al. Liquid biopsies come of age: towards implementation of circulating tumour DNA. *Nat Rev Cancer*. 2017;17(4):223-38. doi: 10.1038/nrc.2017.7
69. Coombes RC, Page K, Salari R, Hastings RK, Armstrong A, Ahmed S, et al. Personalized Detection of Circulating Tumor DNA Antedates Breast Cancer Metastatic Recurrence. *Clinical Cancer Research*. 2019;25(14):4255-63. doi: 10.1158/1078-0432.ccr-18-3663
70. Kelly CJ, Karthikesalingam A, Suleyman M, Corrado G, King D. Key challenges for delivering clinical impact with artificial intelligence. *BMC Med*. 2019;17(1):195. doi: 10.1186/s12916-019-1426-2
71. Bates DW, Auerbach A, Schulam P, Wright A, Saria S. Reporting and Implementing Interventions Involving Machine Learning and Artificial Intelligence. *Ann Intern Med*. 2020;172(11 Suppl):S137-S144. doi: 10.7326/M19-0872
72. Howard FM, Dolezal J, Kochanny S, Schulte J, Chen H, Heij L, et al. The impact of site-specific digital histology signatures on deep learning model accuracy and bias. *Nat Commun*. 2021;12(1):4423. doi: 10.1038/s41467-021-24698-1
73. Hsu W, Hippe DS, Nakhaei N, Wang PC, Zhu B, Siu N, et al. External Validation of an Ensemble Model for Automated Mammography Interpretation by Artificial Intelligence. *JAMA Netw Open*. 2022;5(11):e2242343. doi: 10.1001/jamanetworkopen.2022.42343

How to Cite This Article

Taslimi J, Minagar A. Artificial Intelligence-Driven Multimodal Integration of Genomic, Pathological, and Clinical Data in Breast Cancer: A Narrative Review of Clinical Applications and Translational Challenges. Archives of Breast Cancer. 2026; 13(4):398-411. doi: 10.32768/abc.4829103847-562
Available from: https://www.archbreastcancer.com/index.php/abc/article/view/1337