cancers

MDPI

Review

Beyond pCR Prediction: Subtype-Specific Artificial Intelligence for Treatment Tailoring in Breast Cancer Neoadjuvant Therapy

Junwen Zhou [1], Beibei Xi [2], Kehuan Yan [1] and Linying Chen [2,*]

[1] College of Computer and Data Science, Fuzhou University, Fuzhou 350108, China; 2501010020@fzu.edu.cn (J.Z.); 241010006@fzu.edu.cn (K.Y.)
[2] Department of Pathology, The First Affiliated Hospital of Fujian Medical University, Fuzhou 350005, China; beibeixi@fjmu.edu.cn
* Correspondence: forestchen@fjmu.edu.cn

Simple Summary

Breast cancer comprises biologically distinct subtypes rather than a single disease, and each is treated with a different pre-surgery strategy. Clinicians want to know early whether a patient is likely to respond, because that could guide safer treatment adjustment. Artificial intelligence has therefore been used to estimate response from routinely collected imaging, pathology slides, and laboratory data, using models that range from statistical scores to deep-learning systems and multimodal designs. This review finds such estimates are not interchangeable across subtypes and must be read in their original treatment context. Most models were tested only where they were built, few in independent patients, and fewer still for reliable probabilities or improved decisions. Current evidence is therefore not strong enough for AI to change an individual treatment or surgery decision safely. We map how mature the evidence is, how it differs across subtypes, and what is still required.

Abstract

Neoadjuvant therapy for breast cancer is planned by subtype: pathologic complete response (pCR) differs in frequency, meaning, and surrogate validity across HR+/HER2−, HER2+, and triple-negative breast cancer (TNBC). This review examines whether current evidence supports using artificial intelligence (AI) to move beyond predicting response under a fixed regimen to guiding systemic treatment tailoring. Here, tailoring means model-guided drug omission, switching, escalation, or de-escalation for an individual patient. We appraised 102 full-text studies of AI-based response prediction (2020–2026), organized by subtype and clinical decision, and graded each on an author-defined five-level clinical-readiness ladder (L1–L5) measuring validation and translational maturity rather than accuracy. Risk of bias was assessed with PROBAST and reporting against TRIPOD+AI. Readiness clustered low: 43 studies reached internal validation only (L1), 53 temporal or geographic external validation (L2), and 6 prospective observational validation (L3); none reached workflow integration (L4) or interventional evidence (L5). Most carried high overall risk of bias (89/102); external validation appeared in 52 (51%), fully reported decision-curve analysis in 44 (43%), and calibration in 19 (19%). Strategy comparison and individualized-treatment-effect analyses were essentially absent. Subtype-specific AI currently predicts response under fixed regimens with moderate-to-good discrimination; the evidence does not yet support changing an individual patient's systemic treatment.

Keywords: breast cancer; neoadjuvant therapy; pathologic complete response; artificial intelligence; clinical prediction models

Check for updates

Academic Editor: Marcus Schmidt

Received: 9 August 2026
Revised: 6 September 2026
Accepted: 7 September 2026
Published: 11 September 2026

Copyright: © 2026 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license.

1. Introduction

Neoadjuvant therapy provides a window to observe breast cancer response before surgery and to inform subsequent treatment. Its purpose now extends beyond tumor shrinkage to subtype-specific decisions about systemic therapy and surgical extent. HR+/HER2−, HER2+, and triple-negative breast cancer (TNBC) differ in biology, treatment, and the meaning of response. A pooled response score can obscure these differences. This review asks whether artificial intelligence (AI) can move from predicting response to supporting treatment tailoring, and examines that question separately for each subtype.

Established treatment pathways show why the distinction matters. KATHERINE supported trastuzumab emtansine (T-DM1) for observed residual HER2+ disease [1]; KEYNOTE-522 demonstrated event-free-survival benefit from immunochemotherapy in TNBC [2]. Pathologic complete response (pCR) has different prognostic meaning across subtypes, with a weaker survival association in HR+/HER2− disease [3]. Residual cancer burden (RCB) and recurrence risk may therefore better match some clinical questions. Changing targets alone does not guarantee improvement: deep-learning radiomics failed to improve RCB prediction over simpler measures in the LIMA MRI cohort [4]. Histopathology-based recurrence and chemotherapy-benefit models broaden the questions under study [5]. Table 1 summarizes endpoint definitions and their clinical meaning.

AI approaches include radiomics, deep learning, digital pathology, and multimodal fusion, which combines inputs such as imaging, pathology, and clinical variables. Their predictions require three separate judgments. Discrimination measures how well a model ranks responders above non-responders, commonly summarized by the area under the receiver operating characteristic curve (AUC). Calibration measures agreement between predicted probabilities and observed outcomes. A calibrated 30% response probability means that approximately 30 of 100 comparable patients respond. Clinical utility asks whether acting on a prediction does more good than harm. Decision-curve analysis (DCA) estimates potential net benefit at specified thresholds against treating all or no patients; it does not demonstrate prospective clinical benefit. Appendix A (Table A1) provides the technical glossary.

Table 1. Endpoint definitions and their clinical meaning across the 102 tier A full-text studies. Abbreviations: pCR, pathologic complete response; DCIS, ductal carcinoma in situ; RCB, residual cancer burden; DFS/iDFS, disease-free/invasive disease-free survival; RFS, recurrence-free survival; OS, overall survival; EFS, event-free survival; DCA, decision-curve analysis.

Endpoint Category	Definition and Meaning	Studies	Decision Relevance
pCR definition groups (*n* = 101 studies with a stated or inferable pCR label; full nine-category breakdown in Supplementary Table S2)			
pCR—ypT0/is ypN0 (residual DCIS allowed)	No residual invasive cancer in breast and axillary nodes; residual DCIS permitted.	50	Most common and FDA-recognized pCR definition; the label most models are trained on.
Stricter or narrower variants	Strict ypT0 ypN0 (no residual in situ), breast-only ypT0/Tis, or standalone axillary pCR definitions.	4	Shift the event rate and the meaning of a positive prediction; not interchangeable with the standard label.
Operationalized through another scale	pCR equated to RCB-0, or represented by the top grade of a pathological-regression system (Miller–Payne, Sataloff, Japanese-society).	22	Ties the label to a different grading continuum; semantics are heterogeneous across systems.
Ambiguous or unverifiable	In situ status unstated, no verifiable verbatim definition, or other non-standard wording.	25	Without a verifiable label one cannot tell what the model predicts; flagged as non-comparable.

Table 1. *Cont.*

Endpoint Category	Definition and Meaning	Studies	Decision Relevance
Other endpoints and decision relevance (studies reporting each endpoint in any role, primary or secondary)			
Residual cancer burden (RCB), continuous/graded	MD Anderson index grading residual disease RCB-0/I/II/III.	8	Tracks long-term outcome better than binary pCR and can inform escalation, but reported in a minority.
Axillary/nodal response (as primary or secondary endpoint)	Endpoint involves axillary nodal status (ypN0 conversion, axillary pCR).	22	Directly relevant to axillary surgical de-escalation (SLNB instead of ALND).
Radiological response (RECIST/radiological CR)	RECIST v1.1 or radiological complete response, usually secondary.	6	Imaging response is not pathological response; weak as a standalone decision basis.
Disease-/recurrence-free survival (DFS/iDFS)	DFS or iDFS, usually secondary/prognostic.	21	Direct long-term benefit measure; a few studies use it to test pCR surrogacy.
Recurrence-/relapse-free survival (RFS)	RFS (recurrence/relapse) used as an endpoint.	9	Close to DFS; reflects recurrence risk.
Overall survival (OS)	OS, usually secondary/exploratory.	8	The hardest benefit endpoint; rarely primary given limited follow-up.
Event-free survival (EFS)	EFS (incl. one study testing pCR as an EFS surrogate).	2	Key regulatory endpoint in neoadjuvant trials; very rare here.
Toxicity/adverse events/safety	Treatment toxicity or safety as a predicted endpoint.	0	Escalation trades off toxicity, yet no model addresses it—a key evidence gap.
Decision-curve analysis/net benefit (decision-relevance metric)	DCA/net benefit estimates net benefit under specified thresholds and assumptions; it is not an endpoint or prospective impact evidence.	44	Separates discrimination from possible threshold-dependent net benefit, but cannot replace prospective impact evaluation.

Notes: The pCR-definition rows classify the 101 studies with a stated or inferable pCR label; one study modeled a non-pCR endpoint only and is not assigned a category. The full nine-category breakdown of verbatim definition classes is given in Supplementary Table S2. Counts in the lower section tally any-role reporting (primary or secondary); a study reporting more than one endpoint appears in more than one row. AUCs obtained under different pCR definitions, cohorts, or subtypes are not directly comparable and must not be ranked.

Here, treatment tailoring means using a model to omit, switch, escalate, or de-escalate an individual patient's systemic therapy. Surgical or axillary management, monitoring without a treatment change, and prognostic stratification are related applications with different evidentiary requirements. We identify these applications explicitly rather than equating them with systemic-treatment selection.

Predicting pCR under a fixed regimen is a prognostic claim. Selecting treatment requires evidence about alternatives, through strategy comparison, individualized treatment-effect (ITE) estimation, or prospective interventional testing. ITE denotes the difference a treatment would make for an individual compared with an alternative. We assess models against this distinction and their intended decision, with Figure 1 mapping the subtype-specific clinical pathway.

The clinical application must therefore be specified before interpreting a response score. The same output may support prognosis, motivate surgical research, or suggest an adaptive-treatment hypothesis. Evidence for one application cannot automatically license another. This distinction allows computational advances to be evaluated without assigning them a treatment-selection role that their design cannot test.

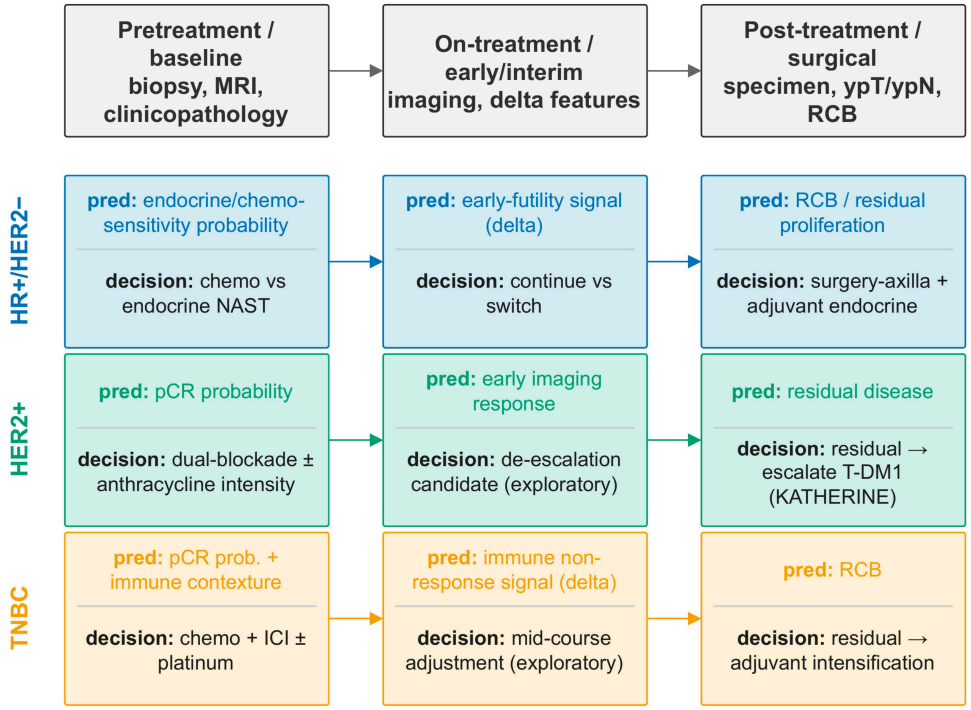


Figure 1. Clinical decision pathway across breast cancer subtypes in the neoadjuvant setting. AI output remains in the prediction layer; the decision available at each timepoint differs by subtype.

2. Methods

2.1. Design and Interpretive Boundary

This critical narrative review combines a subtype-organized synthesis with structured quantitative appraisal; it is not a systematic review. The documented multi-source search was non-exhaustive, and full-text appraisal covered a purposively selected subset. Heterogeneous endpoints, cohorts, and validation designs precluded meaningful pooling of accuracy or direct ranking of models. Instead, we examined what each study validated, how it was evaluated, and which decision it could inform. The same predefined extraction and appraisal criteria were applied throughout. Proportions describe the appraised tier A set only, not the prevalence of practices across the field. The design supports within-set reporting patterns and statements about whether included studies meet specified readiness criteria. Selection and assessment limitations are discussed in Section 10.

2.2. Study Identification

We searched PubMed/PMC, Crossref, OpenAlex, Semantic Scholar, Unpaywall, bioRxiv/medRxiv, and arXiv, supplemented by citation tracking. The search covered English-language primary studies from January 2020 to 18 June 2026. Eligibility required AI or machine-learning prediction of breast-cancer neoadjuvant response with at least one performance metric. Screening or diagnosis outside this setting, segmentation or subtype classification alone, reviews, protocols, and animal or cell-only studies were excluded. Earlier landmark trials and guidelines served as background; preprints and trial-registry records supplied context rather than appraised outcome evidence.

Figure 2 documents identification, core prioritization, coverage-directed full-text additions, and final tier classification; Supplementary Methods S1 retains the complete numerical cascade and selection rules. Prioritization combined relevance with coverage of

subtype, decision timepoint, modality, and treatment theme. Each study was classified by these dimensions. Pooled performance was not extrapolated to an individual subtype.

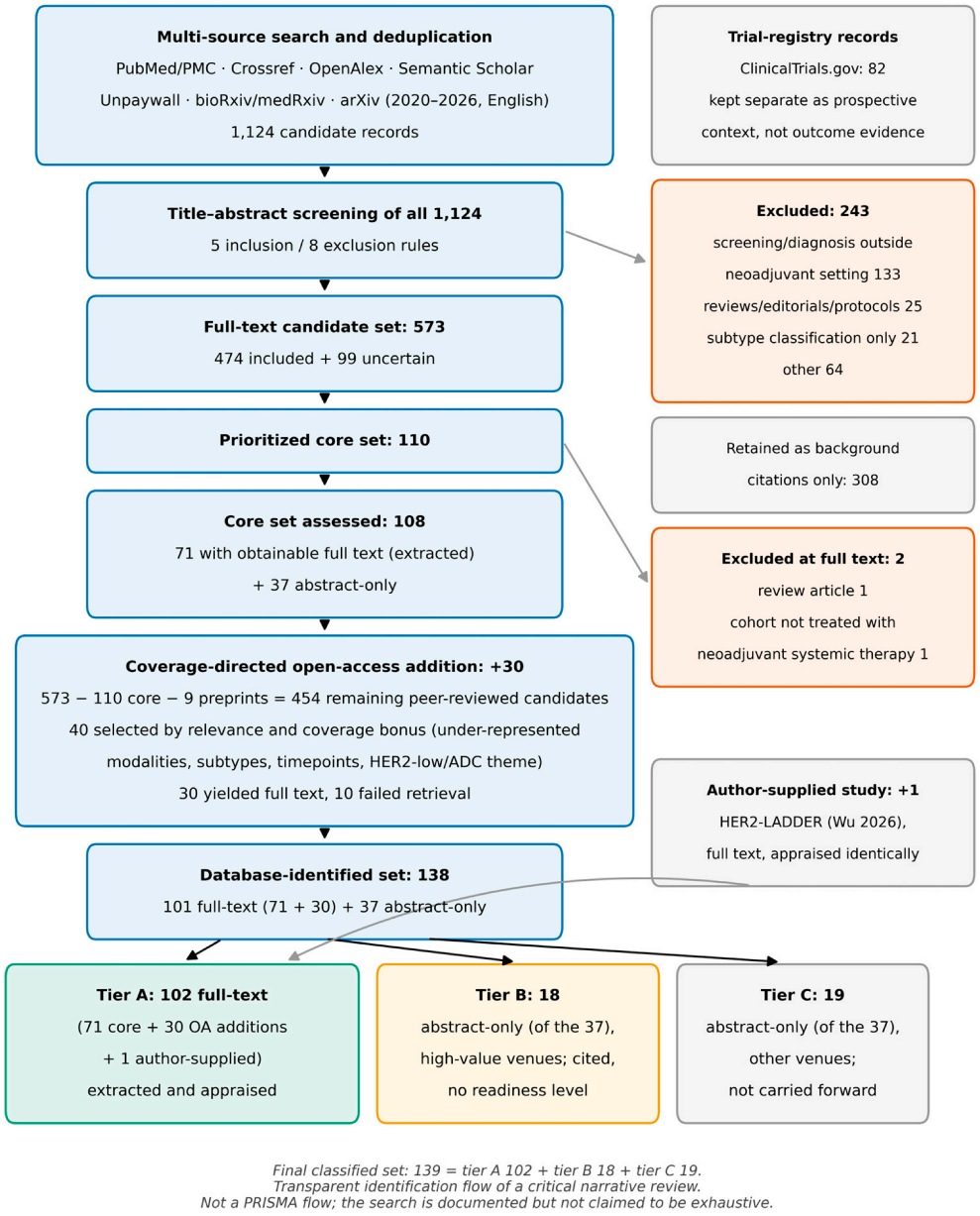


Figure 2. Study identification and tier classification. Multi-source search and deduplication yielded 1124 candidate records (January 2020 to 18 June 2026, English), with 82 trial-registry records held separate as prospective context. Title–abstract screening produced 474 included and 99 uncertain records, 243 exclusions with reasons, and 308 records retained only as background citations. From the full-text candidate set (n = 573), 110 records were prioritized as the core set; 2 were excluded at full-text review, leaving 108 (71 full-text, 37 abstract-only). A coverage-directed open-access addition then selected 40 records from the 454 remaining peer-reviewed candidates, of which 30 yielded full text, giving 138 database-identified records; one author-supplied study completed the set of 139. The final classification comprises tier A (102 full-text studies, extracted and appraised), tier B (18 abstract-only, cited without readiness grading), and tier C (19, not carried forward). The diagram documents a transparent narrative process; it is not a PRISMA flow, and the search is not claimed to be exhaustive [6].

For corpus counts, a named subtype required a cohort restricted to that subtype by design. Mixed cohorts remained mixed even when one subtype predominated. For clinical

synthesis, separately reported subtype results could inform the corresponding decision question. This distinction prevents assigning a pooled outcome to patients for whom performance was never estimated. Full-text prioritization sought coverage across these questions rather than statistical representativeness.

2.3. Evidence Tiers

Tier A comprised 102 full-text studies, including the author-supplied HER2-LADDER report [6], all extracted and appraised in Supplementary Table S1. Tier B comprised 18 abstract-only reports from high-value venues, cited as supporting evidence without readiness grading. Tier C comprised 19 remaining abstract-only records and was not carried forward. Tiers describe access to appraisal material, not scientific quality: a tier B study may be stronger than a tier A study. Venue-based prioritization and full-text accessibility therefore require the selection caveat in Section 10.

An abstract may establish that a model and a reported association exist, but usually cannot verify data splitting, calibration, or the independence of validation cohorts. Abstract-level citations therefore supplement clinical context without contributing to the quantitative appraisal. Conversely, obtaining a full text permits assessment without certifying methodological quality. Keeping these two judgments separate is necessary to interpret the evidence tiers correctly.

2.4. Appraisal and Clinical Readiness

Extraction covered cohort, subtype, regimen, endpoint, validation, sample size, and reported metrics. Risk of bias was assessed with PROBAST, referring to PROBAST+AI [7,8]. Reporting was read against TRIPOD+AI, CLAIM 2024, and CLEAR [9–11].

The author-defined L1–L5 ladder grades validation and translational maturity, not accuracy; it is not an externally validated standard. L1 denotes internal validation, including any random same-source split. L2 requires temporal or geographic separation from development. L3 requires frozen-model prospective observational or silent evaluation without altering care. L4 requires live workflow influence with measured impact. L5 requires prospective comparative model-guided intervention or regulated deployment with prospective outcome surveillance. High AUC, calibration, DCA, and reader comparisons cannot independently raise a level.

Supplementary Methods S2 preserves mandatory criteria, exclusions, promotion rules, validation subtypes, and the prior four-study re-audit. It also explains the distinction from bias appraisal and the relationship to DECIDE-AI [12]. All counts use the grading documented in Supplementary Methods S2.

3. Clinical Decisions and the Model Landscape

3.1. Decision Landscape

The relevant clinical decision determines what a response model must demonstrate. Pretreatment selection, on-treatment adjustment, and post-treatment management require different evidence, even within the same subtype. Figure 1 locates these decisions. A model can estimate the outcome under observed care while leaving the consequences of changing that care unanswered.

In HR+/HER2− disease, the pretreatment question is often whether chemotherapy is needed at all. Low pCR frequency and its weaker survival association limit the relevance of a pCR classifier to that choice. Endocrine sensitivity, residual burden, and recurrence risk are more closely aligned targets. During treatment, detecting endocrine futility would require longitudinal evidence early enough to guide a switch. After treatment, nodal response may inform surgical planning [13]. Recurrence and disease-free-survival models

remain prognostic [14,15]; nodal-response predictors must also be compared with routine factors such as clinical response and Ki-67 [16].

In HER2+ disease, observed residual disease already supports post-neoadjuvant escalation. Pretreatment or early-response estimates instead raise de-escalation questions, including whether anthracyclines or part of chemotherapy could be omitted. The safety requirements differ because omission depends on an outcome that has not yet been observed. Multimodal, histopathological, radiomic, and clinicopathologic models predict fixed-regimen pCR [17–20]. Their discrimination cannot establish the residual-disease risk under a reduced regimen. The missing evidence concerns the proposed treatment change, even where prediction is externally validated.

In TNBC, pCR is more strongly associated with survival, but a high probability under immunochemotherapy does not show that immunotherapy or chemotherapy components are dispensable. Gene-expression and radiomics models provide subtype-restricted predictive evidence [21,22]. Stromal tumor-infiltrating lymphocytes (sTILs), assessable on routine hematoxylin-and-eosin (H&E) slides, provide a clinically accessible comparator. Complex models need to show useful increment beyond this baseline, particularly when proposing a drug-sparing decision.

Longitudinal monitoring spans all subtypes but has a separate evidentiary gap. External multicenter MRI validation, multi-timepoint multimodal prediction, and prospective longitudinal testing show that response estimates can be updated [23–25]. The additional information may identify eventual responders earlier. However, demonstrating that stopping or switching therapy at that time improves outcomes requires a model-guided strategy evaluation. Monitoring studies therefore contribute response evidence without establishing a safe adaptive treatment pathway.

3.2. Models, Endpoints, and Subtype Coverage

Model families largely follow the available data. Regression and classical machine learning use selected clinical, radiomic, or molecular features. Convolutional networks learn imaging representations, while transformers and attention mechanisms support spatial or longitudinal integration. Gradient boosting and stacking combine tabular or model-level information. Table 2 and Figure 3a summarize these families; Supplementary Note S3 retains the backbone census and parameterization detail.

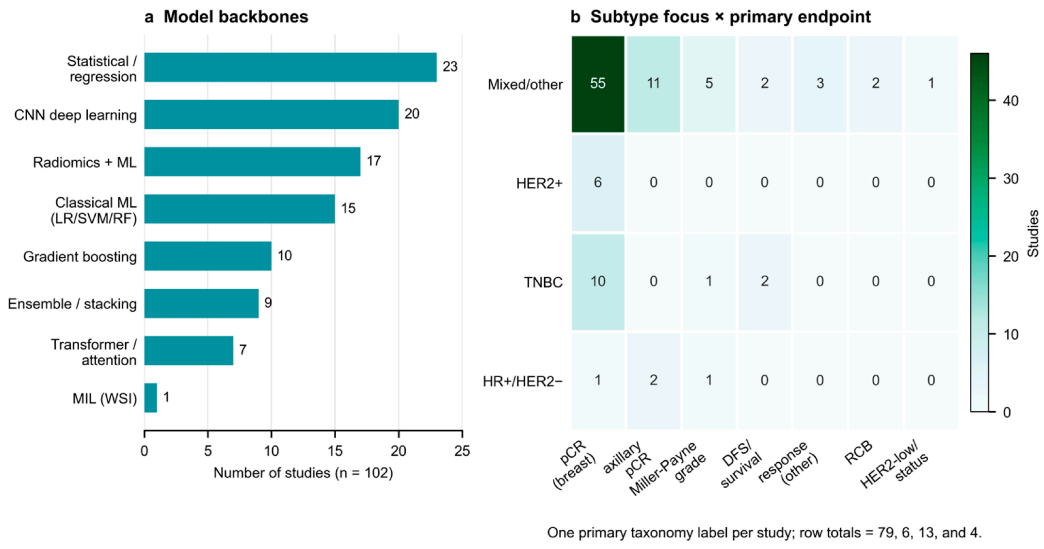


Figure 3. Model taxonomy across the 102 studies: (**a**) model backbones and (**b**) subtype focus by endpoint.

Table 2. Model backbones and their typical parameterization across the 102 tier A studies, with the number of studies using each backbone by endpoint. Abbreviations: CNN, convolutional neural network; ML, machine learning; MIL, multiple-instance learning; pCR, pathologic complete response; RCB, residual cancer burden; MP, Miller–Payne; DFS, disease-free survival; HER2, human epidermal growth factor receptor 2.

Model Backbone	Typical Parameterization	Representative Study	pCR (Breast)	Axillary Node pCR	MP Grade	DFS/ Surv	Resp Other	RCB	HER2 Low	Total
Statistical/regression	LASSO/elastic-net feature selection + logistic regression	Ni et al. 2025 [26]	15	4	1	2	1	0	0	23
CNN deep learning	ResNet/U-Net backbones; segmentation + deep features	Mao et al. 2025 [27]	14	3	3	0	0	0	0	20
Radiomics+ML	LASSO radiomic scores across imaging modalities	Yu et al. 2023 [14]	11	2	1	2	1	0	0	17
Classical ML (LR/SVM/RF)	RFE/LASSO selection + SVM/RF/naive Bayes	Chen et al. 2025 [28]	11	2	1	0	1	0	0	15
Gradient boosting (XGBoost/ LightGBM)	tabular multimodal/ multi-omic fusion	Huang et al. 2025 [23]	8	0	1	0	0	0	1	10
Ensemble/stacking	stacked base models across timepoints/ modalities	Zhu et al. 2024 [25]	7	1	0	0	0	1	0	9
Transformer/ attention DL	cross-modal/ longitudinal attention fusion	Liu et al. 2025 [13]	6	1	0	0	0	0	0	7
Multiple-instance learning (WSI)	attention pooling over WSI patch embeddings	hou et al. 2026 [29]	0	0	0	0	0	1	0	1

Notes: Counts are study numbers; a study may use more than one endpoint, so row totals need not equal the study count. Backbone assignment reflects the primary modeling approach reported in each study.

The distribution in Figure 3b and Table 3 uses mutually exclusive design-based categories: 79 mixed/other studies, 6 HER2+-restricted, 13 TNBC-restricted, and 4 HR+/HER2−-restricted. TNBC-predominant cohorts enrolling other subtypes remain mixed/other. Dedicated subtype designs are uncommon, especially in HR+/HER2− disease. This partition differs from decision-cell mapping, where a mixed study can inform several cells through separately reported subtype results.

MRI is the most frequent single imaging source, and multimodal combinations draw on imaging, pathology, clinical variables, and omics. Recent work extends to spatial–semantic biopsy analysis, explainable multiomics, transcriptomic–histopathological integration, inferred spatial transcriptomics, and ultrasound pretraining [30–34]. These advances broaden representation learning; their clinical significance still depends on the outcome, validation setting, and comparator. Model capacity is therefore an appraisal consideration rather than a basis for ranking technologies. Learned features may improve discrimination while remaining difficult to audit across centers or scanners.

Endpoint variation is more consequential than architectural labels. The pCR definitions grouped in Table 1 differ in whether residual in situ disease or nodal involvement is allowed. Supplementary Table S2 retains the verbatim nine-category breakdown. These differences alter event frequency and the meaning of a positive prediction. Equal AUCs under different definitions do not imply equal predictive performance for the same clinical task. The

distinction also affects threshold selection: a probability calibrated for one definition cannot simply be interpreted as the probability of another.

Table 3. Subtype-conditioned distribution of (a) imaging/data modalities, (b) evaluation endpoints, and (c) model backbones with clinical-readiness ceilings across the 102 tier A studies (study counts). Abbreviations: MRI, magnetic resonance imaging; PET, positron-emission tomography; DBT, digital breast tomosynthesis; pCR, pathologic complete response; RCB, residual cancer burden; MP, Miller–Payne; DFS, disease-free survival; HER2, human epidermal growth factor receptor 2; CNN, convolutional neural network; ML, machine learning; LR, logistic regression; SVM, support vector machine; RF, random forest; L1–L3, clinical-readiness levels.

(a) Subtype × modality

Subtype focus	multimodal	MRI	ultrasound	clinicopath	pathology	omics	PET	liquid biopsy	mammo-DBT	Total
Mixed/other	24	21	12	10	5	2	3	2	0	79
HER2+	1	0	1	1	1	1	1	0	0	6
TNBC	4	3	0	1	3	2	0	0	0	13
HR+/HER2−	2	0	0	1	1	0	0	0	0	4

(b) Subtype × endpoint

Subtype focus	pCR (breast)	axillary node pCR	MP grade	DFS/surv	resp other	RCB	HER2 low	Total
Mixed/other	55	11	5	2	3	2	1	79
HER2+	6	0	0	0	0	0	0	6
TNBC	10	0	1	2	0	0	0	13
HR+/HER2−	1	2	1	0	0	0	0	4

(c) Subtype × backbone and clinical-readiness

Subtype focus	Studies	Dominant backbone(s)	Readiness (L1/L2/L3)	Ceiling
Mixed/other	79	CNN deep learning (17); Statistical/regression (16)	34/40/5	L3
HER2+	6	Statistical/regression (3); CNN deep learning (2)	1/5/0	L2
TNBC	13	Classical ML (6); Statistical/regression (3)	6/6/1	L3
HR+/HER2−	4	No dominant family (one study each)	2/2/0	L2

Notes: Modality and endpoint panels use one primary taxonomy label per study, so each row sums to its mutually exclusive study total. The separate subtype-by-timepoint evidence map presented later is non-exclusive. A study is assigned to a named subtype only when its design is restricted to that subtype; subtype-predominant mixed cohorts remain mixed/other. Readiness (L1/L2/L3) gives the count of studies at each level; ceiling is the highest level reached.

pCR and RCB answer related but distinct questions. pCR records absence of invasive residual disease as a binary event; RCB grades remaining disease on a continuous index with prognostic strata. Axillary pCR addresses nodal management, survival endpoints address longer-term prognosis, and immune or HER2 scores characterize tumor features. A model validated for one endpoint cannot be assumed to predict the others. Table 4 organizes primary endpoints and evaluation metrics without treating them as interchangeable.

Evaluation concentrates on discrimination, whereas calibration and threshold-dependent net benefit are less completely reported. Section 8 quantifies these gaps. A comparison is interpretable only within its cohort, outcome definition, and validation design. Subtype-conditioned reporting should therefore test whether performance persists within the intended clinical population. Adding a subtype indicator to a pooled model does not by itself establish that evidence. The following sections assess each subtype against its decision, appropriate comparator, and remaining validation needs.

Table 4. Evaluation endpoints, dominant metrics, and coverage of external validation, calibration, and decision-curve analysis across the 102 tier A studies, by primary modeled endpoint. Abbreviations: AUC, area under the receiver operating characteristic curve; CI, confidence interval; DCA, decision-curve analysis; pCR, pathologic complete response; RCB, residual cancer burden; TIL, tumor-infiltrating lymphocyte; DFS, disease-free survival; HER2, human epidermal growth factor receptor 2.

Evaluation Endpoint	Studies	Dominant Metric(s)	External Val. (Yes)	Calibration (Yes)	DCA (Yes)	Representative Study
pCR (breast)	72	AUC/ROC; sens/spec; accuracy	38	11	27	Sammut et al. [17]
Axillary/nodal pCR	13	AUC/ROC; sens/spec; accuracy	7	7	11	Li et al. [35]
Miller–Payne/ordinal grade	7	AUC/ROC; sens/spec; accuracy	1	0	4	Chen et al. [36]
Residual cancer burden (RCB)	2	AUC/ROC; sens/spec; accuracy	2	0	1	Li et al. [37]
TILs/immune spatial/other	3	AUC/ROC; accuracy; F1/κ	1	0	0	Peng et al. [38]
DFS/survival	4	AUC/ROC; C-index; sens/spec	3	0	1	Yu et al. [14]
HER2-low status	1	AUC/ROC; sens/spec; accuracy	0	1	0	Huang et al. [39]

Notes: "Studies" assigns each study once, by its primary modeled endpoint; the seven rows partition the 102 studies. Any-role endpoint tallies, in which one study can appear in several rows, are given in the lower section of Table 1. External validation, calibration, and DCA columns give the number of studies reporting each item. Totals: external validation, 52; calibration, 19 (fully reported); DCA, 44 (fully reported). **Abbreviations:** AUC, area under the receiver operating characteristic curve; ROC, receiver operating characteristic; sens/spec, sensitivity/specificity; DCA, decision-curve analysis; pCR, pathologic complete response; RCB, residual cancer burden; DFS, disease-free survival; HER2, human epidermal growth factor receptor 2; C-index, concordance index; F1, F1 score; κ, kappa statistic.

The study-level extraction also distinguishes a model's primary target from outcomes reported in secondary analyses. For example, a paper may model breast pCR while additionally reporting nodal response or survival. Such reporting does not make those outcomes equivalent prediction tasks. The endpoint partition and broader reporting counts consequently answer different questions, as specified in the tables.

Within a backbone family, feature selection, dimensionality, and regularization further distinguish models. Statistical coefficients expose named predictors, whereas deep representations can encode imaging or spatial information without an equally direct interpretation. Those differences help explain what a model can reveal about response biology. They do not determine whether its output has been calibrated, independently tested, or compared with the current clinical decision rule. Supplementary Note S3 retains this computational detail for readers assessing individual pipelines.

4. HR-Positive/HER2-Negative Subtype

4.1. Clinical Questions and Appropriate Targets

Four questions structure this subtype: which tumors can avoid chemotherapy through endocrine sensitivity, what residual burden warrants escalation, when axillary surgery can be reduced, and whether futility can be detected early. The surgical question concerns treatment extent rather than systemic selection. Established benchmarks already include high-risk adjuvant CDK4/6 inhibition from monarchE and NATALEE [40,41], on-treatment Ki-67 from POETIC [42], and PEPI residual-risk assessment [43]. Observational comparisons of neoadjuvant and adjuvant chemotherapy remain confounded by baseline risk [44]. An AI model must add to these decision-relevant benchmarks. Table 5 summarizes the outstanding comparisons.

Table 5. Subtype-specific AI: dominant modalities and model families, what the models demonstrably do, and distinctive insights. Abbreviations: MRI, magnetic resonance imaging; DCE, dynamic contrast-enhanced; DL, deep learning; ML, machine learning; WSI, whole-slide image; TIL, tumor-infiltrating lymphocyte; FTV, functional tumor volume; RCB, residual cancer burden; PPV, positive predictive value; ADC, antibody–drug conjugate; ITE, individualized treatment effect.

Subtype	Modalities & Model Families	What AI Demonstrably Does	Distinctive AI Insight
HR+/HER2−	MRI (incl. functional/DCE), ultrasound, clinicopathologic±genomic/hematologic, longitudinal imaging; radiomics+classical ML/nomograms, MRI deep learning, longitudinal delta models	Moderate pCR discrimination (subtype AUC ~0.78–0.86 [23,45–48]); FTV AUC 0.70 but only 20% sensitivity at PPV $\geq$ 50% [49]; shift to RCB endpoints [37,50,51]; an axillary-pCR nomogram with calibration+DCA [52]; recurrence/DFS stratification [14,15]	pCR is arguably the wrong target (rare ~20%, weak surrogate); the real value lies in RCB/residual, endocrine sensitivity, and whether AI beats the in-use Ki-67/genomic scores (essentially unconfirmed head-to-head); the low event rate suppresses PPV
HER2+	Digital-pathology WSI (H&E+HER2 IHC, spatial), MRI radiomics/DL, dual-tracer PET, clinicopathologic; gene-expression scores, multimodal fusion	HER2-LADDER [6], multi-cohort with calibration, DCA and spatial interpretability; gene-expression pCR score, external AUC 0.80/0.77 [53]; dual-tracer PET prospective cohort [54]; HER2-low AI-TIL scoring κ 0.68–0.72 [55] and MRI HER2-zero/low classification [39]	Closest an AI output comes to a drug decision, via HER2-low scoring for ADC/T-DXd eligibility. Fixed-regimen pCR $\neq$ de-escalation safety, and response under an actual T-DXd regimen remains unstudied
TNBC	Digital pathology (TIL/spatial immune, spatial graphs), MRI radiomics/DL with longitudinal delta, multi-omics/transcriptomics, liquid biopsy (RNA-seq, ctDNA), multimodal PET+genomics	The strongest predictive signal, with one prospective blinded test [56], cycle-4 AUC 0.83 vs. 0.65 baseline; AI-TILs κ 0.68–0.87 [55]; spatial-graph external AUC ~0.83 [57]; longitudinal T-cell composite [58]; TVR+sTIL+Ki-67 beats clinicopathologic-only [59]	Two open questions: whether AI beats the low-incremental-cost sTIL read, which no head-to-head study has tested, and whether longitudinal spatial or dynamic signals can monitor immune non-response. Predicting pCR $\neq$ safe to stop or reduce ICI
Cross-cutting: multimodal/subtype-conditioning	Cross-modal fusion (N = 31, the largest modality class); attention fusion, stacking ensembles, foundation models as frozen feature extractors	Fusion lifts discrimination, not readiness (of 31 studies: L1 = 17/L2 = 13/L3 = 1, none L4/L5); cross-subtype transfer collapses (HER2+ 0.90 $\rightarrow$ TNBC 0.59 [20])	Conditioning on subtype biology $\neq$ stacking modalities, and the conditioning is the hard part. Federated learning and foundation models remain proposals, not evidence

Notes: AUC values come from different cohorts, pCR definitions, and validation designs; they are for within-study reading only and must not be ranked across studies or subtypes (per-study values in Supplementary Table S1). Synthesis of the subtype sections and the multimodal section; citations are the manuscript's existing reference numbers.

The dominant pCR target is often poorly matched to this biology. I-SPY 2 analyses and functional tumor volume prediction illustrate the limitations of a rare endpoint with weak surrogacy [49,60]. RCB, residual proliferation, and recurrence risk offer more defensible alternatives [37,50,51]. Subtype-stratified imaging and multimodal studies report pCR discrimination, but their heterogeneous outcomes and designs preclude a common performance ranking [23,45–48]. Supplementary Table S1 retains the individual metrics. Predicting response within a treated cohort also leaves unresolved whether chemotherapy should have been selected initially.

Evidence of incremental value remains limited. A radiomics nomogram incorporating Ki-67 improved on baseline imaging but not on an on-treatment signature, without an HR+/HER2−-specific increment [51]. A disease-free-survival nomogram showed net benefit over clinical staging, rather than over a genomic recurrence score [15]. Essentially no study compared AI head-to-head with clinically used recurrence scores. These comparator choices matter because outperforming staging or a weak imaging baseline does not demonstrate added value over the tools used for endocrine-versus-chemotherapy decisions.

Low event frequency also affects the clinical meaning of a positive result. Functional tumor volume modeling showed limited responder identification at a clinically constrained positive predictive value [49]. Improving a ranking statistic need not resolve that limitation. The relevant comparison asks whether AI changes a decision after the standard endocrine-

sensitivity or recurrence-risk information is already available. That question remains more informative than another isolated pCR AUC in this subtype.

4.2. Residual Disease, Axillary Management, and Early Futility

Longitudinal imaging and integrated tissue-sequencing, circulating-tumor-DNA, and clinical models reached multicenter external validation for residual outcomes [37,50]. The imaging model reported DCA, but neither study established calibrated prospective deployment. Additional MRI-based RCB classification broadens residual-disease prediction [61]. This evidence supports continued validation of residual targets, without establishing which escalation strategy improves outcome for a predicted high-risk patient.

Axillary models in luminal disease report calibration and net benefit or validated nodal discrimination [52,62]. Conversely, low axillary-pCR frequency forced an FDG-PET/CT study to pool this subtype with others [63]. A mixed-cohort result cannot substitute for a subtype-specific surgical safety estimate. Predictive nodal evidence should also remain separate from endocrine sensitivity and chemotherapy benefit.

On-treatment imaging provides incremental response information but rarely validates an early stop-or-switch rule [37,49,51,64–66]. The three-gene endocrine-resistance score links cell models, retrospective survival cohorts, and eight post-letrozole samples; it remains exploratory [67]. Luminal phenotyping, subtype-specific ultrasound thresholds, and on-treatment MRI extend these signals [68–70]. A TNBC volume-reduction, sTIL, and Ki-67 model is an analogy rather than endocrine-futility validation [59].

4.3. Clinical Bottom Line: HR+/HER2−Disease

AI can stratify pCR or residual burden under observed treatment, with some residual-disease models reaching external validation. Clinically meaningful increment over Ki-67 and genomic recurrence tools remains unestablished. The evidence does not support choosing endocrine therapy over chemotherapy or stopping ineffective endocrine treatment on an AI score. These decisions need appropriately targeted and prospectively tested models.

The residual and axillary applications also differ in when their output becomes useful. A pretreatment estimate anticipates response, whereas a post-treatment nodal estimate addresses the extent of surgery after systemic therapy is complete. Combining these studies under a single maturity verdict would hide the different decisions they evaluate. The subtype summary therefore preserves their clinical separation.

5. HER2-Positive Subtype

HER2+ disease presents a clear distinction between escalation for observed residual disease and de-escalation based on predicted response. The KATHERINE pathway already supports post-neoadjuvant T-DM1, while anthracycline omission or shorter chemotherapy requires evidence of safety under the reduced regimen. Models must also demonstrate added value beyond routine hormone-receptor status and HER2 amplification. Table 5 locates these questions. A high pCR probability under dual blockade and chemotherapy does not answer what would happen if part of that treatment were removed.

MRI, ultrasound, elastography, deep learning, and digital-pathology studies report HER2+ performance, usually in subgroups of mixed cohorts [23,27,35,46,48,65,69,71–76]. HER2 scoring itself has also been linked to response [77]. Single-center optimism, falling external recall, differing clinical baselines, and reuse of public cohorts limit interpretation of apparent gains [17,20,64,78,79]. Individual AUCs remain in Supplementary Table S1 rather than being combined into a cross-study range.

HER2-LADDER derives cellular composition and spatial topology from baseline H&E and HER2 immunohistochemistry whole-slide images [6]. It combines cross-cohort validation, calibration, DCA, and spatial interpretability, making it an informative example of comparatively complete prediction reporting. However, its FASCINATE-N trial slides were assessed retrospectively. The temporal-validation AUC differs between reported locations, and the pCR label lacks a verbatim ypT0/ypN0 criterion. It remains L2 with high overall PROBAST risk, rather than evidence of prospectively successful treatment selection.

De-escalation evidence is still missing. A HER2DX-type gene-expression score was externally validated without calibration or DCA [53]; its probability threshold consequently lacks the evidence needed for a drug-omission decision. More fundamentally, fixed-regimen classifiers do not estimate the counterfactual response after anthracycline omission. ExteNET supplies clinical evidence for extended neratinib in the post-adjuvant setting [80], but the reviewed AI models do not predict differential T-DM1 or neratinib benefit. Observed residual disease, predicted residual disease, and predicted treatment benefit remain separate quantities.

Some models outperform clinicopathologic predictors or manual pathology reads [6,20,78]. HER2-LADDER also reports improved DCA relative to component models. These comparisons support incremental prediction within the evaluated cohorts. They do not establish superiority to externally developed systems, transportability to new practice settings, or prospective clinical utility.

The choice of comparator changes the claim that can be supported. An improvement over routine HR status or HER2 staining can demonstrate added predictive information. A comparison with manual pathology can support a reading aid. Neither evaluates whether the model-guided regimen is preferable to standard treatment. HER2-LADDER's spatial explanations strengthen the interpretability of its response score, but a plausible explanation and a favorable decision curve still do not supply the missing strategy contrast.

Early Response and the Clinical Bottom Line

Early-response studies include dual-tracer PET with internal cross-validation, temporally validated imaging fusion, and mixed-cohort longitudinal MRI [54,81,82]. Prospective enrollment alone does not raise an internally validated model above L1. A multimodal model reached L3 for axillary response, but that endpoint concerns nodal management rather than breast pCR or anthracycline omission [25]. No study linked an early AI-guided treatment reduction to comparative clinical outcomes.

HER2+ AI supports fixed-regimen response prediction, with some externally validated and calibrated models. It has not demonstrated that a predicted responder can safely receive less systemic treatment. Drug omission or shortening therefore requires strategy comparison and prospective evaluation of the proposed decision.

6. Triple-Negative Breast Cancer Subtype

For TNBC, response prediction is interpreted within chemotherapy and immune-checkpoint inhibition, while CREATE-X and OlympiA anchor post-neoadjuvant options for appropriate residual-disease populations [83,84]. Two questions govern AI's contribution: whether it adds to standardized sTIL assessment and whether on-treatment change can identify immune non-response early enough for safe adjustment. Trial-embedded TIL comparison in HER2+ disease, TNBC immunochemotherapy histology, TIL-based prognosis, and longitudinal MRI provide relevant context [85–88]. The HER2+ comparison does not establish equivalent TNBC evidence. Table 5 summarizes the subtype's predictive applications and limits.

6.1. Incremental Value over Stromal TILs

sTILs are an established, pathologist-readable marker with relatively low incremental cost when routine H&E slides and pathology infrastructure are available. The burden of proof therefore rests with the complex model. Added complexity needs to improve discrimination, calibration, or net benefit against a standardized sTIL read, rather than against random classification or a weaker clinical baseline.

Automated TIL scoring, spatial-graph features, transcriptomic nomograms, and biopsy models show heterogeneous predictive signals [55,57,89,90]. External TIL concordance declined in one study, and its spatial immune subtype was not an independent pCR predictor [55]. ARTEMIS provides a more direct incremental signal: tumor-volume reduction, sTILs, and Ki-67 improved discrimination over a clinicopathologic model [59]. However, validation used a random split within one institution and therefore remains L1. Calibration and DCA were absent, and treatment predated the immunotherapy standard. These limitations leave head-to-head value over standardized sTIL assessment unresolved.

6.2. Prediction and De-Escalation

Radiomics, longitudinal MRI, PET-genomics, post-treatment imaging, biology-informed models, and uncertainty-aware approaches extend pCR or residual-disease prediction [22,91–95]. Their positive predictions concern the regimen received. They do not estimate whether an anthracycline, platinum, or immunotherapy component could be removed without compromising outcome. ARTEMIS proposed identifying de-escalation candidates but acknowledged single-institution design, absent prospective validation, and pre-immunotherapy timing [59]. Disease-free-survival and metaplastic-TNBC prognostic models likewise stratify risk without identifying a safe treatment reduction [26,96].

This boundary remains relevant even when a model labels patients as potential de-escalation candidates. Candidate identification is a research hypothesis, with the associated safety threshold still to be tested. In TNBC, treatment context further matters because cohorts treated before routine immunotherapy do not directly validate predictions during contemporary immunochemotherapy. A model may retain prognostic information while requiring new validation for the treatment decision now being proposed.

6.3. On-Treatment Immune Non-Response

Longitudinal imaging gives the most mature on-treatment signal. A 3D convolutional model used baseline and cycle-four MRI in a blinded prospective TNBC test group, reaching L3 [56]. Its prospective AUC increased from 0.65 with baseline-only imaging to 0.83 with longitudinal input. This within-study comparison shows the information gained during treatment, but neither calibration nor DCA established a decision threshold. Volume-reduction models face similar limitations [59]. By the time a late-cycle signal is measured, the remaining opportunity for treatment adjustment also differs from a pretreatment prediction.

Peripheral-blood RNA-sequencing in I-SPY 2 linked early T-cell composite-score changes to TNBC response, with variable external performance attributed to batch, regimen, and overfitting effects [58]. Tryptophan metabolites, exosomal microRNAs, and myeloid-derived suppressor-cell profiles supply biological rationale rather than validated adjustment rules [97]. Molecular change is closely aligned with immune non-response, but its transportability remains fragile.

6.4. Clinical Bottom Line: TNBC

TNBC models provide response discrimination and an isolated prospective longitudinal signal. Increment beyond standardized sTIL assessment and the safety of drug-sparing

decisions remain unproven. Detecting non-response and safely changing therapy must be evaluated as a linked clinical strategy before monitoring becomes treatment tailoring.

The comparator and timing questions must ultimately be addressed together. A delta signal that improves on baseline imaging may still add little after an on-treatment sTIL, volume, or proliferation assessment. The included studies rarely test this complete decision context. Their results justify targeted evaluation of monitoring strategies rather than assuming that additional measurements already identify a safe treatment adjustment.

7. Multimodal Fusion, Subtype Conditioning, and HER2-Low

7.1. Fusion Gains and Readiness

Multimodal fusion can improve discrimination within a study without advancing clinical readiness. Of 31 multimodal studies, 17 reached L1, 13 L2, and 1 L3; none reached workflow or interventional evaluation. Integrated clinical and molecular prediction and MuFi illustrate improvement over component inputs [17,46,98]. Further multiomics, MRI, PET, and radiopathomic studies report fusion gains, frequently without external testing [99–106]. Spatial graphs and longitudinal registration extend these approaches [107,108]; externally or temporally validated pathology, ultrasound, MRI, and TNBC studies broaden the evidence [19,22,48,73,109]. Related single- and multi-source approaches are retained for comparison [110–117].

The added inputs leave familiar appraisal problems. Calibration, Brier scores, and DCA are reported in selected studies, often on internal data [19,52,66,73,98,104,109,118]. Translation of modeled net benefit into absolute numbers remains uncommon [24,98]; it still estimates potential rather than observed treatment benefit. Increasing architectural complexity adds opportunities for overfitting and makes the complete pipeline harder to audit [24,27,46,119]. The prospective MIFAPS test followed model freeze, but the gap between training and external discrimination and remaining leakage concerns constrain interpretation [27]. Its prospective design advances validation, without proving utility from acting on the prediction.

Fusion comparisons are most informative when evaluated against the best single modality and a clinically credible baseline in the same patients. That design isolates the contribution of the added information without conflating it with a change in cohort or endpoint. The examples above support examining complementary inputs, but they should not be interpreted as a universal advantage for adding modalities. External performance of the final fused model matters even when individual components have already been validated.

7.2. Subtype Conditioning and Transportability

Including a binary HER2 variable in a pooled model does not establish subtype-specific performance. Stratified reporting and luminal-restricted designs provide more relevant evidence [48,64,66,68,73,101,102,120]. Yet performance can deteriorate after stratification: one pathology model reported external AUC 0.9005 for HER2+ disease and 0.5865 for TNBC [20]. Another declined from 0.84 in HR+/HER2− disease to 0.70 in TNBC [48]. These are within-study contrasts, not cross-study rankings. They show why a pooled multimodal AUC cannot support a subtype-specific treatment threshold.

Restricted designs reduce biological mixing but may narrow sample size and generalizability [22,52,54,68,92,121]. Reuse of I-SPY, DUKE, TransNEO, and TCGA also means different publications need not represent independent validation populations [64,66,99,122,123]. The required evidence is external discrimination, calibration, and net benefit within the intended subtype. Transportability should be judged for the complete pipeline and treatment context, not inferred from the name of a public dataset or a pretrained backbone.

This distinction is particularly relevant to a model that reports a high pooled AUC while subgroup estimates are sparse. The average may conceal different response frequencies or treatment contexts. Subtype conditioning must therefore be reflected in the target and evaluation, not only in an input column or subgroup label. A dedicated model and a pooled model with subtype-specific evaluation are different designs; neither are automatically sufficient for clinical tailoring.

7.3. Federated Learning and Foundation Models

No appraised study demonstrated clinical deployment supported by federated multicenter training. Foundation models supplied pretrained pathology features within supervised pipelines, as in MuFi and MIFAPS [27,46]; pretraining does not resolve their calibration, domain-shift, or prospective-utility gaps.

7.4. HER2-Low Scoring and ADC Response

DESTINY-Breast04 established the relevance of HER2-low status to trastuzumab deruxtecan (T-DXd) in its studied setting [124]. Within the reviewed neoadjuvant literature, MRI-based HER2-zero versus HER2-low classification and AI-TIL spatial scoring show early standardization signals [39,55]. The former lacks external validation and calibration; external concordance in the latter remains imperfect. These models address reproducibility of scoring, which is distinct from predicting benefit under a particular antibody–drug conjugate (ADC).

An ultrasound-radiomics model predicted response in HER2-low patients receiving conventional neoadjuvant chemotherapy [38]. It supplies no evidence about response to T-DXd. Prediction of changing HER2 status contributes a biological-evolution concept only [125]. Thus, AI shows preliminary HER2-low scoring signals, while no included study predicts response under an actual T-DXd neoadjuvant regimen. Scoring eligibility and predicting ADC benefit must remain separate claims.

8. Evidence Quality and Clinical Readiness

Table 6 summarizes appraisal under the rules of Section 2.4, with study-level evidence in Supplementary Table S1. The coverage includes MRI [36,126–132], ultrasound [133–137], pathology [29,138–141], PET and liquid biopsy [142–144], and clinicopathologic models [145–151].

Table 6. Methodological quality and clinical readiness of the 102 tier A full-text studies, summarized by mutually exclusive subtype focus. Counts are numbers of studies. Per-study detail appears in Supplementary Table S1. Abbreviations: PROBAST, Prediction model Risk Of Bias ASsessment Tool; RoB, risk of bias; DCA, decision-curve analysis; L1–L3, clinical-readiness levels (L1 development/internal validation only; L2 temporal or geographic external validation; L3 frozen prospective observational or silent validation); TNBC, triple-negative breast cancer; HER2, human epidermal growth factor receptor 2; HR, hormone receptor.

Subtype Focus	Representative Studies	Studies (*n*)	PROBAST High RoB	Readiness L1/L2/L3	External Validation	Calibration	DCA	Clinical Comparator
Mixed/other	[25,136,152]	79	68	34/40/5	40	16	38	42
HER2+	[6,53,78]	6	5	1/5/0	4	2	2	4
TNBC	[22,55,91]	13	12	6/6/1	6	0	2	3
HR+/HER2−	[16,52,68]	4	4	2/2/0	2	1	2	2
All studies	—	102	89	43/53/6	52	19	44	51

Notes: A study is assigned to a named subtype only when its design is restricted to that subtype; TNBC-predominant mixed cohorts remain mixed/other. This rule differs from the non-exclusive subtype-cell map presented later in this section. PROBAST overall RoB: high 89, low 8, unclear 5. Readiness is an author-defined ladder, not an official scale; none reached L4 or L5. External validation, calibration, DCA, and clinical comparator columns count fully reported "yes" records; see Supplementary Table S1 for partial/unclear cases.

Readiness remained concentrated at internal validation (L1, 43 studies) and temporal or geographic external validation (L2, 53). Six studies reached frozen prospective observational validation (L3); none reached L4 or L5. Figure 4 shows this distribution. These studies demonstrate prediction under observed treatment, without evidence that exposing clinicians to a model improves routine care. Supplementary Methods S2 distinguishes the validation designs grouped within each level. Readiness does not override risk of bias: an external test can coexist with weaknesses in model development or reporting.

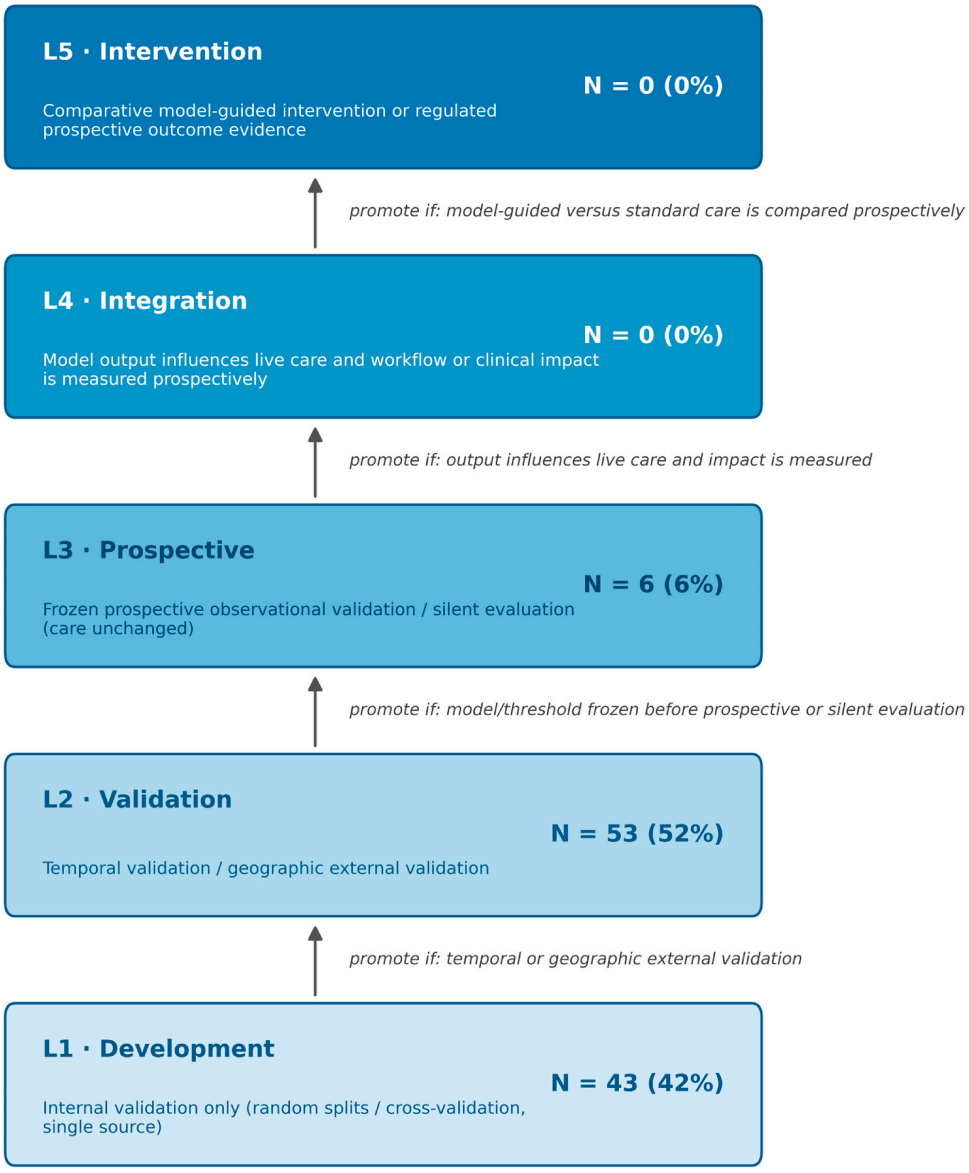


Figure 4. Author-proposed clinical-readiness ladder (L1–L5) across the 102 tier A studies, graded under the frozen L1–L5 rules of Section 2.4. All studies cluster at L1–L3 (L1 = 43, L2 = 53, L3 = 6); none reached L4 or L5.

Independent external validation was clearly reported in 52 of 102 studies (51%). Calibration was fully reported in 19 (19%) and partially in 27 more, totaling 46 (45%) with any reporting. DCA was fully reported in 44 (43%) and partially in one additional study. Figure 5 distinguishes full, partial, and cumulative reporting. Partial calibration reporting does not establish reliable probabilities. Patient-level separation was explicitly verified in

66 studies (65%); otherwise, slice- or lesion-level leakage could not be excluded from the report. Reporting flags and readiness use different criteria: the external-validation flag records a clearly reported item, whereas L2 also includes temporal separation.

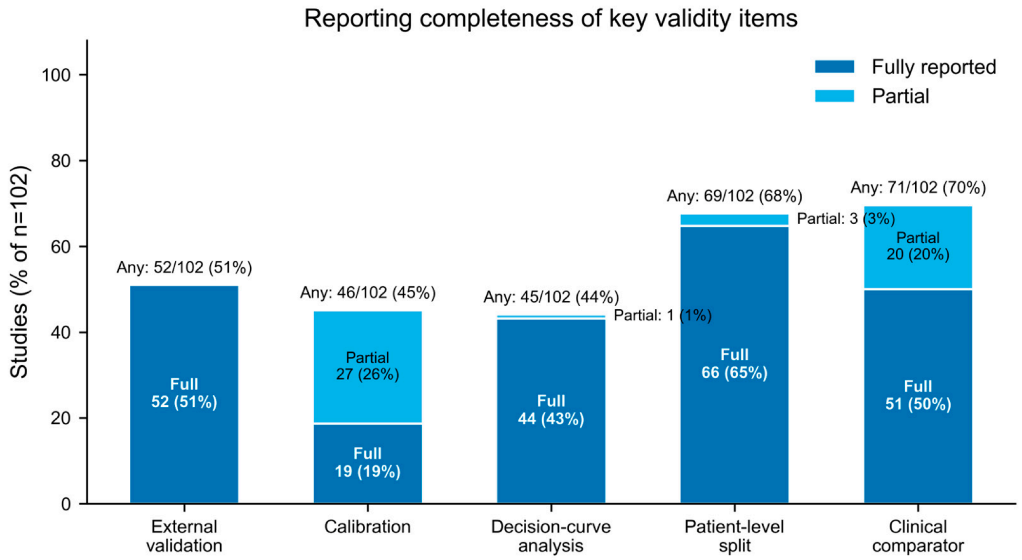


Figure 5. Reporting completeness of key validity items (n = 102). Dark segments show fully reported items, light segments show partially reported items, and labels above the bars show the cumulative number and percentage with any reporting. Full and partial categories are mutually exclusive.

PROBAST rated 89 of 102 studies (87%) at high overall risk, 8 low, and 5 unclear. High risk concentrated in the analysis domain (83 studies), compared with participants (20), outcomes (10), and predictors (3). Figure 6 presents these domain ratings. Small event counts, absent sample-size justification, feature selection outside the training split, and full-sample normalization contributed to concerns about optimism. Leakage risk was high in 13% and moderate in a further 42%. A clinical baseline or clinician comparator was clearly reported in 51 studies (50%), but comparator strength varied.

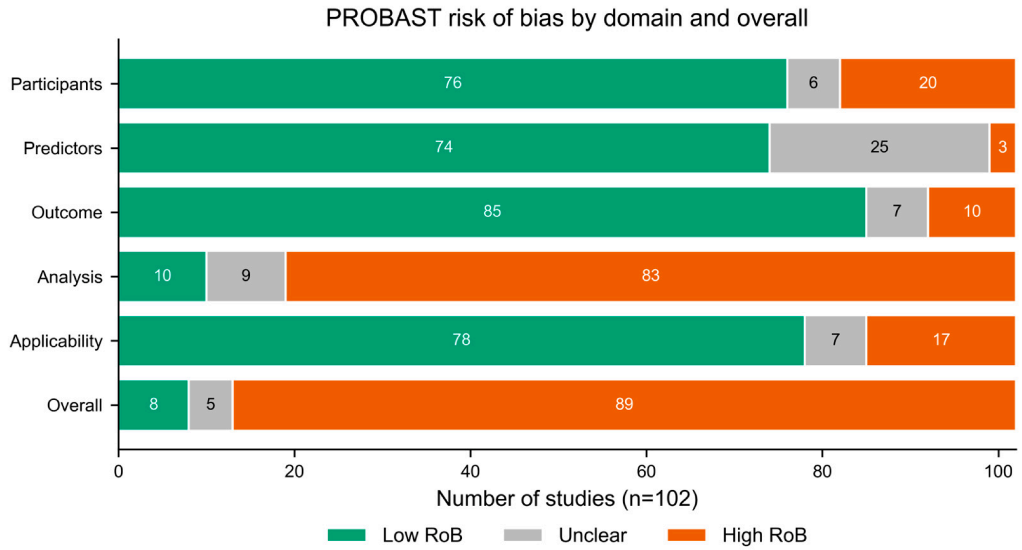


Figure 6. PROBAST risk of bias by domain and overall (n = 102). The analysis domain and overall rating are dominated by high risk of bias.

Architecture does not remove these limitations. Figure 7 shows that every modality remained within the lower readiness levels. Limited access to code, trained models, data,

and preregistration also constrained reproducibility. Subtype-level calibration and decision thresholds are especially consequential when a model is proposed for treatment omission, because ranking patients cannot quantify the risk accepted by that decision.

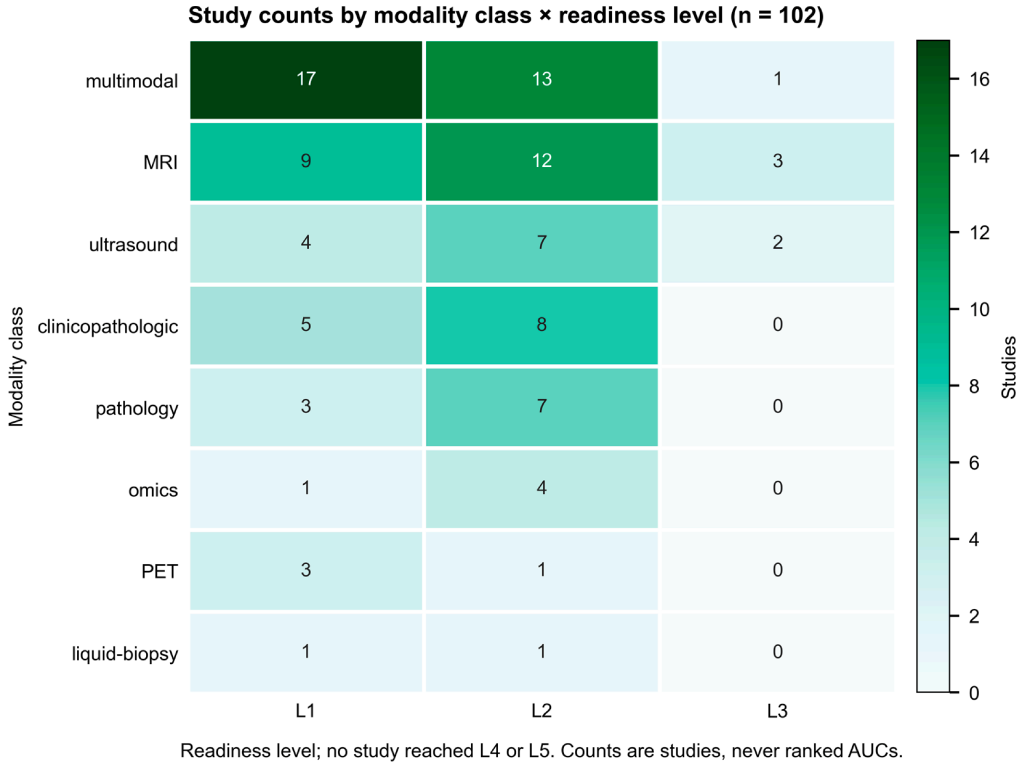


Figure 7. Study counts by modality class and readiness level (n = 102). All modalities cluster at L1–L3; none reached L4 or L5.

Table 7 maps the resulting gaps by subtype and decision timepoint. Its cells are non-exclusive: mixed-cohort studies contribute only where relevant stratified evidence is available. Cell-level maturity summarizes the decision-specific evidence and should not be read as a new per-study grading. The concentration on retrospective response prediction and sparse on-treatment evidence explains why methodological diversification has not yet supplied treatment-selection evidence.

Table 7. Evidence map of subtype by decision timepoint across the 102 tier A studies. Each cell gives the number of studies informing that subtype and timepoint (n), the number with subtype-stratified results (stratified), and the imaging/data modalities represented. Abbreviations: HR, hormone receptor; HER2, human epidermal growth factor receptor 2; TNBC, triple-negative breast cancer; MRI, magnetic resonance imaging; PET, positron-emission tomography.

Subtype	Pretreatment	On-Treatment	Post-Treatment
HR+/HER2−	n = 48 (stratified 45); Modalities: MRI, PET, clinicopathologic, liquid-biopsy, multimodal, omics, pathology, ultrasound	n = 11 (stratified 11); Modalities: MRI, liquid-biopsy, multimodal, ultrasound	n = 22 (stratified 19); Modalities: MRI, clinicopathologic, liquid-biopsy, multimodal, omics, pathology, ultrasound
HER2+	n = 48 (stratified 42); Modalities: MRI, PET, clinicopathologic, multimodal, omics, pathology, ultrasound	n = 11 (stratified 9); Modalities: MRI, PET, multimodal, ultrasound	n = 18 (stratified 17); Modalities: MRI, clinicopathologic, multimodal, omics, pathology, ultrasound
TNBC	n = 56 (stratified 45); Modalities: MRI, PET, clinicopathologic, liquid-biopsy, multimodal, omics, pathology, ultrasound	n = 11 (stratified 10); Modalities: MRI, liquid-biopsy, multimodal, ultrasound	n = 22 (stratified 19); Modalities: MRI, clinicopathologic, liquid-biopsy, multimodal, omics, pathology, ultrasound

Note: A study informing more than one arm or timepoint is counted in each relevant cell, so cell counts sum to more than 102; this map is deliberately non-exclusive and must not be read as the mutually exclusive corpus distribution of Section 2.2.

Several quality dimensions can fail independently. Leakage inflates estimated performance; poor calibration undermines probabilities; weak comparators obscure incremental value; and absent workflow testing leaves consequences of use unknown. A study can address one of these problems without resolving the others. Reporting the dimensions separately prevents an externally validated label or a favorable AUC from standing in for a complete clinical assessment.

The practical reading of this appraisal is therefore decision-specific. A model proposed for risk stratification should establish reliable risk estimates in the intended population. A model proposed for treatment omission must additionally address the outcome under omission and the harm of a wrong decision. The latter claim cannot be inferred from the former, regardless of the model family. The following section brings these requirements together as the review's clinically actionable conclusions.

9. From Prediction to Tailoring: What AI Can, Cannot, and Must Still Demonstrate

9.1. Where the Evidence Sits

The actionability ladder concerns the clinical claim, whereas the readiness ladder concerns validation and translation. Its first rung is predicting pCR or RCB under a fixed regimen. The second compares treatment strategies; the third estimates individualized treatment effects, including outcomes under alternatives. The fourth tests prospectively whether model-guided decisions improve outcomes. These distinctions prevent a prospective prediction cohort from being mistaken for a prospective treatment-selection trial.

The appraised response models remain on the first rung [6,24,25,27,28,35,56,58,152–154]. The prior appraisal found no study performing strategy comparison or ITE estimation. Cross-regimen validation of an I-SPY 2 peripheral-immune model tested prediction across settings rather than selection between treatments [58]. HER2-LADDER's risk groups and retrospective trial-slide validation similarly do not provide a randomized strategy contrast [6]. Naming reduced, standard, or alternative regimens for predicted risk groups does not demonstrate their relative benefit for an individual.

Axillary prediction supplies the closest link to a changeable decision. FAIS-DL, dual-modality ultrasound, and prospective nodal cohorts report simulated reductions in unnecessary axillary lymph node dissection or improved false-negative performance [28,35,152]. These estimates concern surgical extent after fixed neoadjuvant treatment. They warrant prospective evaluation of surgical decisions, without establishing a basis for changing systemic drugs. Prospective pCR or nodal validation elsewhere retains the same distinction between observing predictions and testing model-guided care [25,27,56].

Thus, increasingly rigorous validation can strengthen a fixed-regimen predictor while leaving the treatment-selection question open. A clinically actionable conclusion should specify the decision that the evidence supports and the boundary it has not crossed. For this review, the practical boundary is between identifying likely response and demonstrating the consequences of acting differently because of that prediction.

9.2. What Models Can and Cannot Support

Current models can rank patients by the likelihood of pCR or residual burden under a specified regimen, sometimes with geographic external validation [23,73,78]. They can estimate axillary response for surgical research and refine response estimates when longitudinal imaging is available [28,35,56,152]. They also provide early assistance with inconsistently reproduced reads, including HER2-low and TIL assessment [55,85]. These are useful predictive or measurement tasks within the populations and conditions evaluated.

Their usefulness should be judged against the existing clinical or pathology comparator for that task.

Current evidence cannot justify systemic drug omission, switching, escalation, or de-escalation on an AI output. Uncalibrated probabilities leave the risks at a proposed decision threshold uncertain. More fundamentally, fixed-regimen training cannot establish the outcome under an alternative regimen. Transportability across subtypes is incomplete, and the reviewed models lack measured live-workflow or interventional impact. Even a well-calibrated prediction needs additional causal and prospective evidence before it can support a treatment change.

The immediate clinical implication is to distinguish an investigational response estimate from a validated treatment rule. A research model can help define whom to study, which response features to measure, and which decision deserves prospective testing. It should not be described as established evidence for reducing or switching a drug solely because the predicted pCR probability is high. The proposed rule needs its own comparator, threshold, and outcome evaluation.

9.3. Evidence Required for Treatment Tailoring

First, validate the intended decision setting. Independent multicenter and multivendor evaluations should report discrimination, calibration curves and slopes, and DCA at clinically defensible thresholds within each subtype. Comparators should include the clinical, genomic, or pathology tools already used for the decision. Code, models, and compliantly shared data should permit appraisal of the entire pipeline under TRIPOD+AI, PROBAST, CLAIM, and CLEAR. Better reporting cannot remove bias, but it makes the proposed clinical use assessable.

Second, make on-treatment signals testable. Preregister longitudinal cohorts with fixed measurement times and harmonized delta-feature definitions. Evaluate whether change reflects biology rather than scanner or timepoint shift, and whether the information arrives early enough for the proposed adjustment. Early endocrine futility, HER2+ de-escalation, and TNBC immune non-response require distinct targets. Predicting eventual pCR should not substitute for evaluating the safety of the associated stop-or-switch rule.

Third, estimate outcomes under alternative treatments. Completed randomized neoadjuvant arms offer a source for strategy comparison and treatment-effect heterogeneity analysis. Candidate ITE models require validation of the treatment-effect estimates, not merely validation of outcome discrimination in each arm. HER2-low and ADC regimens are potential prospective testbeds, provided the outcome is response or benefit under the actual regimen being considered.

Fourth, evaluate the consequences of use. Following frozen prospective observational validation, live-workflow studies should measure how model output changes decisions and affects process, safety, or patient outcomes. Comparative interventional evaluation must then test model-guided against standard care, or appropriate regulated deployment must provide prospective outcome evidence. Retrospective reader comparisons and modeled net benefit cannot substitute for these steps. Federated learning and foundation models may help address data scarcity or domain shift, but their contribution must be demonstrated within this evaluation sequence.

These requirements are complementary. External calibration makes a score interpretable at a threshold; strategy comparison addresses the alternative treatment; workflow and intervention studies test what happens when the rule is used. Progress on one requirement does not eliminate the others. Designing studies around the intended treatment decision makes these dependencies explicit and keeps clinical evaluation aligned with the claim.

10. Ethics, Implementation, and Limitations

Clinical implementation requires fairness, interpretable outputs, regulatory assessment, and fit with existing care pathways [155]. Supportive care, including the scalp-cooling study, addresses tolerability rather than AI-based tumor response or systemic-treatment selection [156]. These patient-centered concerns should remain distinct from the decision a response model claims to inform.

Only 41 of 102 studies (40%) reported any subgroup performance. Aggregate AUC therefore cannot establish fair performance across age, race, ethnicity, scanner, institution, or subtype. Relevant populations need both discrimination and calibration assessed separately. Deployment also requires evaluation under applicable medical-device and change-control frameworks; no regulatory clearance was reported for the surveyed models. The absence of measured workflow impact prevents assuming that an accurate research score will improve care when presented to clinicians.

Interpretability remains largely post hoc, through saliency maps or feature attributions. Such displays can aid interrogation but do not establish causal correctness or safe treatment selection. Workflow evaluations should examine how users understand and act on an output, alongside its predictive accuracy. Standalone reader-performance gains leave these implementation questions unanswered.

The review's own limitations constrain its conclusions. Its purposive, English-language narrative search and tier A full-text access favor particular venues and regions. All proportions describe the appraised set, not unbiased field-wide prevalence. Tier B reports were cited from abstracts and excluded from readiness grading. Screening, extraction, and appraisal followed one structured protocol with automated consistency checks and author adjudication, rather than duplicate independent assessment. Traceable records reduce but cannot eliminate misclassification. The readiness ladder is author-defined and has not been externally validated; Supplementary Methods S2 preserves its operational rules.

Most underlying studies are retrospective proof-of-concept reports, often with overlapping public data and optimism in performance estimation. The appraisal identifies limits of the included evidence, without asserting that no relevant unpublished or unselected study exists. Accordingly, the absence of workflow or interventional evidence applies to this appraised set. These boundaries remain essential when interpreting the review's clinically focused conclusions.

11. Conclusions

Subtype-specific AI currently predicts response under observed neoadjuvant regimens more convincingly than it guides treatment selection. HR+/HER2− disease requires targets aligned with endocrine sensitivity and residual risk. HER2+ disease and TNBC offer stronger pCR-related signals, but a predicted response under treatment does not establish safety after reducing that treatment. Surgical-response models address a related decision and should be evaluated on their own terms.

The principal bottleneck is the evidence surrounding a prediction. Discrimination alone cannot establish reliable probabilities, transportability to the intended subtype, or benefit from acting at a particular threshold. Even prospective prediction validation leaves the counterfactual question unresolved when patients continue to receive the same care. Current findings therefore support further development and decision-specific validation, without supporting routine AI-guided changes to systemic drugs.

Progress requires independently validated and calibrated subtype-specific predictions, clinically timed longitudinal measurements, explicit strategy comparisons, and prospective testing of model-guided care. The treatment decision and acceptable risk should determine the model's target, comparator, and evaluation design. This path preserves the value

of response prediction while specifying what must be demonstrated before it can safely inform an individual patient's treatment.

Supplementary Materials: The following supporting information can be downloaded at https://www.mdpi.com/article/10.3390/cancers18182947/s1. Table S1: Characteristics, per-study AI metrics, and PROBAST-based methodological appraisal of the 102 tier A full-text studies (spreadsheet); Table S2: Full breakdown of verbatim pCR-definition categories across the 102 tier A full-text studies; Supplementary Methods S1: Study identification and coverage-directed selection; Supplementary Methods S2: Full readiness rules and prior re-audit; Supplementary Note S3: Model-backbone census and parameterization.

Author Contributions: Conceptualization, J.Z. and L.C.; methodology, J.Z.; validation, J.Z., B.X. and L.C.; formal analysis, J.Z.; investigation (literature identification, data extraction, and quality/readiness appraisal), J.Z.; resources, B.X. and L.C.; data curation, J.Z.; writing—original draft preparation, J.Z.; writing—review and editing, J.Z., K.Y. and L.C.; visualization, J.Z.; supervision, L.C.; project administration, L.C. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Fujian Provincial Natural Science Foundation General Program, grant number 2025J01103. The APC was funded by the corresponding author.

Institutional Review Board Statement: Not applicable. This study is a narrative review of previously published literature and involved no new studies on humans or animals.

Informed Consent Statement: Not applicable. This review did not involve any new collection of patient data; all data discussed derive from previously published, already-consented studies.

Data Availability Statement: No new patient-level data were generated in this review. The study-level extraction and appraisal data are available in Supplementary Tables S1 and S2.

Acknowledgments: During the preparation of this manuscript, the authors used a large language model to correct English grammar and syntax errors and to improve the structure of the text. The authors have reviewed and edited the output and take full responsibility for the content of this publication.

Conflicts of Interest: The authors declare no conflicts of interest. The funders had no role in the design of the study; in the collection, analyses or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

Abbreviations

The following abbreviations are used in this manuscript:

ADC	antibody–drug conjugate
AI	artificial intelligence
ALND	axillary lymph node dissection
AUC	area under the receiver operating characteristic curve
CLAIM	Checklist for Artificial Intelligence in Medical Imaging
CLEAR	CheckList for EvaluAtion of Radiomics research
CNN	convolutional neural network
ctDNA	circulating tumor DNA
DCA	decision-curve analysis
DCE-MRI	dynamic contrast-enhanced magnetic resonance imaging
DCIS	ductal carcinoma in situ
DFS/iDFS	disease-free/invasive disease-free survival
EFS	event-free survival
EPV	events per variable
FTV	functional tumor volume

H&E	hematoxylin and eosin
HER2	human epidermal growth factor receptor 2
HR	hormone receptor
ICI	immune-checkpoint inhibitor
IHC	immunohistochemistry
ITE	individualized treatment effect
MIL	multiple-instance learning
ML	machine learning
MRI	magnetic resonance imaging
NAC	neoadjuvant chemotherapy
OS	overall survival
pCR	pathologic complete response
PET	positron-emission tomography
PPV	positive predictive value
PROBAST	Prediction model Risk Of Bias ASsessment Tool
RCB	residual cancer burden
RECIST	Response Evaluation Criteria in Solid Tumors
RFS	recurrence-free survival
SaMD	software as a medical device
SLNB	sentinel lymph node biopsy
sTIL/TIL	stromal/tumor-infiltrating lymphocytes
SVM	support vector machine
T-DM1	trastuzumab emtansine
T-DXd	trastuzumab deruxtecan
TNBC	triple-negative breast cancer
TRIPOD+AI	Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis, AI extension
WSI	whole-slide image

Appendix A. Glossary of Key Terms

Table A1 gives plain-language definitions of the technical terms used in this review, for readers less familiar with prediction-model methodology.

Table A1. Glossary of key terms used in this review.

Term	Plain-Language Definition
Discrimination	A model's ability to rank patients, so that responders generally receive higher scores than non-responders; summarized by the AUC.
AUC (area under the ROC curve)	The probability that a randomly chosen responder receives a higher model score than a randomly chosen non-responder; 0.5 is chance, 1.0 is perfect ranking.
Calibration	Agreement between predicted and observed risk: when a calibrated model predicts a 30% chance of response, close to 30 of 100 such patients respond.
Decision-curve analysis (DCA)/net benefit	An analysis estimating potential net benefit at specified risk thresholds by comparing the model with treating all or no patients, under stated assumptions; it does not demonstrate prospective clinical benefit.
Clinical utility	Whether acting on a model's output at a chosen risk threshold does more good than harm; assessed, but not proven, by decision-curve analysis.
Internal validation	Testing a model on data from its own development setting, such as cross-validation or a random split within one center.
Temporal validation	Testing on patients from the same center treated in a later period than the development cohort.
Geographic external validation	Testing on patients from centers that took no part in model development.
Prospective (observational) validation	Scoring a registered or prospectively enrolled cohort after the model is fixed, without the model altering care.
Individualized treatment effect (ITE)	The difference a given treatment would make for one particular patient, compared with the alternative.

Table A1. *Cont.*

Term	Plain-Language Definition
Counterfactual evidence	Evidence about what would have happened under the treatment the patient did not receive; required for any claim that a model can select between treatments.
Strategy comparison	Evaluating a model's predictions across two or more regimens, rather than under a single fixed regimen.
Multimodal fusion	Combining more than one data type, such as imaging with pathology or clinical variables, in a single model.
Radiomics	Extraction of large numbers of quantitative features from medical images for use in statistical or machine-learning models.
Deep learning	Machine learning built on multi-layer neural networks that learn features directly from raw data such as images, rather than from hand-engineered variables.
pCR (pathologic complete response)	No residual invasive cancer in the breast and axillary nodes at surgery; most commonly defined as ypT0/is ypN0.
RCB (residual cancer burden)	A validated continuous index grading residual disease from RCB-0 to RCB-III; finer-grained than the binary pCR label.
sTIL (stromal tumor-infiltrating lymphocytes)	A pathologist-readable immune marker in TNBC, prognostic and predictive, assessable on routine slides with relatively low incremental cost when tissue and pathology infrastructure are already available.
Clinical-readiness ladder (L1–L5)	The author-defined validation and translational-maturity scale used in this review: internal only (L1), temporal/geographic external validation (L2), frozen prospective observational or silent validation (L3), live workflow influence with measured impact (L4), and comparative model-guided intervention or regulated prospective outcome evidence (L5); defined in Section 2.4.
PROBAST	The Prediction model Risk Of Bias ASsessment Tool, which grades risk of bias in prediction-model studies across four domains.
TRIPOD+AI	The reporting standard for prediction-model studies involving machine learning.
Data leakage	Any construction that lets information from the evaluation data influence model development, such as feature selection performed outside the train–test split; it inflates reported performance.
Treatment tailoring	Changing an individual patient's systemic therapy (omitting, switching, escalating, or de-escalating a drug) on the basis of a model output.

References

1. von Minckwitz, G.; Huang, C.S.; Mano, M.S.; Loibl, S.; Mamounas, E.P.; Untch, M.; Wolmark, N.; Rastogi, P.; Schneeweiss, A.; Redondo, A.; et al. Trastuzumab Emtansine for Residual Invasive HER2-Positive Breast Cancer (KATHERINE). *N. Engl. J. Med.* **2019**, *380*, 617–628. [CrossRef]
2. Schmid, P.; Cortes, J.; Dent, R.; Pusztai, L.; McArthur, H.; Kümmel, S.; Bergh, J.; Denkert, C.; Park, Y.H.; Hui, R.; et al. Event-free Survival with Pembrolizumab in Early Triple-Negative Breast Cancer (KEYNOTE-522). *N. Engl. J. Med.* **2022**, *386*, 556–567. [CrossRef]
3. Cortazar, P.; Zhang, L.; Untch, M.; Mehta, K.; Costantino, J.P.; Wolmark, N.; Bonnefoi, H.; Cameron, D.; Gianni, L.; Valagussa, P.; et al. Pathological complete response and long-term clinical benefit in breast cancer: The CTNeoBC pooled analysis. *Lancet* **2014**, *384*, 164–172. [CrossRef]
4. Janse, M.H.A.; Janssen, L.M.; Wolters-van der Ben, E.J.M.; Moman, M.R.; Viergever, M.A.; van Diest, P.J.; Gilhuijs, K.G.A. Deep learning-based radiomics does not improve residual cancer burden prediction post-chemotherapy in LIMA breast MRI trial. *Eur. Radiol.* **2026**, *36*, 850–862. [CrossRef]
5. Shamai, G.; Cohen, S.; Binenbaum, Y.; Sabo, E.; Cretu, A.; Mayer, C.; Barshack, I.; Goldman, T.; Bar-Sela, G.; Polónia, A.; et al. Deep learning on histopathological images to predict breast cancer recurrence risk and chemotherapy benefit: A multicentre, model development and validation study. *Lancet Oncol.* **2026**, *27*, 512–526. [CrossRef]
6. Wu, X.R.; Lv, H.; Zhao, S.; Zeng, X.H.; Dai, L.J.; Xu, Y.Z.; Li, Y.W.; Qiu, Z.Y.; Huang, J.T.; Zhang, N.N.; et al. Spatially interpretable artificial intelligence framework to tailored neoadjuvant dual HER2 blockade in HER2-positive breast cancer. *Signal Transduct. Target. Ther.* **2026**, *11*, 241. [CrossRef]
7. Wolff, R.F.; Moons, K.G.M.; Riley, R.D.; Whiting, P.F.; Westwood, M.; Collins, G.S.; Reitsma, J.B.; Kleijnen, J.; Mallett, S.; PROBAST Group. PROBAST: A Tool to Assess the Risk of Bias and Applicability of Prediction Model Studies. *Ann. Intern. Med.* **2019**, *170*, 51–58. [CrossRef]
8. Moons, K.G.M.; Damen, J.A.A.; Kaul, T.; Hooft, L.; Andaur Navarro, C.; Dhiman, P.; Beam, A.L.; Van Calster, B.; Celi, L.A.; Denaxas, S.; et al. PROBAST+AI: An updated quality, risk of bias, and applicability assessment tool for prediction models using regression or artificial intelligence methods. *BMJ* **2025**, *388*, e082505. [CrossRef]

9. Collins, G.S.; Moons, K.G.M.; Dhiman, P.; Riley, R.D.; Beam, A.L.; Van Calster, B.; Ghassemi, M.; Liu, X.; Reitsma, J.B.; van Smeden, M.; et al. TRIPOD+AI statement: Updated guidance for reporting clinical prediction models that use regression or machine learning methods. *BMJ* **2024**, *385*, e078378. [CrossRef]
10. Tejani, A.S.; Klontzas, M.E.; Gatti, A.A.; Mongan, J.T.; Moy, L.; Park, S.H.; Kahn, C.E., Jr. CLAIM 2024 Update Panel. Checklist for Artificial Intelligence in Medical Imaging (CLAIM): 2024 Update. *Radiol. Artif. Intell.* **2024**, *6*, e240300. [CrossRef]
11. Kocak, B.; Baessler, B.; Bakas, S.; Cuocolo, R.; Fedorov, A.; Maier-Hein, L.; Mercaldo, N.; Müller, H.; Orlhac, F.; Pinto Dos Santos, D.; et al. CheckList for EvaluAtion of Radiomics research (CLEAR): A step-by-step reporting guideline for authors and reviewers endorsed by ESR and EuSoMII. *Insights Imaging* **2023**, *14*, 75. [CrossRef]
12. Vasey, B.; Nagendran, M.; Campbell, B.; Clifton, D.A.; Collins, G.S.; Denaxas, S.; Denniston, A.K.; Faes, L.; Geerts, B.; Ibrahim, M.; et al. Reporting Guideline for the Early-Stage Clinical Evaluation of Decision Support Systems Driven by Artificial Intelligence: DECIDE-AI. *Nat. Med.* **2022**, *28*, 924–933. [CrossRef]
13. Liu, Y.; Wang, Y.; Huang, J.; Pei, S.; Wang, Y.; Cui, Y.; Yan, L.; Yao, M.; Wang, Y.; Zhu, Z.; et al. Deep learning-based prediction of axillary pathological complete response in patients with breast cancer using longitudinal multiregional ultrasound. *EBioMedicine* **2025**, *119*, 105896. [CrossRef]
14. Yu, Y.; Ren, W.; He, Z.; Chen, Y.; Tan, Y.; Mao, L.; Ouyang, W.; Lu, N.; Ouyang, J.; Chen, K.; et al. Machine learning radiomics of magnetic resonance imaging predicts recurrence-free survival after surgery and correlation of LncRNAs in patients with breast cancer: A multicenter cohort study. *Breast Cancer Res.* **2023**, *25*, 132. [CrossRef]
15. Wang, X.; Xie, T.; Luo, J.; Zhou, Z.; Yu, X.; Guo, X. Radiomics predicts the prognosis of patients with locally advanced breast cancer by reflecting the heterogeneity of tumor cells and the tumor microenvironment. *Breast Cancer Res.* **2022**, *24*, 20. [CrossRef]
16. Iwamoto, N.; Horiguchi, S.I.; Kuwayama, T.; Toi, M. Low-progesterone receptor expression predicts nodal pathological complete response after neoadjuvant chemotherapy in premenopausal patients with ER-positive/HER2-negative breast cancer. *Breast Cancer* **2026**, *33*, 744–751. [CrossRef]
17. Sammut, S.J.; Crispin-Ortuzar, M.; Chin, S.F.; Provenzano, E.; Bardwell, H.A.; Ma, W.; Cope, W.; Dariush, A.; Dawson, S.J.; Abraham, J.E.; et al. Multi-omic machine learning predictor of breast cancer therapy response. *Nature* **2022**, *601*, 623–629. [CrossRef]
18. He, C.; Yu, T.; Yang, L.; He, L.; Zhu, J.; Chen, J. Clinical prediction of pathological complete response in breast cancer: A machine learning study. *BMC Cancer* **2025**, *25*, 933. [CrossRef]
19. Zhang, J.; Wu, Q.; Yin, W.; Yang, L.; Xiao, B.; Wang, J.; Yao, X. Development and validation of a radiopathomic model for predicting pathologic complete response to neoadjuvant chemotherapy in breast cancer patients. *BMC Cancer* **2023**, *23*, 431. [CrossRef]
20. Huang, Z.; Shao, W.; Han, Z.; Alkashash, A.M.; De la Sancha, C.; Parwani, A.V.; Nitta, H.; Hou, Y.; Wang, T.; Salama, P.; et al. Artificial intelligence reveals features associated with breast cancer neoadjuvant chemotherapy responses from multi-stain histopathologic images. *npj Precis. Oncol.* **2023**, *7*, 14. [CrossRef]
21. Omar, M.; Nuzzo, P.V.; Ravera, F.; Bleve, S.; Fanelli, G.N.; Zanettini, C.; Valencia, I.; Marchionni, L. Notch-based gene signature for predicting the response to neoadjuvant chemotherapy in triple-negative breast cancer. *J. Transl. Med.* **2023**, *21*, 811. [CrossRef]
22. Zhou, J.; Bai, Y.; Zhang, Y.; Wang, Z.; Sun, S.; Lin, L.; Gu, Y.; You, C. A preoperative radiogenomic model based on quantitative heterogeneity for predicting outcomes in triple-negative breast cancer patients who underwent neoadjuvant chemotherapy. *Cancer Imaging* **2024**, *24*, 98. [CrossRef]
23. Huang, Y.H.; Shi, Z.Y.; Zhu, T.; Zhou, T.H.; Li, Y.; Li, W.; Qiu, H.; Wang, S.Q.; He, L.F.; Wu, Z.Y.; et al. Longitudinal MRI-Driven Multi-Modality Approach for Predicting Pathological Complete Response and B Cell Infiltration in Breast Cancer. *Adv. Sci.* **2025**, *12*, e2413702. [CrossRef]
24. Gao, Y.; Ventura-Diaz, S.; Wang, X.; He, M.; Xu, Z.; Weir, A.; Zhou, H.Y.; Zhang, T.; van Duijnhoven, F.H.; Han, L.; et al. An explainable longitudinal multi-modal fusion model for predicting neoadjuvant therapy response in women with breast cancer. *Nat. Commun.* **2024**, *15*, 9613. [CrossRef]
25. Zhu, T.; Huang, Y.H.; Li, W.; Wu, C.G.; Zhang, Y.M.; Zheng, X.X.; Zhang, T.F.; Lin, Y.Y.; Liu, Z.Y.; Ye, G.L.; et al. A non-invasive artificial intelligence model for identifying axillary pathological complete response to neoadjuvant chemotherapy in breast cancer: A secondary analysis to multicenter clinical trial. *Br. J. Cancer* **2024**, *131*, 692–701. [CrossRef]
26. Ni, P.; Wang, Y.; Bai, X.; Yang, Z.; Wu, T.; Gao, C.; Cheng, Y.; Niu, L.; Yan, Y.; Zhou, Y.; et al. TNMpBC-NeoBCSS model: A breast cancer specific survival prediction model for triple-negative metaplastic breast carcinoma patients with neoadjuvant therapy. *Sci. Rep.* **2025**, *15*, 8351. [CrossRef]
27. Mao, N.; Dai, Y.; Zhou, H.; Lin, F.; Zheng, T.; Li, Z.; Yang, P.; Zhao, F.; Li, Q.; Wang, W.; et al. A multimodal and fully automated system for prediction of pathological complete response to neoadjuvant chemotherapy in breast cancer. *Sci. Adv.* **2025**, *11*, eadr1576. [CrossRef]

28. Chen, M.; Hong, T.; Wang, Y.; Li, S.; Zeng, B.; Chen, C.; Zhang, J.; Guo, W.; Chen, L.; Lin, Y.; et al. Effect of a machine learning prediction model on the false-negative rate of sentinel lymph node biopsy for clinically node-positive breast cancer after neoadjuvant chemotherapy. *Breast* **2025**, *83*, 104543. [CrossRef]
29. Zhou, Y.; Shu, X.; Wang, F.; Xu, H.; Tang, H.Q.; Fang, H.; Huang, J.; Wang, Y.W.; Ji, H.L.; Zhang, S.W.; et al. Prediction of neoadjuvant therapy response in breast cancer based on interpretable artificial intelligence. *Int. J. Surg.* **2026**, *112*, 1066–1080. [CrossRef]
30. Hou, W.T.; Pu, Z.F.; Xu, Z.Y.; Wu, A.H.; Liu, Z.H.; Zhao, K.; Duan, C.L.; Guo, J.; Chen, K.; Qiu, S.Q.; et al. Predicting neoadjuvant therapy response in breast cancer from preoperative biopsy via spatial-semantic-differential learning and interpretable clinicopathological-guided fusion. *Med. Image Anal.* **2026**, *111*, 104069. [CrossRef]
31. Azimzade, Y.; Haugen, M.H.; Kristensen, V.N.; Frigessi, A.; Köhn-Luque, A. Explainable machine learning-guided integrated multiomics analysis reveals macrophage-driven immune suppression in breast cancer. *Nat. Commun.* **2026**, *17*, 7193. [CrossRef]
32. Cantore, T.; Hoang, D.T.; Pal, L.R.; Stemmer, A.; Chang, T.G.; Dhruba, S.R.; Shulman, E.D.; Campagnolo, E.; Lee, J.S.; Levy, J.; et al. Predicting neoadjuvant breast cancer therapy response using BRIDGE from tumor transcriptomics and histopathology. *Ann. Oncol.* **2026**, *37*, 1293–1307. [CrossRef]
33. Shulman, E.D.; Campagnolo, E.M.; Lodha, R.; Chung, Y.; Stemmer, A.; Cantore, T.; Ru, B.; Chang, T.G.; Biswas, S.; Dhruba, S.R.; et al. AI-predicted spatial transcriptomics unlocks breast cancer biomarkers from pathology. *Cell* **2026**, *189*, 4225–4240.e25. [CrossRef]
34. Fürböck, C.; Janickova, I.; Langs, G.; Helbich, T.H.; Clauser, P.; Varga, R.; Baltzer, P.; Kapetas, P. Deep-Learning with Domain-Specific Pretraining for Breast Cancer Neoadjuvant Chemotherapy Response Prediction from Pre-Treatment B-Mode Ultrasound. *Cancers* **2026**, *18*, 1345. [CrossRef]
35. Li, Z.; Gao, J.; Zhou, H.; Li, X.; Zheng, T.; Lin, F.; Wang, X.; Chu, T.; Wang, Q.; Wang, S.; et al. Multiregional dynamic contrast-enhanced MRI-based integrated system for predicting pathological complete response of axillary lymph node to neoadjuvant chemotherapy in breast cancer: Multicentre study. *EBioMedicine* **2024**, *107*, 105311. [CrossRef]
36. Chen, X.; Luo, Y.; Xie, Z.; Wen, Y.; Mou, F.; Zeng, W. Prediction of neoadjuvant chemotherapy efficacy in breast cancer: Integrating multimodal imaging and clinical features. *BMC Med. Imaging* **2025**, *25*, 118. [CrossRef]
37. Li, W.; Huang, Y.H.; Zhu, T.; Zhang, Y.M.; Zheng, X.X.; Zhang, T.F.; Lin, Y.Y.; Wu, Z.Y.; Liu, Z.Y.; Lin, Y.; et al. Noninvasive Artificial Intelligence System for Early Predicting Residual Cancer Burden During Neoadjuvant Chemotherapy in Breast Cancer. *Ann. Surg.* **2025**, *281*, 645–654. [CrossRef]
38. Peng, Q.; Ji, Z.; Xu, N.; Dong, Z.; Zhang, T.; Ding, M.; Qu, L.; Liu, Y.; Xie, J.; Jin, F.; et al. Prediction of neoadjuvant chemotherapy efficacy in patients with HER2-low breast cancer based on ultrasound radiomics. *Cancer Imaging* **2025**, *25*, 112. [CrossRef]
39. Huang, X.; Wu, L.; Liu, Y.; Xu, Z.; Liu, C.; Liu, Z.; Liang, C. Development and validation of machine learning models for predicting HER2-zero and HER2-low breast cancers. *Br. J. Radiol.* **2024**, *97*, 1568–1576. [CrossRef]
40. Johnston, S.R.D.; Harbeck, N.; Hegg, R.; Toi, M.; Martin, M.; Shao, Z.M.; Zhang, Q.Y.; Martinez Rodriguez, J.L.; Campone, M.; Hamilton, E.; et al. Abemaciclib Combined With Endocrine Therapy for the Adjuvant Treatment of HR+, HER2-, Node-Positive, High-Risk, Early Breast Cancer (monarchE). *J. Clin. Oncol.* **2020**, *38*, 3987–3998. [CrossRef]
41. Slamon, D.; Lipatov, O.; Nowecki, Z.; McAndrew, N.; Kukielka-Budny, B.; Stroyakovskiy, D.; Yardley, D.A.; Huang, C.S.; Fasching, P.A.; Crown, J.; et al. Ribociclib plus Endocrine Therapy in Early Breast Cancer (NATALEE). *N. Engl. J. Med.* **2024**, *390*, 1080–1091. [CrossRef]
42. Smith, I.; Robertson, J.; Kilburn, L.; Wilcox, M.; Evans, A.; Holcombe, C.; Horgan, K.; Kirwan, C.; Mallon, E.; Sibbering, M.; et al. Long-term outcome and prognostic value of Ki67 after perioperative endocrine therapy in postmenopausal women with hormone-sensitive early breast cancer (POETIC): An open-label, multicentre, parallel-group, randomised, phase 3 trial. *Lancet Oncol.* **2020**, *21*, 1443–1454. [CrossRef]
43. Ellis, M.J.; Tao, Y.; Luo, J.; A'Hern, R.; Evans, D.B.; Bhatnagar, A.S.; Chaudri Ross, H.A.; von Kameke, A.; Miller, W.R.; Smith, I.; et al. Outcome prediction for estrogen receptor-positive breast cancer based on postneoadjuvant endocrine therapy tumor characteristics (PEPI). *J. Natl. Cancer Inst.* **2008**, *100*, 1380–1388. [CrossRef]
44. Wang, J.; Yang, B.; Bian, L.; Zhang, H.; Peng, C.; Zou, Y.; Tang, H. Long-term Outcomes for Neoadjuvant Versus Adjuvant Chemotherapy in ER+/HER2- Locally Advanced Breast Cancer. *Ann. Surg. Oncol.* **2025**, *32*, 4890–4898. [CrossRef]
45. Zhao, F.; Polley, E.; McClellan, J.; Howard, F.; Olopade, O.I.; Huo, D. Predicting pathologic complete response to neoadjuvant chemotherapy in breast cancer using a machine learning approach. *Breast Cancer Res.* **2024**, *26*, 148. [CrossRef]
46. Zhang, W.; Zhang, S.; You, J.; Li, F.; Wu, X.; Lu, X.; Lv, Q.; Huang, J.; Yi, Y.; Bu, H. Attention-based multimodal fusion transformer for predicting the efficacy of neoadjuvant therapy in breast cancer: A cross-institutional retrospective study. *Breast Cancer Res.* **2025**, *28*, 4. [CrossRef]
47. Feng, X.; Shi, Y.; Wu, M.; Cui, G.; Du, Y.; Yang, J.; Xu, Y.; Wang, W.; Liu, F. Predicting the efficacy of neoadjuvant chemotherapy in breast cancer patients based on ultrasound longitudinal temporal depth network fusion model. *Breast Cancer Res.* **2025**, *27*, 30. [CrossRef]

48. Zeng, H.; Qiu, S.; Zhuang, S.; Wei, X.; Wu, J.; Zhang, R.; Chen, K.; Wu, Z.; Zhuang, Z. Deep learning-based predictive model for pathological complete response to neoadjuvant chemotherapy in breast cancer from biopsy pathological images: A multicenter study. *Front. Physiol.* **2024**, *15*, 1279982. [CrossRef]
49. Li, W.; Onishi, N.; Gibbs, J.E.; Wilmes, L.J.; Le, N.N.; Metanat, P.; Price, E.R.; Joe, B.N.; Kornak, J.; Yau, C.; et al. MRI-Based Model for Personalizing Neoadjuvant Treatment in Breast Cancer. *Tomography* **2025**, *11*, 26. [CrossRef]
50. Liu, Z.; Yu, B.; Su, M.; Yuan, C.; Liu, C.; Wang, X.; Song, X.; Li, C.; Wang, F.; Ma, J.; et al. Construction of a risk stratification model integrating ctDNA to predict response and survival in neoadjuvant-treated breast cancer. *BMC Med.* **2023**, *21*, 493. [CrossRef]
51. Yang, M.; Liu, H.; Dai, Q.; Yao, L.; Zhang, S.; Wang, Z.; Li, J.; Duan, Q. Treatment Response Prediction Using Ultrasound-Based Pre-, Post-Early, and Delta Radiomics in Neoadjuvant Chemotherapy in Breast Cancer. *Front. Oncol.* **2022**, *12*, 748008. [CrossRef]
52. Ma, T.; Wang, Z.; Zhou, Z.; Wang, Y.; Ma, T.; Liu, X.; Mao, Y.; Wang, H. Multi-omics prediction of axillary treatment response and tumour microenvironment alterations in lymph node-positive luminal breast cancer. *Cell Death Dis.* **2025**, *16*, 588. [CrossRef]
53. Prat, A.; Guarneri, V.; Pascual, T.; Brasó-Maristany, F.; Sanfeliu, E.; Paré, L.; Schettini, F.; Martínez, D.; Jares, P.; Griguolo, G.; et al. Development and validation of the new HER2DX assay for predicting pathological response and survival outcome in early-stage HER2-positive breast cancer. *EBioMedicine* **2022**, *75*, 103801. [CrossRef]
54. Yang, B.; Chen, M.; Li, P.; Xiao, X.; Xiu, B.; Bai, J.; Sun, Y.; Xu, X.; Zhang, J.; Qiao, Y.; et al. A prospective single-center cohort study: Integrating PET imaging and genomics to identify early biomarkers of response to dual-targeted PH neoadjuvant therapy in breast cancer. *Int. J. Surg.* **2026**, *112*, 3312–3324. [CrossRef]
55. Xie, T.; Ai, S.; Liu, C.; Luo, B.; Yan, H.; Zhang, H.; Chen, S.; Chen, Z.; Jin, T.; Yuan, J. Focal hotspot and diffuse immune subtypes of tumor-infiltrating lymphocytes: AI-powered spatial clustering classification and its clinical relevance to HER2 expression in triple-negative breast cancer. *J. Transl. Med.* **2025**, *24*, 111. [CrossRef]
56. Zhou, Z.; Adrada, B.E.; Candelaria, R.P.; Elshafeey, N.A.; Boge, M.; Mohamed, R.M.; Pashapoor, S.; Sun, J.; Xu, Z.; Panthi, B.; et al. Prediction of pathologic complete response to neoadjuvant systemic therapy in triple negative breast cancer using deep learning on multiparametric MRI. *Sci. Rep.* **2023**, *13*, 1171. [CrossRef]
57. Fisher, T.B.; Saini, G.; Rekha, T.S.; Krishnamurthy, J.; Bhattarai, S.; Callagy, G.; Webber, M.; Janssen, E.A.M.; Kong, J.; Aneja, R. Digital image analysis and machine learning-assisted prediction of neoadjuvant chemotherapy response in triple-negative breast cancer. *Breast Cancer Res.* **2024**, *26*, 12. [CrossRef]
58. Sun, X.; Ocampo, A.A.; Hanna, A.; Taylor, B.C.; Marshall, J.L.; Steele, J.A.; Axelrod, M.L.; Wescott, E.C.; Opalenik, S.R.; DeMichele, A.; et al. Peripheral blood transcriptional profiling predicts tumor subtype and neoadjuvant chemoimmunotherapy outcomes in human breast cancer. *Sci. Transl. Med.* **2026**, *18*, eaec2358. [CrossRef]
59. Adrada, B.E.; Guirguis, M.S.; Huo, L.; Yam, C.; Tripathy, D.; Candelaria, R.; Yang, W.; Patel, M.; Moseley, T.; Whitman, G.J.; et al. Imaging- and Tumor Biomarker-Based Multivariable Model for Early Prediction of Pathologic Complete Response to Neoadjuvant Systemic Therapy in Triple-Negative Breast Cancer. *JCO Precis. Oncol.* **2025**, *9*, e2500410. [CrossRef]
60. Park, J.W.; Liu, M.C.; Yee, D.; Yau, C.; van 't Veer, L.J.; Symmans, W.F.; Paoloni, M.; Perlmutter, J.; Hylton, N.M.; Hogarth, M.; et al. Adaptive Randomization of Neratinib in Early Breast Cancer (I-SPY2). *N. Engl. J. Med.* **2016**, *375*, 11–22. [CrossRef]
61. Janse, M.H.A.; Janssen, L.M.; van der Velden, B.H.M.; Moman, M.R.; Wolters-van der Ben, E.J.M.; Kock, M.C.J.M.; Viergever, M.A.; van Diest, P.J.; Gilhuijs, K.G.A. Deep Learning-Based Segmentation of Locally Advanced Breast Cancer on MRI in Relation to Residual Cancer Burden: A Multi-Institutional Cohort Study. *J. Magn. Reson. Imaging* **2023**, *58*, 1739–1749. [CrossRef]
62. Guo, R.; Su, Y.; Si, J.; Xue, J.; Yang, B.; Zhang, Q.; Chi, W.; Chen, J.; Chi, Y.; Shao, Z.; et al. A nomogram for predicting axillary pathologic complete response in hormone receptor-positive breast cancer with cytologically proven axillary lymph node metastases. *Cancer* **2020**, *126*, 3819–3829. [CrossRef]
63. de Mooij, C.M.; Mitea, C.; Mottaghy, F.M.; Smidt, M.L.; van Nijnatten, T.J.A. Value of ^{18}F-FDG PET/CT for predicting axillary pathologic complete response following neoadjuvant systemic therapy in breast cancer patients: Emphasis on breast cancer subtype. *EJNMMI Res.* **2021**, *11*, 116. [CrossRef]
64. Comes, M.C.; Fanizzi, A.; Bove, S.; Didonna, V.; Diotaiuti, S.; La Forgia, D.; Latorre, A.; Martinelli, E.; Mencattini, A.; Nardone, A.; et al. Early prediction of neoadjuvant chemotherapy response by exploiting a transfer learning approach on breast DCE-MRIs. *Sci. Rep.* **2021**, *11*, 14123. [CrossRef]
65. Huang, Y.; Cao, Y.; Chen, H.; Lan, X.; Tang, S.; Zhang, Z.; Yin, T.; Wang, X.; Zhang, J. Quantifying tumor morphological complexity based on pretreatment MRI fractal analysis for predicting pathologic complete response and survival in breast cancer: A retrospective, multicenter study. *Breast Cancer Res.* **2025**, *27*, 86. [CrossRef]
66. Zhang, X.; Teng, X.; Zhang, J.; Lai, Q.; Cai, J. Enhancing pathological complete response prediction in breast cancer: The role of dynamic characterization of DCE-MRI and its association with tumor heterogeneity. *Breast Cancer Res.* **2024**, *26*, 77. [CrossRef]
67. Ping, L.; Zhu, L.; Chen, N.; Liu, X.; Zhong, J.; Sun, X.; Tang, H.; Zhang, K. Endocrine Resistance Score Based on Three Key Genes Predicts Prognosis and Reveals Potential Therapeutic Targets for ER-positive/HER2-negative Breast Cancer. *Cell Prolif.* **2026**, *59*, e70100. [CrossRef]

68. Liang, Y.; Chen, X.; Lai, P.; Huang, S.; Hong, M.; Jiang, W.; Cai, H.; Ying, Y.; Chen, K.; Hou, K.; et al. Lumi-Guide: An Artificial-Intelligence-Driven Multimodal Framework for Optimizing Personalized Neoadjuvant Therapy Decision-Making in Luminal Breast Cancer. *Research* **2026**, *9*, 1303. [CrossRef]
69. Wang, K.N.; Meng, Y.J.; Yu, Y.; Cai, W.R.; Wang, X.; Cao, X.C.; Ge, J. Predicting pathological complete response after neoadjuvant chemotherapy: A nomogram combining clinical features and ultrasound semantics in patients with invasive breast cancer. *Front. Oncol.* **2023**, *13*, 1117538. [CrossRef]
70. Wu, C.; Wang, L.; Wang, N.; Shiao, S.; Dou, T.; Hsu, Y.C.; Christodoulou, A.G.; Xie, Y.; Li, D. Prediction of Early Neoadjuvant Chemotherapy Response of Breast Cancer through Deep Learning-based Pharmacokinetic Quantification of DCE MRI. *Radiol. Artif. Intell.* **2025**, *7*, e240769. [CrossRef]
71. Sun, Y.; Liao, Q.; Fan, Y.; Cui, C.; Wang, Y.; Yang, C.; Hou, Y.; Zhao, D. DCE-MRI radiomics of primary breast lesions combined with ipsilateral axillary lymph nodes for predicting efficacy of NAT. *BMC Cancer* **2025**, *25*, 589. [CrossRef]
72. Huang, Y.; Zhu, T.; Zhang, X.; Li, W.; Zheng, X.; Cheng, M.; Ji, F.; Zhang, L.; Yang, C.; Wu, Z.; et al. Longitudinal MRI-based fusion novel model predicts pathological complete response in breast cancer treated with neoadjuvant chemotherapy: A multicenter, retrospective study. *EClinicalMedicine* **2023**, *58*, 101899. [CrossRef]
73. Zixuan, Y.; Jie, H.; Taolang, L.; Changdong, L.; Yongsheng, W.; Yu, R.; Wenhe, Z.; Choo, C.C.; Qiang, L.; Liang, X.; et al. A multicenter study of a predictive model for pathological complete response after neoadjuvant therapy in breast cancer using multimodal digital biomarkers. *Chin. J. Cancer Res.* **2025**, *37*, 984–999. [CrossRef]
74. Wu, L.; Ye, W.; Liu, Y.; Chen, D.; Wang, Y.; Cui, Y.; Li, Z.; Li, P.; Li, Z.; Liu, Z.; et al. An integrated deep learning model for the prediction of pathological complete response to neoadjuvant chemotherapy with serial ultrasonography in breast cancer patients: A multicentre, retrospective study. *Breast Cancer Res.* **2022**, *24*, 81. [CrossRef]
75. Aswolinskiy, W.; Munari, E.; Horlings, H.M.; Mulder, L.; Bogina, G.; Sanders, J.; Liu, Y.H.; van den Belt-Dusebout, A.W.; Tessier, L.; Balkenhol, M.; et al. PROACTING: Predicting pathological complete response to neoadjuvant chemotherapy in breast cancer from routine diagnostic histopathology biopsies with deep learning. *Breast Cancer Res.* **2023**, *25*, 142. [CrossRef]
76. Li, B.; Li, F.; Liu, Z.; Xu, F.; Ye, G.; Li, W.; Zhang, Y.; Zhu, T.; Shao, L.; Chen, C.; et al. Deep learning with biopsy whole slide images for pretreatment prediction of pathological complete response to neoadjuvant chemotherapy in breast cancer: A multicenter study. *Breast* **2022**, *66*, 183–190. [CrossRef]
77. Zhang, J.; Li, Y.; Li, Z.; Cui, Z.; Li, P.; Li, J.; Li, Z.; Xie, Y.; Chan, K.W.Y.; Zhang, Q.; et al. Deep-learning-based HER2 status assessment from multimodal breast cancer data predicts neoadjuvant therapy response. *Nat. Biomed. Eng.* **2026**, *10*, 1235–1246. [CrossRef]
78. Liu, Y.; Wang, Y.; Wang, Y.; Xie, Y.; Cui, Y.; Feng, S.; Yao, M.; Qiu, B.; Shen, W.; Chen, D.; et al. Early prediction of treatment response to neoadjuvant chemotherapy based on longitudinal ultrasound images of HER2-positive breast cancer patients by Siamese multi-task network: A multicentre, retrospective cohort study. *EClinicalMedicine* **2022**, *52*, 101562. [CrossRef]
79. Wu, H.Y.; Lin, C.Y.; Tzeng, Y.D.; Hung, C.C.; Liu, S.I.; Yin, C.H.; Chen, J.S.; Chen, Y.S.; Yang, J.R. Preoperative systemic inflammation response index: Clinicopathologic predictor of pathological complete response in HER2-positive breast cancer patients receiving neoadjuvant systemic therapy. *J. Chin. Med. Assoc.* **2024**, *87*, 226–235. [CrossRef]
80. Chan, A.; Delaloge, S.; Holmes, F.A.; Moy, B.; Iwata, H.; Harvey, V.J.; Robert, N.J.; Silovski, T.; Gokmen, E.; von Minckwitz, G.; et al. Neratinib after trastuzumab-based adjuvant therapy in patients with HER2-positive breast cancer (ExteNET): A multicentre, randomised, double-blind, placebo-controlled, phase 3 trial. *Lancet Oncol.* **2016**, *17*, 367–377. [CrossRef]
81. Zeng, T.; He, Y.; Zhang, T.; Ren, C.; Xu, J.; Cheng, J.; Ming, W. Dedicated Breast PET-Based Deep Learning Radiomics for Prediction of Pathologic Complete Response to Neoadjuvant Chemotherapy in HER2-Positive Breast Cancer. *Cancers* **2026**, *18*, 1581. [CrossRef]
82. Kim, S.Y.; Lee, J.; Cho, N.; Kim, Y.G. Deep-learning based discrimination of pathologic complete response using MRI in HER2-positive and triple-negative breast cancer. *Sci. Rep.* **2024**, *14*, 23065. [CrossRef]
83. Masuda, N.; Lee, S.J.; Ohtani, S.; Im, Y.H.; Lee, E.S.; Yokota, I.; Kuroi, K.; Im, S.A.; Park, B.W.; Kim, S.B.; et al. Adjuvant Capecitabine for Breast Cancer after Preoperative Chemotherapy. *N. Engl. J. Med.* **2017**, *376*, 2147–2159. [CrossRef]
84. Tutt, A.N.J.; Garber, J.E.; Kaufman, B.; Viale, G.; Fumagalli, D.; Rastogi, P.; Gelber, R.D.; de Azambuja, E.; Fielding, A.; Balmaña, J.; et al. Adjuvant Olaparib for Patients with BRCA1- or BRCA2-Mutated Breast Cancer (OlympiA). *N. Engl. J. Med.* **2021**, *384*, 2394–2405. [CrossRef]
85. Schlam, I.; Tang, G.; AbdulJabbar, K.; Yan, H.; Gu, X.; Freeman, T.J.; Maley, S.; Harris, B.T.; Rastogi, P.; Hook, I.; et al. AI-powered and manual assessment of tumor-infiltrating lymphocytes in early HER2-positive breast cancer in NSABP B-41. *npj Breast Cancer* **2026**, *12*, 98. [CrossRef]
86. Lu, X.; Luo, B.; Wei, Y.; Zhang, W.; Wu, X.; Chen, J.; Shi, H.; Yuan, J.; Bu, H.; Yi, Y.; et al. Integration of histopathological characteristics by machine learning improves the prediction of neoadjuvant immunochemotherapy response in triple-negative breast cancer. *J. Pathol.* **2026**, *268*, 353–365. [CrossRef]

87. Wang, Y.; Drubay, D.; Jonas, S.F.; Dixon-Douglas, J.; Acs, B.; Adams, S.; André, F.; Badve, S.; Bataillon, G.; Carter, J.M.; et al. Including tumor-infiltrating lymphocytes into the PREDICT prognostic model for triple-negative breast cancer survival. *Ann. Oncol.* **2026**, *37*, 1093–1105. [CrossRef]
88. Yan, L.; Huang, X.; Liu, L.; Wu, A.; Luo, Y.; Li, H.; Yi, S.; Yu, T.; Zeng, Q. Predicting response to neoadjuvant therapy in breast cancer using longitudinal DCE-MRI deep learning integrated with tumor microenvironment data. *Front. Immunol.* **2026**, *17*, 1749668. [CrossRef]
89. Han, Y.; Wang, J.; Xu, B. Novel biomarkers and prediction model for the pathological complete response to neoadjuvant treatment of triple-negative breast cancer. *J. Cancer* **2021**, *12*, 936–945. [CrossRef]
90. Liang, H.; Zhu, H.; Wang, L.; Yu, J.; Feng, R.; Yan, X.; Shi, L.; Xiao, J.; Zhou, H.; Chen, P.; et al. AI-driven spatial analysis of tumor-infiltrating lymphocytes predicts chemo-immunotherapy response in triple-negative breast cancer. *Front. Immunol.* **2026**, *17*, 1797753. [CrossRef]
91. Chen, Y.; Huang, Y.; Li, W.; Zhu, T.; Cheng, M.; Wu, C.; Zhang, L.; Peng, H.; Wang, K. Intratumoral microbiota-aided fusion radiomics model for predicting tumor response to neoadjuvant chemoimmunotherapy in triple-negative breast cancer. *J. Transl. Med.* **2025**, *23*, 352. [CrossRef]
92. Groheux, D.; Ferrer, L.; Vargas, J.; Martineau, A.; Borgel, A.; Teixeira, L.; Menu, P.; Bertheau, P.; Gallinato, O.; Colin, T.; 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*, 1249. [CrossRef]
93. Ramtohul, T.; Lollivier, D.; Spriet, J.; Jin, M.; Djerroudi, L.; Gaillard, T.; Bonneau, C.; Loirat, D.; Bello-Roufai, D.; Kirova, Y.; et al. Posttreatment MRI to Predict Pathologic Complete Response of Triple-Negative Breast Cancer to Neoadjuvant Chemoimmunotherapy. *Radiology* **2025**, *316*, e243824. [CrossRef]
94. Stowers, C.E.; Wu, C.; Xu, Z.; Kumar, S.; Yam, C.; Son, J.B.; Ma, J.; Tamir, J.I.; Rauch, G.M.; Yankeelov, T.E. Combining Biology-based and MRI Data-driven Modeling to Predict Response to Neoadjuvant Chemotherapy in Patients with Triple-Negative Breast Cancer. *Radiol. Artif. Intell.* **2025**, *7*, e240124. [CrossRef]
95. Chen, X.; Deng, X.; Zhou, Z. Evidential reasoning-enabled deep learning for reliable treatment outcome prediction in cancer therapy. *Artif. Intell. Med.* **2026**, *178*, 103445. [CrossRef]
96. Xia, B.; Wang, H.; Wang, Z.; Qian, Z.; Xiao, Q.; Liu, Y.; Shao, Z.; Zhou, S.; Chai, W.; You, C.; et al. A Combined Nomogram Model to Predict Disease-free Survival in Triple-Negative Breast Cancer Patients With Neoadjuvant Chemotherapy. *Front. Genet.* **2021**, *12*, 783513. [CrossRef]
97. Salvador-Coloma, C.; Santaballa, A.; Sanmartín, E.; Calvo, D.; García, A.; Hervás, D.; Cordón, L.; Quintas, G.; Ripoll, F.; Panadero, J.; et al. Immunosuppressive profiles in liquid biopsy at diagnosis predict response to neoadjuvant chemotherapy in triple-negative breast cancer. *Eur. J. Cancer* **2020**, *139*, 119–134. [CrossRef]
98. Cai, L.; Sidey-Gibbons, C.; Nees, J.; Riedel, F.; Schäfgen, B.; Togawa, R.; Killinger, K.; Heil, J.; Pfob, A.; Golatta, M. Can multi-modal radiomics using pretreatment ultrasound and tomosynthesis predict response to neoadjuvant systemic treatment in breast cancer? *Eur. Radiol.* **2024**, *34*, 2560–2573. [CrossRef]
99. 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. [CrossRef]
100. Jing, B.; Wang, K.; Schmitz, E.; Tang, S.; Li, Y.; Zhang, Y.; Wang, J. Prediction of pathological complete response to chemotherapy for breast cancer using deep neural network with uncertainty quantification. *Med. Phys.* **2024**, *51*, 9385–9393. [CrossRef]
101. Jannusch, K.; Dietzel, F.; Bruckmann, N.M.; Morawitz, J.; Boschheidgen, M.; Minko, P.; Bittner, A.K.; Mohrmann, S.; Quick, H.H.; Herrmann, K.; et al. Prediction of therapy response of breast cancer patients with machine learning based on clinical data and imaging data derived from breast ^{18}F-FDG-PET/MRI. *Eur. J. Nucl. Med. Mol. Imaging* **2024**, *51*, 1451–1461. [CrossRef]
102. Choi, J.H.; Kim, H.A.; Kim, W.; Lim, I.; Lee, I.; Byun, B.H.; Noh, W.C.; Seong, M.K.; Lee, S.S.; Kim, B.I.; et al. Early prediction of neoadjuvant chemotherapy response for advanced breast cancer using PET/MRI image deep learning. *Sci. Rep.* **2020**, *10*, 21149. [CrossRef]
103. Yang, L.; Chang, J.; He, X.; Peng, M.; Zhang, Y.; Wu, T.; Xu, P.; Chu, W.; Gao, C.; Cao, S.; et al. PET/CT-based radiomics analysis may help to predict neoadjuvant chemotherapy outcomes in breast cancer. *Front. Oncol.* **2022**, *12*, 849626. [CrossRef]
104. Xu, N.; Guo, X.; Ouyang, Z.; Ran, F.; Li, Q.; Duan, X.; Zhu, Y.; Niu, X.; Liao, C.; Yang, J. Multiparametric MRI-based radiomics combined with pathomics features for prediction of the efficacy of neoadjuvant chemotherapy in breast cancer. *Heliyon* **2024**, *10*, e24371. [CrossRef]
105. Wang, L.; Zhang, X.; Zhang, J.; Liu, J.; Chen, Y.; Huang, W.; Xie, X. Multiomics Machine Learning to Predict Neoadjuvant Chemotherapy Outcome and Relapse of Breast Cancer. *BME Front.* **2026**, *7*, 0212. [CrossRef]
106. Liang, Y.; Xu, H.; Lin, J.; Tang, W.; Liu, X.; Gan, K.; Wan, Q.; Du, X. Multi-modal radiomics model based on four imaging modalities for predicting pathological complete response to neoadjuvant treatment in breast cancer. *BMC Cancer* **2025**, *25*, 985. [CrossRef]

107. Zhou, L.Q.; Wang, X.Y.; Xu, Y.; Zhang, H.X.; Yang, X.X.; Jin, R.Q.; Sang, X.Q.; Zhu, Y.M.; Meng, H.X.; Kuai, Z.X. Radiopathomic Graph Deep Learning for Multiscale Spatial-Contextual Modeling of Intratumoral Heterogeneity to Predict Breast Cancer Response to Neoadjuvant Therapy. *Radiol. Artif. Intell.* **2026**, *8*, e250760. [CrossRef]
108. Han, L.; Tan, T.; Zhang, T.; Gao, Y.; Wang, X.; Longo, V.; Ventura-Díaz, S.; D'Angelo, A.; Sun, Y.; Teuwen, J.; et al. Tumor-preserving Deformable Registration of Longitudinal DCE Breast MRI during Neoadjuvant Chemotherapy. *Radiol. Artif. Intell.* **2026**, *8*, e250789. [CrossRef]
109. Jiang, C.; Zhang, X.; Qu, T.; Yang, X.; Xiu, Y.; Yu, X.; Zhang, S.; Qiao, K.; Meng, H.; Li, X.; et al. The prediction of pCR and chemosensitivity for breast cancer patients using DLG3, RADL and Pathomics signatures based on machine learning and deep learning. *Transl. Oncol.* **2024**, *46*, 101985. [CrossRef]
110. Huang, Y.; Song, X.; Chen, Y.; Qiu, H.; Zhou, T.; Wang, S.; Zhou, Y.; Li, W.; Lin, Y.; Wang, Q.; et al. Intratumoral Microbiome-related MRI Model for Predicting Breast Cancer Shrinkage Pattern Following Neoadjuvant Therapy. *Radiology* **2025**, *316*, e243545. [CrossRef]
111. Hou, J.; Yang, F.; Gao, Y.; Cai, H.; Li, X.; Lin, C.; Hong, M.; He, X.; Shu, Z.; Chen, X.; et al. MR imaging of breast cancer: Interpretable radiomics analysis to assess treatment response and survival prognosis after neoadjuvant therapy. *Int. J. Cancer* **2025**, *157*, 1723–1733. [CrossRef]
112. Saednia, K.; Tran, W.T.; Sadeghi-Naini, A. A hierarchical self-attention-guided deep learning framework to predict breast cancer response to chemotherapy using pre-treatment tumor biopsies. *Med. Phys.* **2023**, *50*, 7852–7864. [CrossRef]
113. Tong, T.; Li, D.; Gu, J.; Chen, G.; Bai, G.; Yang, X.; Wang, K.; Jiang, T.; Tian, J. Dual-Input Transformer: An End-to-End Model for Preoperative Assessment of Pathological Complete Response to Neoadjuvant Chemotherapy in Breast Cancer Ultrasonography. *IEEE J. Biomed. Health Inform.* **2023**, *27*, 251–262. [CrossRef]
114. Yu, F.H.; Miao, S.M.; Li, C.Y.; Hang, J.; Deng, J.; Ye, X.H.; Liu, Y. Pretreatment ultrasound-based deep learning radiomics model for the early prediction of pathologic response to neoadjuvant chemotherapy in breast cancer. *Eur. Radiol.* **2023**, *33*, 5634–5644. [CrossRef]
115. Jiang, M.; Li, C.L.; Luo, X.M.; Chuan, Z.R.; Lv, W.Z.; Li, X.; Cui, X.W.; Dietrich, C.F. Ultrasound-based deep learning radiomics in the assessment of pathological complete response to neoadjuvant chemotherapy in locally advanced breast cancer. *Eur. J. Cancer* **2021**, *147*, 95–105. [CrossRef]
116. Chen, K.; Wang, J.; Li, S.; Zhou, W.; Xu, W. Predictive value of ^{18}F-FDG PET/CT-based radiomics model for neoadjuvant chemotherapy efficacy in breast cancer: A multi-scanner/center study with external validation. *Eur. J. Nucl. Med. Mol. Imaging* **2023**, *50*, 1869–1880. [CrossRef]
117. Li, P.; Wang, X.; Xu, C.; Liu, C.; Zheng, C.; Fulham, M.J.; Feng, D.; Wang, L.; Song, S.; Huang, G. ^{18}F-FDG PET/CT radiomic predictors of pathologic complete response (pCR) to neoadjuvant chemotherapy in breast cancer patients. *Eur. J. Nucl. Med. Mol. Imaging* **2020**, *47*, 1116–1126. [CrossRef]
118. Liu, S.; Du, S.; Gao, S.; Teng, Y.; Jin, F.; Zhang, L. A delta-radiomic lymph node model using dynamic contrast enhanced MRI for the early prediction of axillary response after neoadjuvant chemotherapy in breast cancer patients. *BMC Cancer* **2023**, *23*, 15. [CrossRef]
119. Guo, J.; Chen, B.; Cao, H.; Dai, Q.; Qin, L.; Zhang, J.; Zhang, Y.; Zhang, H.; Sui, Y.; Chen, T.; et al. Cross-modal deep learning model for predicting pathologic complete response to neoadjuvant chemotherapy in breast cancer. *npj Precis. Oncol.* **2024**, *8*, 189. [CrossRef]
120. Valderrama, N.F.; Morel, L.O.; Mweze, D.T.; Derangère, V.; Desmoulins, I.; Mayeur, D.; Kaderbhai, C.; Ilie, S.; Hennequin, A.; Roussot, N.; et al. Chemo-prAIdict Breast: A deep learning solution for predicting residual disease on biopsies of breast cancer patients treated with neoadjuvant chemotherapy. *Eur. J. Cancer* **2026**, *234*, 116222. [CrossRef]
121. Zhang, Y.; You, C.; Pei, Y.; Yang, F.; Li, D.; Jiang, Y.Z.; Shao, Z. Integration of radiogenomic features for early prediction of pathological complete response in patients with triple-negative breast cancer and identification of potential therapeutic targets. *J. Transl. Med.* **2022**, *20*, 256. [CrossRef]
122. Nishizawa, T.; Maldjian, T.; Jiao, Z.; Duong, T.Q. Attention-based multimodal deep learning for interpretable and generalizable prediction of pathological complete response in breast cancer. *J. Transl. Med.* **2025**, *23*, 774. [CrossRef]
123. Comes, M.C.; Fanizzi, A.; Bove, S.; Boldrini, L.; Latorre, A.; Guven, D.C.; Iacovelli, S.; Talienti, T.; Rizzo, A.; Zito, F.A.; et al. Monitoring Over Time of Pathological Complete Response to Neoadjuvant Chemotherapy in Breast Cancer Patients Through an Ensemble Vision Transformers-Based Model. *Cancer Med.* **2024**, *13*, e70482. [CrossRef]
124. Modi, S.; Jacot, W.; Yamashita, T.; Sohn, J.; Vidal, M.; Tokunaga, E.; Tsurutani, J.; Ueno, N.T.; Prat, A.; Chae, Y.S.; et al. Trastuzumab Deruxtecan in Previously Treated HER2-Low Advanced Breast Cancer. *N. Engl. J. Med.* **2022**, *387*, 9–20. [CrossRef]
125. Yan, Y.; Xue, X.; Xie, J.; Liu, J.; Sui, L.; Jiang, T.; Jin, Z.; Ou, D.; Chuan, Z.; Jin, M.; et al. Prediction of HER2 changes post-neoadjuvant therapy based on fusion of ultrasound radiomics and clinicopathological features empowered by explainable AI: A multicenter study. *Eur. J. Cancer* **2026**, *232*, 116158. [CrossRef]

126. Gilad, M.; Partridge, S.C.; Iima, M.; Rakow-Penner, R.M.; Freiman, M. Radiomics-based Machine Learning Prediction of Neoadjuvant Chemotherapy Response in Breast Cancer Using Physiologically Decomposed Diffusion-weighted MRI. *Radiol. Imaging Cancer* **2025**, *7*, e240312. [CrossRef]
127. Guo, L.; Du, S.; Gao, S.; Zhao, R.; Huang, G.; Jin, F.; Teng, Y.; Zhang, L. Delta-Radiomics Based on Dynamic Contrast-Enhanced MRI Predicts Pathologic Complete Response in Breast Cancer Patients Treated with Neoadjuvant Chemotherapy. *Cancers* **2022**, *14*, 3515. [CrossRef]
128. Li, S.; Li, R.; Xu, M.; Liao, J.; Cao, Z.; Gong, B.; Yan, F.; Wang, M.; Tan, H. Deep Learning Radiomics of Multiparametric MRI for Individualized Prediction of Axillary Lymph Node Response After Neoadjuvant Chemotherapy in Breast Cancer. *Breast Cancer* **2026**, *18*, 568337. [CrossRef]
129. Sutton, E.J.; Onishi, N.; Fehr, D.A.; Dashevsky, B.Z.; Sadinski, M.; Pinker, K.; Martinez, D.F.; Brogi, E.; Braunstein, L.; Razavi, P.; et al. A machine learning model that classifies breast cancer pathologic complete response on MRI post-neoadjuvant chemotherapy. *Breast Cancer Res.* **2020**, *22*, 57. [CrossRef]
130. Wang, X.; Huang, Y.; Shi, J.; Cao, Y.; Chen, H.; Li, L.; Wang, L.; Tang, S.; Gong, X.; Huang, H.; et al. Biomechanical parameters quantified by MR elastography for predicting response to neoadjuvant chemotherapy and disease-free survival in breast cancer: A prospective longitudinal study. *Breast Cancer Res.* **2025**, *27*, 72. [CrossRef]
131. Xu, S.; Ying, Y.; Hu, Q.; Li, X.; Li, Y.; Xiong, H.; Chen, Y.; Ye, Q.; Li, X.; Liu, Y.; et al. Fusion model integrating multi-sequence MRI radiomics and habitat imaging for predicting pathological complete response in breast cancer treated with neoadjuvant therapy. *Cancer Imaging* **2025**, *25*, 108. [CrossRef]
132. Zhu, H.; Zhang, B.; Peng, D.; Song, S.; Huang, L.; Zhang, Y.; Wu, X. MRI-Based Habitat Analysis for Predicting Pathologic Complete Response to Neoadjuvant Chemotherapy in Breast Cancer: A Multicenter Study. *J. Magn. Reson. Imaging* **2026**, *63*, 1769–1779. [CrossRef]
133. Fu, Y.; Lei, Y.T.; Huang, Y.H.; Mei, F.; Wang, S.; Yan, K.; Wang, Y.H.; Ma, Y.H.; Cui, L.G. Longitudinal ultrasound-based AI model predicts axillary lymph node response to neoadjuvant chemotherapy in breast cancer: A multicenter study. *Eur. Radiol.* **2024**, *34*, 7080–7089. [CrossRef]
134. Gu, J.; Zhong, X.; Fang, C.; Lou, W.; Fu, P.; Woodruff, H.C.; Wang, B.; Jiang, T.; Lambin, P. Deep Learning of Multimodal Ultrasound: Stratifying the Response to Neoadjuvant Chemotherapy in Breast Cancer Before Treatment. *Oncologist* **2024**, *29*, e187–e197. [CrossRef]
135. Huang, J.X.; Lu, Y.; Tan, Y.T.; Liu, F.T.; Li, Y.L.; Wang, X.Y.; Huang, J.H.; Lin, S.Y.; Huang, G.L.; Zhang, Y.T.; et al. Elastography-based AI model can predict axillary status after neoadjuvant chemotherapy in breast cancer with nodal involvement: A prospective, multicenter, diagnostic study. *Int. J. Surg.* **2025**, *111*, 221–229. [CrossRef]
136. Liu, J.; Leng, X.; Yuan, Z.; Li, Y.; Luo, H.; Chen, J.; Li, R.; Li, X.; He, Z.; Zou, Q.; et al. Predicting breast cancer response to neoadjuvant chemotherapy with ultrasound-based deep learning radiomics models—dual-center study. *BMC Cancer* **2025**, *25*, 1737. [CrossRef]
137. Wang, S.; Liu, J.; Song, L.; Zhao, H.; Wan, X.; Peng, Y. Integrative multimodal ultrasound and radiomics for early prediction of neoadjuvant therapy response in breast cancer: A clinical study. *BMC Cancer* **2025**, *25*, 1156. [CrossRef]
138. Mo, Z.; Yang, M.; Zhu, Z.; Wang, M.; Wang, H.; Zhang, Z.; Lyu, S.; Xu, F.; Shang, H.; Lin, H.; et al. Multiomic integration reveals subtype-specific predictors of neoadjuvant treatment response in breast cancer. *Sci. Adv.* **2025**, *11*, eadu1521. [CrossRef]
139. Rasic, D.; Stovgaard, E.I.S.; Jylling, A.M.B.; Salgado, R.; Hartman, J.; Rantalainen, M.; Lænkholm, A.V. AI assessment of tumor-infiltrating lymphocytes on routine H&E-slides as a predictor of response to neoadjuvant therapy in breast cancer-a real-world study. *Virchows Arch.* **2025**. [CrossRef]
140. Saednia, K.; Lagree, A.; Alera, M.A.; Fleshner, L.; Shiner, A.; Law, E.; Law, B.; Dodington, D.W.; Lu, F.I.; Tran, W.T.; et al. Quantitative digital histopathology and machine learning to predict pathological complete response to chemotherapy in breast cancer patients using pre-treatment tumor biopsies. *Sci. Rep.* **2022**, *12*, 9690. [CrossRef]
141. Yue, J.; Liu, J.; Kang, X.; Yuan, P.; Wang, W.; Wang, Z.; Shang, C.; Shang, Q.; Li, G.; Dong, X.; et al. Development and validation of a pathomics-driven machine learning model for individualized prediction of neoadjuvant chemotherapy response and early recurrence in HR-positive, HER2-negative breast cancer. *Front. Oncol.* **2026**, *16*, 1770037. [CrossRef]
142. Urso, L.; Manco, L.; Cittanti, C.; Adamantiadis, S.; Szilagyi, K.E.; Scribano, G.; Mindicini, N.; Carnevale, A.; Bartolomei, M.; Giganti, M. ^{18}F-FDG PET/CT radiomic analysis and artificial intelligence to predict pathological complete response after neoadjuvant chemotherapy in breast cancer patients. *Radiol. Med.* **2025**, *130*, 543–554. [CrossRef]
143. Wu, Y.; Li, Y.; Chen, B.; Zhang, Y.; Xing, W.; Guo, B.; Wang, W. ^{18}F-FDG PET/CT for early prediction of pathological complete response in breast cancer neoadjuvant therapy: A retrospective analysis. *Oncologist* **2024**, *29*, e1646–e1655. [CrossRef]
144. Yang, X.; Liu, Q.; Guo, Z.; Yang, X.; Li, K.; Han, B.; Zhang, M.; Sun, M.; Huang, L.; Cai, G.; et al. Promoter profiles in plasma CfDNA exhibits a potential utility of predicting the efficacy of neoadjuvant chemotherapy in breast cancer patients. *Breast Cancer Res.* **2024**, *26*, 112. [CrossRef]

145. Cheligeer, K.; Wu, G.; Laws, A.; Quan, M.L.; Li, A.; Brisson, A.M.; Xie, J.; Xu, Y. Validation of large language models for detecting pathologic complete response in breast cancer using population-based pathology reports. *BMC Med. Inform. Decis. Mak.* **2024**, *24*, 283. [CrossRef]
146. Fang, S.; Xia, W.; Zhang, H.; Ni, C.; Wu, J.; Mo, Q.; Jiang, M.; Guan, D.; Yuan, H.; Chen, W. A real-world clinicopathological model for predicting pathological complete response to neoadjuvant chemotherapy in breast cancer. *Front. Oncol.* **2024**, *14*, 1323226. [CrossRef]
147. Fang, S.; Zhang, J.; Han, C.; Kong, M.; Zhang, H.; Zhong, M.; Chen, W.; Yuan, H.; Xia, W.; Zhang, W. Interpretable Machine Learning for Predicting Neoadjuvant Chemotherapy Response in Breast Cancer Using the Baseline Clinical and Pathological Characteristics. *Cancer Med.* **2025**, *14*, e71221. [CrossRef]
148. Häberle, L.; Erber, R.; Gass, P.; Hein, A.; Niklos, M.; Volz, B.; Hack, C.C.; Schulz-Wendtland, R.; Huebner, H.; Goossens, C.; et al. Prediction of pathological complete response after neoadjuvant chemotherapy for HER2-negative breast cancer patients with routine immunohistochemical markers. *Breast Cancer Res.* **2025**, *27*, 13. [CrossRef]
149. Qian, B.; Yang, J.; Zhou, J.; Hu, L.; Zhang, S.; Ren, M.; Qu, X. Individualized model for predicting pathological complete response to neoadjuvant chemotherapy in patients with breast cancer: A multicenter study. *Front. Endocrinol.* **2022**, *13*, 955250. [CrossRef]
150. Qu, F.; Luo, Y.; Peng, Y.; Yu, H.; Sun, L.; Liu, S.; Zeng, X. Construction and validation of a prognostic nutritional index-based nomogram for predicting pathological complete response in breast cancer: A two-center study of 1,170 patients. *Front. Immunol.* **2023**, *14*, 1335546. [CrossRef]
151. Shigematsu, H.; Sasada, S.; Hashizume, J.; Sasada, T.; Ozaki, S.; Noma, M.; Yoshiyama, T.; Noriko, G.; Kawamata, A.; Emi, A.; et al. Development and validation of a nomogram incorporating YpT stage for predicting ypN0 using multicenter data in clinically node-positive breast cancer. *Sci. Rep.* **2025**, *16*, 2344. [CrossRef]
152. Huang, J.X.; Mei, J.S.; Chen, F.; Huang, J.H.; Tan, Y.T.; Wu, Y.W.; Liu, F.T.; Qiu, S.D.; Shi, C.G.; Lu, Y.; et al. The development and validation of a risk stratification system for assessing axillary status after neoadjuvant therapy in node-positive breast cancer: A multicenter, prospective, observational study. *Int. J. Surg.* **2025**, *111*, 3731–3741. [CrossRef]
153. Pfob, A.; Sidey-Gibbons, C.; Rauch, G.; Thomas, B.; Schaefgen, B.; Kuemmel, S.; Reimer, T.; Hahn, M.; Thill, M.; Blohmer, J.U.; et al. Intelligent Vacuum-Assisted Biopsy to Identify Breast Cancer Patients With Pathologic Complete Response (ypT0 and ypN0) After Neoadjuvant Systemic Treatment for Omission of Breast and Axillary Surgery. *J. Clin. Oncol.* **2022**, *40*, 1903–1915. [CrossRef]
154. Lin, R.; He, Z.; Han, J.; Wang, Z.; Chen, Y.; Mao, L.; Peng, Q.; Zhang, Z.; Li, T.; Huang, Z.; et al. A foundation model for predicting outcomes of neoadjuvant chemotherapy in breast cancer. *Int. J. Surg.* **2026**, *112*, 2225–2242. [CrossRef]
155. Ruelle, T.; Grinda, T.; Del Mastro, L.; Lacroix-Triki, M.; Pistilli, B.; Gessain, G. How artificial intelligence applied to digital pathology could guide treatment personalization in breast cancer. *ESMO Real World Data Digit. Oncol.* **2026**, *11*, 100662. [CrossRef]
156. Sántha, D.; Dobi, Á.; Iványi, M.; Válóczi, D.; Varga, Z.; Oláh, J.; Nikolényi, A. Efficacy and tolerability of scalp cooling in preventing alopecia during (neo)adjuvant chemotherapy for breast cancer. *Pathol. Oncol. Res.* **2026**, *32*, 1612403. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.