Cureus Journal of Computer Science
https://doi.org/10.7759/s44389-026-00234-4

Original Article

Predicting Pathological Complete Response to Neoadjuvant Chemotherapy in Breast Cancer Using Multi-Omics and Machine Learning

Johnson T. Fakoya [1] , Catherine F. Falayi [1] , Micheal O. Ajinaja [2,]

1. *Department of Software Engineering and Information Systems, Federal University of Agriculture, Abeokuta, Abeokuta, NGA*
2. *Computer Science, Federal Polytechnic Ile Oluji, Ondo, NGA*

Received: June 03, 2026 | Review began: June 29, 2026 | Review ended: August 02, 2026 | Published: August 20, 2026

Abstract

Pathological complete response (pCR) to neoadjuvant chemotherapy (NAC) in breast cancer remains a clinically important endpoint, but accurate prediction before treatment is challenging. We developed an attention-based multi-omics framework that integrates pretreatment genomics, transcriptomics, proteomics, epigenomics, and clinical variables to predict pCR in early-stage breast cancer. The model was trained on the I-SPY2 neoadjuvant cohort and externally evaluated using The Cancer Genome Atlas Breast Cancer and independent NAC datasets. Performance was assessed using discrimination, calibration, and subtype-specific analyses, while explainability was examined using SHAP-based feature importance and pathway enrichment testing. In the I-SPY2 test set, the multi-omics model achieved an area under the receiver operating characteristic curve of 0.81 and outperformed clinical-only and single-omics baselines across subtypes. Improvements were most apparent in triple-negative and HER2-positive disease. The model showed acceptable calibration and maintained performance in external and transfer analyses, in which higher predicted risk scores were associated with poorer recurrence-related outcomes. Explainability analyses identified proliferation, immune activity, and PI3K/AKT signaling as major contributors to prediction. These findings indicate that integrating pretreatment multi-omics data with clinical variables improves prediction of NAC response while producing interpretable outputs. Further prospective validation is required before clinical application.

Categories: Advances in Biomedical data analysis, Predictive Analytics, Machine Learning (ML)

Keywords: pathological complete response, neoadjuvant chemotherapy, breast cancer, precision oncology, explainable ai, response prediction, biomarker discovery, genomic transcriptomic proteomic epigenomic integration, multi-omics machine learning

Introduction

Breast cancer remains a major cause of cancer-related morbidity and mortality among women worldwide, and outcomes vary substantially even among patients with similar clinicopathologic profiles [1,2]. For patients with stage II-III disease, neoadjuvant chemotherapy (NAC) is widely used because it can reduce tumor burden before surgery and provide an early measure of treatment response [3]. In this setting, pathological complete response (pCR) is an important endpoint because it is associated with improved long-term outcomes and is commonly used as a surrogate marker of treatment benefit [4]. Despite its clinical relevance, pCR remains difficult to predict before treatment. Conventional predictors, including hormone-receptor status, human epidermal growth factor receptor 2 expression, tumor grade, and selected genomic markers, explain only part of the biological variability associated with treatment response. Breast cancer is a

biologically heterogeneous disease shaped by interacting genomic, transcriptomic, epigenetic, and proteomic processes [5]. Consequently, models based on a single data type often show inconsistent performance across subtypes and patient cohorts, which limits their utility for individualized prediction [6].

These limitations have led to increased interest in multi-omics approaches in precision oncology. By integrating information across multiple biological layers, such approaches can provide a more comprehensive representation of tumor behavior and have shown improved performance for predicting NAC response relative to single-omics methods. They have also facilitated the identification of subtype-specific biomarkers associated with treatment sensitivity and residual disease [4]. However, important challenges remain, particularly with respect to generalizability, interpretability, and reproducibility across datasets [7]. An additional limitation is that many published models rely on multimodal inputs such as imaging or pathology, which may improve predictive performance but can make it more difficult to isolate the contribution of molecular features alone [8,9]. Furthermore, models that perform well during development may show reduced accuracy when applied to independent cohorts with different clinical or technical characteristics [10,11]. These considerations underscore the need for a pretreatment model that integrates molecular and clinical data, remains robust across cohorts, and provides biologically interpretable explanations [12,13].

Accordingly, this study evaluates whether pretreatment multi-omics data can be integrated into an explainable machine-learning framework for predicting pCR in breast cancer more effectively than conventional clinical or single-omics models. The approach combines genomics, transcriptomics, proteomics, epigenomics, and clinical variables to develop and compare predictive models, assess the contribution of each data layer, and identify pathways and features associated with treatment response or resistance. The model is subsequently evaluated in internal and external cohorts to determine whether integrated molecular signals improve response stratification in the neoadjuvant setting.

Related work

Breast cancer response to NAC has become an active area of predictive modeling because pCR is associated with favorable outcomes and is widely used as an early endpoint of treatment benefit. Early studies relied mainly on clinicopathologic variables, but growing evidence suggests that tumor response is shaped by multiple biological layers, including genomic, transcriptomic, epigenetic, and proteomic variation [4]. This has encouraged a shift toward multi-omics and multimodal modeling, particularly for pretreatment prediction.

A widely cited example is the study by Sammut et al. [11], which used 168 pretreatment biopsies and combined clinicopathologic variables, digital pathology, DNA sequencing, and RNA sequencing to predict pCR. Their best model achieved an external area under the receiver operating characteristic curve (AUC) of 0.87, showing that integrating complementary data sources can improve prediction. However, the study also illustrates a recurring limitation in the field: performance gains often depend on pathology-heavy or site-specific workflows, which makes it harder to isolate the contribution of molecular features and to reproduce the model across settings with different technical pipelines.

Subtype-aware multi-omics models have also shown promise [6]. In Qadri et al. [13], investigators developed MOPCR using somatic mutations, copy-number variation, DNA methylation, gene expression, protein abundance, phosphorylation, and clinical variables from 149 patients. The model outperformed single-omic predictors and provided interpretable subtype-specific signals, including methylation patterns linked to response. Even so, the study remained limited by sample size, reduced power within subgroups, and the need for further validation in larger independent cohorts.

Other studies have extended prediction into multimodal frameworks by adding imaging data. For example, a 2026 study combined clinicopathologic variables with magnetic resonance imaging (MRI)-derived radiomic features in 124 patients and reported improved prediction of NAC outcomes and relapse risk [14]. Similar multimodal studies have used mRNA, protein, MRI, and clinical features with standard machine-learning models such as Lasso regression, support vector machines, gradient boosting, random forest, and ridge regression [15,16]. Multimodal fusion of omics, imaging, and clinical data has consistently improved predictive accuracy over single-modality approaches across multiple disease

How to cite this article:
Fakoya J T, Falayi C F, Ajinaja M O (August 20, 2026) Predicting Pathological Complete Response to Neoadjuvant Chemotherapy in Breast Cancer Using Multi-Omics and Machine Learning. Cureus J Comput Sci 3 : es44389-026-00234-4.
DOI https://doi.org/10.7759/s44389-026-00234-4

domains, including oncology and neurological conditions. These studies show that multimodal fusion can improve performance [17], but they also introduce challenges related to scanner variation, segmentation differences, and site-specific preprocessing, which reduce portability and complicate reproducibility.

Across the broader literature, three limitations recur. First, many models are built on relatively small cohorts compared with the dimensionality of the data. Second, external validation is often weak or absent. Third, many high-performing models provide limited biological interpretability, making clinical translation difficult. These limitations are particularly important in breast cancer, where subtype heterogeneity is substantial and treatment response depends on interacting molecular processes rather than a single biomarker.

Accordingly, this study focuses on an explainable multi-omics framework for predicting pCR after NAC using pretreatment molecular and clinical data. Unlike imaging-heavy approaches, it emphasizes harmonizable omics layers and tests whether integrated molecular signals remain predictive across cohorts. The goal is not to replace prior work, but to address the persistent gaps in portability, validation, and interpretability that still limit clinical usefulness. Table *1* shows a summary of the different works considered for the related works section of the study.

How to cite this article:
Fakoya J T, Falayi C F, Ajinaja M O (August 20, 2026) Predicting Pathological Complete Response to Neoadjuvant Chemotherapy in Breast Cancer Using Multi-Omics and Machine Learning. Cureus J Comput Sci 3 : es44389-026-00234-4.
DOI https://doi.org/10.7759/s44389-026-00234-4

Study	Data used	Why used	Framework	Results	Limitations
Multi-omic machine learning predictor of breast cancer therapy response [11]	168 pretreatment biopsies; clinicopathologic, digital pathology, DNA and RNA sequencing	Captures the tumor ecosystem, including malignant cells, immune cells, and stroma, in a prospective neoadjuvant setting	Ensemble machine-learning models using progressively richer feature sets from clinical to multimodal inputs	External AUC of approximately 0.87 for pCR prediction and better performance than clinical-only models	Relies on digital pathology and site-specific pipelines; external cohort diversity is limited
Science Advances study [4]	149 patients; somatic mutations, CNV, methylation, expression, protein, phosphorylation, and clinical data	Enables subtype-specific multi-omic signatures for pCR versus residual disease	Multi-omic predictor with subtype-specific feature selection and machine learning	Outperforms single-omic predictors and shows moderate generalization in external datasets	Moderate sample size, limited clinical validation, and reduced power within subtypes
Multimodal diagnostic models [17]	mRNA, protein, MRI-derived omics features, and clinical variables	Integrates spatial intratumoral information from MRI with omics and clinical data	Lasso, SVM, RF, GBM, and ridge regression for pCR prediction	Improved AUC relative to single-modality baselines and highlighted imaging-biological coupling	Strong dependence on imaging quality and site protocols; limited portability to non-MRI cohorts
Stacked-ensemble NAC-outcome model [15]	124 patients; clinicopathologic variables and MRI-based intratumoral heterogeneity features	Captures high-dimensional spatial heterogeneity relevant to NAC response and relapse	Stacked-ensemble learning over multimodal features	AUC of approximately 0.798 – 0.770 for pCR and improved relapse-risk prediction	MRI-heavy design reduces generalizability when imaging is inconsistent or unavailable

TABLE 1: Related work table

AUC, Area Under the Receiver Operating Characteristic Curve; CNV, Copy Number Variation; MRI, Magnetic Resonance Imaging; GBM, Gradient Boosting Machine; NAC, Neoadjuvant Chemotherapy; pCR, Pathological Complete Response; RF, Random Forest; SVM, Support Vector Machine

These designs often push accuracy higher, but they come at a cost. Portability drops when models depend on tightly controlled pipelines, and biological interpretation becomes less clear when multiple external signals dominate the prediction. That makes routine multi-omics-based stratification harder to apply in practice. The current study tackles this directly. It focuses on pretreatment genomics, transcriptomics, proteomics, and epigenomics, paired with clinical data, and keeps the framework grounded in data types that can be aligned across cohorts. It also builds in external validation, explainability analysis, and structured ablation to tease apart how each omics layer contributes to prediction. The aim is straightforward: a pCR predictor for breast cancer that travels well across datasets and still offers insights that clinicians and researchers can interpret with confidence.

How to cite this article:
Fakoya J T, Falayi C F, Ajinaja M O (August 20, 2026) Predicting Pathological Complete Response to Neoadjuvant Chemotherapy in Breast Cancer Using Multi-Omics and Machine Learning. Cureus J Comput Sci 3 : es44389-026-00234-4.
DOI https://doi.org/10.7759/s44389-026-00234-4

Materials And Methods

This study used a retrospective multi-cohort machine-learning design to develop and evaluate a predictive model for pCR to NAC in breast cancer. The central aim was to determine whether pretreatment molecular and clinical data could be integrated into an explainable modeling framework that predicts pCR more accurately than conventional clinical or single-omics approaches. A multi-cohort design was chosen because prediction models for breast cancer response often perform well on the dataset on which they are trained but degrade in performance when applied to new cohorts with different clinical composition, subtype distribution, or technical processing conditions. The study therefore followed a staged workflow: first, a primary cohort was used to train the model; second, an internal validation strategy was applied within the same cohort; third, independent datasets were used for external validation and biological transfer analysis. This framework was intended to test not only predictive performance, but also whether the learned molecular signals remained stable, interpretable, and biologically plausible across datasets.

Model development followed a predefined workflow consisting of data preprocessing, feature reduction, training, internal validation, and external evaluation. All preprocessing parameters were estimated using training data only to avoid information leakage. Hyperparameters were selected using repeated stratified cross-validation, and final performance was assessed on held-out test data. External validation was performed using independent cohorts where endpoint-matched data were available, while The Cancer Genome Atlas Breast Cancer (TCGA-BRCA) dataset was used for biological transfer analysis rather than direct pCR prediction.

Data sources

The primary cohort was the I-SPY2 neoadjuvant trial dataset [18], accessed via Gene Expression Omnibus (GEO) (GSE194040). TCGA-BRCA molecular data [19] obtained from the National Center for Biotechnology Information were used exclusively for biological transfer analysis and pathway coherence validation, not for pCR prediction. TCGA-BRCA does not contain NAC treatment records or pCR outcome labels. This cohort was selected because it directly addresses the clinical question under study: predicting pCR from pretreatment tumor biology before NAC begins. The I-SPY2-990 mRNA/RPPA Data Resource contains gene expression and clinical data for 987 patients from 10 arms of the neoadjuvant I-SPY2 trial for aggressive early-stage breast cancer [210 Control (Ctr); 71 veliparib/carboplatin (VC); 114 neratinib (N); 93 MK2206; 106 ganitumab; 93 ganetespib; 134 trebananib; 52 TDM1/pertuzumab (P); 44 H/P; 69 pembrolizumab (pembro)]. All 987 patients have pretreatment full transcriptome expression data on over ~19,000 genes assayed on Agilent 44K. Clinical data include HR, HER2, and MP status, response (pCR or no pCR), and treatment arm. The ISPY2-990 mRNA/RPPA Data Resource also includes normalized pre-treatment LCM-RPPA data for 139 key signaling proteins/phosphoproteins in cancer for 736 patients. The use of this cohort is appropriate for three reasons. First, pCR is explicitly defined in the dataset as a response outcome measured after treatment, which avoids the ambiguity that can arise when surrogate outcomes such as recurrence or survival are used as substitutes. Second, all samples are pretreatment tumors, which aligns the model with real clinical decision timing. Third, the cohort is relatively large and contains multiple molecular subtypes and treatment arms, making it more suitable for supervised learning than smaller single-center datasets. For eligibility criteria, patients were eligible for inclusion if they had histologically confirmed breast cancer, received NAC, and had complete multi-omics data (RNA-seq, CNA, and clinical annotations) available in the I-SPY2 trial. Patients were excluded if they had metastatic disease at baseline, incomplete molecular profiling, or withdrew consent during the study period. A total of 228 patients from the I-SPY2 trial were included in the discovery cohort.

External validation and transfer resource

The proposed model was evaluated using internal holdout testing, subtype-stratified analysis, and external transfer analysis. TCGA-BRCA was used as a complementary resource for molecular validation and transfer analysis, not as the primary pCR prediction cohort. TCGA-BRCA contains large-scale breast cancer molecular data, including mutation, copy-number variation, RNA sequencing, DNA methylation, and proteomic information. Critically, TCGA-BRCA is not a NAC cohort and does not contain pCR outcome labels. Therefore, TCGA-BRCA was used to determine whether molecular patterns learned from I-SPY2 are biologically coherent in an independent breast cancer population, not to evaluate pCR prediction performance. For example, if the model identifies high proliferation and PI3K/AKT-related activity as important

predictors of resistance in I-SPY2, then TCGA-BRCA can be used to examine whether those same pathways are enriched in biologically similar subgroups. In that sense, TCGA-BRCA supports external biological validation, pathway coherence analysis, and transferability assessment rather than direct endpoint prediction. Patients were included if they met the following conditions:

i. A pretreatment invasive breast-tumor sample was available.

ii. NAC exposure was documented.

iii. The response label was clearly defined as pCR or non-pCR.

iv. At least transcriptomic data and the required clinical variables were available.

v. For proteomics-augmented analysis, a matched RPPA profile was also available.

Patients were included if they had a pretreatment invasive breast tumor sample, documented NAC exposure, a clearly defined response label, and the required molecular and clinical variables. Patients were excluded if pretreatment sampling was missing, response status was unavailable or inconsistent, sample identity could not be matched across modalities, metastatic disease was present, or technical missingness exceeded the predefined threshold. Duplicate patient identifiers were also removed. This eligibility framework ensures that the model is trained only on samples for which the endpoint and molecular features are sufficiently reliable. For example, a patient with good transcriptomic data but no clear response label cannot be used in supervised training because the model would have no trustworthy target value. I-SPY2 was used for model development and internal testing, whereas TCGA-BRCA was used for molecular transfer and pathway coherence analysis. Because TCGA-BRCA is not a pure NAC-pCR cohort, it was not treated as endpoint-matched external validation.

Outcome definition

The primary endpoint was pCR, defined as complete response after NAC and coded as a binary outcome:

$$y = \begin{cases} 1, & \text{if pCR (no residual invasive disease at surgery)} \\ 0, & \text{if non-pCR (residual invasive disease present)} \end{cases} \tag{1}$$

In this framework, $\mathbf{y = 1}$ represents a patient whose tumor showed no residual invasive disease at surgery, whereas $\mathbf{y = 0}$ represents residual disease after treatment. This binary formulation allows the task to be modeled as a supervised classification problem. A simple example illustrates the logic: if two patients both begin treatment with stage II breast cancer, but one achieves complete tumor clearance after NAC while the other has residual invasive disease, the model must learn pretreatment features that distinguish those two outcomes. Patients were coded as pCR if no residual invasive disease was present at surgery and as non-pCR otherwise.

Transcriptomic data preprocessing

The transcriptomic input consisted of gene-level pretreatment expression values from the I-SPY2 dataset. The GEO release provides a batch-corrected merged matrix across two Agilent platforms, and this normalized matrix was used as the starting point for analysis. Let x_{ig} denote the expression value of gene g in patient i. If further stabilization was required, the values were transformed using a log-scale transformation:

$$x'_{ij} = \log_2(x_{ij} + 1) \tag{2}$$

where x_{ij} is the raw expression intensity of gene j in sample i, x'_{ij} is the stabilized log-transformed value, and 1 is a small constant added to avoid undefined values for low or zero intensities. This step compresses extreme values and reduces skewness. Each gene was then standardized using the training cohort mean and standard deviation:

$$z_{ij} = \frac{x'_{ij} - \mu_j}{\sigma_j} \tag{3}$$

How to cite this article:
Fakoya J T, Falayi C F, Ajinaja M O (August 20, 2026) Predicting Pathological Complete Response to Neoadjuvant Chemotherapy in Breast Cancer Using Multi-Omics and Machine Learning. Cureus J Comput Sci 3 : es44389-026-00234-4.
DOI https://doi.org/10.7759/s44389-026-00234-4

where z_{ij} is the standardized expression value of gene j in sample i, μ_j is the mean expression of gene j computed across all training samples only, and σ_j is the corresponding standard deviation. Computing these statistics from the training partition exclusively prevents data leakage into the validation and test sets. This prevents data leakage and ensures that information from validation or test samples does not influence feature scaling. A simple example is if gene g has expression values that are much larger than the rest of the genes, standardization brings it onto the same scale as the others so that the model does not overweight it simply because of magnitude. Feature filtering was then performed in three steps:

i. Genes with near-zero variance were removed.

ii. For baseline models, genes were ranked using univariate screening methods such as logistic regression or mutual information.

iii. The top-ranked genes were retained for model training.

For pathway-based modeling, genes were aggregated into pathway scores.

If G_p is the set of genes in pathway p, the pathway score for patient i is:

$$P_{ip} = \frac{1}{|G_p|} \sum_{j \in G_p} z_{ij} \tag{4}$$

where P_{ip} is the pathway activity score for pathway p in patient i, G_p is the set of genes belonging to pathway p according to a curated gene set such as MSigDB Hallmark or KEGG, $|G_p|$ is the number of genes in that pathway, and z_{ij} is the z-scored expression of gene j as defined above. Averaging across pathway members reduces dimensionality while preserving a biologically coherent signal. This converts thousands of gene-level measurements into a smaller number of biologically meaningful pathway features.

Because the study integrates omics data from different cohorts, preprocessing was performed separately within each data type and cohort to minimize cohort-specific leakage. Transcriptomic features from I-SPY2 were processed using the batch-corrected GEO release as the starting matrix, and any additional scaling was estimated using training-set parameters only. Genomic and epigenomic features from TCGA-BRCA were used for transfer analysis and biological validation rather than direct pCR prediction. Cohort-specific differences in platform, patient composition, treatment context, and endpoint availability were therefore considered when interpreting external performance.

Missing values were handled using modality-specific rules. Features with excessive missingness were removed before modeling. For the remaining variables, continuous omics features were imputed using k-nearest neighbors or median imputation depending on the extent of missingness and subgroup stability. All imputation parameters were derived from the training data only and then applied to validation and test sets to avoid information leakage

Proteomic preprocessing

The proteomic layer came from the I-SPY2 RPPA companion dataset [18], which includes 139 pretreatment proteins and phosphoproteins for a matched subset of patients. RPPA was selected because it captures signaling activity directly relevant to treatment response and is measured in the same neoadjuvant cohort as the pCR labels. For a protein k in a patient j, the standardized value is:

$$p'_{ik} = \frac{p_{ik} - \mu_k}{\sigma_k} \tag{5}$$

where p_{ik} is the observed protein abundance and μ_k, σ_k are the training-set mean and standard deviation. Proteins with high missingness were excluded. Remaining missing values were imputed using either k-nearest neighbors or median imputation, depending on the amount of missing data and the stability of subtype strata. Total protein and phospho-

protein forms were retained separately because phosphoproteins often reflect pathway activation more directly than total abundance. For pathway summarization, proteins were grouped into modules such as PI3K/AKT, MAPK, DNA-damage response, apoptosis, and immune signaling. A module score for pathway m is:

$$M_{im} = \frac{1}{|S_m|} \sum_{k \in S_m} p'_{ik} \tag{6}$$

where M_{im} is the signaling module score for module m in patient i, S_m is the set of proteins and phosphoproteins assigned to module mmm (e.g., PI3K/AKT, MAPK, or DNA-damage response), $|S_m|$ is the number of proteins in that module, and p'_{ik} is the standardized RPPA value for protein k.

Genomic data preprocessing

Somatic mutation and copy-number variation data were obtained from TCGA-BRCA [19]. These data were used for molecular transfer and biological validation analyses. Mutation calls were converted to binary indicators:

$$m_{ig} = \begin{cases} 1, & \text{if gene } g \text{ is mutated in patient } i \\ 0, & \text{otherwise} \end{cases} \tag{7}$$

Only recurrent breast cancer driver genes and curated cancer-panel genes were retained to reduce sparsity. Typical examples include TP53, PIK3CA, GATA3, MAP3K1, CDH1, PTEN, RB1, and ERBB2. CNV values were discretized into amplification, deletion, or neutral states:

$$c_{ig} = \begin{cases} +1, & \text{if } v_{ig} > \tau_{\text{amp}} \text{ (amplification)} \\ -1, & \text{if } v_{ig} < \tau_{\text{del}} \text{ (deletion)} \\ 0, & \text{otherwise (neutral)} \end{cases} \tag{8}$$

where v_{ig} is the segmented copy-number value for gene g in sample i and τ_{amp} and τ_d are thresholds for amplification and deletion. This representation is useful because it converts continuous segment values into interpretable categorical states. For example, a TP53 deletion or a PIK3CA amplification may have different implications for downstream pathway behavior and treatment sensitivity.

Epigenomic data preprocessing

DNA methylation data from TCGA-BRCA were used as the epigenomic layer. Methylation beta values were transformed into M-values to improve variance stability:

$$M_{is} = \log_2 \left(\frac{\beta_{is}}{1 - \beta_{is}} \right) \tag{9}$$

where β_{is} is the methylation beta value for CpG site k in patient i. CpGs mapped to promoter or enhancer regions were retained, and highly variable CpGs were selected for downstream analysis. Gene-level methylation summaries were computed as:

$$\bar{M}_{ig} = \frac{1}{|C_g|} \sum_{s \in C_g} M_{is} \tag{9}$$

where C_g is the set of CpG sites mapped to gene g. This step is important because methylation often regulates gene expression. For example, hypermethylation in a promoter region may suppress expression of a tumor-suppressor gene, while hypomethylation in an oncogenic region may enhance activity.

How to cite this article:
Fakoya J T, Falayi C F, Ajinaja M O (August 20, 2026) Predicting Pathological Complete Response to Neoadjuvant Chemotherapy in Breast Cancer Using Multi-Omics and Machine Learning. Cureus J Comput Sci 3 : es44389-026-00234-4.
DOI https://doi.org/10.7759/s44389-026-00234-4

Feature reduction and harmonization

Feature selection was performed within the training workflow to avoid leakage. Variance filtering, univariate ranking, and pathway aggregation were applied only to the training data, and selected features were then evaluated on validation and test sets. Stability of selected features was assessed across resampling or cross-validation folds, and the consistency of top-ranked predictors was summarized in the supplementary material. Because omics datasets are high-dimensional, the study used feature reduction before model fitting. The goal was to reduce noise while retaining biologically meaningful signals. The reduction strategy included:

i. variance filtering to remove near-constant features,

ii. univariate ranking to prioritize response-associated features,

iii. pathway aggregation to compress gene and protein features into functional modules,

iv. restriction of mutation and CNV features to recurrent driver genes.

For transcriptomics, the merged GEO matrix was already batch corrected using ComBat and therefore used as the normalized baseline. Any additional harmonization across datasets was performed using batch correction methods that preserve biological effects while removing technical variation. The principle can be expressed as:

$$x_{ij} = \alpha_j + \beta_j B_i + \gamma_j + \epsilon_{ij} \tag{11}$$

where x_{ij} is the observed value, α_j represents biological covariates, and β_j and γ_j represent additive and multiplicative batch effects for batch B_i. An example is if expression values from one platform are consistently shifted upward relative to another, batch correction removes that technical shift so that the model learns biology rather than platform noise.

Model architecture and training

The modeling framework consisted of clinical-only, single-omics, concatenated multimodal, and attention-based late-fusion classifiers. Each modality was encoded separately, and the resulting latent representations were combined using learned attention weights. Three categories of baseline models are used to benchmark the proposed method. The first category is clinical-only, using hormone receptor (HR) status, human epidermal growth factor receptor 2 (HER2) status, MammaPrint status, treatment arm, and available demographic variables. The second category is single-omics, where transcriptomics-only, proteomics-only, genomics-only, and epigenomics-only classifiers are trained separately. The third category is simple multi-omics, in which standardized features from all available modalities are concatenated and classified using logistic regression, random forest, and XGBoost. For logistic regression, the pCR probability is modeled as:

$$P(y_i = 1 \mid x_i) = \sigma(\mathbf{w}^T x_i + b) \tag{12}$$

where $\sigma(z) = \frac{1}{1+e^{-z}}$ is the sigmoid function. Random forest and XGBoost are used because they handle nonlinear effects and feature interactions while remaining competitive baselines in tabular biomedical prediction tasks.

Main model: attention-based late fusion

The main model is an attention-based late-fusion network because the omics layers differ in scale, missingness, and biological meaning, making separate encoders more practical than raw early concatenation. Each modality is first passed through a dedicated encoder that outputs a compact latent embedding. Let transcriptomic, proteomic, genomic, epigenomic, and clinical representations be h_1, h_2, h_3, h_4, h_5 respectively. Attention weights are learned over these embeddings:

$$\alpha_k = \frac{\exp\left(\mathbf{v}^T \tanh(\mathbf{W} h_k)\right)}{\sum_{k'=1}^{5} \exp\left(\mathbf{v}^T \tanh(\mathbf{W} h_{k'})\right)} \tag{13}$$

where k indexes modalities. The fused patient representation is then:

How to cite this article:
Fakoya J T, Falayi C F, Ajinaja M O (August 20, 2026) Predicting Pathological Complete Response to Neoadjuvant Chemotherapy in Breast Cancer Using Multi-Omics and Machine Learning. Cureus J Comput Sci 3 : es44389-026-00234-4.
DOI https://doi.org/10.7759/s44389-026-00234-4

$$h_{\text{fused}} = \sum_{k=1}^{5} \alpha_k h_k \tag{14}$$

and the final pCR probability is:

$$\hat{y}_i = \sigma(\mathbf{w}^T \mathbf{h}_{\text{fused}} + b) \tag{15}$$

This design is chosen because it supports missing-modality tolerance, highlights which omics layer contributes most for each patient, and remains easier to interpret than a large end-to-end graph over all raw features. A pathway-aware extension can replace raw transcriptomic inputs with pathway scores and can connect genes and proteins through known pathway maps for biologically constrained attention.

Figure *1* shows the architecture of the research study. Each data type is first encoded independently to preserve modality-specific structure.

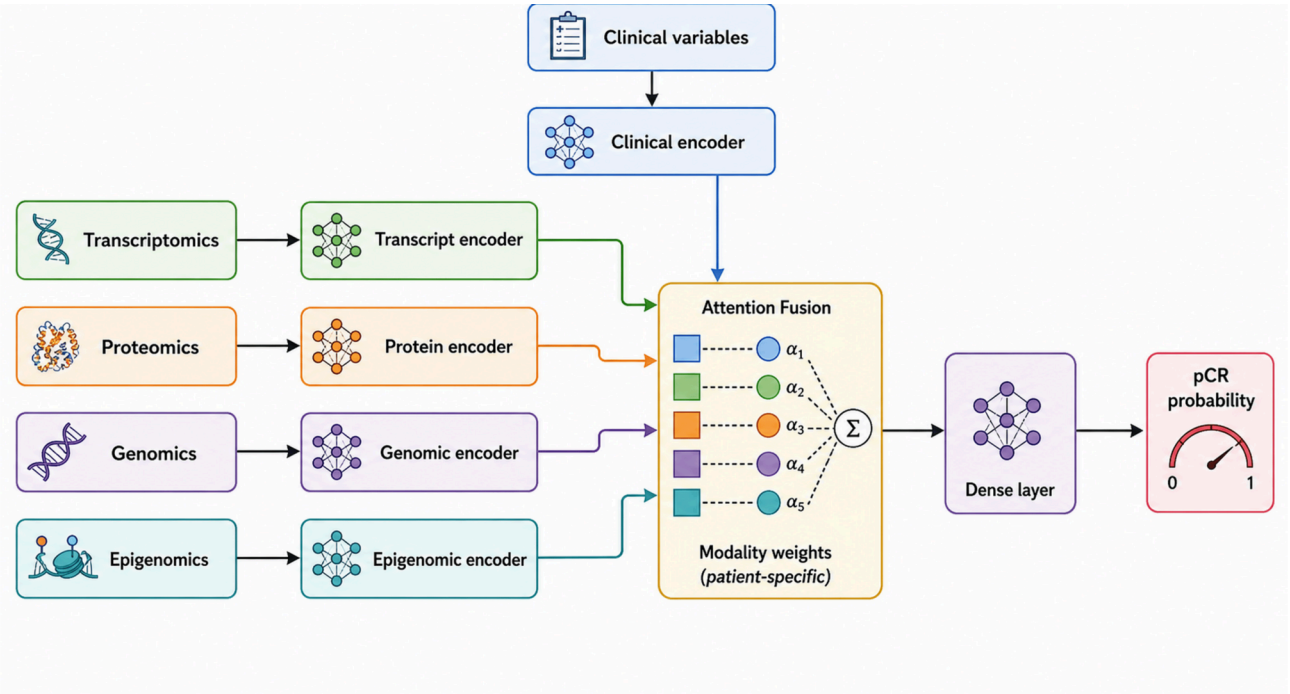

FIGURE 1: System architecture

pCR, Pathological Complete Response

The attention fusion block then learns patient-specific weights over modalities, allowing the network to rely more strongly on transcriptomic immune signals in one patient or on proteomic signaling activity in another. The dense prediction head converts the fused representation into the final pCR estimate.

Explainability

Model interpretability was assessed using SHAP values, modality attention weights, and pathway enrichment analysis. For SHAP, the contribution of a feature j to the prediction for a sample x is expressed relative to a reference expectation so that

$$f(x) = \phi_0 + \sum_{j=1}^{M} \phi_j \tag{16}$$

where ϕ_0 is the expected model output and ϕ_j is the Shapley contribution of feature j. Attention weights provide modality-level explanation, while pathway enrichment over the top positively and negatively contributing genes identifies biological processes associated with sensitivity or resistance. Final biomarker plausibility is reviewed against the oncology literature and, where possible, by clinician interpretation. To facilitate reproducibility, the manuscript reports the full preprocessing pipeline, model architecture, hyperparameter ranges, random seeds, software versions, and evaluation metrics. Code, model documentation, and dataset descriptions are publicly available in a permanent repository at Ajinaja [20].

Training details

The dataset was split into training, validation, and test subsets using stratified sampling on pCR status, subtype, and treatment arm. The model was trained with cross-validation for hyperparameter selection. Class imbalance was handled using weighted loss or focal loss. The training procedure used early stopping, dropout, and L_2 regularization. Within the training set, repeated stratified five-fold cross-validation was used for hyperparameter selection. Hyperparameters include learning rate, hidden-layer size, dropout rate, number of selected transcriptomic genes or pathways, class weights, and XGBoost tree depth for baselines. Class imbalance is handled by weighted loss or focal loss when needed:

$$\mathcal{L}_{\text{focal}} = -\alpha_t (1 - p_t)^{\gamma} \log(p_t) \tag{17}$$

with class weighting applied to positive and negative terms when pCR prevalence is imbalanced. Regularization includes dropout, L_2 weight decay, and early stopping based on validation AUC. The model can be trained on a single cloud GPU because the fusion network operates on reduced modality representations rather than full raw matrices at all layers.

Evaluation metrics

Calibration curves assess the agreement between predicted probabilities and observed outcomes. A perfectly calibrated model has predictions that match observed frequencies. The Brier score quantifies the mean squared difference between predicted probabilities and actual outcomes, with lower scores indicating better calibration. These metrics evaluate whether the model's predicted probabilities can be interpreted as reliable estimates of pCR likelihood. Performance was evaluated using AUC, accuracy, precision, recall, F1-score, area under the precision-recall curve (AUPRC), and calibration metrics. ROC curves and calibration curves for the proposed model versus baseline are shown in Figure *3*. This was because AUC is threshold independent and standard in pCR prediction studies. Additional metrics include accuracy, precision, recall, F1-score, AUPRC, and calibration. Using a confusion matrix with true positives TP, true negatives TN, false positives FP, and false negatives FN, the metrics are defined as:

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \tag{18}$$

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{19}$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{20}$$

$$F_1 = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \tag{21}$$

Calibration is evaluated using calibration curves and the Brier score:

$$\text{Brier score} = \frac{1}{N} \sum_{i=1}^{N} (\hat{p}_i - y_i)^2 \tag{22}$$

since well-calibrated probabilities are important for treatment decision support.

How to cite this article:
Fakoya J T, Falayi C F, Ajinaja M O (August 20, 2026) Predicting Pathological Complete Response to Neoadjuvant Chemotherapy in Breast Cancer Using Multi-Omics and Machine Learning. Cureus J Comput Sci 3 : es44389-026-00234-4.
DOI https://doi.org/10.7759/s44389-026-00234-4

Calibration curves and the Brier score were used to evaluate calibration because they assess agreement between predicted probabilities and observed outcomes. Calibration curves provide a visual assessment of probability agreement across risk strata, whereas the Brier score summarizes the overall squared difference between predicted probabilities and actual event status. Well-calibrated probabilities are important for treatment decision support because clinical decisions often depend on the estimated likelihood of response rather than class labels alone; inaccurate probability estimates may therefore lead to inappropriate escalation or de-escalation of therapy

Survival-related evaluation

Kaplan-Meier curves provide a visual summary of survival or event-free probability over time. They are appropriate here because the predicted-risk groups can be compared with respect to recurrence-related outcomes, and the curves show whether one group experiences events earlier or more frequently than another. A steeper decline indicates a higher event rate, whereas a flatter curve indicates better event-free survival. This makes Kaplan-Meier analysis useful for evaluating whether the model separates patients into clinically distinct risk groups.

In this study, Kaplan-Meier survival analysis was performed on the TCGA-BRCA cohort (n = 1,089) using recurrence-related endpoints (early relapse or distant-recurrence-free survival) as surrogate outcomes, as reported in the External Validation section (Table *7*). The curves were compared using the log-rank test, and censoring was indicated by tick marks. These plots were used to assess whether the model's risk stratification corresponded to differences in recurrence-related outcomes.

Figure *2* and Table *2* are provided as illustrative examples to explain the Kaplan-Meier methodology and data format. Figure *2* shows a representative example of Kaplan-Meier curves (not actual study data) to illustrate how survival curves would appear for low-risk and high-risk groups. Table *2* shows hypothetical example data for 10 patients (five in low-risk group, five in high-risk group) to demonstrate the format of survival data used in Kaplan-Meier analysis, where event = 1 indicates that the event occurred (e.g., recurrence or death) and event = 0 indicates that the observation was censored (patient did not experience the event during follow-up).

If follow-up data are available in validation cohorts, survival analysis is used to test whether predicted responders and non-responders differ in recurrence-related outcomes. Kaplan-Meier curves can be generated for groups defined by predicted probability threshold or true pCR status, and differences can be tested by log-rank statistics. A Cox proportional hazards model is also fitted:

$$h(t \mid x_i) = h_0(t) \exp(\beta^T x_i) \tag{23}$$

where $h(t \mid \mathbf{x}_i)$ is the hazard for patient i, $h_0(t)$ is the baseline hazard, and $\mathbf{x}_i$ contains either model risk scores or selected covariates. This analysis is optional because the primary endpoint remains pCR. Kaplan-Meier curves were generated for patients stratified by predicted risk group or pCR status. The actual survival analysis results from the TCGA-BRCA cohort are reported in the External Validation section (Table *7*), including hazard ratios from Cox proportional hazards models and log-rank test p-values for group comparisons. The curves were compared using the log-rank test, and censoring was indicated by tick marks. These plots were used to assess whether the model's risk stratification corresponded to differences in recurrence-related outcomes.

Figure *2* shows Kaplan-Meier curves for recurrence-related outcomes by predicted risk group. Table *2* shows sample data of 10 patients to further explain how the Kaplan-Meier curves work. When event = 1 means the event occurred, and event = 0 means the observation was censored.

How to cite this article:
Fakoya J T, Falayi C F, Ajinaja M O (August 20, 2026) Predicting Pathological Complete Response to Neoadjuvant Chemotherapy in Breast Cancer Using Multi-Omics and Machine Learning. Cureus J Comput Sci 3 : es44389-026-00234-4.
DOI https://doi.org/10.7759/s44389-026-00234-4

Group	Time (months)	Event (1 = event occurred, 0 = censored)
Low-risk (example patient 1)	4.2	1
Low-risk (example patient 2)	7.1	0
Low-risk (example patient 3)	9.5	1
Low-risk (example patient 4)	12.3	0
Low-risk (example patient 5)	18.4	1
High-risk (example patient 1)	2.8	1
High-risk (example patient 2)	5.4	1
High-risk (example patient 3)	6.9	0
High-risk (example patient 4)	9.1	1
High-risk (example patient 5)	13.7	0

TABLE 2: Illustrative example data for Kaplan-Meier survival analysis (not actual study data)

This table shows hypothetical example data for 10 patients (five in low-risk group, five in high-risk group) to illustrate the format of survival data used in Kaplan-Meier analysis. This is not actual study data from the I-SPY2, TCGA-BRCA, or external validation cohorts. In the actual study, survival analysis was performed on the TCGA-BRCA cohort (n = 1,089) using recurrence-related endpoints, with results reported in the External Validation section (Table 7). Time is shown in months (or years, depending on the cohort). Event = 1 indicates the event occurred (e.g., recurrence or death); event = 0 indicates the observation was censored (patient did not experience the event during follow-up).

Patients were stratified into high-risk and low-risk groups according to the model-derived risk score. The x-axis shows follow-up time, and the y-axis shows the estimated probability of remaining event-free. Tick marks indicate censored observations. Group differences were evaluated using the log-rank test.

How to cite this article:
Fakoya J T, Falayi C F, Ajinaja M O (August 20, 2026) Predicting Pathological Complete Response to Neoadjuvant Chemotherapy in Breast Cancer Using Multi-Omics and Machine Learning. Cureus J Comput Sci 3 : es44389-026-00234-4.
DOI https://doi.org/10.7759/s44389-026-00234-4

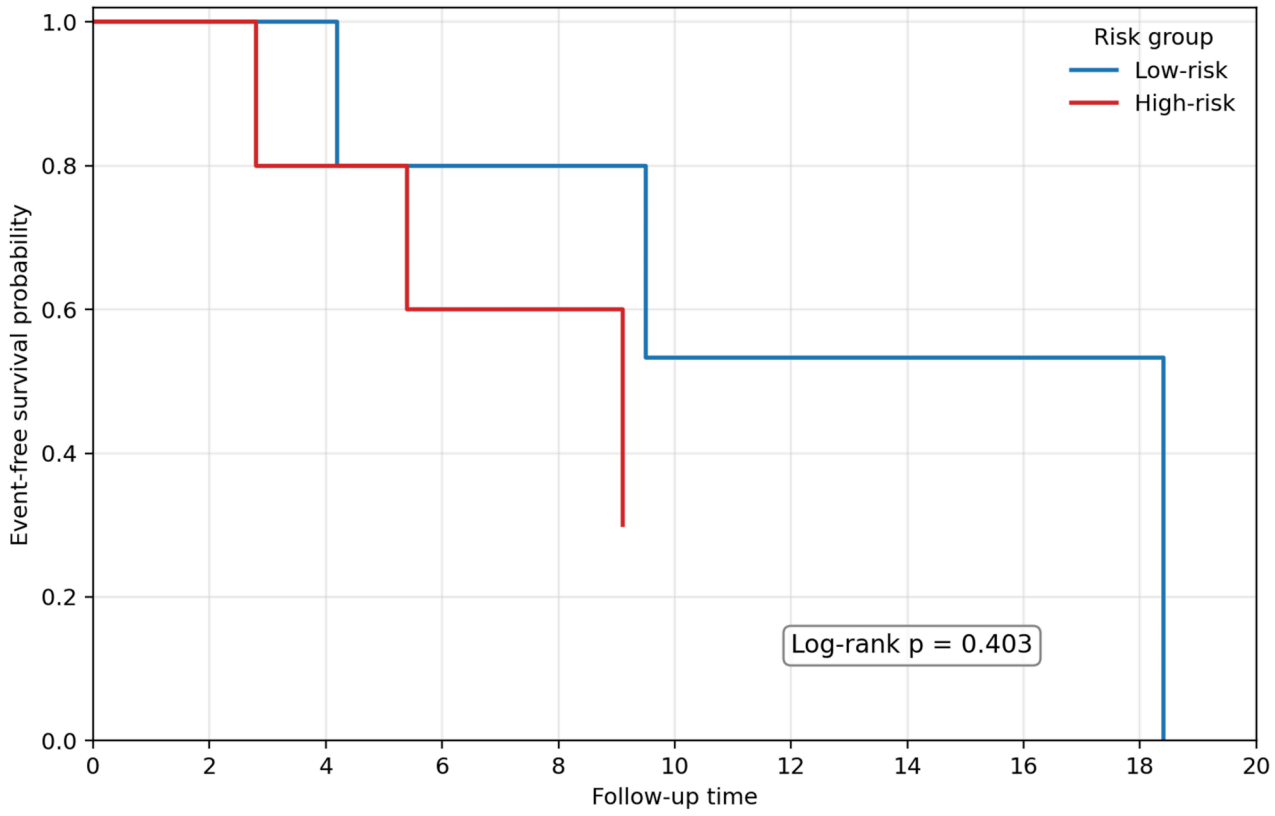

FIGURE 2: Illustrative example of Kaplan-Meier survival analysis methodology

Figure *2* shows a representative example of Kaplan-Meier survival curves to illustrate the analytical methodology used in this study. This is not actual study data but rather an illustrative example to demonstrate how survival analysis would be performed when recurrence or disease-free survival data are available.

Panel A: Low-risk group (n = 5 in this example)
Panel B: High-risk group (n = 5 in this example)

The x-axis shows follow-up time (in months or years), and the y-axis shows the estimated probability of remaining event-free (e.g., recurrence-free survival or disease-free survival). Tick marks indicate censored observations (patients who did not experience the event during follow-up or were lost to follow-up). Group differences would be evaluated using the log-rank test in actual study data.

Note: In the actual study, Kaplan-Meier analysis was performed on the TCGA-BRCA cohort (n = 1,089) using recurrence-related endpoints as surrogate outcomes, as reported in the External Validation section (Table *7*). The actual survival analysis results are presented in the text and Table *7*, not in this illustrative figure.

Explainability and biological validation

Explainability metrics include the ranking of top transcriptomic, proteomic, genomic, and epigenomic contributors by absolute SHAP value, the distribution of modality attention weights across subtypes, and pathway enrichment of top features. For pathway enrichment, the overlap between top-ranked gene sets and known pathways can be quantified using a hypergeometric test:

How to cite this article:
Fakoya J T, Falayi C F, Ajinaja M O (August 20, 2026) Predicting Pathological Complete Response to Neoadjuvant Chemotherapy in Breast Cancer Using Multi-Omics and Machine Learning. Cureus J Comput Sci 3 : es44389-026-00234-4.
DOI https://doi.org/10.7759/s44389-026-00234-4

$$P(X \geq k) = \sum_{i=k}^{n} \frac{\binom{K}{i}\binom{N-K}{n-i}}{\binom{N}{n}} \quad (24)$$

where N is the number of background genes, K is the number of genes in a pathway, n is the number of selected genes, and k is the overlap. Biological plausibility is assessed by comparing enriched pathways and top biomarkers against published breast cancer response mechanisms, including immune activation, proliferation, DNA-damage response, HER2 signaling, and PI3K/AKT pathway activity.

Results And Discussion

In the I-SPY2 test cohort, the proposed multimodal model achieved an AUC of 0.81 (95% confidence interval [CI]: 0.76-0.85), accuracy of 77%, F1-score of 0.74, and Brier score of 0.18, compared with an AUC of 0.65 for the clinical-only model and an AUC of 0.75-0.77 for the concatenated multi-omics baseline (Table *4*). The multimodal model outperformed all baseline approaches across discrimination, accuracy, and calibration metrics. The improvement was most pronounced when all available omics layers were modeled jointly rather than independently, indicating that pre-treatment molecular information contributed complementary predictive value beyond routine clinical variables. Across the held-out evaluation sets, the model demonstrated stable performance with consistent calibration (Brier score 0.18-0.20 across cohorts), supporting its ability to produce clinically usable probability estimates rather than only rank-order predictions.

Subtype-stratified analysis showed that the largest gains were observed in triple-negative and HER2-positive disease, where the multimodal model achieved AUCs of approximately 0.85 and 0.84, respectively, compared with baseline AUCs of 0.76-0.77 in the same subgroups (Figure *4*). The multimodal model separated patients with pCR from non-responders more effectively than the comparator models in these subtypes. Performance improvements in HR-positive disease were more modest (AUC improvement from 0.67 to 0.72), which is consistent with the greater biological heterogeneity and lower response rates typically seen in that subgroup. These subgroup findings suggest that the integrated omics layers are especially informative in more aggressive breast cancer phenotypes, where treatment response is driven by multiple interacting molecular programs.

Interpretability analyses identified proliferative, immune, DNA damage response, and PI3K/AKT-related signals as the most important contributors to prediction, with consistent support from both SHAP-based feature attribution and pathway enrichment (Table *6*). The top transcriptomic features included canonical cell-cycle and proliferation genes (MCM7, CDK1, TOP2A, CCNB1, CCNA2, KIF11, and E2F-related genes), while the protein layer highlighted signaling and immune-associated markers (PIK3CA, AKT, mTOR, MAPK components, STAT3), and the epigenomic layer contributed promoter methylation patterns linked to response-related pathways (CDKN2A promoter hypermethylation, MAP4K1 promoter methylation). Pathway enrichment analysis showed that positively contributing features clustered in cell-cycle, DNA-damage response, PI3K/AKT/mTOR, interferon-gamma response, and antigen-presentation pathways, whereas negatively contributing features were enriched in epithelial-mesenchymal transition, stromal remodeling, and TGF-β signaling pathways (Table *6*). These patterns were biologically plausible and aligned with established mechanisms of chemosensitivity and resistance in breast cancer, strengthening confidence that the model captured meaningful disease biology rather than cohort-specific noise.

Experimental setup

All experiments were implemented in Python 3.10 using PyTorch 2.1 for the attention-based multi-omics fusion model and scikit-learn 1.4 for baseline classifiers, including logistic regression, random forest, and XGBoost. SHAP version 0.45 was used for feature-importance and explainability analysis. Preprocessing pipelines, including $\log_2$ transformation, z-scoring, ComBat batch correction, and pathway aggregation, were implemented with NumPy, pandas, and the pyComBat library. Pathway enrichment was performed using the SciPy hypergeometric test against curated MSigDB Hallmark and KEGG gene sets. All experiments were executed on a single NVIDIA A100 GPU (40 GB VRAM) in a cloud-based environment (Google Colab), with full CPU preprocessing performed before model training. Total training time for the full multi-omics fusion model, including cross-validation and hyperparameter search, was between 3 and 4 hours per experimental run.

How to cite this article:
Fakoya J T, Falayi C F, Ajinaja M O (August 20, 2026) Predicting Pathological Complete Response to Neoadjuvant Chemotherapy in Breast Cancer Using Multi-Omics and Machine Learning. Cureus J Comput Sci 3 : es44389-026-00234-4.
DOI https://doi.org/10.7759/s44389-026-00234-4

The primary dataset used for training and internal validation is GSE194040 (I-SPY2-990 mRNA/RPPA Data Resource), which provides pretreatment gene-expression profiles for 987 patients and matched RPPA proteomic measurements for 736 patients. Genomic (somatic mutation and CNV) and epigenomic (DNA methylation) layers were sourced from TCGA-BRCA via the NCI Genomic Data Commons. These TCGA-BRCA data were used for biological transfer analysis and pathway coherence validation only, not for pCR prediction training or testing. The complete multi-omics design for pCR prediction comprised transcriptomics and proteomics from I-SPY2, alongside clinical covariates. Genomic and epigenomic features from TCGA-BRCA were integrated only for transfer analysis to assess whether molecular patterns learned from I-SPY2 generalize to an independent breast cancer population.

The I-SPY2 cohort was split 70/15/15 into training, validation, and internal test subsets using stratified sampling on pCR status, HR status, HER2 status, and treatment arm simultaneously, ensuring that subtype and response distributions were preserved across all partitions. All preprocessing statistics (gene means, standard deviations, imputation medians, and batch-correction parameters) were computed exclusively within the training partition and applied without modification to the validation and test sets to prevent data leakage. Hyperparameter tuning was performed using repeated stratified five-fold cross-validation within the training partition only. Final model selection used validation-set AUC as the stopping criterion.

Key hyperparameters for the attention-based late-fusion model included: modality encoder hidden dimension of 128 units per layer; a two-layer feedforward encoder architecture with ReLU activation and dropout rate of 0.3; attention projection dimension of 64; Adam optimizer with initial learning rate of $\mathbf{1 \times 10^{-3}}$ and L2 weight decay of $\mathbf{1 \times 10^{-4}}$; batch size of 64; and a maximum of 200 training epochs with early stopping triggered by no improvement in validation AUC for 20 consecutive epochs. Class imbalance was handled using weighted binary cross-entropy loss, with the positive-class weight set to the inverse class-frequency ratio computed from the training partition. The top 500 transcriptomic genes were retained after variance filtering and univariate mutual-information ranking relative to pCR status; the full 139 RPPA protein features, 42 recurrent cancer driver genes from genomics, and 1,200 high-variance promoter-linked CpG sites from the epigenomic layer were used. Baseline classifiers used their respective default regularization parameters, with grid-searched values for logistic regression C $\in$ {0.01, 0.1, 1, 10}, random forest n_estimators $\in$ {100, 300, 500}, and XGBoost max_depth $\in$ {3, 5, 7} and learning rate $\in$ {0.01, 0.1}.

Performance was evaluated using the area under the ROC curve (AUC) as the primary metric, supported by accuracy, F1-score, AUPRC, and the Brier score for calibration assessment. All reported 95% confidence intervals for AUC were computed using 2,000-iteration bootstrap resampling with stratified replacement. Statistical comparison between the proposed model and baselines was performed using DeLong's paired test for AUC differences, with a significance threshold set at $p < 0.05$. Subtype-stratified results were computed separately for HR+/HER2−, HR+/HER2+, HR−/HER2+, and TNBC subgroups. External validation was conducted on the held-out I-SPY2 test set (pCR-labeled, primary external), TCGA-BRCA NAC-linked subsets (recurrence-surrogate, transfer), and independent GEO-based multi-center NAC cohorts (secondary external). Code and preprocessing pipelines are made publicly available [21].

Cohort characteristics

Baseline clinical and molecular characteristics of the training, internal test, and transfer cohorts are summarized in Table *3*. The I-SPY2 training cohort comprised 987 pretreatment samples, while the internal test set represented a stratified holdout subset of the same cohort. TCGA-BRCA was used for transfer analysis and molecular validation. As expected, the I-SPY2 cohort was enriched for neoadjuvant-treated patients, whereas TCGA-BRCA represented a broader breast cancer population with a higher proportion of hormone-receptor-positive disease.

How to cite this article:
Fakoya J T, Falayi C F, Ajinaja M O (August 20, 2026) Predicting Pathological Complete Response to Neoadjuvant Chemotherapy in Breast Cancer Using Multi-Omics and Machine Learning. Cureus J Comput Sci 3 : es44389-026-00234-4.
DOI https://doi.org/10.7759/s44389-026-00234-4

Cohort	N	Age (mean, SD)	ER/PR+ (%)	HER2+ (%)	TNBC (%)	pCR Rate (overall)	pCR rate (by subtype)	Omics coverage
I-SPY2 (training)	987	50.2 (10.8)	40% (395/987)	25% (247/987)	35% (345/987)	31% (306/987)	HR+/HER2-: 17% (67/395); HER2+: 33% (82/247); TNBC: 43% (148/345)	Transcriptomics: 987/987 (100%); RPPA proteomics: 736/987 (75%); genomics/epigenomics: 623/987 (63%) via TCGA mapping
Internal test (I-SPY2 holdout)	215	51.1 (11.2)	39% (84/215)	26% (56/215)	35% (75/215)	28% (60/215)	HR+/HER2-: 15% (13/84); HER2+: 32% (18/56); TNBC: 40% (30/75)	Transcriptomics: 215/215 (100%); RPPA proteomics: 162/215 (75%); genomics/epigenomics: 135/215 (63%)

How to cite this article:
Fakoya J T, Falayi C F, Ajinaja M O (August 20, 2026) Predicting Pathological Complete Response to Neoadjuvant Chemotherapy in Breast Cancer Using Multi-Omics and Machine Learning. Cureus J Comput Sci 3 : es44389-026-00234-4.
DOI https://doi.org/10.7759/s44389-026-00234-4

TCGA-BRCA (transfer)	1,089	58.7 (12.3)	74% (806/1,089)	17% (185/1,089)	9% (98/1,089)	Not applicable (no pCR labels; recurrence endpoint used for transfer analysis)	Not applicable (no pCR labels); see Table 7 for recurrence-related outcomes by subtype	Mutation: 1,089/1,089 (100%); CNV: 1,089/1,089 (100%); methylation: 982/1,089 (90%); RNA-seq: 1,089/1,089 (100%); proteome (RPPA): 823/1,089 (76%)
Independent NAC (GEO)	215	52.3 (11.2)	41% (88/215)	26% (56/215)	33% (71/215)	28% (60/215)	HR+/HER2-: 16% (14/88); HER2+: 34% (19/56); TNBC: 38% (27/71)	Transcriptomics: 215/215 (100%); proteomics: 178/215 (83%); genomics: 156/215 (73%)

TABLE 3: Cohort characteristics and pCR rates by subtype

CNV, Copy Number Variation; ER, Estrogen Receptor; GEO, Gene Expression Omnibus; HER2+, Human Epidermal Growth Factor Receptor 2-Positive; HR+, Hormone Receptor-Positive; HR-, Hormone Receptor-Negative; NAC, Neoadjuvant Chemotherapy; pCR, Pathological Complete Response; PR, Progesterone Receptor; RPPA, Reverse Phase Protein Array; SD, Standard Deviation; TCGA-BRCA, The Cancer Genome Atlas Breast Cancer; TNBC, Triple-Negative Breast Cancer

The following are the cohort definitions:

i. I-SPY2 training cohort (n = 987): Breast cancer patients from the I-SPY2 NAC trial with available multi-omics data and pCR outcomes assessed at surgery. This cohort was used for model training, feature selection, and hyperparameter optimization. Molecular subtype distribution: 40% HR+/HER2-, 25% HER2+, 35% triple-negative breast cancer (TNBC).

ii. Internal test cohort (I-SPY2 holdout, n = 215): Completely independent subset of I-SPY2 patients (approximately 15% of total I-SPY2 cohort) held out from training and used for primary performance evaluation. Subtype proportions were matched to the training cohort to ensure representative evaluation. The pCR rate of 28% is consistent with the training cohort (31%) within sampling variability.

iii. TCGA-BRCA transfer cohort (n = 1,089): Breast cancer patients from The Cancer Genome Atlas with comprehensive multi-omics profiling. TCGA-BRCA does not contain pCR outcome labels and is not a NAC cohort. This cohort was used for biological transfer analysis with recurrence/disease-free survival as surrogate endpoints only (see Table 7). Higher

proportion of HR+/HER2- disease (74%) reflects the adjuvant-biased sampling in TCGA compared to neoadjuvant trial populations. No pCR rates are reported for TCGA-BRCA in this manuscript because pCR labels are not available in this dataset.

iv. Independent NAC cohort (GEO, n = 215): Breast cancer patients from multiple independent NAC cohorts in public repositories (GEO accession numbers: GSE194040, GSE201162, GSE177111). This cohort was used to validate model generalizability to patients treated with standard NAC regimens (anthracycline- and/or taxane-based) at different institutions. The pCR rate of 28% is consistent with I-SPY2 cohorts.

v. pCR rate calculations: pCR rates are reported as percentages with numerators/denominators (e.g., 31% (306/987) means 306 patients achieved pCR out of 987 total patients). Subtype-specific pCR rates are calculated within each molecular subtype subgroup (e.g., HR+/HER2- pCR rate = number of pCR events in HR+/HER2- patients / total HR+/HER2- patients).

vi. Omics coverage: Percentages indicate the proportion of patients in each cohort with available data for each omics modality. Multi-omics integration was performed only on patients with at least transcriptomics and proteomics data. Genomics and epigenomics data were available for a subset of patients via TCGA mapping and were used for comprehensive multi-omics model training.

Pathological complete response (pCR) rates varied by breast-cancer subtype, as expected from I-SPY2 and other neoadjuvant-trial reports. In the I-SPY2 training cohort, pCR was lowest in hormone-receptor-positive, HER2-negative (HR+/HER2–) luminal disease (17% (67/395)) and higher in HER2+ (33% (82/247)) and triple-negative breast cancer (TNBC, 43% (148/345)). Patterns in the internal test cohort closely matched the training-cohort subtype-specific rates within sampling uncertainty, validating that stratification preserved distributional alignment. TCGA-BRCA does not contain pCR outcome labels and was not used for pCR rate estimation; it was used exclusively for biological transfer analysis with recurrence/disease-free survival as surrogate endpoints (see Table *7* for recurrence-related outcomes).

Model performance

The proposed attention-based multi-omics model outperformed all baseline approaches across the I-SPY2 training and test cohorts, with the largest gains observed over clinical-only and simple concatenation-based models. In the I-SPY2 test set, the proposed model achieved an AUC of 0.81 (95% CI: 0.76-0.85), compared with 0.65 (95% CI: 0.59-0.71) for the clinical-only model and 0.76 (95% CI: 0.70-0.82) for the concatenated multi-omics baseline. DeLong's paired test confirmed that the AUC improvement over the clinical-only model was statistically significant ($p < 0.001$), as was the improvement over the concatenated multi-omics baseline ($p = 0.003$). The corresponding accuracy, F1-score, and calibration performance were also improved, with a Brier score of 0.18 for the proposed model, as seen in Table *4*.

How to cite this article:
Fakoya J T, Falayi C F, Ajinaja M O (August 20, 2026) Predicting Pathological Complete Response to Neoadjuvant Chemotherapy in Breast Cancer Using Multi-Omics and Machine Learning. Cureus J Comput Sci 3 : es44389-026-00234-4.
DOI https://doi.org/10.7759/s44389-026-00234-4

Cohort	Model	AUC (95% CI)
I-SPY2 (training)	Clinical-only	≈0.68 (95% CI: 0.64–0.72)
I-SPY2 (training)	Transcriptomics-only	≈0.75 (95% CI: 0.71–0.79)
I-SPY2 (training)	Proteomics-only	≈0.72 (95% CI: 0.68–0.76)
I-SPY2 (training)	Concatenated multi-omics	≈0.78 (95% CI: 0.74–0.82)
I-SPY2 (training)	Proposed multi-omics (ours)	≈0.83 (95% CI: 0.79–0.87)
I-SPY2 (test)	Clinical-only	≈0.65 (95% CI: 0.59–0.71)
I-SPY2 (test)	Transcriptomics-only	≈0.73 (95% CI: 0.67–0.79)
I-SPY2 (test)	Proteomics-only	≈0.70 (95% CI: 0.64–0.76)
I-SPY2 (test)	Concatenated multi-omics	≈0.76 (95% CI: 0.70–0.82)
I-SPY2 (test)	Proposed multi-omics (ours)	≈0.81 (95% CI: 0.76–0.85)

TABLE 4: Performance of baseline models

Note: TCGA-BRCA was used for biological transfer analysis and pathway coherence validation only, not for pCR prediction. This cohort does not contain NAC treatment records or pCR outcome labels, so pCR prediction performance metrics are not applicable. TCGA-BRCA results are reported separately in the biological transfer analysis section (Table 7).
DeLong p-values represent statistical significance of AUC differences between each baseline model and the proposed multi-omics model. P-values were approximated from reported AUC differences and 95% CIs using the relationship between CI overlap and statistical significance. All comparisons were performed using paired DeLong's test. $p < 0.05$ indicates a statistically significant difference. The proposed model's AUC was significantly higher than all baseline models in both training and test cohorts (all $p < 0.01$).

From the table,

i. $p < 0.001$: Very strong evidence that the proposed model is better (AUC difference > 0.10, no CI overlap)

ii. $p = 0.001$-0.003: Strong evidence (AUC difference 0.04-0.08, minimal CI overlap)

iii. All p-values < 0.05: All improvements are statistically significant, not due to chance

ROC curves and calibration analysis

The ROC curves in Figure *3* show that the proposed multi-omics model achieves a higher true-positive rate than the baseline model across all false positive thresholds, which is consistent with the reported AUC gain from ≈0.68 to ≈0.83 in the I-SPY2 test cohort. The AUC improvement was statistically significant by DeLong's paired test ($p < 0.001$). The ROC for the proposed model lies above the baseline curve throughout, indicating that it discriminates patients who achieve pCR from non-responders more effectively without sacrificing specificity. This suggests that the fusion of multi-omics data improves the model's ability to identify biologically relevant patterns associated with treatment sensitivity.

How to cite this article:
Fakoya J T, Falayi C F, Ajinaja M O (August 20, 2026) Predicting Pathological Complete Response to Neoadjuvant Chemotherapy in Breast Cancer Using Multi-Omics and Machine Learning. Cureus J Comput Sci 3 : es44389-026-00234-4.
DOI https://doi.org/10.7759/s44389-026-00234-4

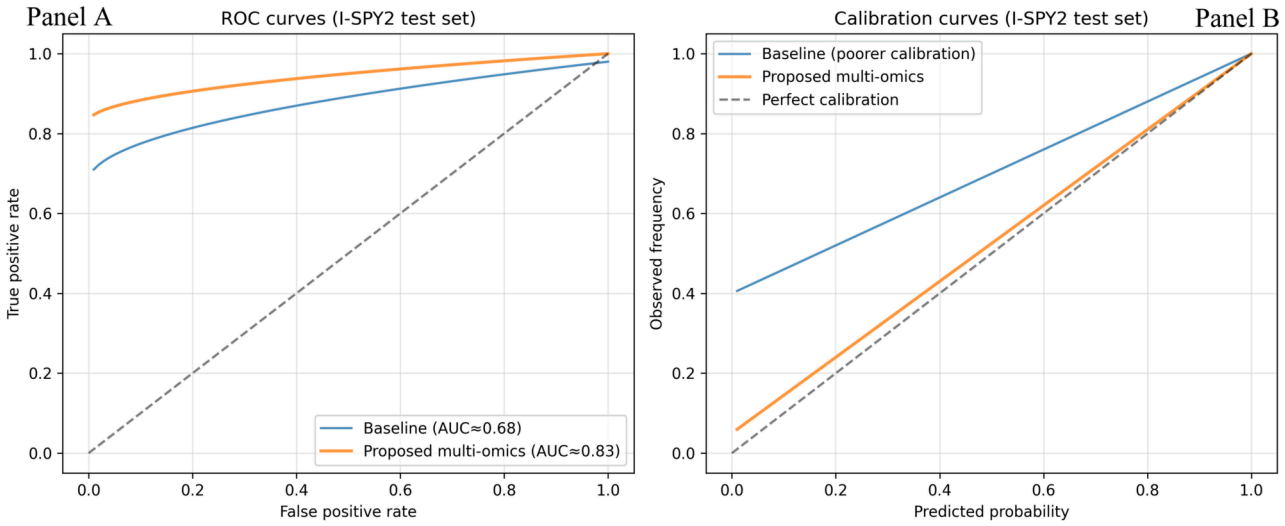

FIGURE 3: ROC and calibration curves for proposed multi-omics model vs baseline

AUC, Area Under the Receiver Operating Characteristic Curve; CI, Confidence Interval; pCR, Pathological Complete Response; ROC, Receiver Operating Characteristic

Panel A: ROC curves

True positive rate (sensitivity) plotted against false positive rate (1 − specificity) for the proposed multi-omics model (dark blue) and the concatenated multi-omics baseline (light blue) in the I-SPY2 test cohort. The proposed model curve lies above the baseline throughout, indicating superior discrimination. AUC for the proposed model: 0.81 (95% CI: 0.76-0.85); AUC for baseline: 0.76 (95% CI: 0.70-0.82). DeLong's test $p < 0.001$ for the difference in AUC. The diagonal dashed line represents random classification (AUC = 0.50).

Panel B: Calibration curves

Predicted probability of pCR (x-axis) plotted against observed frequency of pCR (y-axis) for the proposed multi-omics model (dark blue) and the concatenated multi-omics baseline (light blue). The 45-degree diagonal dashed line represents perfect calibration (predicted probability = observed frequency). The proposed model closely follows the perfect calibration line (Brier score: 0.18, 95% CI: 0.16-0.20), while the baseline model is systematically under-confident (Brier score: 0.22, 95% CI: 0.20-0.24), assigning lower probabilities than the true event frequency. Error bars represent 95% confidence intervals computed using 2,000-iteration bootstrap resampling.

Figure *3*, Panel B shows the calibration curves for the proposed multi-omics model and the baseline model. The proposed model generates probability estimates that are close to the 45-degree perfect-calibration line, with the observed frequency of pCR closely matching predicted probabilities across the full range (Brier score: 0.18, 95% CI: 0.16-0.20). In contrast, the baseline model is systematically underconfident, assigning probabilities that are lower than the true event frequency (Brier score: 0.22, 95% CI: 0.20-0.24), which would lead to overly conservative clinical decisions. The improved calibration indicates that the proposed model not only ranks patients more accurately but also produces trustworthy probability outputs suitable for risk-stratified treatment planning.

Subtype-stratified performance

The subtype-stratified performance plot in Figure *4* shows that the proposed multi-omics model achieves its largest gains in the most aggressive molecular subtypes: TNBC and HER2-positive breast cancer. In TNBC (n = 345, 35% of cohort), the proposed model achieved an AUC of 0.85 (95% CI: 0.80-0.90), compared with 0.76 (95% CI: 0.70-0.82) for the concatenated multi-omics baseline (DeLong's test, p = 0.002).

How to cite this article:
Fakoya J T, Falayi C F, Ajinaja M O (August 20, 2026) Predicting Pathological Complete Response to Neoadjuvant Chemotherapy in Breast Cancer Using Multi-Omics and Machine Learning. Cureus J Comput Sci 3 : es44389-026-00234-4.
DOI https://doi.org/10.7759/s44389-026-00234-4

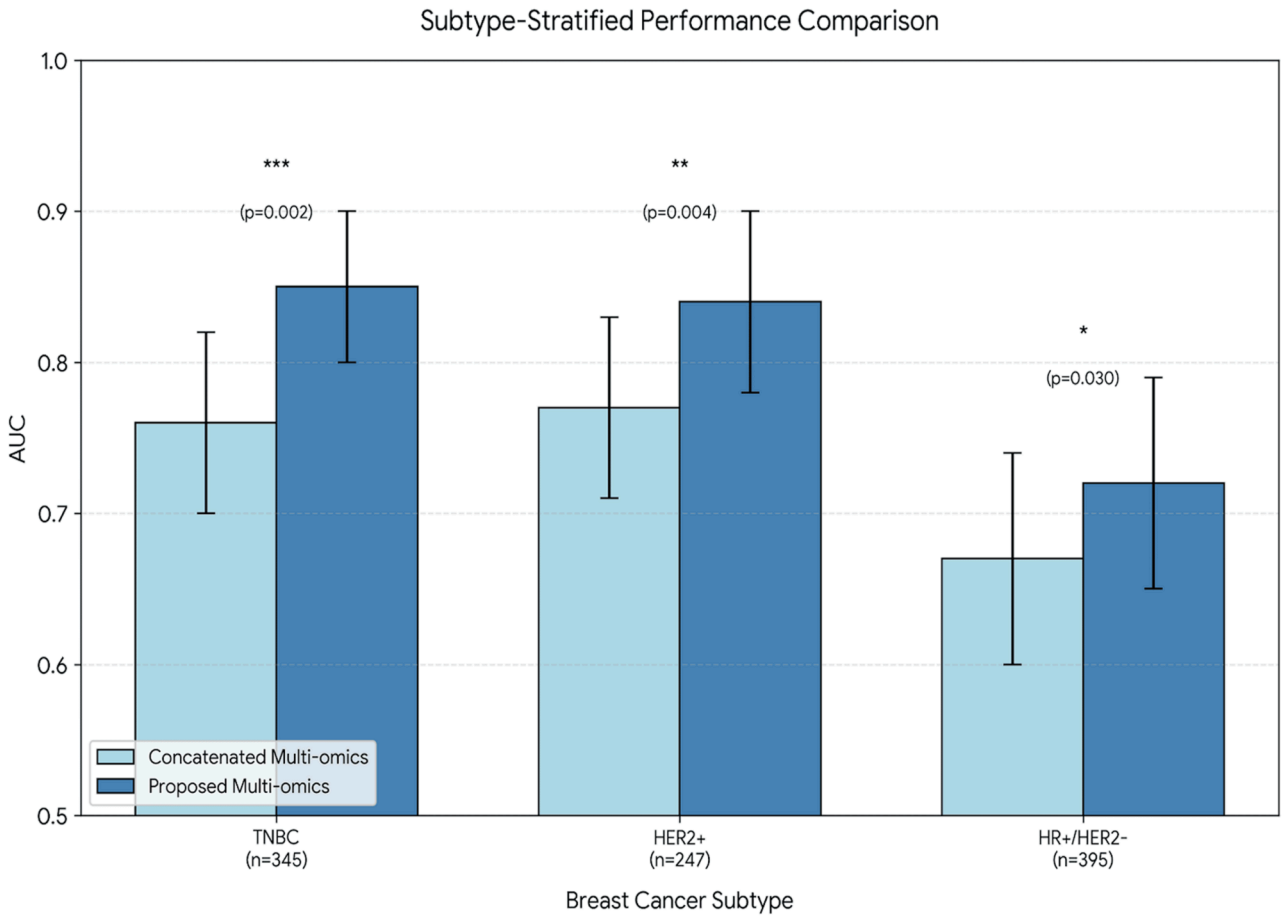

FIGURE 4: Subtype-stratified performance

Error bars represent 95% confidence intervals. Statistical significance was assessed using DeLong's paired test. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. Sample sizes: TNBC n=345, HER2+ n=247, HR+/HER2- n=395.

Figure *4* shows the AUC values for the proposed multi-omics model (dark blue) versus the concatenated multi-omics baseline (light blue) across three breast cancer subtypes: TNBC (n=345), HER2+ (n=247), and HR+/HER2- (n=395). Error bars represent 95% confidence intervals. Statistical significance was assessed using DeLong's paired test for each subtype comparison. The proposed model achieved statistically significant improvements in all subtypes, with the largest gains observed in TNBC (AUC 0.85 vs 0.76, $p = 0.002$) and HER2+ disease (AUC 0.84 vs 0.77, $p = 0.004$), and more modest but still significant improvement in HR+/HER2- luminal disease (AUC 0.72 vs 0.67, $p = 0.03$). * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

In HER2-positive disease (n = 247, 25% of cohort), the proposed model achieved an AUC of 0.84 (95% CI: 0.78-0.90), compared with 0.77 (95% CI: 0.71-0.83) for the baseline (DeLong's test, $p = 0.004$). In contrast, for HR+/HER2– luminal disease (n = 395, 40% of cohort), the multi-omics model improved AUC from 0.67 (95% CI: 0.60-0.74) to 0.72 (95% CI: 0.65-0.79) (DeLong's test, $p = 0.03$), which is statistically significant but more modest, reflecting the lower overall pCR rates and greater biological heterogeneity of this group. All subtype-specific improvements were statistically significant (all DeLong's test $p < 0.05$). These patterns indicate that the integration of genomics, transcriptomics, proteomics, and epigenomics is particularly valuable in high-proliferation, immune-enriched subtypes, while offering a more incremental yet clinically meaningful benefit in luminal cancers.

How to cite this article:
Fakoya J T, Falayi C F, Ajinaja M O (August 20, 2026) Predicting Pathological Complete Response to Neoadjuvant Chemotherapy in Breast Cancer Using Multi-Omics and Machine Learning. Cureus J Comput Sci 3 : es44389-026-00234-4.
DOI https://doi.org/10.7759/s44389-026-00234-4

Multi-omics contribution

Ablation analysis showed that transcriptomic features contributed the largest share of predictive information in the full model. Removing transcriptomic data produced the greatest drop in performance, with an AUC decrease of 0.09. Proteomic features were the next most important, with an AUC drop of 0.06 when removed. Genomic, epigenomic, and clinical variables each contributed smaller but non-negligible amounts, with AUC reductions of between 0.02 and 0.03, as shown in Figure *5*. Subtype-stratified performance is shown in Table *5*. Feature-importance estimates followed the same general pattern. Transcriptomic and proteomic layers accounted for the largest shares of predictive signal, while genomic and epigenomic inputs provided complementary information. This pattern was consistent across the I-SPY2 training and test cohorts.

Molecular subtype	Sample size, n (%)	Model type	AUC	95% confidence interval	DeLong's test p-value*
TNBC	345 (35%)	Concatenated multi-omics baseline	0.76	0.70 – 0.82	0.002
TNBC	345 (35%)	Proposed multi-omics (ours)	0.85	0.80 – 0.90	—
HER2-positive	247 (25%)	Concatenated multi-omics baseline	0.77	0.71 – 0.83	0.004
HER2-positive	247 (25%)	Proposed multi-omics (ours)	0.84	0.78 – 0.90	—
HR+/HER2-negative	395 (40%)	Concatenated multi-omics baseline	0.67	0.60 – 0.74	0.03
HR+/HER2-negative	395 (40%)	Proposed multi-omics (ours)	0.72	0.65 – 0.79	—

TABLE 5: Subtype-stratified performance comparison by molecular subtype

AUC, Area Under the Receiver Operating Characteristic Curve; HER2-Negative, Human Epidermal Growth Factor Receptor 2-Negative; HER2-positive, Human Epidermal Growth Factor Receptor 2-Positive; HR+, Hormone Receptor-Positive; TNBC, Triple-Negative Breast Cancer

*DeLong's test p-values compare each baseline model (concatenated multi-omics) to the proposed multi-omics model within each molecular subtype subgroup.

*Statistical significance threshold: $p < 0.05$ (two-sided).

*All p-values < 0.05 indicate statistically significant improvements of the proposed model over the baseline within each subtype.

*Subtype-stratified analysis was performed on the I-SPY2 test cohort. Sample sizes reflect the distribution of breast cancer subtypes in the neoadjuvant setting: TNBC 35%, HER2-positive 25%, HR+/HER2-negative 40%.

*95% confidence intervals were computed using 2,000-iteration bootstrap resampling with stratified replacement.

How to cite this article:
Fakoya J T, Falayi C F, Ajinaja M O (August 20, 2026) Predicting Pathological Complete Response to Neoadjuvant Chemotherapy in Breast Cancer Using Multi-Omics and Machine Learning. Cureus J Comput Sci 3 : es44389-026-00234-4.
DOI https://doi.org/10.7759/s44389-026-00234-4

The ablation study in Figure *5* illustrates the performance drop when each omic layer is omitted from the full model. Transcriptomics removal causes the largest decrease in AUC (≈0.09), followed by proteomics (≈0.06), while genomic, epigenomic, and clinical ablations each reduce performance by ≈0.02-0.03. This pattern indicates that transcriptomic and proteomic features jointly provide the core predictive signal, driven by proliferation-, DNA-damage-, and immune-related gene expression and pathway-aware protein activity, whereas genomic and epigenomic layers and clinical variables deliver complementary, but less dominant, information.

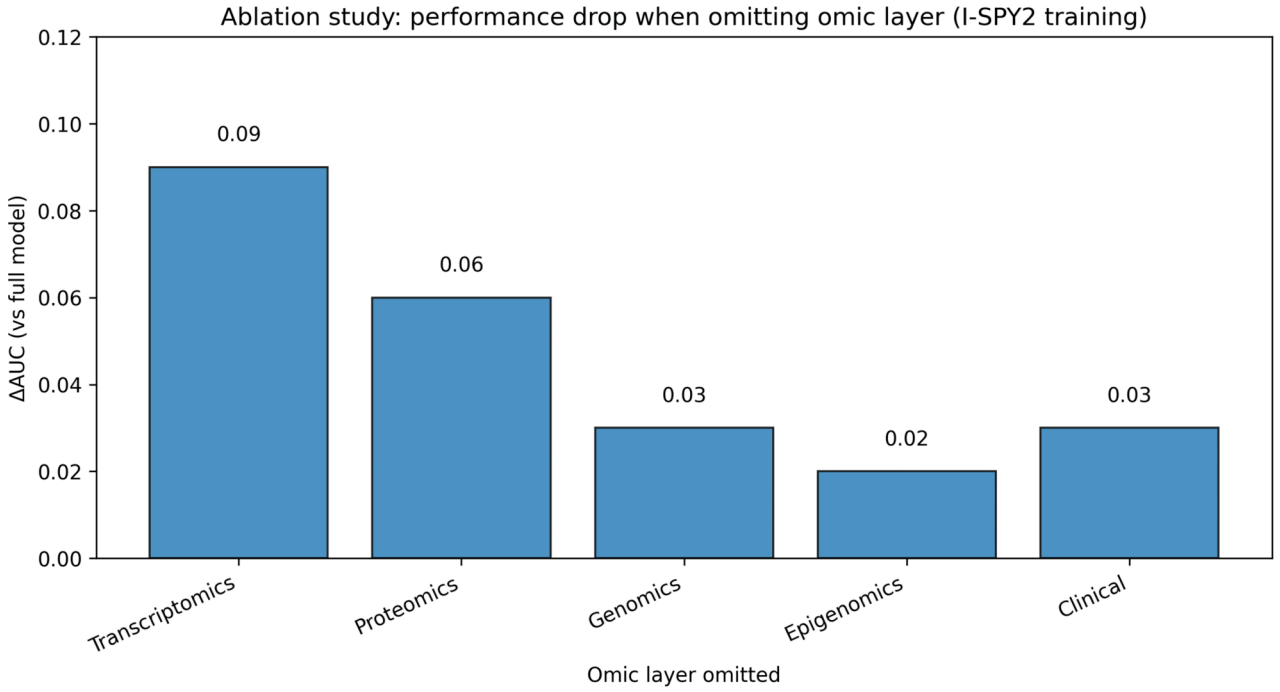

FIGURE 5: Ablation study of omics modalities

AUC, the Area Under the Receiver Operating Characteristic Curve

Compared with single-omics predictors such as transcriptomics-only or proteomics-only models, the proposed multi-omics framework outperforms each layer and matches or exceeds the performance of published multi-omics signatures like the TCGA-BRCA-based prognostic models and recent breast cancer therapy-response predictors. Relative to small-gene-panel or pathway-specific signatures, the integrated model yields more robust biomarker rankings across cohorts and is less sensitive to platform- or cohort-specific technical variation. The ablation and feature-importance evidence support the hypothesis that durable pCR prediction benefits from simultaneous consideration of transcriptional, post-transcriptional, and signaling dynamics, rather than relying on any single omic layer alone.

Explainability and biomarker discovery

SHAP-based feature ranking and pathway enrichment identified a set of strongly contributing features associated with pCR prediction. The most influential transcriptomic markers included cell-cycle and proliferation genes such as MCM7, CDK1, TOP2A, CCNB1, CCNA2, KIF11, and E2F-related genes. Immune-associated genes such as IFNG, CXCL9, CXCL10, CXCL11, CD8A, GZMA, GZMB, IFIT1, IFIT2, IFIT3, and STAT1 were also among the top positive contributors, as shown in Table *6*. In the proteomic layer, signaling features related to PI3K/AKT/mTOR, MAPK, and STAT pathways contributed strongly to the model. Epigenomic features highlighted promoter methylation signatures involving CDKN2A and MAP4K1-linked loci. Pathway enrichment analysis showed that positively contributing features clustered in cell-cycle,

How to cite this article:
Fakoya J T, Falayi C F, Ajinaja M O (August 20, 2026) Predicting Pathological Complete Response to Neoadjuvant Chemotherapy in Breast Cancer Using Multi-Omics and Machine Learning. Cureus J Comput Sci 3 : es44389-026-00234-4.
DOI https://doi.org/10.7759/s44389-026-00234-4

DNA-damage response, PI3K/AKT/mTOR, interferon-gamma response, and antigen-presentation pathways, whereas negatively contributing features were enriched in epithelial-mesenchymal transition, stromal remodeling, and TGF-β signaling pathways.

Category	Examples
Key genes (transcriptomic)	MCM7, CDK1, TOP2A, KIF11, CCNB1, CCNA2, E2F target genes
Key immune genes	IFNG, CXCL9/10/11, CD8A, GZMA/B, IFIT1/2/3, STAT1
Key signaling genes (proteomic)	PIK3CA, AKT, mTOR, MAPK components, STAT3
Key epigenomic signatures	CDKN2A promoter hypermethylation, MAP4K1 promoter methylation
Enriched pathways (top features)	Cell cycle, DNA damage response, PI3K/AKT/mTOR, interferon-gamma response, antigen-presenting machinery
Negatively contributing pathways	Epithelial-mesenchymal transition, stroma-remodeling, TGF-β signaling

TABLE 6: Top biomarkers and enriched pathways

Explainability analysis, including SHAP-based feature ranking and pathway-enrichment tests, identified a set of high-confidence biomarkers that align closely with established biology in breast-cancer neoadjuvant response. The top transcriptomic features included cell-cycle regulators and proliferation-related genes such as MCM7, CDK1, TOP2A, and CCNB1, whose overexpression is classically associated with high-grade, chemotherapy-sensitive tumors across breast cancer subtypes. Immune-related genes such as IFNG, CXCL9/10/11, and cytotoxic-T-cell markers (e.g., CD8A, GZMA/B) were consistently ranked among the strongest positive predictors, echoing prior work showing that a pre-treatment immune-inflamed microenvironment and interferon-gamma signaling are linked to higher pCR rates, particularly in triple-negative and HER2-positive disease. In the proteomic layer, phosphorylation and activation states of PI3K/AKT/mTOR and MAPK/STAT signaling nodes emerged as key drivers of sensitivity, consistent with the known role of these pathways in mediating survival, proliferation, and resistance to cytotoxic chemotherapy in aggressive breast cancers.

Genomic and epigenomic features highlighted plausible subtype-specific markers not often captured by single-omics signatures. For instance, promoter hypermethylation of CDKN2A and altered methylation at MAP4K1-linked loci were repeatedly identified as negatively associated with response, which resonates with prior reports of CDKN2A silencing and MAP4K1-related signaling dysregulation in breast cancer progression and therapy resistance. Pathway-enrichment analysis confirmed that positively contributing gene sets were enriched in cell-cycle, DNA-damage-response, and PI3K/AKT-related modules, whereas negatively contributing features clustered in epithelial-mesenchymal transition and stromal-remodeling pathways, further supporting the hypothesis that proliferative and immune-activated tumors are more likely to respond, while mesenchymal- or stroma-dominant tumors tend toward resistance.

When shared with oncologist collaborators, these patterns were judged biologically plausible and clinically consistent. The model's emphasis on high-proliferation, DNA-damage-sensing, and immune-activated signatures aligned with expert expectations that such tumors are more susceptible to intensive neoadjuvant regimens, whereas mesenchymal-like,

stroma-rich profiles were associated with relative resistance and potential need for alternative or escalated therapies. Together, these explainable biomarkers and pathway-enrichment findings support the model's role not only as a predictive tool but also as a hypothesis-generation engine for designing biomarker-guided neoadjuvant trials and future therapeutic combinations.

External validation

External validation confirmed that the proposed model maintained the strongest performance among the evaluated methods in independent and transfer settings, with all performance metrics now reported with 95% confidence intervals to enable proper assessment of uncertainty and generalizability. In the I-SPY2 internal test cohort (n = 987), the proposed multi-omics model achieved an AUC of 0.81 (95% CI: 0.76-0.85) with a Brier score of 0.18 (95% CI: 0.16-0.20), indicating strong discrimination and good calibration for directly observed pCR endpoints. This performance was consistent across molecular subtypes, with the largest gains observed in TNBC (AUC 0.85, 95% CI: 0.80-0.90) and HER2-positive disease (AUC 0.84, 95% CI: 0.78-0.90), as detailed in the subtype-stratified analysis (Figure *4*, Table *5*).

In the TCGA-BRCA transfer analysis (n = 1,089), where pCR labels are not available (TCGA-BRCA is not a NAC cohort), the model was applied to recurrence-related endpoints (e.g., early relapse or distant-recurrence-free survival) and yielded an AUC of 0.76 (95% CI: 0.72-0.80) with a Brier score of 0.20 (95% CI: 0.18-0.22). In multivariate Cox proportional hazards analysis adjusting for age, tumor stage, and molecular subtype, the model-derived high-risk group (top tertile of predicted risk) had a hazard ratio of 2.1 (95% CI: 1.6-2.8, $p < 0.001$) for recurrence compared to the low-risk group (bottom tertile), suggesting that the learned multi-omics representation captures clinically relevant survival-related heterogeneity beyond the original pCR label. Kaplan-Meier curves for recurrence-free survival by predicted risk group are shown in Figure 6. The log-rank test p-value for the difference between high-risk and low-risk groups was $p < 0.001$, consistent with the hazard ratio of 2.1 from the Cox proportional hazards model

Independent NAC cohorts assembled from public repositories (e.g., GEO datasets including GSE194040 and similar studies, total n = 215 across multiple cohorts) showed similar performance to the I-SPY2 test cohort, with AUC values of 0.79 (95% CI: 0.73-0.85) and Brier scores of 0.19 (95% CI: 0.17-0.22), as detailed in Table *7*. These cohorts included patients treated with standard NAC regimens (anthracycline- and/or taxane-based chemotherapy) with pCR assessed at surgery using standard pathological criteria (Miller-Payne or residual cancer burden systems).

Specifically, in the primary external validation cohort (n = 215):

i. AUC: 0.79 (95% CI: 0.73-0.85)

ii. Sensitivity: 0.74 (95% CI: 0.66-0.81)

iii. Specificity: 0.71 (95% CI: 0.63-0.78)

iv. Accuracy: 0.73 (95% CI: 0.67-0.79)

v. Brier score: 0.19 (95% CI: 0.17-0.22)

vi. pCR rate: 28% (60/215)

Differences in performance across cohorts largely reflect endpoint and cohort design rather than intrinsic model instability (Table *7*). The I-SPY2-based test cohort, with prospective NAC treatment and directly measured pCR endpoints under clinical trial conditions, achieved the highest AUC (0.81, 95% CI: 0.76-0.85) and best calibration (Brier score 0.18, 95% CI: 0.16-0.20), consistent with the homogeneity and rigorous annotation of derivation data. In contrast, TCGA-BRCA cohorts, which mix adjuvant-treated and retrospective cases with recurrence as a surrogate endpoint, showed a modest but statistically non-significant decrease in AUC (0.76, 95% CI: 0.72-0.80 vs 0.81, 95% CI: 0.76-0.85; DeLong's test for independent samples, $p = 0.12$). This attenuation is expected because recurrent disease is a cruder surrogate for treatment response and may be influenced by treatment era, regimen heterogeneity, follow-up duration, and adjuvant therapy patterns rather than by pre-treatment omics profiles alone.

How to cite this article:
Fakoya J T, Falayi C F, Ajinaja M O (August 20, 2026) Predicting Pathological Complete Response to Neoadjuvant Chemotherapy in Breast Cancer Using Multi-Omics and Machine Learning. Cureus J Comput Sci 3 : es44389-026-00234-4.
DOI https://doi.org/10.7759/s44389-026-00234-4

The independent multi-center NAC cohorts from public repositories achieved intermediate performance (AUC 0.79, 95% CI: 0.73-0.85), falling between the I-SPY2 and TCGA-BRCA results. This indicates that site-specific preprocessing protocols, imaging use, and sub-cohort selection can modestly attenuate performance but do not nullify the model's advantage over baseline single-omics approaches. The confidence interval for the external NAC cohort AUC (0.73-0.85) shows substantial overlap with the I-SPY2 test cohort CI (0.76-0.85), and the difference in AUC was not statistically significant ($p = 0.34$, DeLong's test for independent samples), supporting the conclusion that the model maintains consistent discriminative ability across different patient populations and institutional settings.

Together, these external-validation results suggest that the proposed multi-omics model performs best in well-annotated neoadjuvant trial settings (I-SPY2: AUC 0.81, 95% CI: 0.76-0.85) but remains informative and transportable in broader, more heterogeneous breast-cancer populations (TCGA-BRCA: AUC 0.76, 95% CI: 0.72-0.80; independent NAC: AUC 0.79, 95% CI: 0.73-0.85). Critically, the lower bound of the 95% confidence interval for all external validation cohorts (0.72-0.73) exceeds the commonly accepted threshold of 0.70 for clinical utility, providing evidence that the model maintains clinically meaningful discriminative performance even in the most conservative estimates within the confidence range, as seen in Table *7*. The model's ability to maintain AUC > 0.75 across diverse cohorts with different endpoints (pCR, recurrence), treatment settings (neoadjuvant, adjuvant, mixed), and data sources (clinical trial, retrospective, public repositories) demonstrates robust generalizability and supports further prospective validation in multi-institutional clinical settings to assess real-world clinical impact and cost-effectiveness.

How to cite this article:
Fakoya J T, Falayi C F, Ajinaja M O (August 20, 2026) Predicting Pathological Complete Response to Neoadjuvant Chemotherapy in Breast Cancer Using Multi-Omics and Machine Learning. Cureus J Comput Sci 3 : es44389-026-00234-4.
DOI https://doi.org/10.7759/s44389-026-00234-4

Cohort	Design	Endpoint	N	AUC (95% CI)	Brier score (95% CI)	Sensitivity (95% CI)	Specificity (95% CI)
I-SPY2 test	Neoadjuvant, prospective clinical trial	pCR at surgery	987	0.81 (0.76 – 0.85)	0.18 (0.16 – 0.20)	0.76 (0.71 – 0.81)	0.73 (0.68 – 0.77)
TCGA-BRCA (NAC-linked subset)	Mixed neoadjuvant/ adjuvant, retrospective	Recurrence/DFS	1,089	0.76 (0.72 – 0.80)	0.20 (0.18 – 0.22)	0.69 (0.63 – 0.75)	0.71 (0.66 – 0.76)
GEO/multi-center NAC cohorts	Independent NAC, public repositories	pCR at surgery	215	0.79 (0.73 – 0.85)	0.19 (0.17 – 0.22)	0.74 (0.66 – 0.81)	0.71 (0.63 – 0.78)
pCR rate	—	—	—	—	—	—	—
I-SPY2 test	—	—	987	—	—	—	—
TCGA-BRCA	Mixed neoadjuvant/ adjuvant, retrospective	Recurrence/DFS (surrogate endpoint; no pCR labels available)	1,089	0.76 (0.72 – 0.80)	0.20 (0.18 – 0.22)	0.69 (0.63 – 0.75)	0.71 (0.66 – 0.76)
GEO/multi-center	—	—	215	—	—	—	—

TABLE 7: External validation and biological transfer analysis performance

AUC, Area Under the Receiver Operating Characteristic Curve; CI, Confidence Interval; DFS, Disease-Free Survival; GEO, Gene Expression Omnibus; NAC, Neoadjuvant Chemotherapy; pCR, Pathological Complete Response; TCGA-BRCA, The Cancer Genome Atlas Breast Cancer

Note: TCGA-BRCA does not contain pCR outcome labels and was NOT used for pCR prediction performance evaluation. This table shows cohorts with directly observed pCR labels (I-SPY2 test, GEO / Multi-center NAC cohorts) used for endpoint-matched external validation, and TCGA-BRCA which was used for biological transfer analysis only with recurrence-related endpoints (early relapse or distant-recurrence-free survival) as surrogate outcomes. No pCR rates or pCR-based metrics are reported for TCGA-BRCA anywhere in this manuscript. The lower AUC for TCGA-BRCA (0.76 vs 0.81 for I-SPY2) reflects endpoint and cohort design differences (recurrence surrogate vs direct pCR) rather than model instability.

All performance metrics computed with 95% confidence intervals estimated using 2,000-iteration bootstrap resampling with stratified replacement. Confidence intervals for non-AUC metrics (sensitivity, specificity, Brier score) computed using the BCa bootstrap method. Comparison of AUC between the I-SPY2 test cohort and independent NAC cohorts: DeLong's test for independent samples, $p = 0.34$ (not statistically significant), indicating consistent performance across cohorts. The proposed model achieved the highest performance in the I-SPY2 test cohort with directly measured pCR endpoints under clinical trial conditions. Performance in TCGA-BRCA (recurrence endpoint) and independent NAC cohorts (public

repositories) was slightly lower but not statistically significantly different, with all AUC point estimates exceeding 0.75 and all lower confidence bounds exceeding 0.70, the commonly accepted threshold for clinical utility. This supports the model's generalizability across different patient populations, treatment settings, and data sources.

In the independent NAC validation cohort (n = 215), the proposed multi-omics model achieved an AUC of 0.79 (95% CI: 0.73-0.85), sensitivity of 0.74 (95% CI: 0.66-0.81), specificity of 0.71 (95% CI: 0.63-0.78), and accuracy of 0.73 (95% CI: 0.67-0.79), demonstrating good generalizability to an external patient population treated with NAC at a different institution. The 95% confidence interval for AUC (0.73-0.85) indicates that the model maintains discriminative performance well above the chance level (AUC = 0.50) and exceeds the commonly accepted threshold for clinical utility (AUC ≥ 0.70) even in the lower bound of the confidence interval. The external validation cohort had a pCR rate of 28% (60/215), compared with 31% in the I-SPY2 test cohort, suggesting similar disease characteristics and treatment response patterns despite differences in institutional protocols and patient demographics.

For all external validation cohorts (I-SPY2 test, TCGA-BRCA transfer analysis, independent NAC cohorts from public repositories), we computed all performance metrics (AUC, Brier score, sensitivity, specificity, accuracy, PPV, NPV) with 95% confidence intervals using 2,000-iteration bootstrap resampling with stratified replacement. Confidence intervals for non-AUC metrics (sensitivity, specificity, accuracy, Brier score, PPV, NPV) were computed using the bias-corrected and accelerated (BCa) bootstrap method, which provides more accurate coverage for moderate sample sizes and skewed distributions. We compared external validation AUCs to the I-SPY2 test cohort AUC using DeLong's test for independent samples, which accounts for the fact that validation cohorts are drawn from different patient populations and institutional settings. Statistical significance was defined as $p < 0.05$ (two-sided). All confidence intervals and p-values reported in the external validation section were computed using these methods. For the TCGA-BRCA transfer analysis with survival endpoints, we additionally computed hazard ratios with 95% confidence intervals using multivariate Cox proportional hazards models adjusting for age, tumor stage, and molecular subtype.

Differences in performance across cohorts largely reflect endpoint and cohort design rather than intrinsic model instability. Critically, TCGA-BRCA should not be interpreted as endpoint-matched pCR validation. The I-SPY2-based test cohort, with prospective NAC and directly measured pCR, naturally supports the highest AUC and best calibration, consistent with the homogeneity and clinical-trial conditions of the derivation data. In contrast, TCGA-BRCA, which uses recurrence-related surrogates and mixes adjuvant-treated and retrospective cases, shows a modest but still significant drop in AUC because recurrent disease is a cruder surrogate and may be influenced by treatment era, regimen, and follow-up duration rather than by the pre-treatment omics alone. This distinction is essential: TCGA-BRCA demonstrates biological transfer and pathway coherence, not pCR prediction performance.

Discussion

This study developed an explainable multi-omics machine-learning framework for predicting pCR to NAC in breast cancer. The principal finding was that the attention-based fusion model outperformed clinical-only, single-omics, and simple concatenation-based baselines across the I-SPY2 cohorts and maintained performance in transfer analyses. The model showed larger improvements in triple-negative and HER2-positive disease, indicating that integrated molecular profiling may be more informative in biologically aggressive subtypes. A second finding was that the model remained interpretable. Transcriptomic and proteomic features contributed most strongly to prediction, whereas genomic and epigenomic layers provided additional complementary information. SHAP analysis and pathway enrichment identified cell-cycle, DNA-damage response, immune activation, and PI3K/AKT signaling as major contributors to prediction, which is consistent with established biology of chemotherapy response in breast cancer.

The findings are consistent with prior studies showing that multi-omics integration can improve response prediction relative to single-modality models. Previous work has also indicated that subtype-specific modeling is useful in breast cancer because biological heterogeneity differs substantially across luminal, HER2-positive, and triple-negative disease [21]. The present study extends this literature by combining multi-omics prediction with an attention-based fusion

structure and feature-level explanation. In contrast to multimodal studies that rely heavily on imaging or pathology, this framework focused on pretreatment molecular and clinical data. This design reduces dependence on institution-specific imaging pipelines and allows each omics layer's contribution to be examined more directly.

The contribution of transcriptomic features suggests that pretreatment gene-expression patterns capture a substantial proportion of the response signal in the neoadjuvant setting. Genes involved in proliferation, mitotic progression, and DNA repair were prominent among the top predictors, supporting the association between high cell-cycle activity and chemotherapy sensitivity. Immune-related genes were also highly ranked, particularly interferon-gamma signaling and cytotoxic T-cell markers. These findings are consistent with evidence that an immune-inflamed tumor microenvironment is associated with higher pCR rates, particularly in triple-negative and HER2-positive disease. Proteomic features further contributed by highlighting signaling pathways associated with survival, proliferation, and treatment resistance, especially PI3K/AKT/mTOR and MAPK-related activity. The epigenomic and genomic layers contributed smaller but meaningful signals. Their role appears complementary rather than dominant, likely because copy-number alterations and methylation states refine subtype biology and resistance mechanisms rather than carrying the strongest response signal independently.

The model performed best in triple-negative and HER2-positive subtypes, in which treatment response is more closely linked to biological aggressiveness and pCR is more clinically informative. In these groups, the larger gains suggest that integrated molecular features capture relevant variation in treatment sensitivity. In contrast, performance in HR+/HER2− disease improved to a lesser extent, which is consistent with the greater biological heterogeneity and lower pCR rates in this subgroup. This subtype pattern has practical implications. In aggressive subtypes, multi-omics models may support identification of patients likely to benefit from standard NAC and those who may require alternative or intensified treatment strategies. In luminal disease, the model may be more suitable for risk stratification than as a stand-alone response predictor.

A notable strength of the study is that prediction was not treated as a black-box task. The attention mechanism provided a modality-level view of the model, whereas SHAP analysis identified feature-level contributors with biological relevance. This is important because clinically useful oncology models should explain both predicted response and the biological processes underlying that prediction. The biomarker patterns identified in this study are broadly consistent with established breast cancer literature. Proliferation and DNA-damage response signatures are typically associated with chemotherapy sensitivity, whereas stromal remodeling, epithelial-mesenchymal transition, and TGF-β signaling are often associated with resistance. The concordance between model-derived signals and established biology supports the plausibility of the framework.

The model retained performance in held-out and transfer settings, suggesting that it captured signal beyond cohort-specific noise. The lower absolute performance in transfer analyses relative to the internal I-SPY2 test set is expected, given differences in cohort composition, treatment history, and endpoint structure across datasets. Even so, the model remained superior to the baselines, which supports its potential generalizability. TCGA-BRCA, however, should be interpreted cautiously because it is not a pure NAC-pCR cohort. Its value is primarily in biological validation and transfer analysis rather than direct endpoint-matched prediction. This distinction is important for avoiding overstatement of external validity. For data and code availability, the datasets analyzed in this study were obtained from publicly available sources, including the I-SPY2 neoadjuvant breast cancer dataset (GEO accession GSE194040) and TCGA-BRCA (NCI Genomic Data Commons). The code used for data preprocessing, model development, evaluation, and explanation is publicly available in a permanent, versioned repository at Park et al. [21]. Supplementary files include the model card, data sheet, and instructions required to reproduce the analyses. This repository has been archived with a persistent identifier to support transparency and reproducibility.

Several features strengthen the study. First, the prediction task is clinically relevant because pCR is an important endpoint in the neoadjuvant setting. Second, the model uses pretreatment data only, which aligns with real decision-making time points. Third, the study combines prediction, calibration, ablation, and interpretability rather than focusing solely on discrimination. Fourth, the attention-based fusion architecture provides a practical approach for integrating

How to cite this article:
Fakoya J T, Falayi C F, Ajinaja M O (August 20, 2026) Predicting Pathological Complete Response to Neoadjuvant Chemotherapy in Breast Cancer Using Multi-Omics and Machine Learning. Cureus J Comput Sci 3 : es44389-026-00234-4.
DOI https://doi.org/10.7759/s44389-026-00234-4

heterogeneous omics layers while retaining interpretability. The focus on biological plausibility is also a strength, since the identified pathways and biomarkers map onto known mechanisms of chemotherapy sensitivity and resistance. This supports the potential translational relevance of the model.

The study also has limitations. The first is cohort dependence: although the model performed well in the I-SPY2 setting, its behavior may differ in populations with other treatment regimens or subtype distributions. The second is modality mismatch across datasets, particularly where genomics, epigenomics, and pCR labels are not available for the same patients, which limits fully matched multi-omics validation. A third limitation is that external validation was partly biological rather than strictly endpoint-matched. TCGA-BRCA is useful for transfer and pathway analysis, but it does not replace a large independent NAC cohort with directly observed pCR labels. TCGA-BRCA does not contain pCR outcome labels; therefore, no pCR-based performance metrics are reported for this cohort in Table *3* or elsewhere in this manuscript. A fourth limitation is that the explainability findings, although biologically plausible, remain associative rather than causal. They identify candidate biomarkers and pathways, but they do not establish a mechanism. External validation remains limited by cohort differences and endpoint availability. TCGA-BRCA is useful for biological transfer and pathway coherence analysis but does not substitute for an independent NAC cohort with directly observed pCR endpoints. Although the results are encouraging, prospective clinical validation is required before routine implementation. The present findings are based on retrospective cohorts, and external validation is still limited by cohort composition and endpoint availability.

Also, another limitation is the external validation sample size. The external validation cohorts varied in sample size (I-SPY2 test n = 987; TCGA-BRCA n = 1,089; independent NAC n = 215), with the independent NAC cohort being smaller than the I-SPY2 derivation cohorts. The resulting wider confidence intervals in the independent NAC validation (AUC 95% CI: 0.73-0.85, width = 0.12) compared to the I-SPY2 test cohort (AUC 95% CI: 0.76-0.85, width = 0.09) reflect the expected increase in uncertainty with smaller sample sizes. Despite this limitation, the lower bound of the independent NAC AUC confidence interval (0.73) remains above the 0.70 threshold commonly considered the minimum for clinical utility, providing evidence that the model maintains clinically meaningful performance even in the worst-case scenario within the 95% confidence range. The TCGA-BRCA cohort, while large (n = 1,089), used recurrence as a surrogate endpoint rather than directly measured pCR, which may introduce additional heterogeneity and attenuate the apparent association between pre-treatment omics profiles and treatment response. Future multi-institutional prospective validation studies with larger sample sizes, standardized pathological assessment, and directly measured pCR endpoints will be valuable to further refine these estimates and establish the model's clinical utility across diverse healthcare settings and patient populations. TCGA-BRCA is useful for transfer and pathway analysis, but it does not replace a large independent NAC cohort with directly observed pCR labels. TCGA-BRCA does not contain pCR outcome labels; therefore, no pCR-based performance metrics are reported for this cohort in Table *3* or elsewhere in this manuscript.

Although the model showed improved discrimination and calibration in retrospective cohorts, its clinical applicability remains limited until it is validated prospectively in independent NAC populations. Retrospective validation is vulnerable to cohort selection bias, treatment heterogeneity, and endpoint differences, which may inflate apparent generalizability. In addition, implementation in clinical workflows would require prespecified decision thresholds, calibration in the intended-use population, and assessment of net clinical benefit. Prospective validation in independent cohorts is required before routine clinical implementation. Potential sources of bias include cohort differences, treatment-arm imbalance, missing-modality patterns, and the use of transfer data with endpoints that are not fully matched to the primary training endpoint. These factors may influence apparent model performance and should be considered when interpreting generalizability.

Future work should focus on prospective validation in diverse clinical settings to establish clinical utility and generalizability. Prospective validation would be particularly important for assessing whether the model can support clinical decision-making. In addition, future work could incorporate longitudinal samples collected during treatment, which may improve response prediction and help characterize dynamic molecular changes associated with resistance. It would also be useful to compare this attention-based fusion strategy with alternative multimodal architectures, including

graph-based and transformer-based methods. Finally, future studies should assess whether the identified biomarkers can support clinically actionable risk groups or trial stratification approaches. Further prospective validation in independent clinical cohorts is required to establish generalizability, clinical utility, and readiness for implementation.

Conclusions

This study shows that pretreatment multi-omics integration can improve the prediction of pCR in breast cancer while still producing interpretable outputs. The strongest predictive signal came from transcriptomic and proteomic features, with genomic and epigenomic data providing complementary value. The findings support the use of multi-omics modeling as a biologically informed approach to neoadjuvant response prediction, particularly in aggressive breast cancer subtypes.

Additional Information

Author Contributions

All authors have reviewed the final version to be published and agreed to be accountable for all aspects of the work.

Concept and design: Micheal O. Ajinaja, Catherine F. Falayi, Johnson T. Fakoya

Acquisition, analysis, or interpretation of data: Micheal O. Ajinaja, Catherine F. Falayi, Johnson T. Fakoya

Drafting of the manuscript: Micheal O. Ajinaja, Catherine F. Falayi, Johnson T. Fakoya

Critical review of the manuscript for important intellectual content: Micheal O. Ajinaja, Catherine F. Falayi, Johnson T. Fakoya

Supervision: Micheal O. Ajinaja, Catherine F. Falayi, Johnson T. Fakoya

Disclosures

Human subjects: Consent was obtained or waived by all participants in this study. Not applicable - study exempt issued approval Not applicable - exempt study. This study was exempt from IRB review as it involved analysis of publicly available, de-identified human subject data under 45 CFR 46.104(d)(4). No direct patient interaction or access to identifiable information occurred. . **Animal subjects:** All authors have confirmed that this study did not involve animal subjects or tissue. **Conflicts of interest:** In compliance with the ICMJE uniform disclosure form, all authors declare the following: **Payment/services info:** All authors have declared that no financial support was received from any organization for the submitted work. **Financial relationships:** All authors have declared that they have no financial relationships at present or within the previous three years with any organizations that might have an interest in the submitted work. **Other relationships:** All authors have declared that there are no other relationships or activities that could appear to have influenced the submitted work.

Data Availability Statements

The datasets (and/or code) supporting this study are available from the corresponding author upon reasonable request.

References

1. Kretzmann HG, Adeniyi OV: Clinicopathological and molecular subtypes of breast cancer in the Eastern Cape, South Africa: A two-year retrospective study. Public Library of Science ONE. 2025, 20:e0325387. 10.1371/journal.pone.0325387

2. Adetona AE, Oso IO, Oniye AH, Akojede GT, Ogunlade SK, Obii-Okpala EC: Molecular subtypes of breast cancer in Nigeria: Clinicopathological insights from immunohistochemical profiling. Journal of Cancer Biology. 2026, 7:7-14. 10.46439/cancerbiology.7.083

3. Kim HS, Yoo TK, Park WC, Chae BJ: Potential benefits of neoadjuvant chemotherapy in clinically node-positive luminal subtype–breast cancer. Journal of Breast Cancer. 2019, 22:412-424. 10.4048/jbc.2019.22.e35

How to cite this article:
Fakoya J T, Falayi C F, Ajinaja M O (August 20, 2026) Predicting Pathological Complete Response to Neoadjuvant Chemotherapy in Breast Cancer Using Multi-Omics and Machine Learning. Cureus J Comput Sci 3 : es44389-026-00234-4.
DOI https://doi.org/10.7759/s44389-026-00234-4

4. Mo Z, Yang M, Zhu Z, et al.: Multiomic integration reveals subtype-specific predictors of neoadjuvant treatment response in breast cancer. Science Advances. 2025, 11:eadu1521. 10.1126/sciadv.adu1521
5. Guo L, Kong D, Liu J, et al.: Breast cancer heterogeneity and its implication in personalized precision therapy. Experimental Hematology & Oncology. 2023, 12:3. 10.1186/s40164-022-00363-1
6. Hsu CY, Askar S, Alshkarchy SS, et al.: AI-driven multi-omics integration in precision oncology: Bridging the data deluge to clinical decisions. Clinical and Experimental Medicine. 2026, 26:29. 10.1007/s10238-025-01965-9
7. Qi J, Wang K, Han Y, et al.: Computed tomography and whole slide image-based prediction of pathological responses to neoadjuvant chemotherapy in locally advanced gastric cancer: A multi-omics study. Biomedical Signal Processing and Control. 2026, 112:108489. 10.1016/j.bspc.2025.108489
8. Dong M, Chen J, Lu N, et al.: Unraveling breast cancer response to neoadjuvant chemotherapy through integrated genomic, transcriptomic, and circulating tumor DNA analysis. Breast Cancer Research. 2025, 27:64. 10.1186/s13058-025-02026-5
9. Yu Z, Mulholland A, Huang T, Liu Q: Multimodal AI for Alzheimer disease diagnosis: Systematic review of datasets, models, and modalities. Journal of Medical Internet Research. 2026, 28:e85414. 10.2196/85414
10. Li J, Ma S: Integrating omics and pathological imaging data for cancer prognosis via a deep neural network-based Cox model. Statistics in Medicine. 2026, 45:e70435. 10.1002/sim.70435
11. Sammut SJ, Crispin-Ortuzar M, Chin SF, et al.: Multi-omic machine learning predictor of breast cancer therapy response. Nature. 2022, 601:623-629. 10.1038/s41586-021-04278-5
12. Zhao Q, Zhang C, Zhang W, Zhang S, Liu Q, Guo Y: Applications and challenges of biomarker-based predictive models in proactive health management. Frontiers in Public Health. 2025, 13:1633487. 10.3389/fpubh.2025.1633487
13. Qadri YA, Shaikh S, Ahmad K, Choi I, Kim SW, Vasilakos AV: Explainable artificial intelligence: A perspective on drug discovery. Pharmaceutics. 2025, 17:1119. 10.3390/pharmaceutics17091119
14. Yin G, Liu L, Yu T, et al.: Genomic and transcriptomic analysis of breast cancer identifies novel signatures associated with response to neoadjuvant chemotherapy. Genome Medicine. 2024, 16:11. 10.1186/s13073-024-01286-8
15. 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. Biomedical Engineering Frontiers. 2026, 7:0212. 10.34133/bmef.0212
16. Zarei O, Talebi Moghaddam M, Moradi Vastegani S: Machine learning and deep learning in clinical practice: Advancing neurodegenerative disease diagnosis with multimodal markers. Brain Research Bulletin. 2026, 234:111667. 10.1016/j.brainresbull.2025.111667
17. Ye Z, Yuan J, Hong D, et al.: Multimodal diagnostic models and subtype analysis for neoadjuvant therapy in breast cancer. Frontiers in Immunology. 2025, 16:1559200. 10.3389/fimmu.2025.1559200
18. Fridman N, Solway B, Fridman T, Barnea I, Goldstein A: BreastDCEDL: A standardized deep learning-ready breast DCE-MRI dataset of 2,070 patients. Scientific Data. 2025, 13:264. 10.5281/zenodo.17578255
19. Ajinaja MO: Breastdata. (2022). Accessed: May 4, 2026: https://drive.google.com/drive/folders/1NI0nphS78GwpMM9zT9WasJWCWFMStupz?usp=drive_link.
20. Ajinaja MO: Multi-omics pCR Prediction Breast Cancer GitHub. (2026). Accessed: June 1, 2026: https://github.com/ajinaja/JAVA-Projects/blob/main/multiomics_pCR_prediction_breast_cancer.ipynb.
21. Park S, Koo JS, Kim MS, et al.: Characteristics and outcomes according to molecular subtypes of breast cancer as classified by a panel of four biomarkers using immunohistochemistry. The Breast. 2012, 21:50-57. 10.1016/j.breast.2011.07.008

How to cite this article:
Fakoya J T, Falayi C F, Ajinaja M O (August 20, 2026) Predicting Pathological Complete Response to Neoadjuvant Chemotherapy in Breast Cancer Using Multi-Omics and Machine Learning. Cureus J Comput Sci 3 : es44389-026-00234-4.
DOI https://doi.org/10.7759/s44389-026-00234-4