

MRI-based Prognostic Model in Early-Stage HR+/HER2- Breast Cancer

Poornima Saha

Endeavor Health

Frederick Howard

University of Chicago

Erica Stringer-Reasor

University of Alabama at Birmingham

Yara Abdou

University of North Carolina Hospitals

Jennifer McMahon

Des Moines University Osteopathic Medical Center

Katharine Yao

Endeavor Health

Yuhan Zhang

SimBioSys (United States)

Michelle Weitz

SimBioSys (United States)

Vignesh Kannan

SimBioSys (United States)

John Cole

SimBioSys (United States)

Joseph Peterson

SimBioSys (United States)

Daniel Cook

SimBioSys (United States)

Bradley Feiger

SimBioSys (United States)

Judy Boughey

Mayo Clinic Cancer Center

Matthew Goetz

Goetz.Matthew@mayo.edu

Mayo Clinic Cancer Center

Article

Keywords: Adjuvant therapy, breast cancer, MRI, prognostic biomarker, spatial biology

Posted Date: June 28th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-9772767/v1>

License:  This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Additional Declarations: Competing interest reported. JRP, JAC, DC, and VK declare a financial interest in SimBioSys, Inc. JRP, JAC, DC, BF, YZ, BF, and VK are employees of SimBioSys, Inc. FMH reports consulting fees from Novartis and Leica Biosystems, advisory board fees from Exact Sciences, and stock ownership in Celcuity. JCB declare funding paid to institution from SimBioSys, Inc. and Eli Lilly. JCB declares honorarium from Up-to-Date and sits on DSMC for CairnSurgical ESR reports consulting fees from AstraZeneca, Daiichi Sankyo, Gilead, SeaGen, Novartis, Pfizer, and Lilly. Dr. Goetz reports personal fees for CME activities from AXIS, BroadcastMed, DAVA Oncology, IDEOlogy Health, MJH Life Sciences, PeerView, Physicians' Education Resource, Research to Practice, Total Health Conferencing; Consulting fees to Mayo Clinic from 858 Therapeutics, Inc., Amneal Pharmaceuticals, AstraZeneca, AstraZeneca Pharmaceuticals LP, AstraZeneca UK Ltd., Ataraxis AI, BeiGene USA, Biotheranostics, Biotheryx, eChinaHealth, EcoR1, Eli Lilly and Company, Engage Health Media, Genentech, Incyclix Bio, Lilly, Laekna, Novartis, Pfizer, Puma, RNA Diagnostics, Seattle Genetics, Sermonix Pharmaceuticals, Stemline Therapeutics, TerSera Therapeutics/ Amplity Health; Grant funding to Mayo Clinic from AstraZeneca, ATOSSA Therapeutics, Biotheryx, Lilly, Loxo, Pfizer, Puma, Sermonix, SimBioSys; travel support from Lilly and Biotheryx

Abstract

Current genomic assays provide prognostic information and guide chemotherapy decisions in early-stage hormone receptor–positive/human epidermal growth factor receptor 2–negative (HR+/HER2–) breast cancer, but limited tissue sampling may incompletely capture tumor heterogeneity. We developed and validated an image-based artificial intelligence (AI) model integrating dynamic contrast-enhanced magnetic resonance imaging (DCE-MRI) and clinicopathologic features to predict recurrence risk in early-stage HR+/HER2 – breast cancer. In this retrospective multi-institutional study, the model was trained using pre-treatment imaging and clinicopathologic data from 522 patients and independently validated in 550 additional patients treated between 2010 and 2023. By analyzing spatial tumor characteristics, the model generated individualized recurrence risk scores that outperformed clinicopathologic features alone. In the validation cohort, median age was 53 years (range, 24–90), 15.6% of patients self-identified as African American, 89.7% had clinical T1–T2 disease, and 34.4% had 1–3 positive lymph nodes. Five-year recurrence-free survival was 93.8% (95% CI, 90.0–96.2%) among patients classified as low risk compared with 82.6% (95% CI, 74.8–88.1%) among those classified as high risk, corresponding to an adjusted hazard ratio of 2.28. Prognostic performance was maintained across clinically relevant subgroups, including age and nodal status. These findings support the potential of a noninvasive image-based AI approach for real-time prognostication and treatment personalization in early-stage HR+/HER2 – breast cancer.

Introduction

Breast cancer is the most common malignancy diagnosed in women worldwide, with the majority of cases being hormone receptor positive (HR+), human epidermal growth factor receptor 2 negative (HER2–) and treated with adjuvant endocrine therapy with or without chemotherapy¹. Several commercially available genomic assays exist to assess risk of recurrence and guide decision making for adjuvant systemic therapy². For example, the OncotypeDX® (ODX) test, is a 21-gene expression assay that is commonly used in early-stage HR+/HER2– breast cancer and is both prognostic and predictive of chemotherapy benefit.

Despite their widespread use, genomic assays require lengthy turnaround time and because they are generated following microdissection of tumor cells from an unstained tumor slide, they may not fully capture tumor heterogeneity when performed on a single tissue block or core needle biopsy². The size, shape, and location of the tumor relative to surrounding healthy tissues and vasculature and interactions therewith are all well-known “hallmarks of cancer”³.

Given these limitations, additional biomarkers are needed to individualize therapy. Biophysical models using detailed MRI information that capture macroscopic tumor heterogeneity including blood flow, morphology, and the spatial relationship with nearby healthy tissues have been utilized to predict response to therapy and rate of pathologic complete response in the neoadjuvant setting^{4–7}. Whether similar approaches can be used to predict recurrence remains unknown.

In this work, we developed and validated an image-based AI model (hereafter MRI-RS) incorporating spatial features extracted from pre-treatment dynamic contrast enhanced breast magnetic resonance imaging (DCE-MRI) along with clinicopathologic characteristics to predict recurrence risk in HR+/HER2- breast cancer (Fig. 1). This model utilizes readily available data which will allow clinicians to personalize treatment decision making in real time for women with HR+/HER2- breast cancer.

Methods

Study Population

Patient demographic and pathologic data (hereafter “clinicopathologic data”) and breast DCE-MRI scans were retrospectively collected and analyzed from five cohorts: I-SPY 2 (a randomized clinical trial^{8–10}), University of Chicago Hospital (UCH), University of Alabama (UAB), Endeavor Health Cancer Institute (EHCI), and Duke University (DU). As this is a multi-institution study, this study was performed in adherence with a program level application approved by the Western Institutional Review Board (no. 1379652) and approval of each individual institutional review board. The requirement for informed consent was waived by the respective review boards owing to the retrospective nature of the study and the use of de-identified data. All procedures were conducted in accordance with institutional guidelines and the Declaration of Helsinki.

Altogether, 1072 eligible patients treated between January 1, 2010, and December 31, 2023, were included in the study (Table 1). Eligible patients for this study were women between 18 and 90 years old diagnosed with early-stage HR+/HER2- breast cancer who had undergone pre-operative breast MRI or pre-treatment MRI if neoadjuvant therapy was administered. Patients with Stage I-III A HR+/HER2- invasive breast cancer, with up to 3 axillary lymph nodes were included. Patients with bilateral disease, cN2 or higher clinical nodal status, distant metastasis at diagnosis, inflammatory breast cancer, or recurrent cancer were excluded. DCE-MRIs were required to be acquired with protocols complying with American College of Radiology (ACR) recommendations¹¹. Patients with MRIs that had severe imaging artifacts were excluded from the study.

Table 1
Patient characteristics in the discovery and testing datasets

Age (years)	Discovery (Training/Tuning) Dataset (N = 522)	Testing Dataset (N = 550)	P-value
			0.163
Mean (SD)	55.1 (12.7)	54.1 (11.0)	
Median [Min, Max]	54.0 [24.0, 90.0]	53.0 [24.0, 89.0]	
Race			< 0.001
African American	130 (24.9%)	86 (15.6%)	
Caucasian	345 (66.1%)	418 (76.0%)	
Other	47 (9.0%)	46 (8.4%)	
Cancer Subtype			< 0.001
IDC	386 (74.0%)	353 (64.2%)	
ILC	63 (12.1%)	53 (9.6%)	
Other	73 (14.0%)	5 (1.0%)	
Missing	0 (0.0%)	139 (25.3%)	
Clinical T-category			< 0.001
T1	241 (46.2%)	239 (43.5%)	
T2	174 (33.3%)	254 (46.2%)	
T3/T4	61 (11.7%)	55 (10.0%)	
Missing	46 (8.8%)	2 (0.4%)	
Clinical N-category			0.386
N0	332 (63.6%)	360 (65.5%)	
N1	152 (29.1%)	189 (34.4%)	
Missing	38 (7.3%)	1 (0.2%)	
Tumor Grade			< 0.001
I	67 (12.8%)	69 (12.5%)	
* Therapy sequence reports number of patient receiving that therapy in the specified phase rather than proportions, therefore no statistical test was performed			

Age (years)	Discovery (Training/Tuning) Dataset (N = 522)	Testing Dataset (N = 550)	P-value
			0.163
II	248 (47.5%)	164 (29.8%)	
III	156 (29.9%)	304 (55.3%)	
Missing	51 (9.8%)	13 (2.4%)	
Endocrine Therapy			0.305
Yes	465 (89.1%)	507 (92.2%)	
No	0 (0.0%)	29 (5.3%)	
Missing	57 (10.9%)	14 (2.5%)	
Chemotherapy			< 0.001
Yes	464 (88.9%)	309 (56.2%)	
No	2 (0.4%)	241 (43.8%)	
Missing	56 (10.7%)	0 (0.0%)	
Therapy Sequence*			N/A
Neoadjuvant chemotherapy	178	167	
Adjuvant chemotherapy	286	155	
Neoadjuvant endocrine therapy	0	57	
Adjuvant endocrine	466	481	
Follow-up years			< 0.001
Mean (SD)	6.02 (3.80)	4.68 (2.31)	
Median [Min, Max]	5.26 [0.099, 15.0]	4.28 [0.131, 13.2]	
Recurrence event			< 0.001
No	431 (82.6%)	509 (92.5%)	
Yes	91 (17.4%)	41 (7.5%)	
* Therapy sequence reports number of patient receiving that therapy in the specified phase rather than proportions, therefore no statistical test was performed			

Discovery and Testing Datasets

The five cohorts were assigned to either a discovery dataset or an independent testing dataset (Fig. 1C). Patients from the I-SPY 2 trial and UCH were assigned to the discovery dataset, while the independent testing dataset comprised patients from EHCI, DU, and UAB. The discovery dataset was used to train the prognostic model and further subdivided by randomly assigning patients in a 75:25 ratio to training (N = 391, events = 68) and tuning (N = 131, events = 23) datasets. Once the model was trained, the independent testing dataset was used to establish an unbiased estimate of the model's performance (N = 550, events = 42).

Study Objective

The primary objective of this study was to develop and validate an image-based AI prognostic risk score (hereafter MRI-RS) and associated risk category (hereafter MRI-RC). Recurrence-free survival (RFS) was selected as the study endpoint, defined per STEEP2.0¹² as the time from breast cancer diagnosis to last follow-up or one of the following events: ipsilateral breast tumor recurrence, local-regional invasive recurrence, distant recurrence, invasive contralateral breast cancer, second primary invasive cancer (non-breast), or death from any cause. Patients without an event were censored at the date of last follow-up.

Model Design

Segmentation Model

Pre-treatment fat suppressed, T1-weighted DCE-MRI scans were acquired as 3D volumes. MRIs were processed by a FDA cleared, fully automated, AI-powered segmentation tool (TumorSight Viz, version 1.2; SimBioSys, Inc.)^{13,14} to identify and annotate various tissues in the 3D volume (Fig. 1A). The resulting 3D segmentation assigned each voxel uniquely as tumor, gland, fat, vascular, skin, chest, or air. A region of interest (ROI) was delineated from the fully segmented MRI by creating a bounding box containing the entirety of the voxels segmented as tumor. Radiomics¹⁵ and shaped-based features were derived from the ROI and the overall 3D segmentation.

Feature Selection and Preparation: Feature selection was performed to identify the image-based characteristics that were most prognostic of breast cancer recurrence (Fig. 1A). Features were first z-score normalized, then eliminated according to the following procedure: a) features with low variability were removed, as characterized by those with a low kurtosis measuring parameter ($k = (X_{99\%ile} - median) / (median - X_{1\%ile})$); b) only one feature was kept when two or more features were highly correlated (> 0.8 correlation coefficient); c) features were ranked based on their univariate association with RFS in Cox Proportional Hazard (CPH) models when trained with the top-ranking five to fifteen features.

To limit biases and facilitate generalizability to novel datasets, the discovery dataset was randomly split into a training set and a tuning set. A random survival forest (RSF) ¹⁶ model was calibrated with the training dataset and assessed on the tuning dataset. The tuning dataset was used to determine the best combination of imaging features to use and to select the best model for MRI-RS.

A RSF model trained on clinical variables and the top five image-based features outperformed all the other models by achieving the highest F1, area under the receive operating characteristic (AUC), and concordance index for 5-year RFS prediction. The final MRI-RS model was parameterized as follows: the number of trees in the forest was set to 100, the minimum number of samples required to split an internal node was set to 10, and the minimum number of samples required to be at a leaf node was set to five. Clinical features included age, race, grade, T- and N-stage. The top 5 MRI features selected for the final model included longest dimension of the tumor, shortest axis-aligned tumor length, solidity, compactness, and vascular density of the surrounding region (Fig. 1B). Solidity reflects the convexity of the tumor, while compactness measures how closely the tumor mass resembles a sphere.

Development of Prediction Model

The best-performing model trained using the discovery dataset was locked prior to further analysis. An optimal threshold for the probabilities between high- and low-risk was specified by maximizing the concordance index from the discovery cohort. Specifically, univariate CPH models were fit with varying probability thresholds, and the threshold associated with the maximum C-index was selected as the optimal threshold. Patients predicted to have higher risk than the threshold were labeled as high-risk, and patients whose predicted risk was equal to or lower were labeled as low risk. Features extracted in the independent testing dataset were z-score normalized based on corresponding distributions in the discovery cohort. The locked model was run on the testing dataset to generate risk scores that approximate the probability of a patient experiencing an event within 5 years. Statistical analyses were performed on the resulting MRI-RS model.

Statistical Analyses

Prognostic efficacy was assessed by first assigning an MRI-RS to each patient and subsequently using the pre-specified cutoff to assign patients to either low- or high-risk groups. Kaplan-Meier (KM) estimators were fit to RFS for each group and hazard ratio (HR) between groups computed. Hazard ratios were estimated using Cox proportional hazards models, and statistical significance was assessed using two-sided p-values and 95% confidence intervals ($p < 0.05$). Additionally, the RFS along with 95% confidence intervals (CI), were calculated directly from KM estimators at the five-year post-diagnosis timepoint.

The primary analysis assessed the capability of the MRI-RS to stratify low- and high-risk HR+/HER2- patients overall. Exploratory analyses in clinically relevant subgroups were also performed, including

assessing the prognostic capabilities of MRI-RS by lymph node status, age, and by administration of adjuvant therapy versus neoadjuvant therapy.

The MRI-RS was compared directly to a multivariate CPH model fitted with just clinical factors to demonstrate the added value of the image-based AI features using a likelihood approach (see **Supplemental Methods: Likelihood**). Finally, a propensity score matching technique was used to estimate the impact of chemotherapy benefit for patients in each MRI-RS category¹⁷. (see **Supplemental Methods: Propensity**).

Results

Patient Characteristics

The discovery dataset included 522 women with a median age of 54 (range: 24–90). The majority (66.1%) were Caucasian and 24.9% were African American. The majority had clinical T1 (46.2%) or T2 tumors (33.3%) and 29.1% were clinically node positive (Table 1). Chemotherapy was administered in 88.9% of patients and 89.1% received endocrine therapy. Ninety-one out of 522 patients (17.4%) in the discovery cohort had recurrence events with a median follow-up time of 6.02 years. Thirty-two percent (32.1%) of the discovery cohort were predicted to be high-risk and were associated with higher grade, larger tumors, and nodal involvement.

In the testing dataset of 550 women, patients predicted as high-risk had significantly larger tumors, nodal involvement, and were more likely to be African American (Table 1, Table 2). In the testing data set, 56.2% received chemotherapy and 92.2% received endocrine therapy. With a median follow up of 4.58 years, 41 (9.3%) out of 550 patients developed recurrence events.

Table 2
Patient characteristics in the discovery and testing datasets stratified by risk groups

	Training/Tuning Dataset			Testing Dataset		
	Low-risk (N = 354)	High-risk (N = 168)	P-value	Low-risk (N = 372)	High-risk (N = 178)	P-value
Age (years)	0.0163			0.0366		
Mean (SD)	56.2 (11.4)	53.0 (15.0)		54.4 (9.43)	51.3 (14.4)	
Median [Min, Max]	54.0 [37.0, 90.0]	52.5 [24.0, 83.0]		53.0 [36.0, 89.0]	50.6 [24.0, 83.0]	
Race	0.231			< 0.001		
African American	84 (23.7%)	46 (27.4%)		50 (12.9%)	30 (28.0%)	
Caucasian	242 (68.4%)	103 (61.3%)		306 (78.7%)	70 (65.4%)	
Other	28 (7.9%)	19 (11.3%)		33 (8.5%)	7 (6.5%)	
Cancer Subtype	0.140			0.594		
IDC	248 (70.9%)	138 (80.2%)		238 (64.0%)	115 (64.6%)	
ILC	46 (13.1%)	17 (9.9%)		37 (9.9%)	16 (9.0%)	
Other	56 (1.06%)	17 (9.9%)		2 (0.5%)	3 (1.7%)	
Missing	0 (0.0%)	0 (0.0%)		95 (25.6%)	44 (24.7%)	
Clinical T-category	< 0.001			< 0.001		
T1	201 (56.8%)	40 (23.8%)		207 (53.2%)	18 (16.8%)	
T2	91 (25.7%)	83 (49.4%)		157 (40.4%)	66 (61.7%)	
T3/T4	31 (8.8%)	30 (17.9%)		24 (6.2%)	23 (21.5%)	
Missing	31 (8.8%)	15 (8.9%)		1 (0.3%)	0 (0%)	
Clinical N-category	< 0.001			< 0.001		
N0	252 (71.2%)	80 (47.6%)		272 (69.9%)	50 (46.7%)	
* Therapy sequence reports number of patient receiving that therapy in the specified phase rather than proportions, therefore no statistical test was performed						

	Training/Tuning Dataset			Testing Dataset		
	Low-risk (N = 354)	High-risk (N = 168)	P-value	Low-risk (N = 372)	High-risk (N = 178)	P-value
N1	74 (20.9%)	78 (46.4%)		116 (29.8%)	57 (53.3%)	
Missing	28 (7.9%)	10 (6.0%)		1 (0.3%)	0 (0%)	
Grade			< 0.001			0.00321
I	59 (16.7%)	8 (4.8%)		52 (13.4%)	5 (4.7%)	
II	181 (51.1%)	67 (39.9%)		101 (26.0%)	45 (42.1%)	
III	88 (24.9%)	68 (40.5%)		226 (58.1%)	54 (50.5%)	
Missing	26 (7.3%)	25 (14.9%)		10 (2.6%)	3 (2.8%)	
Endocrine Therapy			0.129			0.239
Yes	305 (87.1%)	160 (93.0%)		345 (92.7%)	162 (91.0%)	
No	0 (0.0%)	0 (0.0%)		16 (4.3%)	13 (7.3%)	
Missing	45 (12.9%)	12 (7.0%)		11 (3.0%)	3 (1.7%)	
Chemotherapy			0.0185			< 0.001
Yes	306 (87.4%)	158 (91.9%)		169 (45.4%)	140 (78.7%)	
No	0 (0.0%)	2 (1.1%)		203 (54.6%)	38 (21.3%)	
Missing	44 (12.6%)	12 (7.0%)		0 (0.0%)	0 (0.0%)	
Therapy Sequence*			N/A			N/A
Neoadjuvant chemotherapy	81	97		81	86	
Adjuvant chemotherapy	225	61		95	60	
Neoadjuvant endocrine therapy	0	0		30	27	
Adjuvant endocrine	306	160		330	151	
* Therapy sequence reports number of patient receiving that therapy in the specified phase rather than proportions, therefore no statistical test was performed						

	Training/Tuning Dataset			Testing Dataset		
	Low-risk (N = 354)	High-risk (N = 168)	P-value	Low-risk (N = 372)	High-risk (N = 178)	P-value
RFS years	< 0.001			0.811		
Mean (SD)	6.76 (3.87)	4.47 (3.11)		4.59 (2.13)	4.53 (2.46)	
Median [Min, Max]	5.99 [0.397, 15.0]	3.30 [0.0985, 14.6]		4.24 [0.131, 13.0]	4.19 [0.515, 13.0]	
RFS event	< 0.001			< 0.001		
No	332 (93.8%)	99 (58.9%)		369 (94.9%)	87 (81.3%)	
Yes	22 (6.2%)	69 (41.1%)		20 (5.1%)	20 (18.7%)	
* Therapy sequence reports number of patient receiving that therapy in the specified phase rather than proportions, therefore no statistical test was performed						

Prognostic value of MRI-RS

The direct outputs of MRI-RS are a numerical value indicating a patient's risk of recurrence at 5 years and an indicator of whether the patient is considered high- or low-risk. In the RSF model, the ensemble prediction is the average across all scores, and the computed risk score represents the expected number of events for one node in one decision tree. This score was shown to be correlated with the empirical probability of recurrence in the testing dataset (Fig. 1). The score was then classified as either "High Risk" or "Low Risk" based on a threshold of 32.9 (see **Methods: Model Design - Development of Prediction Model**). Risk scores that are higher than 32.9 were labeled high risk, while those equal or lower were labeled low risk. In the discovery and testing datasets, the selected cutoffs distributed 32.2% and 19.5% of subjects to the high-risk group, respectively. In the testing dataset, probability of recurrence in the low-risk group is 6.2% (95%CI: 3.8–10.0%) and is 17.4% (95%CI: 11.9–25.2%) in the high-risk group, while in the discovery cohort recurrence probabilities were 0.7% (95%CI: 0.2–2.7%) and 38.7% (95%CI: 30.9–47.8%), for low- and high-risk, respectively (**Supplemental Fig. 1**).

Univariate CPH models were built with the categorical prediction of 5-year RFS as the independent variable. Binary MRI-RC were significantly associated with RFS in the independent testing data with an unadjusted hazard ratio of 3.09 (95% CI: 1.66–5.76; $p < 0.005$) (Fig. 2). MRI-RC remained significantly associated with RFS after adjusting for age, clinical T/N-category, grade, and whether the patient was treated with neoadjuvant, endocrine, and chemotherapy with an adjusted hazard ratio of 2.28 (95%CI: 1.18–4.41; $p < 0.01$). The cumulative 5-year recurrence free survival (RFS) for patients predicted as low risk by the MRI-RS was 93.8% (95% CI: 90.0–96.2%) compared to 82.6% (95% CI: 74.8–88.1%) in patients

predicted as high risk. Positive and negative predictive value of the test at 5 years were 17.4% (95% CI: 11.9–25.2%) and 93.8% (95% CI: 90.0–96.2%), respectively. The numerical score was found to correlate with RFS (Fig. 3). A linear fit between the numerical score and RFS shows a positive, statistically significant ($p < 0.036$; weighted least squares¹⁸) association between the AI risk score and empirical recurrence rate in the holdout test dataset. Low- and high-risk patients, stratified by the score cutoff, have recurrence rates that are statistically different ($p < 0.00018$; logrank test¹⁹).

Performance of risk model across subgroups

Subgroup analyses indicate MRI-RS was prognostic regardless of age or nodal status. In patients aged less than 50 years, the 5-year RFS for low-risk group was 90.8% versus 80.0% in the high-risk group (uHR 3.08; 95% CI: 1.23–7.73; aHR 3.96; 95% CI: 1.33–11.76). In patients over 50 years of age, the 5-year RFS was 95.2% in the low-risk group vs 84.4% in the high-risk group (uHR 2.89; 95% CI: 1.23–6.81; aHR 1.73; 95% CI: 0.71–4.21). The model showed similar directional prognostic separation in both node-negative (uHR 3.63; 95% CI: 1.60–8.23; aHR 2.82; 95% CI: 1.16–6.83) and node-positive (uHR 2.65; 95% CI: 0.93–7.54; aHR 1.85; 95% CI: 0.68–5.05) cohorts, although estimates in the node-positive subgroup were less precise (Fig. 4).

Overall, the performance of the MRI-RS outweighed a “clinical-only” risk model trained in a similar fashion as demonstrated by multi-variate CPH modeling. To establish a clinical risk score (CRS) baseline, a random survival forest model with identical hyperparameter as MRI-RS was trained with clinical factors (T- and N-category, grade, age, race/ethnicity). The CRS was then combined with MRI-RS to assess the added value of the imaging features. Specifically, (1) the MRI-RS was added to the CRS as a continuous variable and (2) the MRI-RC was added to the CRS as risk groups. In both cases, a null model consisting of CRS alone was compared to an alternate model using a likelihood ratio (LR) test. The MRI-RS was found to be more prognostic than clinical variables alone (Table 3). Explainability analysis using the SHAP technique²⁰ demonstrated the importance of image-derived which all ranked in the top half of model features (**Supplemental Methods: Explainability** and **Supplemental Fig. 2**).

Table 3

Summary of comparison between a clinical factors model and the AI Risk Score. The image-based risk score was compared to a Cox proportional hazards model fitted with tumor size, nodal status, race/ethnicity, grade, and age as covariates. When the image-based score (MRI-RS) or categorical prediction (MRI-RC) was combined with the predicted clinical risk score (CRS), the predictive capabilities were significantly enhanced relative to the clinical factors model alone. Adding imaging features either as a continuous variable (MRI-RS) or as a risk category (MRI-RC) to the clinical features yielded significantly more prognostic estimation of recurrence free survival (RFS) than clinical factors alone ($p < 0.01$).

Model	Log-likelihood	Likelihood Ratio	P-value	C-Index
Clinical Risk Score (CRS)	-259.3	N/A	N/A	0.58
CRS + MRI-RS	-250.1	18.4	< 0.01	0.71
CRS + MRI-RC	-253.7	11.1	< 0.01	0.69

The secondary objective of this study was to investigate the clinical utility of MRI-RS and prediction of chemotherapy benefit. Since this was not a randomized control trial, a causal comparison cannot be made. Instead, a propensity score matching approach was used to simulate a fair trial with a “control arm” in the testing dataset²¹. We found that propensity score matching led to patients in the MRI-RS low-risk group with higher RFS with endocrine therapy alone as compared to the chemoendocrine therapy group (**Supplemental Fig. 3**). Conversely, patients identified as high-risk group treated with chemoendocrine therapy were less likely to recur at 5 years. While suggesting that the MRI-RS could be used to tailor chemotherapy use, this analysis is underpowered with a limited number of events.

Discussion

Genomic assays are frequently used to guide adjuvant therapy decisions for patients with HR+/HER2- breast cancer but are limited by the need for sufficient tissue and lengthy processing time. Clinical uncertainty also exists regarding the applicability of these genomic assays in specific high-risk subgroups, such as premenopausal women, racial and ethnic minorities and those with larger tumors, where tumor heterogeneity is more likely. To address this clinical gap, we developed and validated an image-based AI prognostic risk score (MRI-RS) using readily available clinicopathologic data and spatial characteristics of the tumor and tumor microenvironment that queries both the tumor and surrounding breast tissue. Since MRI is frequently performed in the early-stage setting, this tool could allow clinicians and patients to further personalize treatment decision making for HR+/HER2- breast cancer using a modality that is already readily available.

In this validation study, we found the MRI-RS was significantly associated with RFS demonstrating an adjusted HR of 2.28 (95% CI: 1.18–4.41; $p < 0.01$). Patients identified as high-risk were 11.2% more likely to have a recurrence than low-risk patients. Notably, subgroup analyses indicated the image-based prediction model was prognostic regardless of age or nodal status. The performance of the combined imaging and clinical risk score was superior to that of the clinical only prediction model. Together, these

results demonstrate that the MRI can be used as an orthogonal biomarker for recurrence in HR+/HER2- breast cancer.

MRI-RS is a novel, non-invasive, prognostic tool that could complement established tissue-based genomic testing. It offers actionable, real-time risk stratification for personalized treatment planning across diverse clinical subgroups. While our ability to assess chemotherapy benefit is constrained by the study's retrospective design, our study indicates MRI-RS has the potential to aid in optimizing adjuvant therapy, guiding chemotherapy escalation or de-escalation based on recurrence risk. Furthermore, whereas genomic- and pathology-based assays identify intrinsic drivers of recurrence, MRI-RS evaluates extrinsic, spatial factors within the tumor macroenvironment. Future research should investigate whether integrating intrinsic and extrinsic drivers of recurrence enhances prognostic accuracy.

Key strengths of our analysis lie in the robust study design and diverse data sources, which enhance the generalizability of the MRI-RS. The study included a large and diverse patient cohort from multiple centers, specifically incorporating a substantial population of pre-menopausal and Black or African American women, groups often underrepresented in AI studies. Unlike many single-site image-based AI studies, the generalizability of the MRI-RS was rigorously assessed using independent training, tuning, and external testing datasets, ensuring the testing population was drawn from different institutions than those used for model development. Images were acquired from over a dozen different MRI machines, decreasing potential scanner-specific biases and increasing the likelihood that the MRI-RS will translate effectively to new clinical cohorts.

This was a retrospective study with several limitations and potential unmeasured confounding variables. The analysis was conducted in an early-stage HR+/HER2- breast cancer population with relatively short follow-up, resulting in a low number of events and differing underlying patient populations between cohorts. The prediction of chemotherapy benefit could not be adequately assessed in this retrospective analysis and thus warrants further investigation through a prospective study. While the test is based on pre-treatment imaging, patients treated with both adjuvant and neoadjuvant chemotherapy were included in the study, which may result in hidden confounders due to differences in downstream treatment.

Conclusions

In this large, retrospective study, we developed and validated an image-based AI model that accurately predicted 5-year RFS in women with early-stage HR+/HER2- breast cancer. The model demonstrated robust prognostic value, regardless of age or lymph node status, which significantly exceeded the performance of traditional clinical-pathologic features alone. By providing real-time, non-invasive risk stratification without the need for tissue sampling or lengthy turnaround times, this tool may enable personalized treatment optimization. Future investigations will focus on assessing the model's ability to predict chemotherapy response to refine personalized therapeutic strategies.

Abbreviations

AUC
Area under the receiver operating characteristic
CI
Confidence interval
CPH
Cox proportional hazards model
CRS
Clinical risk score
DCE
MRI–Dynamic contrast enhanced MRI
aHR
Adjusted hazard ratio
HR
Hazard ratio
uHR
Unadjusted hazard ratio
HR+
Hormone receptor positive
HER2
–Human epidermal growth factor receptor 2 negative
LR
Likelihood ratio
MRI
Magnetic resonance imaging
MRI
RS–MRI–risk score
MRI
RC–MRI–risk category
SHAP
Shapley additive explanations
ODX
OncotypeDX®
RFS
Recurrence free survival
RSF
Random survival forest

Declarations

Data Availability

The datasets generated and/or analyzed during the current study are not publicly available due to institutional data governance policies, and Institutional Review Board (IRB) restrictions at participating institutions. De-identified data may be made available from the corresponding author upon reasonable request and subject to institutional review, data use agreements, and IRB approval where applicable.

Code Availability

Custom code supporting the findings of this study is available from the corresponding author upon reasonable request, subject to institutional approvals and data governance requirements.

Ethics approval and consent to participate

As this is a multi-institution study, this study was performed in adherence with a program level application approved by the Western Institutional Review Board (no. 1379652) and approval of each individual institutional review board/ethics committee. The requirement for informed consent was waived by the respective review boards owing to the retrospective nature of the study and the use of de-identified data. All procedures were conducted in accordance with institutional guidelines and the Declaration of Helsinki.

Author Contributions: PS, JP, and BF had full access to all of the data in the study and take responsibility for the integrity of the data and the accuracy of the data analysis.

Concept and design: MG, PS, JC, BF, JP

Acquisition, analysis, or interpretation of data: PS, FH, ESR, YZ, DC, BFP

Drafting of the manuscript: PS, JP, BF, DC

Critical revision of the manuscript for important intellectual content: All authors.

Statistical analysis: JP, YZ, MW, VK, DC, BF

Acknowledgements

This study received no funding.

Conflict of Interest Disclosures:

JRP, JAC, DC, and VK declare a financial interest in SimBioSys, Inc.

JRP, JAC, DC, BF, YZ, BF, and VK are employees of SimBioSys, Inc.

FMH reports consulting fees from Novartis and Leica Biosystems, advisory board fees from Exact Sciences, and stock ownership in Celcuity.

JCB declare funding paid to institution from SimBioSys, Inc. and Eli Lilly.

JCB declares honorarium from Up-to-Date and sits on DSMC for CairnSurgical

ESR reports consulting fees from AstraZeneca, Daiichi Sankyo, Gilead, SeaGen, Novartis, Pfizer, and Lilly.

Dr. Goetz reports personal fees for CME activities from AXIS, BroadcastMed, DAVA Oncology, IDEOlogy Health, MJH Life Sciences, PeerView, Physicians' Education Resource, Research to Practice, Total Health Conferencing;

Consulting fees to Mayo Clinic from 858 Therapeutics, Inc., Amneal Pharmaceuticals, AstraZeneca, AstraZeneca Pharmaceuticals LP, AstraZeneca UK Ltd., Ataraxis AI, BeiGene USA, Biotheranostics, Biotheryx, eChinaHealth, EcoR1, Eli Lilly and Company, Engage Health Media, Genentech, Incyclix Bio, Lilly, Laekna, Novartis, Pfizer, Puma, RNA Diagnostics, Seattle Genetics, Sermonix Pharmaceuticals, Stemline Therapeutics, TerSera Therapeutics/ Amplity Health;

Grant funding to Mayo Clinic from AstraZeneca, ATOSSA Therapeutics, Biotheryx, Lilly, Loxo, Pfizer, Puma, Sermonix, SimBioSys; travel support from Lilly and Biotheryx

Meeting Presentation: This study was presented during the San Antonio Breast Cancer Symposium (SABCS) 2024 Poster Spotlight Session; December 2024; San Antonio, Texas

References

1. Howlader N, Altekruse SF, Li CI, et al. US Incidence of Breast Cancer Subtypes Defined by Joint Hormone Receptor and HER2 Status. *JNCI: Journal of the National Cancer Institute*. 2014;106(5). doi:10.1093/jnci/dju055
2. Ehsani S, Wisinski KB. Genomic Testing in the Management of Early-Stage Breast Cancer. *J Clin Outcomes Manag*. 2017;24(5):229–238.
3. Hanahan D. Hallmarks of Cancer: New Dimensions. *Cancer Discov*. 2022;12(1):31–46. doi:10.1158/2159-8290.CD-21-1059
4. Cole JA, Peterson JR, Earnest TM, et al. SimBioSys TumorScope: Spatio-temporal modeling of the tumor microenvironment to predict chemotherapeutic response. *Journal of Clinical Oncology*. 2020;38(15_suppl):e12650-e12650. doi:10.1200/JCO.2020.38.15_suppl.e12650
5. Howard FM, He G, Peterson JR, et al. Highly accurate response prediction in high-risk early breast cancer patients using a biophysical simulation platform. *Breast Cancer Res Treat*. 2022;196(1):57–66. doi:10.1007/s10549-022-06722-0
6. Peterson JR, Cole JA, Pfeiffer JR, et al. Novel computational biology modeling system can accurately forecast response to neoadjuvant therapy in early breast cancer. *Breast Cancer Research*. 2023;25(1):54. doi:10.1186/s13058-023-01654-z
7. Whitman J, Adhikarla V, Tumyan L, et al. Validation of Clinical Dynamic Contrast-Enhanced Magnetic Resonance Imaging Perfusion Modeling and Neoadjuvant Chemotherapy Response Prediction in

- Breast Cancer Using ^{18}F FDG and ^{64}Cu -DOTA-Trastuzumab Positron Emission Tomography Studies. *JCO Clin Cancer Inform.* 2025;(9). doi:10.1200/CCI.23.00248
8. Li W, Newitt DC, Gibbs J, et al. Predicting breast cancer response to neoadjuvant treatment using multi-feature MRI: results from the I-SPY 2 TRIAL. *NPJ Breast Cancer.* 2020;6(1):63. doi:10.1038/s41523-020-00203-7
 9. Li W, Newitt DC, Wilmes LJ, et al. I-SPY 2 Breast Dynamic Contrast Enhanced MRI Trial (ISPY2) (Version 1) [Data set]. *The Cancer Imaging Archive.* Preprint posted online 2022. doi:10.7937/TCIA.D8Z0-9T85
 10. Newitt DC, Partridge SC, Zhang Z, et al. ACRIN 6698/I-SPY2 Breast DWI [Data set]. *The Cancer Imaging Archive.* Preprint posted online 2021.
 11. Price R, Allison J, Clarke G, et al. *2015 ACR MR QC Manual.* 2015.
 12. Tolaney SM, Garrett-Mayer E, White J, et al. Updated Standardized Definitions for Efficacy End Points (STEEP) in Adjuvant Breast Cancer Clinical Trials: STEEP Version 2.0. *Journal of Clinical Oncology.* 2021;39(24):2720–2731. doi:10.1200/JCO.20.03613
 13. Weitz M, Pfeiffer JR, Patel S, et al. Performance of an AI-powered visualization software platform for precision surgery in breast cancer patients. *NPJ Breast Cancer.* 2024;10(1):98. doi:10.1038/s41523-024-00696-6
 14. Pekis A, Kannan V, Kaklamanos E, Antony A, Patel S, Earnest T. Seeing beyond cancer: multi-institutional validation of object localization and 3D semantic segmentation using deep learning for breast MRI. In: Astley SM, Chen W, eds. *Medical Imaging 2024: Computer-Aided Diagnosis.* SPIE; 2024:14. doi:10.1117/12.3009341
 15. van Timmeren JE, Cester D, Tanadini-Lang S, Alkadhi H, Baessler B. Radiomics in medical imaging –“how-to” guide and critical reflection. *Insights Imaging.* 2020;11(1):91. doi:10.1186/s13244-020-00887-2
 16. Ishwaran H, Kogalur UB, Blackstone EH, Lauer MS. Random survival forests. *Ann Appl Stat.* 2008;2(3). doi:10.1214/08-AOAS169
 17. Xu Z, Kalbfleisch JD. Propensity Score Matching in Randomized Clinical Trials. *Biometrics.* 2010;66(3):813–823. doi:10.1111/j.1541-0420.2009.01364.x
 18. Strutz T. *Data Fitting and Uncertainty: A Practical Introduction to Weighted Least Squares and Beyond.* 2nd ed. Springer Fachmedien; 2015.
 19. Peto R, Peto J. Asymptotically Efficient Rank Invariant Test Procedures. *J R Stat Soc Ser A.* 1972;135(2):185. doi:10.2307/2344317
 20. Štrumbelj E, Kononenko I. Explaining prediction models and individual predictions with feature contributions. *Knowl Inf Syst.* 2014;41(3):647–665. doi:10.1007/s10115-013-0679-x
 21. ROSENBAUM PR, RUBIN DB. The central role of the propensity score in observational studies for causal effects. *Biometrika.* 1983;70(1):41–55. doi:10.1093/biomet/70.1.41

22. Kline A, Luo Y. PsmPy: A Package for Retrospective Cohort Matching in Python. In: 2022 44th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC). IEEE; 2022:1354–1357. doi:10.1109/EMBC48229.2022.9871333

23. Rubin DB. The Use of Matched Sampling and Regression Adjustment to Remove Bias in Observational Studies. *Biometrics*. 1973;29(1):185. doi:10.2307/2529685

Figures

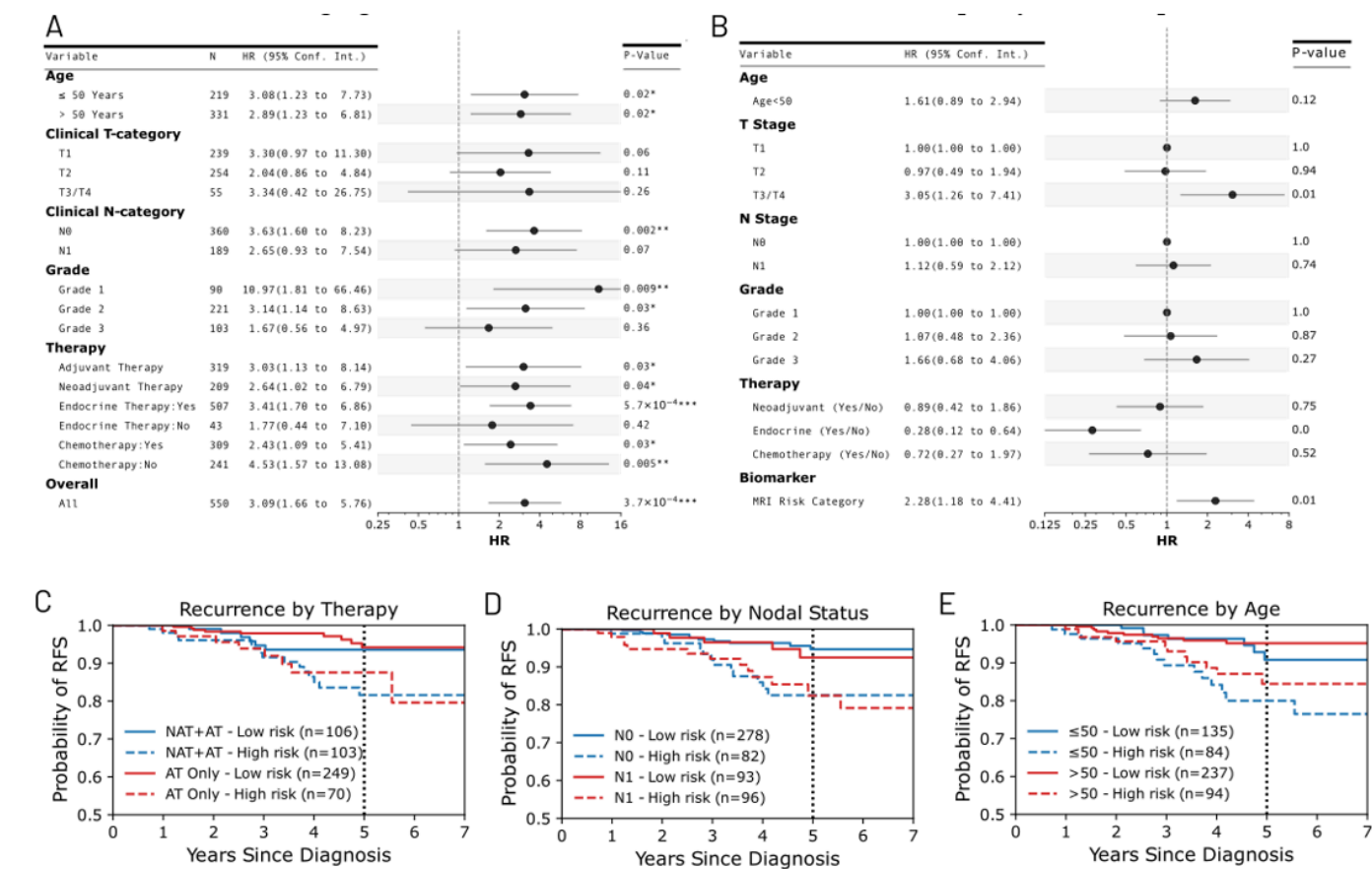


Figure 1

Image-Based AI Risk Score. A) An image-based risk score was developed using features derived from standard of care dynamic contrast-enhanced breast MR images acquired for patients diagnosed with early-stage or locally advanced breast cancer. Spatial and radiomic features were extracted from the MRI via a validated multi-tissue segmentation algorithm^{13,14} and subsequently used to train a machine learning model after iterative feature selection. AI model accuracy at predicting probability of recurrence free survival was assessed in an independent, multi-institution holdout testing dataset. B) Explainable features related to vascularity proximal to the tumor, and tumor size and morphology were found to be predictive of risk of recurrence. C) Over 1000 subjects from over five institutions were included in the study. Patients were stratified into training, tuning (together the discovery dataset) or testing datasets, with no overlap of institution between discovery and testing datasets.

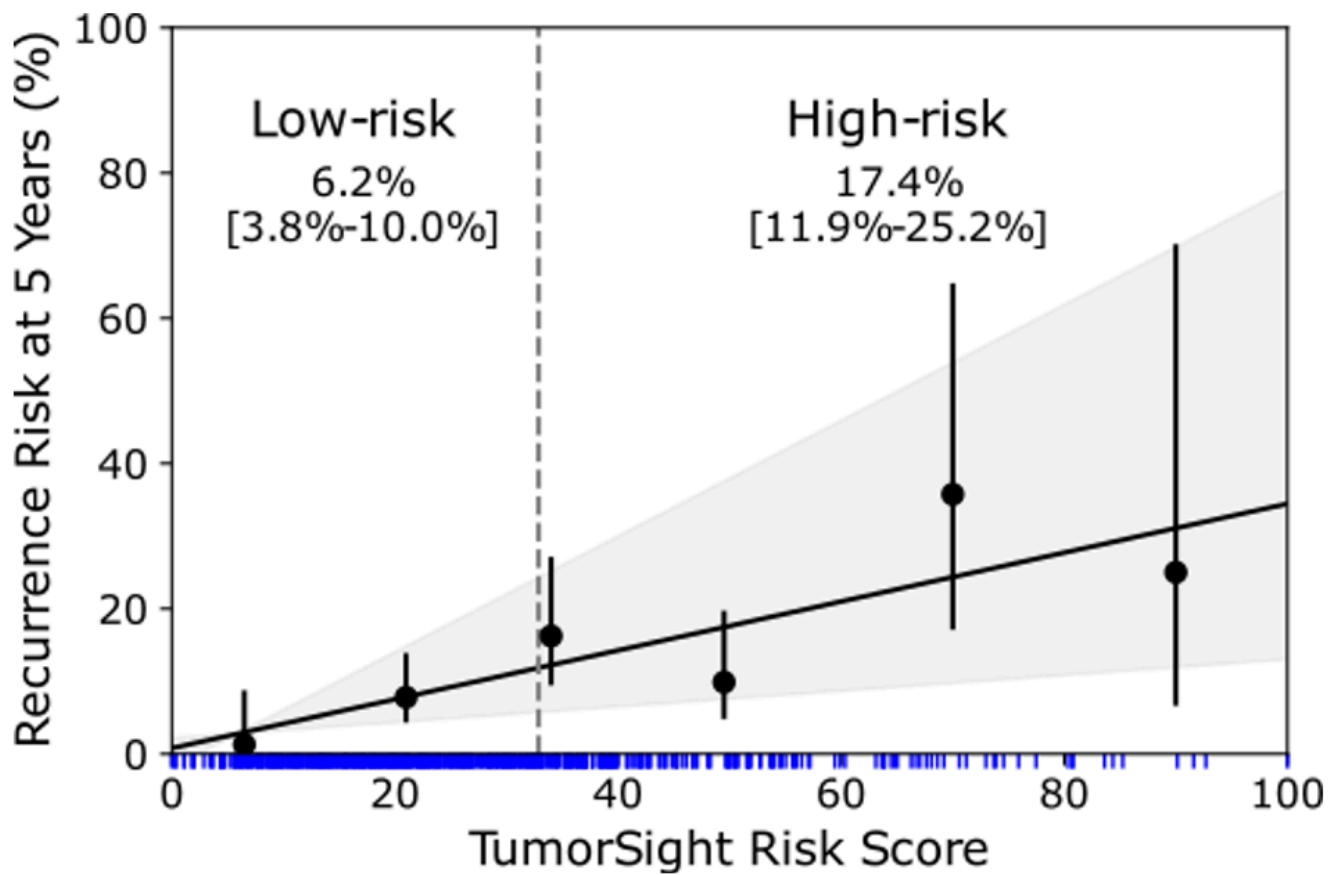


Figure 2

Recurrence-free survival by AI Risk groups. Kaplan-Meier survival curves showing the probability of recurrence-free survival for low- and high-risk groups in the independent, multi-institution holdout test dataset, as predicted by the AI Risk model. Log-rank p-value is reported for 5-year risk recurrence.

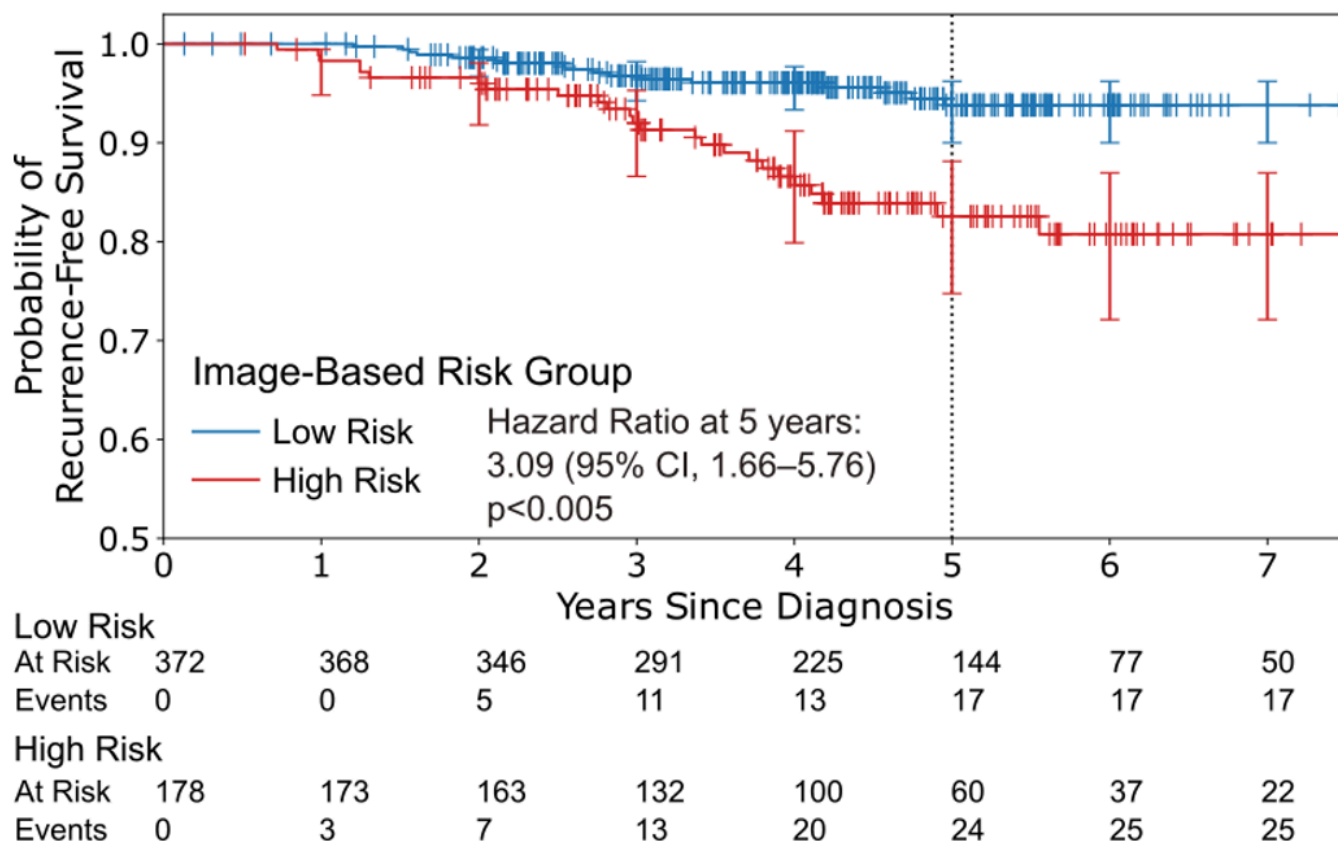


Figure 3

MRI Risk Score and Probability of Recurrence. Probability of recurrence risk evaluated five years post-diagnosis in the independent, multi-institution holdout testing dataset is highly correlated with the image-based AI risk score. Density of patients in each bucket is depicted as a rugplot showing the distribution across the entire score range. Points with error bars depict the empirically determined recurrence rates for patients in each risk score range. A linear fit shows a positive, statistically significant ($p < 0.036$; weighted least squares¹⁸) association between the AI risk score and empirical recurrence rate in the holdout test dataset. Low- and high-risk patients, stratified by the score cutoff (depicted by the vertical dashed line), have the recurrence rates that are statistically different ($p < 0.00018$; logrank test¹⁹).

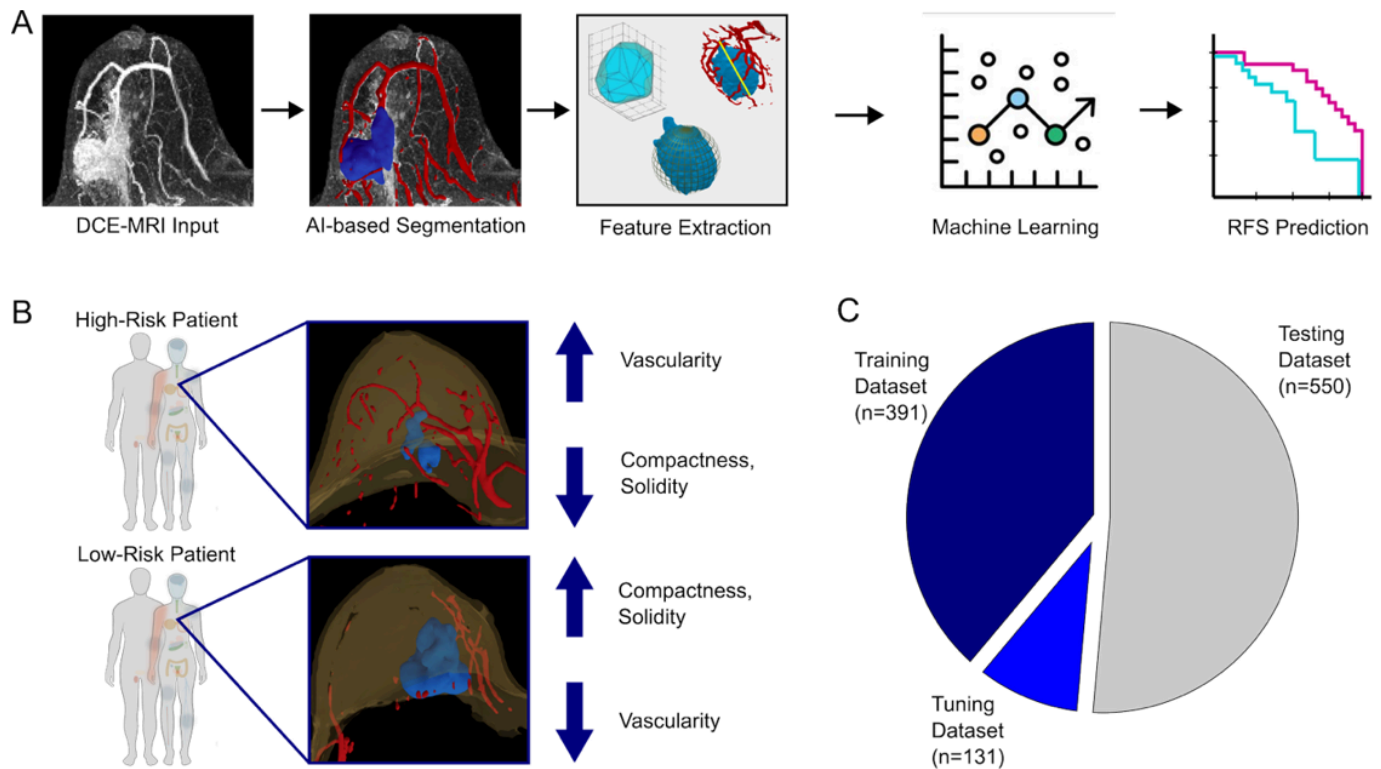


Figure 4

A) Forest plot showing univariate analyses for the unadjusted hazard ratio (HR) between low- and high-risk categories determined by the AI risk score across relevant clinical subgroups. While not statistically significant in every category, the results demonstrate that the hazard is similar magnitude (~2-3), and directionally the same across all clinical subgroups. B) Forest plot showing a multivariate analysis accounting for demographic, staging, pathology, and treatment factors, demonstrating that the MRI-RC is independently prognostic of survival. Kaplan-Meier curves for low- and high-risk groups for selected subgroups including: (C) those treated with neoadjuvant therapy (NAT) or adjuvant therapy (AT), (D) nodal status, and (E) age greater than or less than 50, which is used as a proxy for menopausal status.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [SupplementalFigures.docx](#)