Diagn Interv Radiol 2025; DOI: 10.4274/dir.2025.253376

BREAST IMAGING

ORIGINAL ARTICLE

Magnetic resonance imaging-based artificial intelligence model predicts neoadjuvant therapy response in triple-negative breast cancer

Raşit Eren Büyüktoka[1]
Zehra Hilal Adıbelli[2,9]
Murat Sürücü[3]
Özlem Özdemir[4]
Yalçın İşler[5]
Ali Murat Koç[6]
Aslı Dilara Büyüktoka[2]
Demet Kocatepe Çavdar[7]
Özge Aslan[8]
Serhat Değer[2]
Ayşenur Oktay[8]

[1]İzmir Foça State Hospital, Clinic of Radiology, İzmir, Türkiye

[2]University of Health Sciences Türkiye, İzmir City Hospital, Department of Radiology, İzmir, Türkiye

[3]Burdur Mehmet Akif Ersoy University Faculty of Bucak Computer and Informatics, Department of Software Engineering, Burdur, Türkiye

[4]University of Health Sciences Türkiye, İzmir City Hospital, Department of Medical Oncology, İzmir, Türkiye

[5]Alanya Alaaddin Keykubat University Faculty of Engineering and Architecture, Department of Electrical and Electronics Engineering, Antalya, Türkiye

[6]İzmir Katip Çelebi University Faculty of Medicine, Department of Radiology, İzmir, Türkiye

[7]University of Health Sciences Türkiye, İzmir City Hospital, Department of Pathology, İzmir, Türkiye

[8]Ege University Faculty of Medicine, Department of Radiology, İzmir, Türkiye

[9]University of Health Sciences Türkiye, İzmir Faculty of Medicine, İzmir, Türkiye

Corresponding author: Raşit Eren Büyüktoka

E-mail: rasiterenbuyuktoka@hotmail.com

Received 04 June 2025; revision requested 10 August 2025; last revision received 01 September 2025; accepted 05 October 2025.

Epub: 27.10.2025

Publication date: 02.03.2026

DOI: 10.4274/dir.2025.253376

PURPOSE

Triple-negative breast cancer (TNBC) is an aggressive subtype of breast cancer with limited treatment options and poorer overall survival than other subtypes. Neoadjuvant chemotherapy (NACT) is often used to reduce tumor size and improve surgical outcomes. However, predicting patients' response to NACT remains challenging, and non-responding patients risk unnecessary chemotoxicity. This study aimed to develop a deep learning-based artificial intelligence (AI) model using pre-treatment magnetic resonance imaging (MRI) to predict pathological complete response (pCR) in patients with TNBC undergoing NACT.

METHODS

This retrospective, double-centered study included 49 lesions from 43 patients with TNBC. Data from MRI, including T2-weighted, T1-weighted, and diffusion-weighted imaging, were segmented and processed to train a residual convolutional neural network model.

RESULTS

The AI model achieved an accuracy of 0.82 and an area under the receiver operating characteristic curve of 0.75 in differentiating pCR from non-pCR cases. The model's performance was validated through intra- and inter-reader agreement metrics, with Dice similarity coefficients ranging from 0.821 to 0.915.

CONCLUSION

Our results demonstrate that AI models can effectively predict NACT responses in patients with TNBC using only pre-treatment MRI data.

CLINICAL SIGNIFICANCE

This proof-of-concept study supports the potential for AI-based tools to aid clinical decision-making and reduce the risks associated with ineffective therapies. Future research with larger datasets and additional imaging modalities is needed to improve model generalizability and clinical applicability.

KEYWORDS

Breast cancer, artificial intelligence, neoadjuvant chemotherapy, magnetic resonance imaging, residual convolutional neural network

Breast cancer (BC) is a common health problem worldwide and remains the most common cancer type among women. Despite its high incidence, mortality rates have consistently decreased over the last decades due to technological advancements in imaging and novel therapeutic options.[1] BC has different subtypes, and each subtype has a different prognosis. It is crucial to evaluate the tumor molecularly to assess the patient's treatment options and clinical outcomes.[2]

Triple-negative BC (TNBC) is characterized by the lack of estrogen receptors, progesterone receptors, and expression of human epidermal growth factor receptor 2. It is the most aggressive subtype and has the least favorable overall survival (OS); it is diagnosed in almost

You may cite this article as: Büyüktoka RE, Adıbelli ZH, Sürücü M, et al. Magnetic resonance imaging-based artificial intelligence model predicts neoadjuvant therapy response in triple-negative breast cancer. *Diagn Interv Radiol*. 2026;32(2):191-199.

15%–20% of all patients with BC. In contrast to other subtypes, TNBC has limited hormonal and target-specific treatment options.[2-4]

Neoadjuvant chemotherapy (NACT) for BC is increasingly used to decrease the tumor volume and to downstage the disease to create a bridge to surgery.[5] Early TNBC is commonly treated with surgery and adjuvant chemotherapy.[6] Furthermore, unresectable and locally advanced TNBC treatment is mainly based on NACT.[6,7] Compared with adjuvant chemotherapy, preoperative systemic therapy for BC has no advantages in disease-free survival or OS.[8,9] However, there is a survival advantage in patients who achieve pathological complete response (pCR) after NACT compared with those with residual disease.[10,11] With NACT becoming the standard treatment, clinicians have focused on patients who do not achieve pCR. This is because patients without pCR show poorer survival outcomes than those with pCR, and post-NACT has been applied to patients without pCR to achieve long-term survival outcomes.[11] Imaging studies and physical exams have provided early response assessments, helping distinguish non-responders. This allows for alternative treatments to overcome resistance, aiming to improve pCR rates and forming the basis for post-neoadjuvant treatment strategies.[12-14]

Assessment of disease stage is mainly based on radiological examinations. Imaging modalities include mammography, ultrasound (US), magnetic resonance imaging (MRI) of the breast, and positron emission tomography/computed tomography.[15] Evaluation of the response after completion of NACT is based on radiological examinations. Mammography and US are routinely used to assess the response to NACT.[16] However, after the initiation of NACT, it is impossible to predict the patient's response status with conventional radiological methods.[17,18]

Main points

- Triple-negative breast cancer (TNBC) is the most aggressive and least common breast cancer subtype.
- Pre-treatment magnetic resonance imaging (MRI) may contain helpful information for artificial intelligence (AI) models to predict neoadjuvant chemotherapy response in advance to individualize treatment.
- Our AI model predicts therapy response in TNBC using pre-treatment MRI data and achieved accuracy of 0.82 and area under the curve of 0.75 in predicting pathological complete response (pCR) compared with non-pCR.

In this proof-of-concept study, we introduce a deep learning-based artificial intelligence (AI) model using pre-treatment MRIs to predict the NACT response status before the initialization of NACT. Convolutional neural networks (CNNs) are artificial neural networks composed of multiple layers, specifically designed to evaluate datasets that contain grid-like (coordinate-based) information such as radiological images.[19-21] We hypothesized that tumor appearances in different MRI sequences, as reflected by different gray-level pixel presentations and tumor features, can be deciphered by a residual CNN-based AI model using pre-treatment MRIs.

Methods

Study design and patient population

Our study was a retrospective double-center study conducted in accordance with the Declaration of Helsinki, and this retrospective study was approved by the University of Health Sciences Türkiye, İzmir Bozyaka Training and Research Hospital Clinical Research Ethics Committee (decision number: 2023/19, date: 08.02.2023). Due to the retrospective design of the study, informed consent was waived by the local ethics committee.

Patients with biopsy-proven TNBC underwent and completed NACT between 2018 and 2023. These patients had pathology data at the time of initial diagnosis and underwent breast MRI before NACT. A flowchart of the patient selection, inclusion, and exclusion criteria is presented in Figure 1.

Magnetic resonance imaging acquisition

MRIs of the patients were acquired at two different centers using 1.5 Tesla MRI units (Magnetom Amira and Symphony, Siemens Healthineers, Erlangen, Germany / Philips Achieva, Philips Medical Systems, Drachten, Netherlands) and a 3-Tesla MRI unit (Magnetom Verio, Siemens Healthineers, Erlangen, Germany). All patients were imaged in the prone position using a breast coil. The MRI sequences included fast spin echo (FSE) T2-weighted images (T2WIs), b800 diffusion-weighted images (DWIs), and fat-suppressed pre- and post-contrast images at 180 seconds, which were used for segmentation. For contrast-enhanced images, 0.1 mmoL/kg of gadobutrol (Gadovist®, Bayer, Germany) or gadoteric acid (Clariscan®, GE Healthcare, Norway) was injected as a rapid bolus, followed by a 10-mL saline flush at 2-mL/s. The 180-second post-contrast images were used to feed the AI algorithm.

Definition of pathological complete response

After completion of NACT, pathological response data from surgical specimens were classified as pCR and non-pCR. Pathological classifications were made according to the Miller-Payne grading system, with Grade 5 classified as a complete response and Grade 4 or below classified as no pCR.[22]

Lesion segmentation

During data collection, the leading researcher (R.E.B.) included 49 lesions from 43 patients based on the inclusion and exclusion criteria. The images were anonymized using local software, all image labels were removed, and new patient numbers were assigned post-anonymization. After anonymization, the FSE-T2WIs, DWIs, and pre- and post-contrast fat-suppressed T1-weighted images (T1WIs) were selected for annotation. The researcher evaluated the images along with pathological data. The lesions were segmented volumetrically in three-dimensional (3D) polygon mode using ITK-SNAP 4.x open-source software in FSE-T2WIs, DWIs, and post-contrast images.

After the initial segmentation, following an interval of at least 1 month, the researcher randomly selected 20% of the lesions from each sequence for re-segmentation to calculate intra-observer agreement using different metrics. Moreover, 20% of the lesions in each sequence were re-segmented by a second radiologist (A.D.B.) with similar experience, and inter-observer agreement was calculated.

Artificial intelligence model

Data preprocessing

Before entering the data into the deep learning network, several preprocessing steps were applied.

- Segmentation: Lesions identified by radiologists were annotated on the imaging sequences.
- Image cropping: Images were cropped to include only the annotated lesions.
- Image scaling: Lesions were resized to a fixed 50 × 50 × 50 scale, with zero padding used for any gaps.
- Normalization: The pixel values of the 3D tumor slices were normalized between 0 and 1.

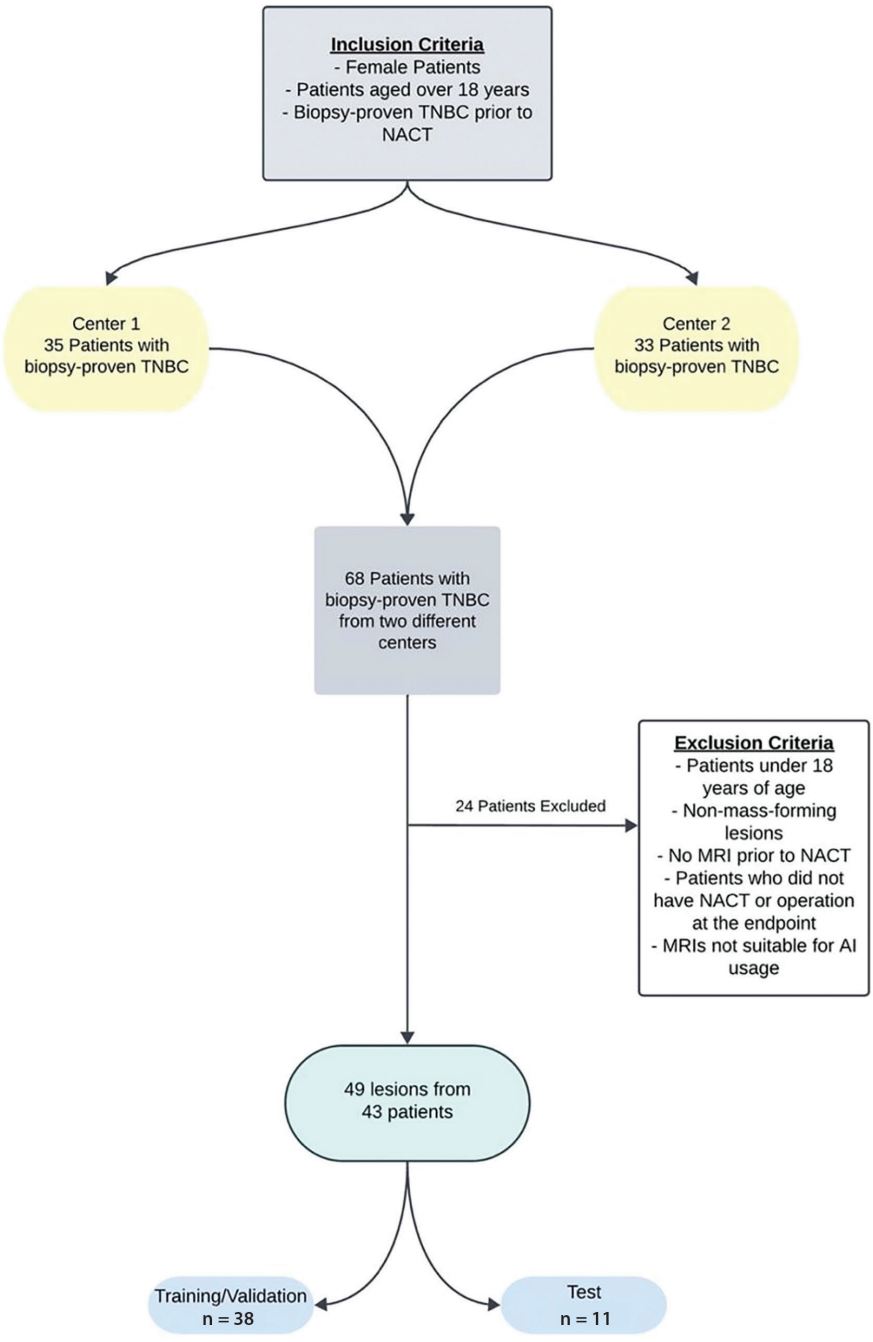

Figure 1. Flowchart of patient selection. TNBC, triple-negative breast cancer; NACT, neoadjuvant chemotherapy; MRI, magnetic resonance imaging; AI, artifical intelligence.

• Data splitting: The normalized tumor slices were randomly split into "Training," "Validation," and "Test" sets (random state: 42). After splitting, 30 tumor slices were selected for training, 8 for validation, and 11 for testing.

Data augmentation

Due to the relatively small dataset and imbalanced data distribution, data augmentation was applied to the training set. There were 13 lesions in the "pCR" class and 17 in the "non-pCR" class. To address this imbalance, data augmentation was first applied to underrepresented classes. Each class was then further augmented by randomly rotating the 3D MRI slices on the two-dimensional (2D; x, y) axis.

Deep learning model

Residual CNNs were used for their advantages in processing limited data and achieving better generalization. A residual CNN layer was designed in accordance with the ResNet architecture (Figure 2). The network input consisted of a 50 × 50 × 50 lesion image. A 2D CNN layer with 64 channels, followed by batch normalization and MaxPooling (MP) layers, reduced the data to 25 × 25 × 64. After two residual blocks and a 128-channel 2D CNN layer, the data were further reduced to 12 × 12 × 128 through another MP layer. Finally, a flattening layer produced a feature pool of 18,432 attributes. Similar processes were applied to other imaging sequences, and the features were combined after they had passed through the residual CNN layers. Although lesions were segmented volumetrically, the implemented architecture functions as a 2D CNN, operating on individual axial slices.

Due to their low count, T2WI sequences were excluded from the study. The features extracted from the pre-contrast T1WI, post-contrast T1WI, and DWI sequences were combined for each lesion, and classification was performed through a fully connected network with 1,792,896, and 256 neurons, respectively, in three dense layers (Figure 3). However, we evaluated multiple input configurations: (i) post-contrast T1WIs alone and (ii) multi-sequence inputs (pre-contrast T1WIs, post-contrast T1WIs, DWIs) using the same backbone. Due to sequence availability and performance on the test set, the final model reported in the Results section utilizes post-contrast T1WIs only. Despite the limited training data, accuracy values comparable to those in the literature were achieved. Residual CNNs offer key advantages, such as easier learning, robustness to model complexity, and training efficiency.

Statistical analysis

All statistical analyses were performed using R statistical software (version 3.6.0, Posit Software, PBC). Descriptive statistics were calculated to summarize patient and lesion characteristics. Continuous variables were expressed as mean ± standard deviation (SD) for normally distributed data and median with quartile values (Q1, Q3) for non-normally distributed data. Categorical variables were presented as absolute frequencies and percentages. Between-group comparisons for continuous variables were conducted using the Student's t-test or the Mann–Whitney U test according to distributional assumptions,

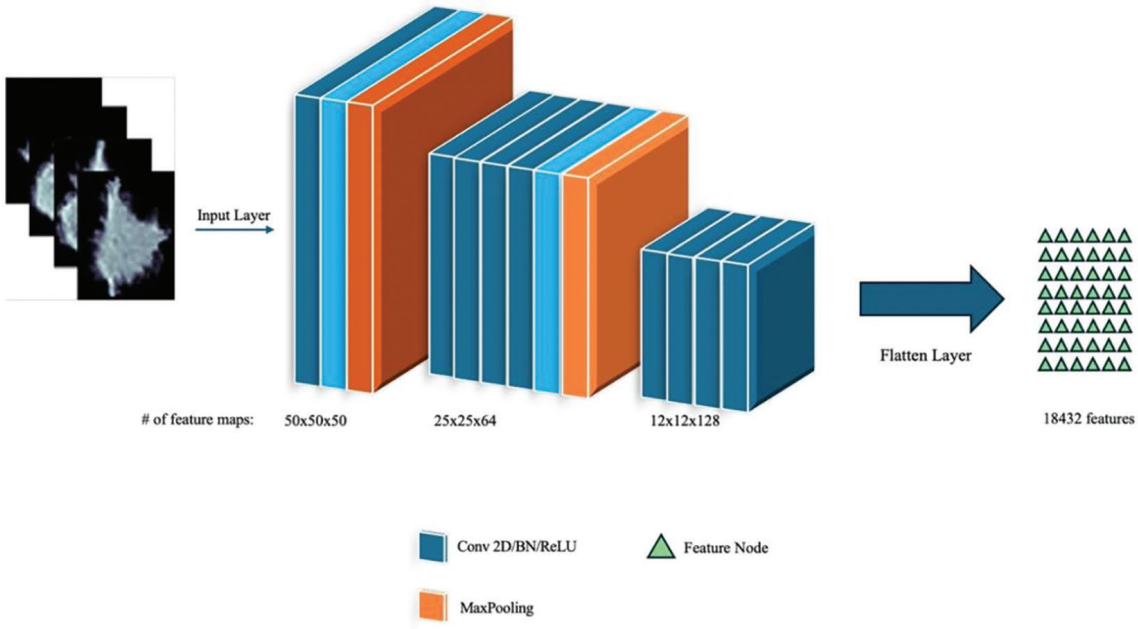

Figure 2. Residual convolutional neural network (CNN) layers. The network input consisted of a 50 × 50 × 50 lesion image. A two-dimensional (2D) CNN layer with 64 channels, followed by batch normalization and MaxPooling (MP) layers, reduced the data to 25 × 25 × 64. After two residual blocks and a 128-channel 2D CNN layer, the data were further reduced to 12 × 12 × 128 through further MP layer. Finally, a flattening layer produced a feature pool of 18,432 attributes.

whereas categorical variables were compared using the Fisher–Freeman–Halton test, as appropriate. It was considered statistically significant when $P < 0.05$. Variables found to be statistically significant in univariable analyses were included in a multivariable logistic regression model to identify independent predictors of pCR following NACT. The results were expressed as odds ratios (ORs) with 95% confidence intervals (CIs). The overall model fit was assessed using the likelihood ratio test, and predictive performance was evaluated with Nagelkerke's pseudo R^2. Model calibration was tested using the Hosmer–Lemeshow goodness-of-fit test. Model performance on the test set was evaluated by calculating accuracy from the confusion matrix (true-positive, true-negative, false-positive, and false-negative counts). Receiver operating characteristic (ROC) curves were plotted, and the area under the curve (AUC) was computed directly from the classification results. Intra-reader agreement was assessed by comparing repeated segmentations from the same reader on the same dataset using the Dice similarity coefficient formula. Pairwise Dice values were computed between all readers, and the mean (± SD) Dice value was reported to summarize inter- and intra-reader agreement.

All analyses were performed using functions from the readxl, dplyr, compareGroups, broom, and ResourceSelection packages in R.

Results

Descriptive results

The patient and lesion characteristics of the 43 patients included in the study are summarized in Table 1.

The study includes a total of 49 lesions, with 20 (40.82%) achieving pCR and 29 (59.18%) not achieving pCR. The mean age of patients in the pCR group was 50.1 ± 10.9 years, slightly older than the non-pCR group, which had a mean age of 48.9 ± 13.3 years. Patients who achieved pCR had significantly smaller median tumor sizes at baseline ($P = 0.034$), with a median of 28.5 mm (Q1–Q3: 22.5–32.0), than those who did not achieve pCR, who had a median tumor size of 35.5 mm (Q1–Q3: 24.25–56.5). The median tumor volume on post-contrast T1WIs was significantly less ($P = 0.045$) in the pCR group [median 9,243 mm^3 (Q1–Q3 3,714–13,665 mm^3)] than in the non-pCR group [median 19,453 mm^3 (Q1–Q3: 5,029–58,595 mm^3)].

In the multivariable logistic regression analysis, tumor volume measured on post-contrast T1WIs and the Ki-67 proliferation index were found to be independent predictors of achieving pCR after NACT. Tumor volume was associated with pCR (adjusted OR: 1.00; 95% CI: 1.00–1.00; $P = 0.040$), and higher Ki-67 levels were significantly associated with increased odds of pCR (adjusted OR: 1.04; 95% CI: 1.01–1.07; $P = 0.018$). Tumor size did not reach statistical significance (adjusted OR: 1.05; 95% CI: 0.96–1.17; $P = 0.299$). The overall model demonstrated a good fit (Nagelkerke's pseudo $R^2 = 0.422$, Hosmer–Lemeshow test $P = 0.702$) and was statistically significant according to the likelihood ratio test ($P < 0.001$). These results are summarized in Table 2.

Intra-reader and inter-reader agreement results

The AI algorithm was fed with 3D volumetric segmentations, and its reliability was evaluated by assessing intra-reader agreement using different scores for each sequence. Accordingly, the average Dice coefficient for segmentations performed on DWIs was 0.841 ± 0.075, and for segmentations performed on post-contrast T1WI sequences, the average Dice coefficient was 0.915 ± 0.046. Considering the inter-reader agreement between the radiologists based on different segmentations, the average Dice coefficient was 0.821 ± 0.050 for segmentations performed on DWIs and 0.890 ± 0.059 for segmentations performed on post-contrast T1WI sequences. These data demonstrate that the segmentations performed by the primary researcher at different times and those performed by the second researcher were highly consistent.

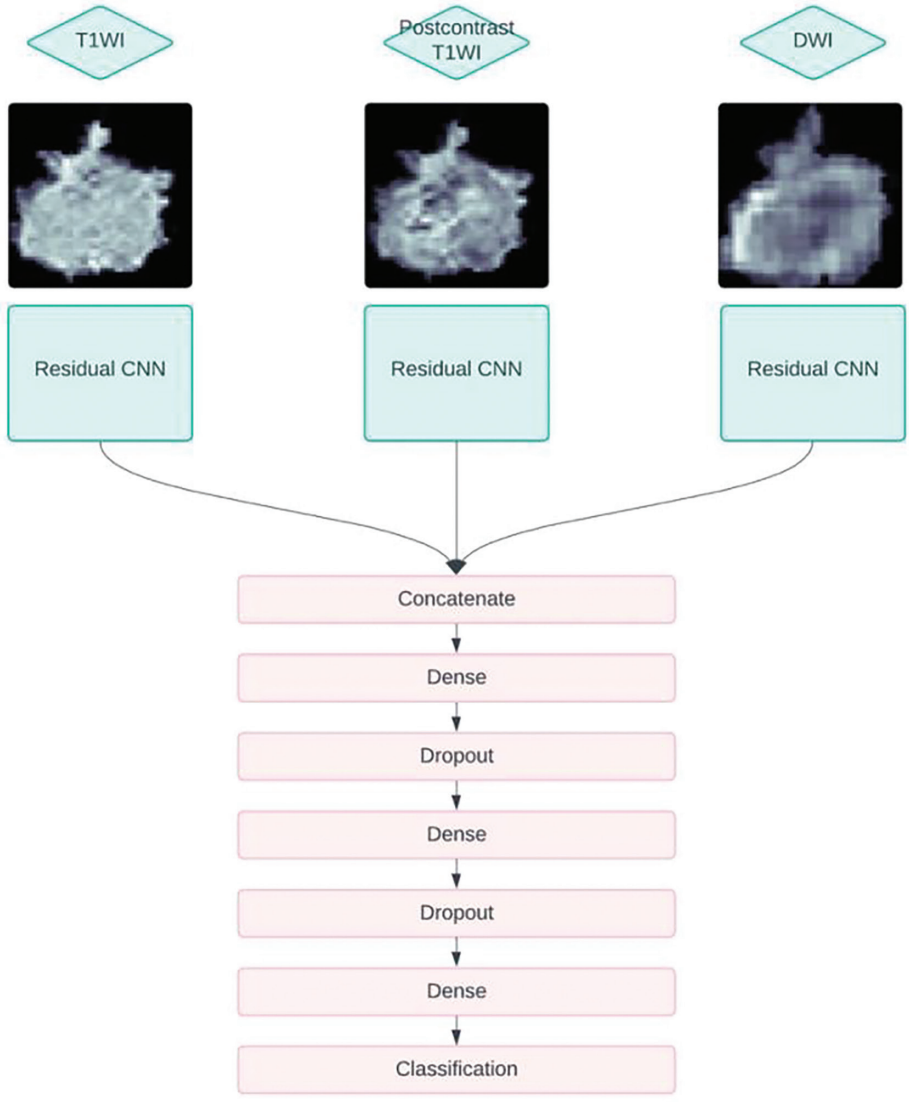

Figure 3. Deep learning model. The features extracted from the pre-contrast T1-weighted image (T1WI), post-contrast T1WI, and diffusion-weighted image (DWI) sequences were combined for each lesion, and classification was performed through a fully connected network with 1,792,896, and 256 neurons, respectively, in three dense layers.

*Note: We tested alternative model configurations using multiple sequences (pre-T1WI, post-T1WI, DWI); however, the final reported model uses post-contrast T1WI only. CNN, convolutional neural network.

Artificial intelligence model results

The best AI model for differentiating pCRs from non-pCRs on the test set revealed an accuracy of 0.82 (95% CI: 0.545–1.000) and AUC ROC of 0.75 (Figure 4). The best-performing model used only post-contrast T1WI data. These results demonstrate that the CNN-based AI model can predict response status with high performance. True-positive and true-negative examples predicted by the model are presented in Figure 5.

Discussion

TNBC is a rare but aggressive subtype of BC with a higher risk of metastasis than other subtypes.[2-4] Neoadjuvant therapy improves surgical outcomes, but its success is still unpredictable. If patients do not respond to this therapy, they face unnecessary toxicities. Therefore, predicting NACT response would help optimize treatment, reduce chemotoxicity risks, and improve clinical decision-making.

Despite advances in radiology, there is still a lack of data for accurately predicting NACT outcomes. Although AI is increasingly applied in radiology, few studies have focused on predicting NACT response in BC, especially for the TNBC subtype. This study aims to fill this gap by using only pre-treatment MRIs to predict responses in patients with TNBC. Our AI model, based on a CNN, achieved an accuracy of 0.82 in distinguishing patients who achieved pCR. Our model has several advantages over the previous studies that tried to predict or detect the response status of NACT in patients with BC. First, our model used only pre-treatment MRIs to classify patients as pCR or non-pCR. This extends the period for clinicians to modify the treatment plan and enhances their decision-making when concluding NACT early. Second, our model tried to predict responses using more sequences than previously used, which may add additional value to the AI model by using the different features of the various sequences. However, the best-performing model was identified as that using only post-contrast T1WI data to predict NACT response status. This might be because tumor heterogeneity is best determined in this sequence, and other sequences, such as DWIs and pre-contrast T1WIs, might lack sufficient data for the AI model to extract. Therefore, we also segmented the tumors in 3D volumetrically, enhancing the information that is acquired from the tumors. Moreover, selecting only a few slices of the tumor might create selection bias. Finally, unlike in earlier studies,[23-26] our model is based on the biopsy-proven TNBC subtype. This is because different types of BC behave differently to NACT, and studies including various types of BC might have heterogeneity that influences the results of the AI model in the future.

In terms of conventional analysis, after logistic regression analysis, we found that the parameters that might help identify NACT predictors in TNBC were the tumor proliferation index (Ki-67) and volume. Previous studies have shown that tumor Ki-67 values can predict response to NACT.[27] In a study by Penault-Llorca et al.,[28] which examined the predictive performance of various pathological markers for NACT in different types of BC with 710 patients, high Ki-67 values were found to be significant in predicting complete response, consistent with our findings. Similarly, MacGrogan et al.[29] identified high Ki-67 as an independent predictor of NACT response in patients with BC (n = 128). By contrast, Petit et al.[30] observed higher Ki-67 values in the complete response group but reported that the difference was not statistically significant. Additionally, studies by Bottini et al.[31] and Estévez et al.[32] found that Ki-67 was not a key predictor of NACT response. These different results are thought to arise from variations in patient groups and treatment protocols.

Table 1. Descriptive results of the dataset indicate that tumor size (mm), tumor volume (mm^3), and proliferation index (Ki67-) are significantly different between pCR and non-pCR groups

		pCR	Non-pCR	
Total		20	29	
Age (years) mean (± SD)		50.1 (10.9)	48.9 (13.3)	$P = 0.732$
Mammographic density BIRADS	A (%)	2 (50%)	2 (50%)	$P = 0.556$
	B (%)	10 (50%)	10 (50%)	
	C (%)	5 (27.7%)	13 (72.3%)	
	D (%)	3 (42.8%)	4 (57.2%)	
Tumor size (mm) median (Q1; Q3)		28.5 (22.5; 32.0)	35.5 (24.25; 56.5)	**$P = 0.041$**
Tumor volume post-contrast T1WI (mm^3) median (Q1; Q3)		9,243 (3,714; 13,665)	19,453 (5,029; 58,595)	**$P = 0.030$**
Background parenchymal enhancement BIRADS	Minimal (%)	13 (44.8%)	16 (55.2%)	$P = 0.782$
	Mild (%)	5 (38.5%)	8 (61.5%)	
	Moderate (%)	2 (40%)	3 (60%)	
	Marked (%)	0 (0%)	2 (100%)	
Proliferation (Ki-67) median (Q1; Q3)		80 (50; 80)	50 (30; 77.5)	**$P = 0.027$**

SD, standard deviation; pCR, pathological complete response; T1WI, T1-weighted image.

Table 2. Multivariable logistic regression analysis of predictors for pathological complete response

	Adjusted OR (95% CI)	*P*
Tumor size (mm)	1.05 (0.96–1.17)	0.299
Tumor volume, post-contrast T1WI (mm^3)	1.00 (1.00–1.00)	0.040
Proliferation (Ki-67)	1.04 (1.01–1.07)	0.018

Pseudo R^2 (Nagelkerke): 0.422, Hosmer–Lemeshow $P = 0.702$, likelihood ratio test $P < 0.001$; OR, odds ratio; CI, confidence interval; T1WI, T1-weighted image.

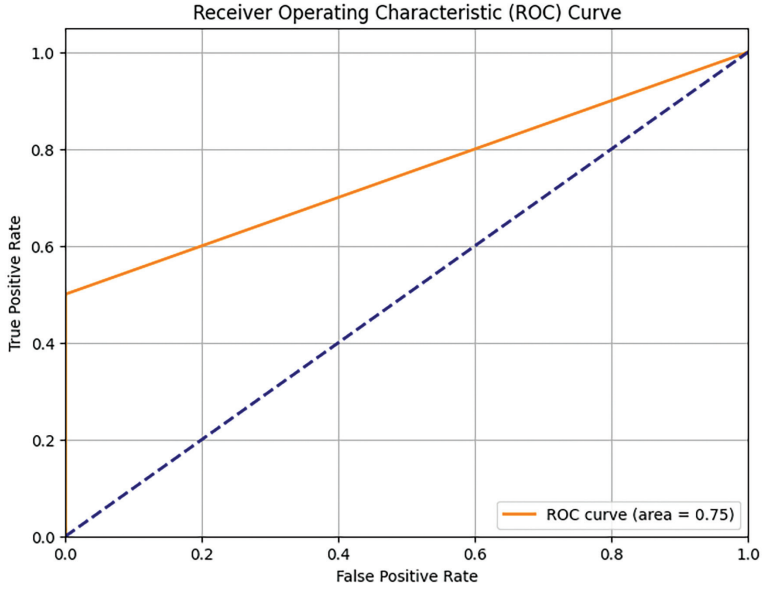

Figure 4. Receiver operating characteristic curve for the best-performing convolutional neural network model using post-contrast T1-weighted imaging to differentiate pathological complete response (pCR) from non-pCR in the test set (n = 11).

Besides conventional analysis, several studies support the potential of AI in predicting NACT outcomes. However, most of these studies used either one imaging method or one sequence, both pre-treatment and post-treatment images in combination, or all subtypes of the BC for the dataset. For instance, Herrero Vicent et al.[24] combined multiparametric MRIs and clinical data to create a machine learning model. This study, conducted on a small group of 58 patients, achieved an accuracy of 0.87 using only radiological imaging features.[24] Similarly, our study achieved high accuracy despite using a small patient group, demonstrating that AI models can perform well even with limited data.

Skarping et al.[23] developed an AI model using pre-treatment digital mammograms to predict pCR in all BC subtypes. This model was applied to 453 lesions, and an AUC score of 0.71[23] was achieved. Although their model used pre-treatment mammographic data, ours focused on MRI, demonstrating the versatility of imaging modalities in AI applications. In addition, we used different sequences to better understand the information in each sequence. In another study, Qu et al.[26] tested deep learning models on different imaging sets, including pre- and post-neoadjuvant T1WIs. Their model using only pre-treatment images had a lower AUC score of 0.55, but the combined model achieved a high AUC of 0.97.[26] This suggests that combining imaging datasets could substantially improve prediction accuracy; however, our study achieved stronger performance by only using pre-treatment images. Ha et al.[25] developed a CNN using pre-neoadjuvant MRI data from 141 patients, achieving an impressive AUC score of 0.98. This high accuracy highlights the potential of deep learning methods in predicting therapy responses. These findings suggest that with more data and re-

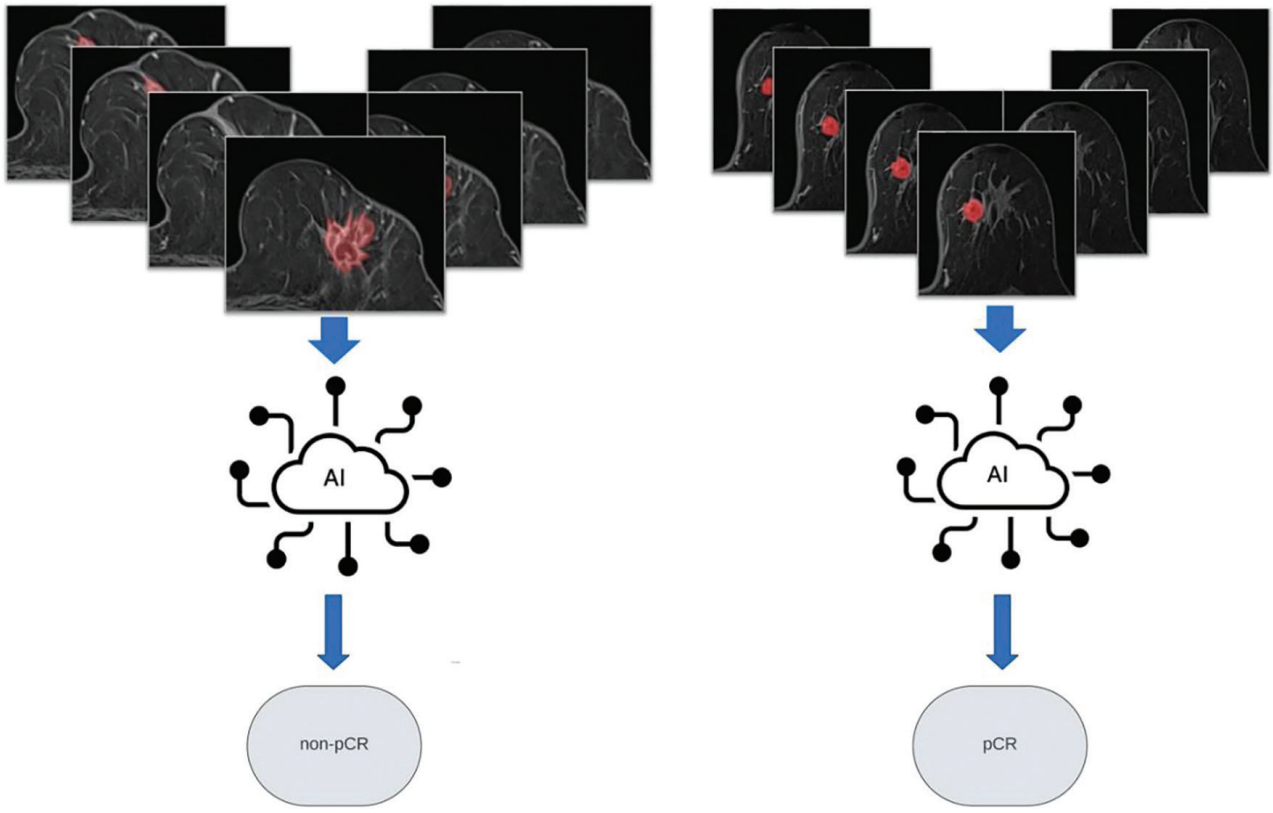

Figure 5. Examples of the artificial intelligence (AI) model's prediction. The upper-left tumor has a median size of 70 mm, a volume of 68,870 mm^3, and a Ki-67 value of 35, with no response. By contrast, the upper-right tumor has a median size of 25 mm, a volume of 3,714 mm^3, and a Ki-67 value of 50, with a complete response. pCR, pathological complete response.

fined methodologies, the performance of AI models such as that presented in this study could be further enhanced. Zhou et al.[33] developed an AI model focusing solely on TNBC, using MRI datasets collected before and after four cycles of NACT. This study achieved an accuracy of 0.77, and using open-source data allowed them to expand their patient group to 162. Furthermore, this study used both pre-treatment and post-treatment images to increase the performance, but they failed to note whether they tried only pre-treatment images for any model.[33] Previous studies are summarized in Table 3.

Overall, these studies highlight the promise of AI-based models in predicting NACT responses. As seen in the literature and our study, AI can provide high accuracy in predicting therapy outcomes, although larger patient groups and refined methodologies are necessary to enhance performance. Integrating clinical and radiological data and AI can substantially aid clinical decision-making processes.

This study has several limitations. First, the cohort size and small test set constrain statistical power and widen uncertainty around performance estimates. Second, clinical staging at diagnosis was not consistently available across centers, which precluded stage-stratified analyses and may introduce clinical heterogeneity. Third, although we initially evaluated multiple MRI sequences, T2WIs were excluded because complete, high-quality series were insufficiently available across patient cohorts and centers; moreover, in other models within our sample, adding DWIs and/or pre-contrast T1WIs did not improve discrimination over post-contrast T1WIs alone. These factors may limit generalizability and should be addressed in larger, prospectively curated, multi-institutional cohorts. Moreover, patients with non-mass enhancement were excluded due to difficulties in tumor segmentation. Although rare, this exclusion limits the model's applicability to specific patient subgroups. Finally, this study evaluates a single residual CNN backbone without head-to-head comparisons against alternative deep learning architectures or classic machine learning approaches using hand-crafted radiomics, which restricts the scope for architectural comparison.

Future research focusing on external validation across multiple institutions and scanners, prospective enrollment to ensure complete clinical staging and acquisition protocols, and development of multimodal models that fuse imaging-derived representations with clinical biomarkers (e.g., Ki-67) might improve discrimination, calibration, and decision utility. With larger datasets and more complete sequence availability, we will revisit multi-sequence inputs and explore end-to-end 3D architectures and other architectural designs to test whether additional sequences and different architectures (e.g., T2WIs, DWIs) provide incremental value beyond post-contrast T1WIs.

In conclusion, AI-based models hold considerable potential in predicting NACT responses, particularly for aggressive subtypes such as TNBC. These models can improve clinical outcomes by optimizing treatment plans and personalizing care. However, expanding research with larger, multicenter datasets is necessary to enhance the models' generalizability and ensure broader clinical application. With continued advancements, AI can play a crucial role in the future of personalized BC treatment.

Table 3. Comparison of the results of different studies, indicating that artificial intelligence-assisted models can predict the neoadjuvant therapy response status in different categories for patients with breast cancer

Study	Data used	Method	Total number	Test number	Subtypes of breast cancer	Performance result
Herrero Vicent et al.[24]	Multiparametric MRI and clinical data	Machine learning	58	24	All types of breast cancer	0.87 of accuracy (only with radiological images)
Skarping et al.[23]	Digital mammograms	Deep learning	453	53	All types of breast cancer	0.71 AUC
Qu et al.[26]	Pre-treatment and post-treatment post-contrast T1WI	Deep learning	302	58	All types of breast cancer	0.55 AUC (only with pre-treatment images) 0.97 AUC (combined)
Ha et al.[25]	Pre-treatment post-contrast T1WI	Deep learning	141	28	All types of breast cancer	0.98 AUC
Zhou et al.[33]	Pre-treatment and post-treatment post-contrast T1WI and DWI	Deep learning	162	32	Triple-negative breast cancer	0.77 of accuracy
This study	**Pre-treatment post-contrast T1WI**	**Deep learning**	**49**	**11**	**Triple-negative breast cancer**	**0.82 of accuracy**

MRI, magnetic resonance imaging; T1WI, T1-weighted image; AUC, area under the curve; DWI, diffusion-weighted image.

Footnotes

Conflict of Interest

The authors declared no conflicts of interest.

References

1. Siegel RL, Giaquinto AN, Jemal A. Cancer statistics, 2024. *CA Cancer J Clin*. 2024;74(1):12-49. [Crossref]
2. Loibl S, Poortmans P, Morrow M, Denkert C, Curigliano G. Breast cancer. *Lancet*. 2021;397(10286):1750-1769. [Crossref]
3. Cho N. Imaging features of breast cancer molecular subtypes: state of the art. *J Pathol Transl Med*. 2021;55(1):16-25. [Crossref]
4. Yin L, Duan JJ, Bian XW, Yu SC. Triple-negative breast cancer molecular subtyping and treatment progress. *Breast Cancer Res*. 2020;22(1):61. [Crossref]
5. Curigliano G, Burstein HJ, Gnant M, et al. Understanding breast cancer complexity to improve patient outcomes: The St Gallen International Consensus Conference for the Primary Therapy of Individuals with Early Breast Cancer 2023. *Ann Oncol*. 2023;34(11):970-986. [Crossref]
6. Cardoso F, Kyriakides S, Ohno S, et al. Electronic address: clinicalguidelines@esmo.org. Early breast cancer: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up†. *Ann Oncol*. 2019;30(8):1194-1220. [Crossref]
7. Golshan M, Loibl S, Wong SM, et al. Breast conservation after neoadjuvant chemotherapy for triple-negative breast cancer: surgical results from the brightness randomized clinical trial. *JAMA Surg*. 2020;155(3):e195410. [Crossref]
8. Bear HD, Anderson S, Brown A, et al. The effect on tumor response of adding sequential preoperative docetaxel to preoperative doxorubicin and cyclophosphamide: preliminary results from National Surgical Adjuvant Breast and Bowel Project Protocol B-27. *J Clin Oncol*. 2003;21(22):4165-4174. [Crossref]
9. Fisher B, Bryant J, Wolmark N, et al. Effect of preoperative chemotherapy on the outcome of women with operable breast cancer. *J Clin Oncol*. 1998;16(8):2672-2685. [Crossref]
10. Rastogi P, Anderson SJ, Bear HD, et al. Preoperative chemotherapy: updates of National Surgical Adjuvant Breast and Bowel Project Protocols B-18 and B-27. *J Clin Oncol*. 2008;26(5):778-785. [Crossref]
11. Cortazar P, Zhang L, Untch M, et al. Pathological complete response and long-term clinical benefit in breast cancer: the CTNeoBC pooled analysis. *Lancet*. 2014;384(9938):164-172. [Crossref]
12. Caparica R, Lambertini M, Pondé N, Fumagalli D, de Azambuja E, Piccart M. Post-neoadjuvant treatment and the management of residual disease in breast cancer: state of the art and perspectives. *Ther Adv Med Oncol*. 2019;11:1758835919827714. [Crossref]
13. Matuschek C, Jazmati D, Bölke E, et al. Post-neoadjuvant treatment strategies in breast cancer. *Cancers (Basel)*. 2022;14(5):1246. [Crossref]
14. Bozer A, Yilmaz C, Çetin Tunçez H, Kocatepe Çavdar D, Adıbelli ZH. Correlation of histopathological and radiological response patterns and their prognostic implications in breast cancer after neoadjuvant chemotherapy. *Breast Cancer (Dove Med Press)*. 2024;16:1005-1017. [Crossref]
15. Expert Panel on Breast Imaging:; Slanetz PJ, Moy L, et al. ACR appropriateness criteria® monitoring response to neoadjuvant systemic therapy for breast cancer. *J Am Coll Radiol*. 2017;14(11S):S462-S475. [Crossref]
16. Ollivier L, Balu-Maestro C, Leclère J. Imaging in evaluation of response to neoadjuvant breast cancer treatment. *Cancer Imaging*. 2005;5(1):27-31. [Crossref]
17. Kong X, Moran MS, Zhang N, Haffty B, Yang Q. Meta-analysis confirms achieving pathological complete response after neoadjuvant chemotherapy predicts favourable prognosis for breast cancer patients. *Eur J Cancer*. 2011;47(14):2084-2090. [Crossref]
18. Houssami N, Macaskill P, von Minckwitz G, Marinovich ML, Mamounas E. Meta-analysis of the association of breast cancer subtype and pathologic complete response to neoadjuvant chemotherapy. *Eur J Cancer*. 2012;48(18):3342-3354. [Crossref]
19. Li J, Jiang P, An Q, Wang GG, Kong HF. Medical image identification methods: a review. *Comput Biol Med*. 2024;169:107777. [Crossref]
20. Torres AD, Yan H, Aboutalebi AH, Das A, Duan L, Rad P. Patient facial emotion recognition and sentiment analysis using secure cloud with hardware acceleration. *Computational Intelligence for Multimedia Big Data on the Cloud with Engineering Applications. Elsevier*. 2018:61-89. [Crossref]
21. Surucu M, Isler Y, Kara R. Diagnosis of paroxysmal atrial fibrillation from thirty-minute heart rate variability data using convolutional neural networks. *Turkish Journal of Electrical Engineering and Computer Sciences*. 2021;29:2886-2900. [Crossref]
22. Wang W, Liu Y, Zhang H, et al. Prognostic value of residual cancer burden and Miller-Payne system after neoadjuvant chemotherapy for breast cancer. *Gland Surg*. 2021;10(12):3211-3221. [Crossref]
23. Skarping I, Larsson M, Förnvik D. Analysis of mammograms using artificial intelligence

to predict response to neoadjuvant chemotherapy in breast cancer patients: proof of concept. *Eur Radiol.* 2022;32(5):3131-3141. [Crossref]

24. Herrero Vicent C, Tudela X, Moreno Ruiz P, et al. Machine learning models and multiparametric magnetic resonance imaging for the prediction of pathologic response to neoadjuvant chemotherapy in breast cancer. *Cancers (Basel).* 2022;14(14):3508. [Crossref]
25. Ha R, Chin C, Karcich J, et al. Prior to initiation of chemotherapy, can we predict breast tumor response? Deep learning convolutional neural networks approach using a breast MRI tumor dataset. *J Digit Imaging*. 2019;32(5):693-701. [Crossref]
26. Qu YH, Zhu HT, Cao K, Li XT, Ye M, Sun YS. Prediction of pathological complete response to neoadjuvant chemotherapy in breast cancer using a deep learning (DL) method. *Thorac Cancer*. 2020;11(3):651-658. [Crossref]
27. Yerushalmi R, Woods R, Ravdin PM, Hayes MM, Gelmon KA. Ki67 in breast cancer: prognostic and predictive potential. *Lancet Oncol*. 2010;11(2):174-183. [Crossref]
28. Penault-Llorca F, Abrial C, Raoelfils I, et al. Changes and predictive and prognostic value of the mitotic index, Ki-67, cyclin D1, and cyclo-oxygenase-2 in 710 operable breast cancer patients treated with neoadjuvant chemotherapy. *Oncologist*. 2008;13(12):1235-1245. [Crossref]
29. MacGrogan G, Mauriac L, Durand M, et al. Primary chemotherapy in breast invasive carcinoma: predictive value of the immunohistochemical detection of hormonal receptors, p53, c-erbB-2, MiB1, pS2 and GST pi. *Br J Cancer.* 1996;74(9):1458-1465. [Crossref]
30. Petit T, Wilt M, Velten M, et al. Comparative value of tumour grade, hormonal receptors, Ki-67, HER-2 and topoisomerase II alpha status as predictive markers in breast cancer patients treated with neoadjuvant anthracycline-based chemotherapy. *Eur J Cancer.* 2004;40(2):205-211. [Crossref]
31. Bottini A, Berruti A, Bersiga A, et al. Relationship between tumour shrinkage and reduction in Ki67 expression after primary chemotherapy in human breast cancer. *Br J Cancer*. 2001;85(8):1106-1112. [Crossref]
32. Estévez LG, Cuevas JM, Antón A, et al. Weekly docetaxel as neoadjuvant chemotherapy for stage II and III breast cancer: efficacy and correlation with biological markers in a phase II, multicenter study. *Clin Cancer Res*. 2003;9(2):686-692. [Crossref]
33. Zhou J, Lu J, Gao C, et al. Predicting the response to neoadjuvant chemotherapy for breast cancer: Wavelet transforming radiomics in MRI. *BMC Cancer.* 2020;20:100. [Crossref]