

## Article

# AI-CAD-Guided Mammographic Assessment of Tumor Size and T Stage: Concordance with MRI for Clinical Staging in Breast Cancer Patients Considered for NAC

Ga Eun Park <sup>1,†</sup>, Kabsoo Shin <sup>2,†</sup> , Han Song Mun <sup>1,\*</sup>  and Bong Joo Kang <sup>1</sup> 

<sup>1</sup> Department of Radiology, Seoul St. Mary's Hospital, College of Medicine, The Catholic University of Korea, Seoul 06591, Republic of Korea; hoonhoony@naver.com (G.E.P.); lionmain@catholic.ac.kr (B.J.K.)

<sup>2</sup> Division of Medical Oncology, Department of Internal Medicine, Seoul St. Mary's Hospital, College of Medicine, The Catholic University of Korea, Seoul 06591, Republic of Korea; kabsoo.shin@catholic.ac.kr

\* Correspondence: im\_hsm@catholic.ac.kr; Tel.: +82-2-2258-1455

† These authors contributed equally to this work.

## Abstract

**Objectives:** To evaluate the agreement between AI-CAD-guided mammographic and MRI measurements of tumor size and T stage in breast cancer patients being considered for neoadjuvant chemotherapy (NAC). **Methods:** This retrospective study included 144 women (mean age,  $52 \pm 11$  years) with invasive breast cancer who subsequently received NAC and underwent both AI-CAD mammography (score  $\geq 10$ ) and pre-treatment MRI. Tumor sizes from AI-CAD contours were compared with MRI using Pearson correlation, intraclass correlation coefficients (ICCs), and Bland–Altman analysis. Concordance was defined as a  $\pm 0.5$  cm difference. The contour showing the highest agreement was used to compare T stage with MRI using weighted kappa. **Results:** The mean AI-CAD abnormality score was  $86.3 \pm 22.2$ . Tumor sizes on mammography were  $3.0 \pm 1.2$  cm (inner),  $3.8 \pm 1.5$  cm (middle), and  $4.8 \pm 2.2$  cm (outer), while the MRI-measured tumor size was  $4.0 \pm 1.9$  cm. The middle contour showed the strongest correlation with MRI ( $r = 0.897$ ;  $ICC = 0.866$ ), the smallest mean difference ( $-0.19$  cm; limits of agreement,  $-1.87$  to  $1.49$ ), and the highest concordance (61.1%). Agreement was higher in mass-only lesions than in NME-involved lesions ( $ICC = 0.883$  vs.  $0.775$ ; concordance, 70.9% vs. 46.6%). T stage comparison using the middle contour showed substantial agreement with MRI ( $\kappa = 0.743$  [95% CI, 0.634–0.852]; agreement, 88.2%), with higher concordance in mass-only lesions (93.0%) than NME-involved lesions (81.0%) and more frequent understaging in the latter (17.2% vs. 2.3%). **Conclusions:** AI-CAD-guided mammographic assessment using the middle contour demonstrated good agreement with MRI for tumor size and T stage, indicating its value as a supportive tool for clinical staging in MRI-limited settings.

**Keywords:** breast neoplasms; neoadjuvant therapy; artificial intelligence; mammography; neoplasm staging



Academic Editors: Ronnie Wirestam and Emilio Quaia

Received: 23 May 2025

Revised: 23 June 2025

Accepted: 23 June 2025

Published: 24 June 2025

**Citation:** Park, G.E.; Shin, K.; Mun, H.S.; Kang, B.J. AI-CAD-Guided Mammographic Assessment of Tumor Size and T Stage: Concordance with MRI for Clinical Staging in Breast Cancer Patients Considered for NAC. *Tomography* **2025**, *11*, 72. <https://doi.org/10.3390/tomography11070072>

**Copyright:** © 2025 by the authors. Licensee MDPI, Basel, Switzerland.

This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

Breast cancer is among the most prevalent malignancies in women, accounting for nearly 12% of all cancer cases globally [1,2]. Accurate assessment of tumor size and extent is a key factor in breast cancer evaluation, guiding therapeutic decisions and offering prognostic insight into overall survival [3,4]. This assessment also plays a significant role

in determining whether patients should undergo upfront surgery or receive neoadjuvant chemotherapy (NAC) prior to surgery.

NAC is commonly indicated for patients with locally advanced breast cancer (T2 or greater, node-positive disease), as well as for those with triple-negative or HER2-positive subtypes, or when downstaging is necessary to enable breast-conserving surgery [5]. Inflammatory breast cancer and cases where the extent of residual disease may influence subsequent treatment decisions are also appropriate candidates for NAC [6]. By reducing tumor burden before surgery, NAC facilitates surgical planning and provides important prognostic information [7,8].

In this context, breast magnetic resonance imaging (MRI) is widely regarded as the most accurate imaging modality for evaluating tumor extent and treatment response, and is, therefore, a reasonable comparative modality for evaluating other imaging approaches [9–11]. MRI findings often play a decisive role in determining whether patients should be considered for NAC, especially when assessing tumor size and clinical T stage.

Mammography, a fundamental imaging modality for breast cancer diagnosis, has long served as the standard screening tool [12]. The incorporation of artificial intelligence-based computer-aided detection (AI-CAD) into mammographic interpretation has led to significant advancements. AI-CAD has been shown to improve cancer detection rates by 0.7–1.6 per 1000 screened women and increase the positive predictive value (PPV) by 0.1–1.9% when used as a supplementary reader, with minimal increases in recall rates (0.16–0.30%) [13]. Large multi-center studies have demonstrated that commercially available AI-CAD systems achieve comparable or even superior sensitivity and specificity compared to single or double readings by radiologists [14,15]. Moreover, AI-CAD enhances reading efficiency, reduces inter-reader variability, and decreases radiologists' workload. In dual reading settings, AI-CAD has shown performance sufficient to replace one reader, with additional readings needed in only 4–6% of cases [13,15,16].

Although AI-CAD systems have been extensively evaluated for cancer detection in screening settings, few studies have explored their diagnostic potential beyond detection—particularly in the context of tumor size estimation or clinical staging. To our knowledge, no prior study has quantitatively examined the agreement between AI-CAD-based lesion size on mammography and MRI measurements in breast cancer patients being considered for NAC. Given the clinical importance of tumor size in staging and treatment planning, along with the widespread availability of mammography even in resource-limited settings, such an investigation may support broader diagnostic applications of AI-CAD and offer supportive information in situations where MRI is unavailable or limited.

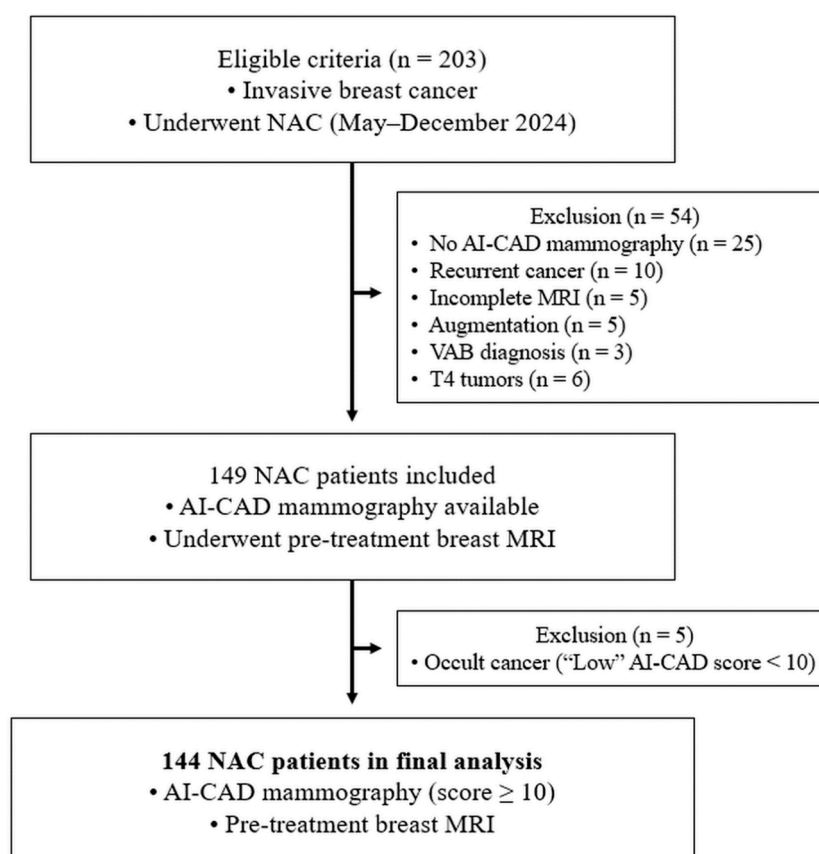
Therefore, the aim of this study was to evaluate the agreement between AI-CAD-guided mammographic and MRI measurements of tumor size and clinical T stage, in order to determine whether AI-CAD mammography can provide supportive information for clinical staging in breast cancer patients being considered for NAC.

## 2. Materials and Methods

### 2.1. Study Population and Clinical Data Collection

We identified 203 consecutive patients diagnosed with invasive breast cancer who subsequently underwent NAC between May and December 2024 at our tertiary referral center. Among them, 54 patients were excluded for the following reasons: mammography not processed with AI-CAD or AI-CAD output was unavailable ( $n = 25$ ), treatment for recurrent breast cancer ( $n = 10$ ), incomplete or suboptimal pre-treatment breast MRI ( $n = 5$ ), augmentation mammoplasty for cosmetic purposes ( $n = 5$ ), diagnosis based on vacuum-assisted biopsy ( $n = 3$ ), and T4 tumors ( $n = 6$ ), which were excluded because their classification is

based on chest wall or skin invasion rather than tumor size. After exclusions, 149 patients with both pre-treatment mammography and breast MRI were initially included. Of these, 5 patients with occult breast cancer—defined as an AI-CAD abnormality score < 10 (which was presented as “low”) and interpreted as a test-negative result—were further excluded. Consequently, a total of 144 patients (mean age,  $52 \pm 11$  years) were included in the final analysis to evaluate the agreement between AI-CAD-guided mammographic and MRI-based tumor size and T stage assessments, which were performed prior to NAC initiation and used to inform clinical staging (Figure 1).



**Figure 1.** Patient selection flowchart for an AI-based mammography study. NAC = neoadjuvant chemotherapy; AI-CAD = artificial intelligence-based computer-aided detection; VAB = vacuum-assisted biopsy.

Clinical data were retrieved from electronic medical records and included age, menopausal status (premenopausal or postmenopausal), family history of cancer (none, breast cancer, or other cancer), and BRCA1/2 mutation status (yes, no, or unknown [not tested]).

## 2.2. Pathologic and Biomarker Assessment

The histologic type (invasive ductal, invasive lobular, or others), histologic grade (I–II or III), hormone receptor (HR) status—including estrogen receptor (ER) and progesterone receptor (PR)—HER2 expression, and Ki-67 expression were determined from the initial pathology reports of core biopsies obtained before NAC. HR status was defined as positive if tumor cells showed ER and/or PR expression, with a cut-off of an Allred score  $\geq 3$  [17]. HER2 positivity was defined as an immunohistochemistry (IHC) score of 3+ or 2+ with a positive silver fluorescence in situ hybridization (SISH) result, in accordance with the

2018 ASCO/CAP guidelines [18]. Ki-67 expression was categorized as high ( $\geq 20\%$ ) or low ( $< 20\%$ ) based on the 2013 St. Gallen consensus [19].

Based on ER, PR, and HER2 status, tumors were further classified into molecular subtypes: HR+/HER2−, HR+/HER2+, HR−/HER2+, and HR−/HER2− (triple-negative breast cancer, TNBC).

### 2.3. MRI Acquisition and Interpretation

Breast MRI was performed using 3T scanners (Verio and Vida, Siemens Healthcare, Erlangen, Germany) with patients in a prone position using a dedicated breast coil. The protocol included axial T2-weighted imaging, diffusion-weighted imaging (DWI), and T1-weighted dynamic contrast-enhanced (DCE) imaging. DCE-MRI was obtained before and after intravenous injection of a gadolinium-based contrast agent (Gadovist, 0.1 mmol/kg). The Verio scanner acquired images at 10, 70, 130, 190, 250, and 310 s using a 3D volumetric interpolated breath-hold examination (VIBE) sequence, while the Vida scanner acquired images at 10, 93, 176, 259, 342, and 425 s using a 3D fast low-angle shot (FLASH) sequence. All DCE sequences were acquired with fat suppression. DICOM data from DCE-MRI were processed using syngo.via VB80 (Siemens Healthcare, Erlangen, Germany) to generate maximum intensity projection (MIP) images and subtracted images for all post-contrast phases.

An experienced breast radiologist (B.J.K., with 19 years of experience) independently reviewed all pre-treatment breast MRI scans in a blinded manner. The radiologist was informed only of the laterality (right or left) of the breast cancer but were blinded to detailed pathological findings, including histologic subtype, biomarker status, and the use of neoadjuvant chemotherapy. Background parenchymal enhancement (BPE) was first assessed and categorized as minimal, mild, moderate, or marked. Lesions were then classified based on their morphological appearance into one of the following categories: mass, non-mass enhancement (NME), or mass with NME. For each case, tumor size assessment was performed with reference to three MRI sequences: axial T2-weighted imaging, early-phase subtracted T1-weighted DCE imaging, and MIP images [20]. The greatest tumor dimension among the three sequences was measured using electronic calipers. When multiple lesions were present, the most suspicious lesion was used for size measurement. Based on the recorded size, MRI-based T staging was subsequently determined by a medical oncologist (K.S.).

### 2.4. AI-Based Mammographic Analysis

Bilateral mammograms were obtained using a dedicated digital mammography system (Selenia Dimensions; Hologic, Marlborough, MA, USA), including standard cranio-caudal (CC) and mediolateral oblique (MLO) views. All images were analyzed using a commercially available AI-CAD program (Lunit INSIGHT MMG; <https://insight.lunit.io>, version 1.1.8.2, accessed on 23 June 2025), which has been validated in multiple international studies [21–25]. The system provides breast parenchymal composition (categorized as A to D: fatty, scattered, heterogeneously dense, or extremely dense) and a corresponding breast density score (0–10) for all cases. It also provides abnormality scores (0–100) for each breast, along with grayscale contour maps and the type of abnormality (mass, calcification, or mass with calcification), if present. An abnormality score  $\geq 10$  was considered positive, and patients with scores  $< 10$  ( $n = 5$ ) were excluded as described in the inclusion criteria.

Lunit INSIGHT MMG (Lunit, Seoul, Republic of Korea) is an AI-CAD software for breast cancer detection in mammography, developed using deep convolutional neural networks (CNNs) [23]. It employs ResNet-34, a widely used CNN architecture, as the backbone network [26]. The algorithm was trained on more than 200,000 mammographic

cases collected from Korea, the United States, and the United Kingdom, and has received regulatory approval from the Korean Ministry of Food and Drug Safety.

For each case, the lesion with the highest abnormality score was selected from the more suspicious side (in bilateral cases,  $n = 4$ ) and from the projection view (CC or MLO) with the higher score. Two board-certified breast radiologists (G.E.P. and H.S.M., with 9 and 14 years of experience, respectively), who were blinded to MRI findings, evaluated lesion size by consensus based on AI-generated contours. When a lesion was detected, the AI-CAD system generated up to three concentric contours according to abnormality score thresholds: inner ( $\geq 90$ ), middle (50–89), and outer (10–49). Each contour represented a distinct level of algorithm-predicted probability of malignancy. Tumor size was measured along the longest axis between the edges of each contour line, excluding line thickness to avoid overestimation. When fewer than three contours were present, the following approach was applied: if only one contour was present, the same value was assigned to all three contours; if two contours were present, the middle value was interpolated as the average of the inner and outer contours. Pixel spacing (0.1670 mm/pixel) was validated in 20 randomly selected cases and was used to convert all sizes measured by radiologists using AI-CAD contours from pixels to centimeters. Tumor sizes derived from each contour were used to assign three separate mammography-based T stages per case (K.S.). Tumor laterality was shared with MRI reviewers to ensure consistency in lesion matching across modalities. As a result, a total of 144 representative lesions from 144 patients were included in the final analysis.

### 2.5. Statistical Analysis

Descriptive statistics were used to summarize clinicopathologic and imaging characteristics. Continuous variables were reported as mean  $\pm$  standard deviation, and categorical variables as counts and percentages.

Pearson correlation coefficients were calculated to assess linear agreement between AI-CAD-guided and MRI tumor sizes. Intraclass correlation coefficients (ICCs) were computed using a two-way random effects model (ICC[2,1]) to evaluate absolute agreement. ICC values were interpreted as follows:  $<0.50$ , poor;  $0.50$ – $0.75$ , moderate;  $0.75$ – $0.90$ , good;  $>0.90$ , excellent [27]. Bland–Altman analysis was also performed to assess mean differences and 95% limits of agreement (LoA). Concordance between tumor sizes obtained from each modality was defined as a difference within  $\pm 0.5$  cm, reflecting a clinically acceptable threshold that is widely used in breast imaging and supported by prior studies [20,28–30].

For T staging, AI-CAD-guided stages derived from different contour levels were compared with MRI-based stages. The best-performing contour was used for the final comparison. Quadratic weighted kappa statistics, appropriate for ordinal data, were used to assess agreement, with interpretation as follows:  $<0.00$ , poor;  $0.00$ – $0.20$ , slight;  $0.21$ – $0.40$ , fair;  $0.41$ – $0.60$ , moderate;  $0.61$ – $0.80$ , substantial;  $0.81$ – $1.00$ , almost perfect [31]. Rates of agreement, understaging (AI-CAD-guided T stage  $<$  MRI T stage), and overestaging (AI-CAD-guided T stage  $>$  MRI T stage) were also calculated.

All statistical analyses were performed using SPSS (version 28.0; IBM Corp., Armonk, NY, USA), with significance set at  $p < 0.05$ .

## 3. Results

### 3.1. Baseline Clinicopathologic and MRI Characteristics

A total of 144 representative lesions from 144 patients (mean age,  $51.7 \pm 10.8$  years) were included in the final analysis (Table 1). Of these, 86.1% were aged 40 years or older. A family history of breast cancer was reported in 13.9% of patients, and BRCA1/2 mutations were confirmed in 4.2%.

**Table 1.** Baseline clinicopathologic and MRI characteristics of the study population.

Category	Variable	Value	n (%)
Demographics	Age (years)	Mean $\pm$ SD	51.7 $\pm$ 10.8
		<40 years	20 (13.9%)
		$\geq$ 40 years	124 (86.1%)
	Menopausal state	Premenopause	76 (52.8%)
		Postmenopause	68 (47.2%)
	Family history of cancer	None	80 (55.6%)
		Breast cancer	20 (13.9%)
		Other cancer	44 (30.6%)
	BRCA1/2 mutation status	No	44 (30.6%)
		Yes	6 (4.2%)
		Unknown	94 (65.3%)
Pathology	Histologic type	Invasive ductal	134 (93.1%)
		Invasive lobular	5 (3.5%)
		Others	5 (3.5%)
	Histologic grade	I–II	80 (55.6%)
		III	63 (43.8%)
		Unknown	1 (0.7%)
	Estrogen receptor	No	63 (43.8%)
		Yes	81 (56.2%)
	Progesterone receptor	No	87 (60.4%)
		Yes	57 (39.6%)
	HER2 receptor	No	79 (54.9%)
		Yes	65 (45.1%)
	Molecular subtype	HR+/HER2–	35 (24.3%)
		HR+/HER2+	22 (15.3%)
		HR–/HER2+	43 (29.9%)
		HR–/HER2– (TNBC)	44 (30.6%)
	Ki-67 expression	<20%	23 (16.0%)
		$\geq$ 20%	121 (84.0%)
MRI Characteristics	MRI BPE	Minimal	58 (40.3%)
		Mild	51 (35.4%)
		Moderate	19 (13.2%)
		Marked	16 (11.1%)
	MRI lesion type	Mass only	86 (59.7%)
		NME only	14 (9.7%)
		Mass with NME	44 (30.6%)
	MRI tumor size (cm)	Mean $\pm$ SD (range)	4.0 $\pm$ 1.9 (1.7–10.8)
	MRI-based T stage	T1	9 (6.2%)
		T2	100 (69.4%)
		T3	35 (24.3%)

MRI = magnetic resonance imaging; SD = standard deviation; HER2 = human epidermal growth factor receptor 2; HR = hormone receptor; TNBC = triple-negative breast cancer; BPE = background parenchymal enhancement; NME = non-mass enhancement.

Most tumors were invasive ductal carcinoma (93.1%), while invasive lobular and other histologic types each accounted for 3.5%. High-grade tumors (grade III) were observed in 43.8% of cases. ER and PR positivity were identified in 56.2% and 39.6% of patients, respectively, and HER2 positivity was observed in 45.1% of cases based on IHC/SISH. According to HR and HER2 status, the most common molecular subtype was HR–/HER2– (TNBC, 30.6%), followed by HR–/HER2+ (29.9%), HR+/HER2– (24.3%), and HR+/HER2+ (15.3%). High Ki-67 expression ( $\geq$ 20%) was found in 84.0% of tumors.

Regarding MRI characteristics, minimal and mild BPE were the most frequently observed patterns (40.3% and 35.4%, respectively). In terms of lesion type, mass lesions only were present in 59.7% of patients, mass with NME in 30.6%, and NME only in 9.7%. The mean tumor size on MRI was 4.0  $\pm$  1.9 cm (range, 1.7–10.8 cm). Based on the MRI-



derived T stage, T2 was the most prevalent (100/144, 69.4%), followed by T3 (35/144, 24.3%) and T1 (9/144, 6.2%).

### 3.2. AI-CAD-Guided Mammographic Tumor Assessment

Among the 144 cases, the most common breast composition was heterogeneously dense (66.0%), followed by scattered (18.1%) and extremely dense (15.3%). The mean breast density score (0–10) was  $6.9 \pm 1.6$  (Table 2).

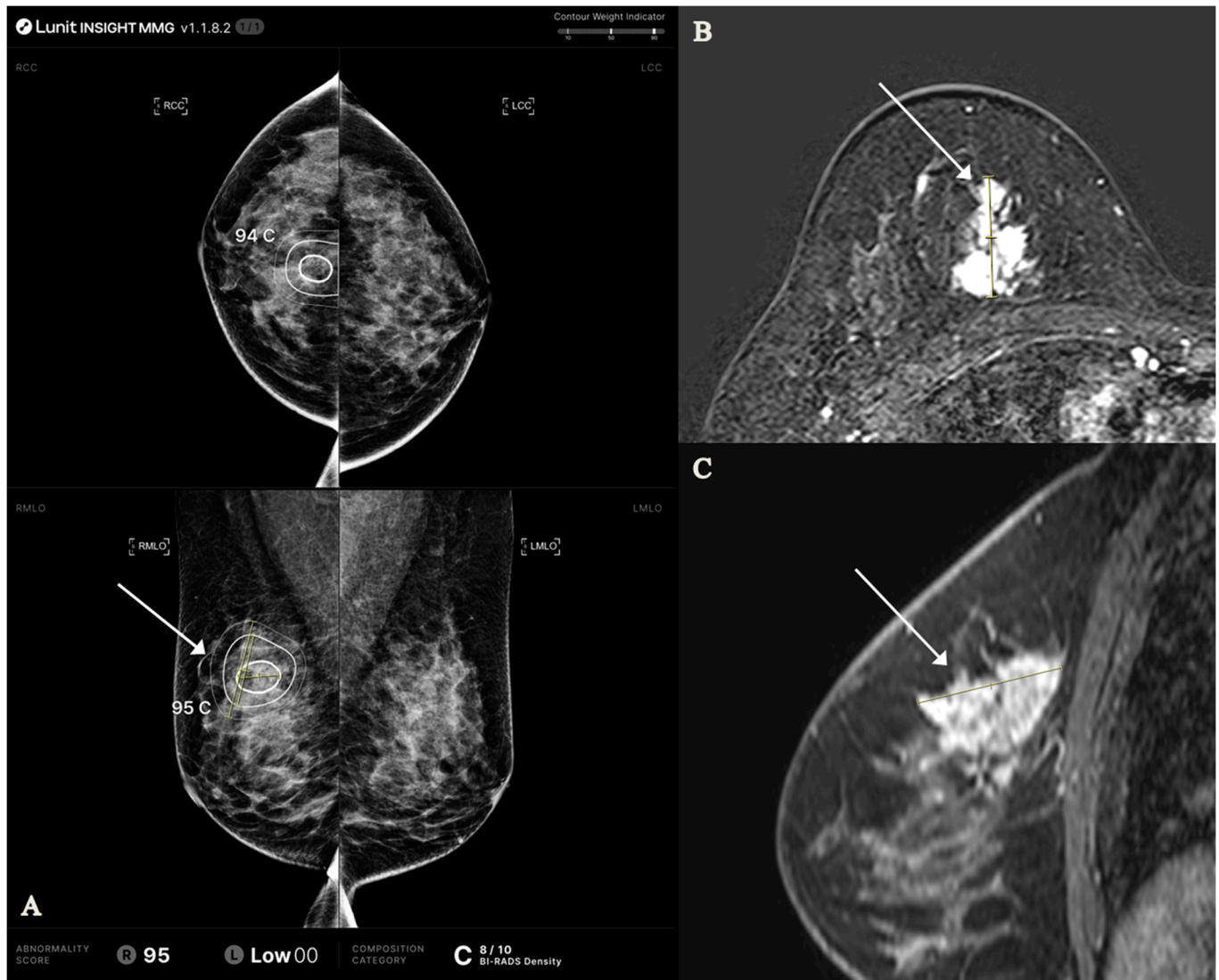
**Table 2.** AI-CAD-guided mammographic lesion assessment.

Category	Variable	Value	n (%)
Breast Density	Breast composition	Fatty	1 (0.7%)
		Scattered	26 (18.1%)
		Heterogeneously dense	95 (66.0%)
		Extremely dense	22 (15.3%)
AI Abnormality Detection	Breast density score (0–10)	Mean $\pm$ SD (range)	$6.9 \pm 1.6$ (3–10)
	Abnormality score (0–100)	Mean $\pm$ SD (range)	$86.3 \pm 22.2$ (10–99)
	Abnormality type	Mass	58 (40.3%)
		Calcification	23 (16.0%)
		Mass with calcification	63 (43.8%)
	Detection view	CC	58 (40.3%)
		MLO	86 (59.7%)
	Number of abnormality contours	1	15 (10.4%)
		2	31 (21.5%)
		3	98 (68.1%)
Contour-based Tumor Size (cm)	Inner size	Mean $\pm$ SD (range)	$3.0 \pm 1.2$ (1.2–7.9)
	Middle size	Mean $\pm$ SD (range)	$3.8 \pm 1.5$ (1.2–9.2)
	Outer size	Mean $\pm$ SD (range)	$4.8 \pm 2.2$ (1.2–13.5)
AI-CAD-guided T stage	Inner T stage	T1	34 (23.6%)
		T2	100 (69.4%)
		T3	10 (6.9%)
	Middle T stage	T1	5 (3.5%)
		T2	115 (79.9%)
		T3	24 (16.7%)
	Outer T stage	T1	2 (1.4%)
		T2	95 (66.0%)
		T3	47 (32.6%)

AI-CAD = artificial intelligence-based computer-aided detection; SD = standard deviation; CC = craniocaudal; MLO = mediolateral oblique.

For the AI-CAD-detected lesion, the mean abnormality score (0–100) was  $86.3 \pm 22.2$  (Figure 2). The most frequently detected abnormality type was mass with calcification (43.8%), followed by mass (40.3%) and calcification alone (16.0%). The MLO view was more frequently selected for analysis (59.7%) compared to the CC view (40.3%). In terms of the number of abnormality contours, three contours were identified in the majority of cases (68.1%).

The mean tumor sizes derived from the inner, middle, and outer contours were  $3.0 \pm 1.2$  cm,  $3.8 \pm 1.5$  cm, and  $4.8 \pm 2.2$  cm, respectively. Based on AI-CAD-guided mammographic T staging, T2 was the most frequent stage across all three contours (inner: 69.4%; middle: 79.9%; outer: 66.0%). The proportion of T3 classification increased with contour size, whereas T1 classification was more common in the inner contour than in the middle or outer contours.



**Figure 2.** AI-CAD mammography and MRI findings in a 48-year-old woman with invasive ductal carcinoma, grade II (HR+/HER2+, Ki-67 index 34%). (A) AI-CAD mammography shows heterogeneously dense breasts (category C, density score 8/10) and a high abnormality score (95) on the right mediolateral oblique view. The lesion was classified as calcification. Tumor size along the inner, middle, and outer contours was 2.0, 3.5, and 4.9 cm, respectively (T1, T2, T2). (B,C) Contrast-enhanced MRI reveals a heterogeneously enhancing mass with non-mass enhancement in the corresponding area. The lesion measured 3.3 cm on axial and 3.7 cm on sagittal views; the maximal diameter (3.7 cm) was used for T2 staging. Arrows indicate the lesion; yellow lines represent caliper-based measurements. AI-CAD = artificial intelligence-based computer-aided detection; HR = hormone receptor; HER2 = human epidermal growth factor receptor 2.

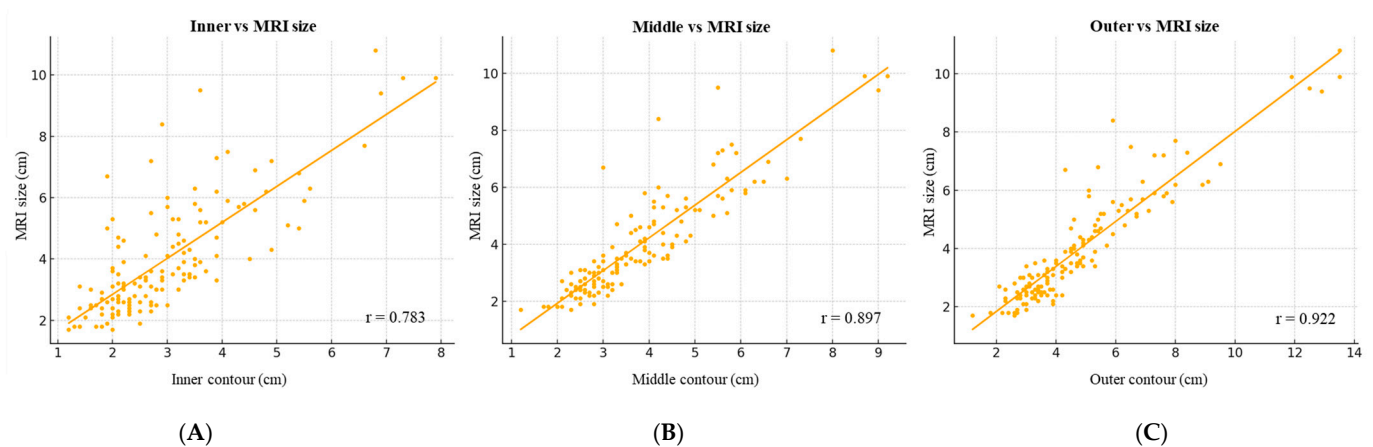
### 3.3. Agreement Between MRI and AI-CAD-Guided Mammographic Tumor Size Measurements

Pearson correlation analysis revealed strong linear associations between MRI-measured tumor size and AI-CAD-guided measurements from all contour levels (inner:  $r = 0.783$ ,  $p < 0.001$ ; middle:  $r = 0.897$ ,  $p < 0.001$ ; outer:  $r = 0.922$ ,  $p < 0.001$ ) (Figure 3).

ICC analysis demonstrated good agreement for the middle and outer contours (ICC = 0.866 [95% CI, 0.815–0.902] and 0.847 [95% CI, 0.446–0.936], respectively), whereas the inner contour showed only moderate agreement (ICC = 0.602 [95% CI, 0.153–0.794]) (Table 3). When analyzed by MRI lesion type, ICCs were consistently higher in mass-only lesions compared to NME-involved lesions, which included both NME-only lesions and masses with associated non-mass enhancement. The highest agreement was observed in



mass-only lesions using the middle contour (ICC = 0.883 [95% CI, 0.827–0.922]), whereas the lowest was found in NME-involved lesions using the inner contour (ICC = 0.443 [95% CI, 0.000–0.729]).



**Figure 3.** Scatter plots showing Pearson correlations between MRI and AI-CAD-guided tumor measurements based on (A) inner, (B) middle, and (C) outer contours ( $r = 0.783$ ,  $0.897$ , and  $0.922$ ; all  $p < 0.001$ ). AI-CAD = artificial intelligence-based computer-aided detection.

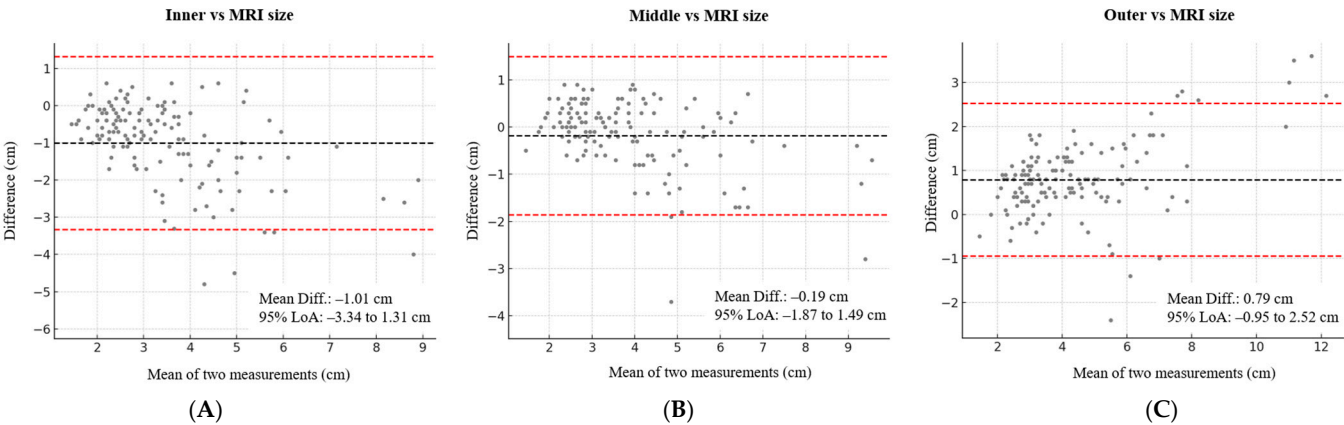
**Table 3.** ICC, Bland–Altman, and concordance analysis of tumor sizes by MRI lesion type and AI-CAD-guided contours.

MRI Lesion Type	AI-CAD-Guided	ICC (95% CI)	Mean Diff. (AI–MRI) (cm)	SD of Diff. (cm)	Lower LoA (cm)	Upper LoA (cm)	Within $\pm 0.5$ cm, $n$ (%)
Total ( $n = 144$ )	Inner	0.602 (0.153, 0.794)	−1.01	1.18	−3.34	1.31	58 (40.3%)
	Middle	0.866 (0.815, 0.902)	−0.19	0.85	−1.87	1.49	88 (61.1%)
	Outer	0.847 (0.446, 0.936)	0.79	0.88	−0.95	2.52	46 (31.9%)
Mass only ( $n = 86$ )	Inner	0.714 (0.246, 0.870)	−0.54	0.62	−1.77	0.68	47 (54.7%)
	Middle	0.883 (0.827, 0.922)	0.07	0.50	−0.91	1.06	61 (70.9%)
	Outer	0.775 (0.133, 0.916)	0.60	0.54	−0.45	1.65	35 (40.7%)
NME-involved * ( $n = 58$ )	Inner	0.443 (0.000, 0.729)	−1.71	1.45	−4.56	1.14	11 (19.0%)
	Middle	0.780 (0.580, 0.880)	−0.57	1.10	−2.73	1.58	27 (46.6%)
	Outer	0.783 (0.301, 0.911)	1.07	1.18	−1.25	3.38	11 (19.0%)

\* NME-involved includes both NME-only lesions and masses with associated NME. ICC = intraclass correlation coefficient; MRI = magnetic resonance imaging; AI-CAD = artificial intelligence-based computer-aided detection; CI = confidence interval; Diff. = difference; SD = standard deviation; LoA = limits of agreement; NME = non-mass enhancement.

Bland–Altman analysis supported these findings (Figure 4). The mean difference ranged from −1.01 to 0.79 cm across all contours. The middle contour showed the smallest mean difference relative to MRI across lesion types: −0.19 cm (LoA, −1.87 to 1.49) in total cases, 0.07 cm (LoA, −0.92 to 1.07) in mass-only lesions, and −0.57 cm (LoA, −2.73 to 1.58) in NME-involved lesions. Across all contours, NME-involved lesions showed consistently

greater mean differences than mass-only lesions, with mean differences ranging from −1.71 to 1.07 cm.



**Figure 4.** Bland–Altman plots comparing MRI and AI-CAD-guided tumor size based on (A) inner, (B) middle, and (C) outer contours. The middle contour showed the smallest mean difference (−0.19 cm; 95% limits of agreement −1.87 to 1.49 cm). AI-CAD = artificial intelligence-based computer-aided detection.

Concordance analysis ranged from 31.9% to 61.1%, with the highest rate observed for the middle contour (61.1%, 88/144). In mass-only lesions, the concordance rates were 70.9% (61/86) for the middle contour, 54.7% (47/86) for the inner contour, and 40.7% (35/86) for the outer contour. In NME-involved lesions, the middle contour also showed the highest concordance rate at 46.6% (27/58), while the inner and outer contours both showed identical rates of 19.0% (11/58 each).

Overall, the middle contour demonstrated the most consistent agreement with MRI tumor size measurements. Based on these results, the middle contour was used for AI-CAD-guided mammographic T stage comparisons in the following Section 3.4.

3.4. Concordance Between MRI and AI-CAD-Guided Mammographic T Staging

As shown in Table 4, the confusion matrix for all cases demonstrated a quadratic weighted  $\kappa$  of 0.743 [95% CI, 0.634–0.852], indicating substantial agreement between MRI and AI-CAD-guided mammographic T staging (middle contour), with an overall agreement rate of 88.2% (127 of 144 cases) (Figure 5). Subgroup analysis showed higher concordance in mass-only lesions ( $\kappa = 0.725$  [95% CI, 0.579–0.871]; agreement, 93.0%) compared to NME-involved lesions ( $\kappa = 0.624$  [95% CI, 0.423–0.825]; agreement, 81.0%), with understaging occurring more frequently in the NME-involved group (17.2% vs. 2.3%). Confusion matrices for each subgroup are presented in Supplementary Table S1.

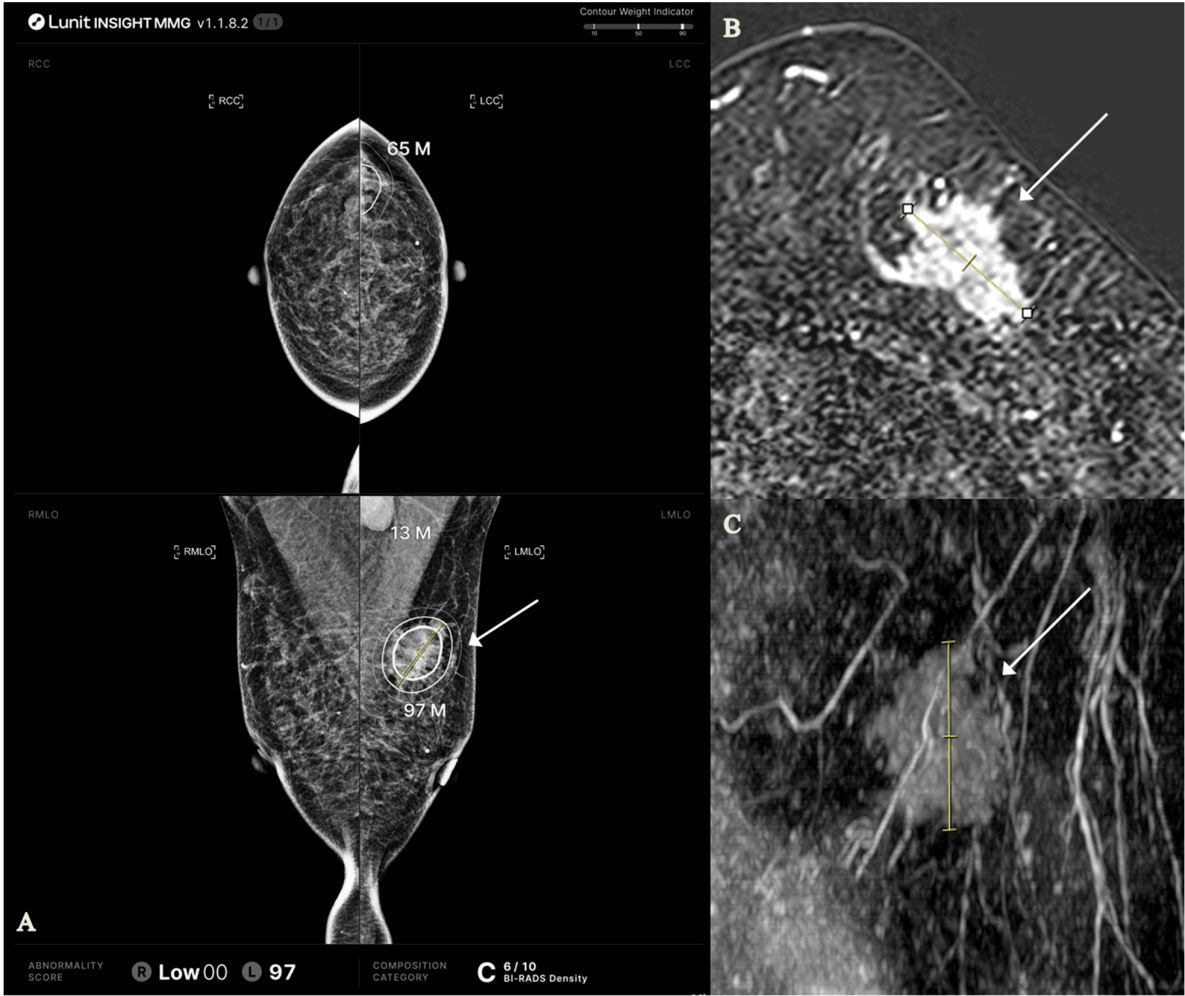
**Table 4.** Confusion matrix for all cases and concordance metrics between MRI and AI-CAD-guided T staging: overall and subgroup analysis.

MRI T stage	AI-Based T Stage (Middle Contour)				Metrics	Total (n = 144)	Mass only (n = 86)	NME-involved * (n = 58)
	T1	T2	T3	Total				
T1	5	4	0	9	Quadratic weighted $\kappa$ (95% CI)	0.743 (0.634, 0.852)	0.725 (0.579, 0.871)	0.624 (0.423, 0.825)
T2	0	99	1	100	Agreement rate (n, %)	127 (88.2%)	80 (93.0%)	47 (81.0%)

Table 4. Cont.

AI-Based T Stage (Middle Contour)								
T3	0	12	23	35	Understaging rate (n, %)	12 (8.3%)	2 (2.3%)	10 (17.2%)
Total	5	115	24	144	Overstaging rate (n, %)	5 (3.5%)	4 (4.7%)	1 (1.7%)

\* NME-involved: includes both NME-only and mass with NME lesions. MRI = magnetic resonance imaging; AI-CAD = artificial intelligence-based computer-aided detection; NME = non-mass enhancement; CI = confidence interval.



**Figure 5.** AI-CAD mammography and MRI findings in a 53-year-old woman with invasive ductal carcinoma, grade III (HR+/HER2−, Ki-67 index 85%). **(A)** AI-CAD mammography shows heterogeneously dense breasts (category C, density score 6/10) and a high abnormality score (97) on the left mediolateral oblique view. The lesion was classified as a mass. Tumor size along inner, middle, and outer contours was 2.6, 3.8, and 4.6 cm, respectively (T2 stage for all). A partially visualized enlarged axillary lymph node was also noted. **(B,C)** Contrast-enhanced MRI reveals an irregularly enhancing mass in the corresponding area. The lesion measured 2.8 cm on the axial view and 3.8 cm on the maximum intensity projection image; the maximal diameter (3.8 cm) was used for T2 staging. Arrows indicate the lesion; yellow lines represent caliper-based measurements. The outer contour in A was omitted due to visual overlap with other lines. AI-CAD = artificial intelligence-based computer-aided detection; HR = hormone receptor; HER2 = human epidermal growth factor receptor 2; MIP = maximum intensity projection.

To further explore the concordance across clinicopathologic subgroups, a molecular subtype-stratified analysis was performed (Table 5). The highest concordance was observed in TNBC cases, with a  $\kappa$  of 0.902 [95% CI, 0.749–1.000] and an agreement rate of 95.4%, indicating almost perfect agreement. Other subtypes showed  $\kappa$  values ranging from 0.629 [95% CI, 0.316–0.854] to 0.704 [95% CI, 0.448–0.892], with agreement rates between 82.9% and 86.4%.

**Table 5.** Concordance metrics by molecular subtype ( $n = 144$ ).

Molecular Subtype	$n$	Quadratic Weighted $\kappa$ (95% CI)	Agreement ( $n$ , %)	Understaging ( $n$ , %)	Overstaging ( $n$ , %)
HR+/HER2−	35	0.629 (0.316, 0.854)	29 (82.9%)	5 (14.3%)	1 (2.8%)
HR+/HER2+	22	0.660 (0.282, 0.942)	19 (86.4%)	2 (9.1%)	1 (4.5%)
HR−/HER2+	43	0.704 (0.448, 0.892)	37 (86.0%)	4 (9.3%)	2 (4.7%)
HR−/HER2− (TNBC)	44	0.902 (0.749, 1.000)	42 (95.4%)	1 (2.3%)	1 (2.3%)

HR = hormone receptor; HER2 = human epidermal growth factor receptor 2; TNBC = triple-negative breast cancer; CI = confidence interval.

#### 4. Discussion

In this study, we evaluated the concordance between AI-CAD-guided mammographic measurements and MRI-based assessments of tumor size and clinical T stage in 144 patients with breast cancer. Among the three AI-CAD-generated contours, the middle contour consistently demonstrated the highest agreement with MRI across all evaluation metrics. It showed good absolute agreement (ICC = 0.866), the smallest mean difference (−0.19 cm), and the highest concordance rate (61.1%). Based on these findings, the middle contour was selected as the representative contour for T stage comparison with MRI, showing substantial agreement ( $\kappa = 0.743$ ; agreement, 88.2%).

The diagnostic performance of radiologists in mammographic interpretation can vary considerably [32–34]. To support consistent image evaluation, CAD systems were introduced, with the first approved by the U.S. Food and Drug Administration (FDA) in 1998 [35]. However, the effectiveness of conventional CAD has remained controversial due to low specificity, frequent false-positive markings, and high recall rates [36–39]. More recently, deep learning and CNNs have enabled substantial improvements in AI-based mammographic interpretation [40,41].

Lee et al. demonstrated that abnormality scores generated by the AI-CAD algorithm correlated well with PPVs, meeting BI-RADS recommendations [21]. Additionally, prior studies have shown that when radiologists could not identify corresponding imaging findings for AI-CAD detections, the likelihood of malignancy was extremely low [21,42]. Nonetheless, some argue that the abnormality score—ranging from 0 to 100—reflects only the degree of suspicion rather than a true probability of cancer [43].

While AI-CAD systems now offer lesion contours to assist with size estimation, their validity for tumor size measurement and T staging remains underexplored. To our knowledge, no prior studies have directly compared AI-CAD-generated measurements with established imaging standards. In this study, we assessed the agreement between AI-CAD-guided mammographic measurements and MRI tumor size determined by experienced radiologists.

Our Pearson correlation analysis demonstrated a linear relationship between tumor sizes measured by AI and those measured by MRI, although this did not imply perfect

agreement in absolute values. While the middle and outer contours showed similar ICCs overall, the middle contour demonstrated a particularly high ICC of 0.883 in mass-only lesions. It also showed the smallest mean difference from MRI measurements across lesion types:  $-0.19$  cm (LoA,  $-1.87$  to  $1.49$ ) in total cases and  $0.07$  cm (LoA,  $-0.92$  to  $1.07$ ) in mass-only lesions. These findings align with a previous study evaluating dual-layer CT (DLCT) with virtual monochromatic imaging (VMI), which reported an ICC of 0.840 and a mean difference of  $-0.05$  cm compared to MRI (LoA,  $-1.29$  to  $1.19$ ) [29].

A recent study using a Res-UNet model demonstrated high concordance between AI-generated and radiologist-derived MRI segmentations in mass lesions, with a T staging accuracy of 93% [44]. In contrast, our study focused on AI-CAD-guided mammographic measurements derived from contour levels, and showed substantial agreement with MRI-based T staging ( $\kappa = 0.743$ ; agreement, 88.2%), particularly in mass-only lesions ( $\kappa = 0.725$ ; agreement, 93.0%). Although our AI-CAD system was not designed to assign T stage directly, the radiologist-assigned staging based on AI-guided size showed promising concordance, suggesting potential utility in pre-treatment assessment.

Our subgroup analysis by molecular subtype revealed that HR-/HER2- (TNBC) cases showed a 95.4% agreement rate, likely due to their tendency to present themselves as mass lesions on MRI [45]. In contrast, NME-involved cases showed an 81.0% agreement rate and a 17.2% understaging rate. Given that HER2-positive cancers frequently present with NME, our findings should be interpreted with caution in this subgroup [46].

The BREAST-MRI randomized trial showed that preoperative MRI did not significantly improve local relapse-free or overall survival, nor did it reduce the reoperation rate [47]. In this context, our findings—demonstrating good concordance between AI-CAD-guided mammographic T staging and MRI—highlight the potential clinical utility of AI-assisted evaluation, particularly in settings where MRI is unavailable or impractical. Given that mammography and ultrasound are more widely accessible, AI-enhanced interpretation of conventional imaging may support appropriate treatment planning in resource-limited environments. In particular, AI-CAD-guided mammographic assessment may be clinically valuable in settings where MRI is not readily available due to cost, examination time, or limited imaging infrastructure, such as in community hospitals or rural clinics. In these contexts, AI-CAD can assist in identifying patients at higher risk of malignancy, thereby helping to triage cases and prioritize the use of advanced imaging modalities such as MRI for those with greater clinical need. Additionally, in situations where MRI is contraindicated—for example, patients with renal impairment, severe claustrophobia, or non-MRI-compatible implants—AI-CAD may provide tumor size estimates and diagnostic support.

Beyond detection and tumor sizing, AI is also being explored for risk stratification in breast cancer by analyzing mammographic features like breast density, with improved prediction over traditional models [16,48]. This expanding role of AI underscores its potential not only in detection but also in future clinical decision-making.

Nonetheless, our study has several limitations. First, this study was conducted at a single institution without the inclusion of a large, multi-center cohort. Although the internal consistency of our findings supports the validity of AI-CAD-guided lesion size assessment within this setting, the generalizability of the results to broader clinical environments may be limited due to potential variations in imaging protocols and patient characteristics across institutions. Second, our study included only patients who received NAC, and therefore those with ductal carcinoma in situ (DCIS) were not represented. As NAC is not indicated for DCIS, such patients were inherently excluded from the study cohort, which reflects current clinical practice. However, as AI-CAD performance may differ in preinvasive lesions or in patients not receiving NAC, caution is warranted in extrapolating our findings to such populations. Further prospective, multi-center studies



involving more diverse patient cohorts are warranted to validate the broader applicability of AI-CAD in clinical staging. Third, we used only one commercially available AI system. Therefore, the applicability of our findings to other platforms with different algorithms remains uncertain. Fourth, a fixed AI-CAD abnormality score threshold ( $\geq 10$ ) was applied as part of our inclusion criteria, based on the classification rule set by the Lunit INSIGHT MMG system. Scores below 10 are interpreted as negative, and no lesion contours are generated. The abnormality score represents the probability of malignancy generated by an algorithm-specific AI model and is influenced by the distribution and characteristics of its training data. Although the rationale for this cutoff is not publicly disclosed, it has been consistently used in prior studies employing the same software [21,22,42]. Finally, since our study included only patients who subsequently received NAC, pathological tumor size—altered by treatment-induced changes—could not be used as the reference standard. MRI, performed before NAC, is widely recognized as the most accurate imaging modality for evaluating tumor extent and, in patients receiving NAC, for assessing treatment response [9–11]. Accordingly, we considered MRI a reasonable and clinically relevant reference for evaluating the concordance of AI-CAD-guided mammographic tumor size estimation and clinical T staging. However, the absence of pathological confirmation limits the ability to assess how accurately AI-CAD reflects tumor extent, and this should be considered when interpreting the findings.

## 5. Conclusions

AI-CAD-guided mammographic tumor assessment, particularly when using the middle contour, demonstrated good agreement with MRI in estimating tumor size and clinical T stage. Given the limited accessibility of MRI in many clinical settings, this study was motivated by the need to evaluate the potential utility of AI-CAD-enhanced mammography as a supportive tool in clinical staging. These findings suggest that AI-CAD-enhanced mammography provides supportive information for clinical staging in breast cancer patients being considered for NAC, especially in settings where MRI is unavailable.

**Supplementary Materials:** The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/tomography11070072/s1>, Table S1: Confusion matrices for mass-only ( $n = 86$ ) and NME-involved ( $n = 58$ ) subgroups\*.

**Author Contributions:** Conceptualization, H.S.M.; methodology, H.S.M.; validation, B.J.K.; formal analysis, K.S.; investigation, H.S.M., G.E.P. and B.J.K.; resources, K.S.; data curation, K.S.; writing—original draft preparation, H.S.M. and G.E.P.; writing—review and editing, H.S.M. and K.S.; visualization, H.S.M.; supervision, B.J.K.; project administration, H.S.M. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research received no external funding.

**Institutional Review Board Statement:** The study was conducted in accordance with the Declaration of Helsinki and was approved by the Institutional Review Board of Seoul St. Mary's Hospital (approval no. KC25RASI0339; date of approval: 16 May 2025).

**Informed Consent Statement:** Patient consent was waived due to the retrospective nature of the study.

**Data Availability Statement:** The datasets generated or analyzed during the current study are not publicly available but may be made available by the corresponding author upon reasonable request.

**Conflicts of Interest:** The authors declare no conflicts of interest.

## Abbreviations

The following abbreviations are used in this manuscript:

NAC	Neoadjuvant chemotherapy
MRI	Magnetic resonance imaging
AI-CAD	Artificial intelligence-based computer-aided detection

## References

1. Sung, H.; Ferlay, J.; Siegel, R.L.; Laversanne, M.; Soerjomataram, I.; Jemal, A.; Bray, F. Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA Cancer J. Clin.* **2021**, *71*, 209–249. [\[CrossRef\]](#)
2. World Health Organization. *Breast Cancer: Fact Sheet*; World Health Organization: Geneva, Switzerland, 2022; Available online: <https://www.who.int/news-room/fact-sheets/detail/breast-cancer> (accessed on 18 May 2025).
3. Morrow, M.; Strom, E.A.; Bassett, L.W.; Dershaw, D.D.; Fowble, B.; Giuliano, A.; Harris, J.R.; O'Malley, F.; Schnitt, S.J.; Singletary, S.E.; et al. Standard for Breast Conservation Therapy in the Management of Invasive Breast Carcinoma. *CA Cancer J. Clin.* **2002**, *52*, 277–300. [\[CrossRef\]](#)
4. Veronesi, U.; Galimberti, V.; Zurrida, S.; Pigatto, F.; Veronesi, P.; Robertson, C.; Paganelli, G.; Sciascia, V.; Viale, G. Sentinel lymph node biopsy as an indicator for axillary dissection in early breast cancer. *Eur. J. Cancer* **2001**, *37*, 454–458. [\[CrossRef\]](#)
5. Gradishar, W.J.; Moran, M.S.; Abraham, J.; Abramson, V.; Aft, R.; Agnese, D.; Allison, K.H.; Anderson, B.; Bailey, J.; Burstein, H.J.; et al. Breast Cancer, Version 3.2024, NCCN Clinical Practice Guidelines in Oncology. *J. Natl. Compr. Cancer Netw.* **2024**, *22*, 331–357. [\[CrossRef\]](#) [\[PubMed\]](#)
6. Korde, L.A.; Somerfield, M.R.; Carey, L.A.; Crews, J.R.; Denduluri, N.; Hwang, E.S.; Khan, S.A.; Loibl, S.; Morris, E.A.; Perez, A.; et al. Neoadjuvant Chemotherapy, Endocrine Therapy, and Targeted Therapy for Breast Cancer: ASCO Guideline. *J. Clin. Oncol.* **2021**, *39*, 1485–1505. [\[CrossRef\]](#) [\[PubMed\]](#)
7. Cortazar, P.; Zhang, L.; Untch, M.; Mehta, K.; Costantino, J.P.; Wolmark, N.; Bonnefoi, H.; Cameron, D.; Gianni, L.; Valagussa, P.; et al. Pathological complete response and long-term clinical benefit in breast cancer: The CTNeoBC pooled analysis. *Lancet* **2014**, *384*, 164–172. [\[CrossRef\]](#)
8. Spring, L.M.; Fell, G.; Arfe, A.; Sharma, C.; Greenup, R.; Reynolds, K.L.; Smith, B.L.; Alexander, B.; Moy, B.; Isakoff, S.J.; et al. Pathologic Complete Response after Neoadjuvant Chemotherapy and Impact on Breast Cancer Recurrence and Survival: A Comprehensive Meta-analysis. *Clin. Cancer Res.* **2020**, *26*, 2838–2848. [\[CrossRef\]](#)
9. Panico, C.; Ferrara, F.; Woitek, R.; D'Angelo, A.; Di Paola, V.; Bufi, E.; Conti, M.; Palma, S.; Cicero, S.L.; Cimino, G.; et al. Staging Breast Cancer with MRI, the T. A Key Role in the Neoadjuvant Setting. *Cancers* **2022**, *14*, 5786. [\[CrossRef\]](#)
10. Marinovich, M.L.; Houssami, N.; Macaskill, P.; Sardanelli, F.; Irwig, L.; Mamounas, E.P.; von Minckwitz, G.; Brennan, M.E.; Ciatto, S. Meta-analysis of magnetic resonance imaging in detecting residual breast cancer after neoadjuvant therapy. *J. Natl. Cancer Inst.* **2013**, *105*, 321–333. [\[CrossRef\]](#)
11. Lobbes, M.B.; Prevos, R.; Smidt, M.; Tjan-Heijnen, V.C.; van Goethem, M.; Schipper, R.; Beets-Tan, R.G.; Wildberger, J.E. The role of magnetic resonance imaging in assessing residual disease and pathologic complete response in breast cancer patients receiving neoadjuvant chemotherapy: A systematic review. *Insights Imaging* **2013**, *4*, 163–175. [\[CrossRef\]](#)
12. Tabár, L.; Vitak, B.; Chen, T.H.; Yen, A.M.; Cohen, A.; Tot, T.; Chiu, S.Y.; Chen, S.L.; Fann, J.C.; Rosell, J.; et al. Swedish two-county trial: Impact of mammographic screening on breast cancer mortality during 3 decades. *Radiology* **2011**, *260*, 658–663. [\[CrossRef\]](#) [\[PubMed\]](#)
13. Ng, A.Y.; Oberije, C.J.G.; Ambrózay, É.; Szabó, E.; Serfőző, O.; Karpati, E.; Fox, G.; Glocker, B.; Morris, E.A.; Forrai, G.; et al. Prospective implementation of AI-assisted screen reading to improve early detection of breast cancer. *Nat. Med.* **2023**, *29*, 3044–3049. [\[CrossRef\]](#)
14. Yoon, J.H.; Strand, F.; Baltzer, P.A.T.; Conant, E.F.; Gilbert, F.J.; Lehman, C.D.; Morris, E.A.; Mullen, L.A.; Nishikawa, R.M.; Sharma, N.; et al. Standalone AI for Breast Cancer Detection at Screening Digital Mammography and Digital Breast Tomosynthesis: A Systematic Review and Meta-Analysis. *Radiology* **2023**, *307*, e222639. [\[CrossRef\]](#)
15. Elhakim, M.T.; Stougard, S.W.; Graumann, O.; Nielsen, M.; Lång, K.; Gerke, O.; Larsen, L.B.; Rasmussen, B.S.B. Breast cancer detection accuracy of AI in an entire screening population: A retrospective, multicentre study. *Cancer Imaging* **2023**, *23*, 127. [\[CrossRef\]](#)
16. Lamb, L.R.; Lehman, C.D.; Gastounioti, A.; Conant, E.F.; Bahl, M. Artificial Intelligence (AI) for Screening Mammography, From the AJR Special Series on AI Applications. *Am. J. Roentgenol.* **2022**, *219*, 369–380. [\[CrossRef\]](#)
17. Harvey, J.M.; Clark, G.M.; Osborne, C.K.; Allred, D.C. Estrogen receptor status by immunohistochemistry is superior to the ligand-binding assay for predicting response to adjuvant endocrine therapy in breast cancer. *J. Clin. Oncol.* **1999**, *17*, 1474–1481. [\[CrossRef\]](#) [\[PubMed\]](#)

18. Wolff, A.C.; Hammond, M.E.H.; Allison, K.H.; Harvey, B.E.; Mangu, P.B.; Bartlett, J.M.S.; Bilous, M.; Ellis, I.O.; Fitzgibbons, P.; Hanna, W.; et al. Human Epidermal Growth Factor Receptor 2 Testing in Breast Cancer: American Society of Clinical Oncology/College of American Pathologists Clinical Practice Guideline Focused Update. *J. Clin. Oncol.* **2018**, *36*, 2105–2122. [\[CrossRef\]](#)
19. Goldhirsch, A.; Winer, E.P.; Coates, A.S.; Gelber, R.D.; Piccart-Gebhart, M.; Thürlimann, B.; Senn, H.-J. Personalizing the treatment of women with early breast cancer: Highlights of the St Gallen International Expert Consensus on the Primary Therapy of Early Breast Cancer 2013. *Ann. Oncol.* **2013**, *24*, 2206–2223. [\[CrossRef\]](#)
20. Choi, W.J.; Cha, J.H.; Kim, H.H.; Shin, H.J.; Chae, E.Y. The Accuracy of Breast MR Imaging for Measuring the Size of a Breast Cancer: Analysis of the Histopathologic Factors. *Clin. Breast Cancer* **2016**, *16*, e145–e152. [\[CrossRef\]](#)
21. Lee, S.E.; Hong, H.; Kim, E.-K. Positive Predictive Values of Abnormality Scores From a Commercial Artificial Intelligence-Based Computer-Aided Diagnosis for Mammography. *Korean J. Radiol.* **2024**, *25*, 343–350. [\[CrossRef\]](#)
22. Choi, W.J.; An, J.K.; Woo, J.J.; Kwak, H.Y. Comparison of Diagnostic Performance in Mammography Assessment: Radiologist with Reference to Clinical Information Versus Standalone Artificial Intelligence Detection. *Diagnostics* **2022**, *13*, 117. [\[CrossRef\]](#)
23. Kim, H.-E.; Kim, H.H.; Han, B.-K.; Kim, K.H.; Han, K.; Nam, H.; Lee, E.H.; Kim, E.-K. Changes in cancer detection and false-positive recall in mammography using artificial intelligence: A retrospective, multireader study. *Lancet Digit. Health* **2020**, *2*, e138–e148. [\[CrossRef\]](#) [\[PubMed\]](#)
24. Kim, E.-K.; Kim, H.-E.; Han, K.; Kang, B.J.; Sohn, Y.-M.; Woo, O.H.; Lee, C.W. Applying Data-driven Imaging Biomarker in Mammography for Breast Cancer Screening: Preliminary Study. *Sci. Rep.* **2018**, *8*, 2762. [\[CrossRef\]](#)
25. Salim, M.; Wåhlin, E.; Dembrower, K.; Azavedo, E.; Foukakis, T.; Liu, Y.; Smith, K.; Eklund, M.; Strand, F. External Evaluation of 3 Commercial Artificial Intelligence Algorithms for Independent Assessment of Screening Mammograms. *JAMA Oncol.* **2020**, *6*, 1581–1588. [\[CrossRef\]](#) [\[PubMed\]](#)
26. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep Residual Learning for Image Recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016; IEEE: Piscataway, NJ, USA, 2016; pp. 770–778.
27. Koo, T.K.; Li, M.Y. A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research. *J. Chiropr. Med.* **2016**, *15*, 155–163. [\[CrossRef\]](#) [\[PubMed\]](#)
28. Park, J.Y. Evaluation of Breast Cancer Size Measurement by Computer-Aided Diagnosis (CAD) and a Radiologist on Breast MRI. *J. Clin. Med.* **2022**, *11*, 1172. [\[CrossRef\]](#)
29. Yeo, B.; Shin, K.M.; Park, B.; Kim, H.J.; Kim, W.H. Clinical Feasibility of Dual-Layer CT With Virtual Monochromatic Image for Preoperative Staging in Patients With Breast Cancer: A Comparison With Breast MRI. *Korean J. Radiol.* **2024**, *25*, 798–806. [\[CrossRef\]](#)
30. Azcona Sáenz, J.; Molero Calafell, J.; Román Expósito, M.; Vall Foraster, E.; Comerma Blesa, L.; Alcántara Souza, R.; Vernet Tomás, M.D.M. Preoperative estimation of the pathological breast tumor size in architectural distortions: A comparison of DM, DBT, US, CEM, and MRI. *Eur. Radiol.* **2025**. [\[CrossRef\]](#)
31. Landis, J.R.; Koch, G.G. The measurement of observer agreement for categorical data. *Biometrics* **1977**, *33*, 159–174. [\[CrossRef\]](#)
32. Nelson, H.D.; Fu, R.; Cantor, A.; Pappas, M.; Daeges, M.; Humphrey, L. Effectiveness of Breast Cancer Screening: Systematic Review and Meta-analysis to Update the 2009 U.S. Preventive Services Task Force Recommendation. *Ann. Intern. Med.* **2016**, *164*, 244–255. [\[CrossRef\]](#)
33. Sprague, B.L.; Arao, R.F.; Miglioretti, D.L.; Henderson, L.M.; Buist, D.S.; Onega, T.; Rauscher, G.H.; Lee, J.M.; Tosteson, A.N.A.; Kerlikowske, K.; et al. National Performance Benchmarks for Modern Diagnostic Digital Mammography: Update from the Breast Cancer Surveillance Consortium. *Radiology* **2017**, *283*, 59–69. [\[CrossRef\]](#)
34. Kim, S.H.; Lee, E.H.; Jun, J.K.; Kim, Y.M.; Chang, Y.W.; Lee, J.H.; Kim, H.W.; Choi, E.J.; the Alliance for Breast Cancer Screening in Korea (ABCS-K). Interpretive Performance and Inter-Observer Agreement on Digital Mammography Test Sets. *Korean J. Radiol.* **2019**, *20*, 218–224. [\[CrossRef\]](#)
35. U.S. Food and Drug Administration. Premarket Approval (PMA). Available online: <https://www.accessdata.fda.gov/scripts/cdrh/cfdocs/cfpma/pma.cfm?id=P970058> (accessed on 19 May 2025).
36. Brem, R.F.; Baum, J.; Lechner, M.; Kaplan, S.; Souders, S.; Naul, L.G.; Hoffmeister, J. Improvement in sensitivity of screening mammography with computer-aided detection: A multiinstitutional Trial. *Am. J. Roentgenol.* **2003**, *181*, 687–693. [\[CrossRef\]](#) [\[PubMed\]](#)
37. Chan, H.-P.; Samala, R.K.; Hadjiiski, L.M. CAD and AI for breast cancer—Recent development and challenges. *Br. J. Radiol.* **2020**, *93*, 20190580. [\[CrossRef\]](#)
38. Mayo, R.C.; Kent, D.; Sen, L.C.; Kapoor, M.; Leung, J.W.T.; Watanabe, A.T. Reduction of False-Positive Markings on Mammograms: A Retrospective Comparison Study Using an Artificial Intelligence-Based CAD. *J. Digit. Imaging* **2019**, *32*, 618–624. [\[CrossRef\]](#) [\[PubMed\]](#)

39. Lehman, C.D.; Wellman, R.D.; Buist, D.S.M.; Kerlikowske, K.; Tosteson, A.N.A.; Miglioretti, D.L.; Breast Cancer Surveillance Consortium. Diagnostic Accuracy of Digital Screening Mammography With and Without Computer-Aided Detection. *JAMA Intern. Med.* **2015**, *175*, 1828–1837. [[CrossRef](#)] [[PubMed](#)]
40. Hosny, A.; Parmar, C.; Quackenbush, J.; Schwartz, L.H.; Aerts, H. Artificial intelligence in radiology. *Nat. Rev. Cancer* **2018**, *18*, 500–510. [[CrossRef](#)]
41. Yoon, J.H.; Kim, E.-K. Deep Learning-Based Artificial Intelligence for Mammography. *Korean J. Radiol.* **2021**, *22*, 1225–1239. [[CrossRef](#)]
42. Yoon, J.H.; Han, K.; Suh, H.J.; Youk, J.H.; Lee, S.E.; Kim, E.-K. Artificial intelligence-based computer-assisted detection/diagnosis (AI-CAD) for screening mammography: Outcomes of AI-CAD in the mammographic interpretation workflow. *Eur. J. Radiol. Open* **2023**, *11*, 100509. [[CrossRef](#)]
43. Park, S.H.; Hwang, E.J. Caveats in Using Abnormality/Probability Scores from Artificial Intelligence Algorithms: Neither True Probability nor Level of Trustworthiness. *Korean J. Radiol.* **2024**, *25*, 328–330. [[CrossRef](#)]
44. Yue, W.; Zhang, H.; Zhou, J.; Li, G.; Tang, Z.; Sun, Z.; Cai, J.; Tian, N.; Gao, S.; Dong, J.; et al. Deep learning-based automatic segmentation for size and volumetric measurement of breast cancer on magnetic resonance imaging. *Front. Oncol.* **2022**, *12*, 984626. [[CrossRef](#)] [[PubMed](#)]
45. Schopp, J.G.; Polat, D.S.; Arjmandi, F.; Hayes, J.C.; Ahn, R.W.; Sullivan, K.; Sahoo, S.; Porembka, J.H. Imaging Challenges in Diagnosing Triple-Negative Breast Cancer. *Radiographics* **2023**, *43*, e230027. [[CrossRef](#)] [[PubMed](#)]
46. Chen, P.; Zhao, S.; Guo, W.; Shao, G. Dynamic contrast-enhanced magnetic resonance imaging features and apparent diffusion coefficient value of HER2-positive/HR-negative breast carcinoma. *Quant. Imaging Med. Surg.* **2023**, *13*, 4816–4825. [[CrossRef](#)] [[PubMed](#)]
47. Mota, B.S.; Reis, Y.N.; de Barros, N.; Cardoso, N.P.; Mota, R.M.S.; Shimizu, C.; Tucunduva, T.C.d.M.; Ferreira, V.C.C.d.S.; Goncalves, R.; Doria, M.T.; et al. Effects of preoperative magnetic resonance image on survival rates and surgical planning in breast cancer conservative surgery: Randomized controlled trial (BREAST-MRI trial). *Breast Cancer Res. Treat.* **2023**, *198*, 447–461. [[CrossRef](#)]
48. Damiani, C.; Kalliatakis, G.; Sreenivas, M.; Al-Attar, M.; Rose, J.; Pudney, C.; Lane, E.F.; Cuzick, J.; Montana, G.; Brentnall, A.R. Evaluation of an AI Model to Assess Future Breast Cancer Risk. *Radiology* **2023**, *307*, e222679. [[CrossRef](#)]

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