







Article

# Robustness Evaluation of a Deep Learning Model on Sagittal and Axial Breast DCE-MRIs to Predict Pathological Complete Response to Neoadjuvant Chemotherapy

Raffaella Massafra <sup>1,†</sup>, Maria Colomba Comes <sup>1,†</sup>, Samantha Bove <sup>1</sup>, Vittorio Didonna <sup>1</sup>, Gianluca Gatta <sup>2</sup> , Francesco Giotta <sup>3</sup>, Annarita Fanizzi <sup>1,\*</sup>, Daniele La Forgia <sup>4,\*</sup> , Agnese Latorre <sup>2</sup> , Maria Irene Pastena <sup>5</sup>, Domenico Pomarico <sup>1</sup> , Lucia Rinaldi <sup>6</sup>, Pasquale Tamborra <sup>1</sup> , Alfredo Zito <sup>5</sup>, Vito Lorusso <sup>3,‡</sup> and Angelo Virgilio Paradiso <sup>7,‡</sup> 

- <sup>1</sup> Struttura Semplice Dipartimentale di Fisica Sanitaria, I.R.C.C.S. Istituto Tumori “Giovanni Paolo II”, Viale Orazio Flacco 65, 70124 Bari, Italy; r.massafra@oncologico.bari.it (R.M.); m.c.comes@oncologico.bari.it (M.C.C.); s.bove@oncologico.bari.it (S.B.); v.didonna@oncologico.bari.it (V.D.); d.pomarico@oncologico.bari.it (D.P.); p.tamborra@oncologico.bari.it (P.T.)
  - <sup>2</sup> Dipartimento di Medicina di Precisione Università della Campania “Luigi Vanvitelli”, 80131 Naples, Italy; gianluca.gatta@unicampania.it (G.G.); a.latorre@oncologico.bari.it (A.L.)
  - <sup>3</sup> Unità Operativa Complessa di Oncologia Medica, I.R.C.C.S. Istituto Tumori “Giovanni Paolo II”, Viale Orazio Flacco 65, 70124 Bari, Italy; f.giotta@oncologico.bari.it (F.G.); vitororusso@me.com (V.L.)
  - <sup>4</sup> Struttura Semplice Dipartimentale di Radiologia Senologica, I.R.C.C.S. Istituto Tumori “Giovanni Paolo II”, Viale Orazio Flacco 65, 70124 Bari, Italy
  - <sup>5</sup> Unità Operativa Complessa di Anatomia Patologica, I.R.C.C.S. Istituto Tumori “Giovanni Paolo II”, Viale Orazio Flacco 65, 70124 Bari, Italy; m.pastena@oncologico.bari.it (M.I.P.); a.zito@oncologico.bari.it (A.Z.)
  - <sup>6</sup> Struttura Semplice Dipartimentale di Oncologia Per la Presa in Carico Globale del Paziente, I.R.C.C.S. Istituto Tumori “Giovanni Paolo II”, Viale Orazio Flacco 65, 70124 Bari, Italy; l.rinaldi@oncologico.bari.it
  - <sup>7</sup> Oncologia Sperimentale e Biobanca, I.R.C.C.S. Istituto Tumori “Giovanni Paolo II”, Viale Orazio Flacco 65, 70124 Bari, Italy; a.paradiso@oncologico.bari.it
- \* Correspondence: a.fanizzi@oncologico.bari.it (A.F.); d.laforgia@oncologico.bari.it (D.L.F.)  
† These authors contributed equally to this work.  
‡ These authors contributed equally to this work.



**Citation:** Massafra, R.; Comes, M.C.; Bove, S.; Didonna, V.; Gatta, G.; Giotta, F.; Fanizzi, A.; La Forgia, D.; Latorre, A.; Pastena, M.I.; et al. Robustness Evaluation of a Deep Learning Model on Sagittal and Axial Breast DCE-MRIs to Predict Pathological Complete Response to Neoadjuvant Chemotherapy. *J. Pers. Med.* **2022**, *12*, 953. <https://doi.org/10.3390/jpm12060953>

Academic Editor: Pierluigi Maria Rinaldi

Received: 25 March 2022

Accepted: 7 June 2022

Published: 10 June 2022

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2022 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/>).

**Abstract:** To date, some artificial intelligence (AI) methods have exploited Dynamic Contrast-Enhanced Magnetic Resonance Imaging (DCE-MRI) to identify finer tumor properties as potential earlier indicators of pathological Complete Response (pCR) in breast cancer patients undergoing neoadjuvant chemotherapy (NAC). However, they work either for sagittal or axial MRI protocols. More flexible AI tools, to be used easily in clinical practice across various institutions in accordance with its own imaging acquisition protocol, are required. Here, we addressed this topic by developing an AI method based on deep learning in giving an early prediction of pCR at various DCE-MRI protocols (axial and sagittal). Sagittal DCE-MRIs refer to 151 patients (42 pCR; 109 non-pCR) from the public I-SPY1 TRIAL database (DB); axial DCE-MRIs are related to 74 patients (22 pCR; 52 non-pCR) from a private DB provided by Istituto Tumori “Giovanni Paolo II” in Bari (Italy). By merging the features extracted from baseline MRIs with some pre-treatment clinical variables, accuracies of 84.4% and 77.3% and AUC values of 80.3% and 78.0% were achieved on the independent tests related to the public DB and the private DB, respectively. Overall, the presented method has shown to be robust regardless of the specific MRI protocol.

**Keywords:** pathological complete response; early prediction; magnetic resonance imaging; deep learning

## 1. Introduction

Neoadjuvant chemotherapy (NAC) is a form of oncological therapy that is used to reduce the size of the tumor and the infiltration of extra-glandular structures, consequently enabling a more conservative surgery [1,2]. Over the last few years, the adoption of NAC

is growing, especially against some molecular subtypes, such as triple-negative tumors and HER2+ tumors of any size, as there is concrete evidence of greater clinical efficacy [3–5]. The achievement of the pathological Complete Response (pCR) at the end of NAC, namely, the absence of residual invasive disease or metastatic lymph nodes at the end of the entire course of the therapy, which is assessed by pathologists on the excised tissue after surgery, leads to a more favorable prognosis than traditional treatments, with an increase in disease-free survival and overall survival [3,6–8]. The prediction of NAC outcome before the beginning of the therapy (early prediction) in terms of pCR is a hot topic in the current clinical research since an earlier identification of patients who will effectively respond to NAC is crucial to improve and change treatment planning during the course of treatment, optimize costs and spare such patients from potentially ineffective or toxic chemotherapy [9,10]. Not less relevant, patients who are identified as responders to NAC since the beginning of the therapy, are more likely to take advantage from breast conserving surgery, avoiding a full mastectomy [1]. Dynamic Contrast-Enhanced Magnetic Resonance Imaging (DCE-MRI) is one of the medical imaging techniques that have come to play a prominent role in the radiomic framework for breast cancer. The possibility of evaluating the same findings with several sequences and in several planes (axial, sagittal, and coronal) allows us to better identify the characteristics of the lesion and the relationships with the contiguous structures [11,12]. Breast MRIs are mostly performed and interpreted in axial and sagittal planes, since coronal projections require more slices and are affected by a greater number of motion artifacts due to breathing [13]. Axial acquisition (a horizontal plane with respect to the human body, dividing it into upper and lower parts) is usually faster than sagittal acquisition (a longitudinal plane with respect to the human body, dividing it into left and right parts). Axial acquisition also provides a better overview of both breasts simultaneously [14]. However, there is no universal guideline for slice direction selection in breast MRI scanning: MRI acquisition protocols can vary across cancer institutions worldwide, with some institutions not conducting the acquisition of both axial and sagittal projections, but instead only one of the two. To date, the evaluation of the NAC response has been successfully investigated through the analysis of DCE-MR images [9,10] since the application of computerized algorithms on MRIs has proven to be crucial in highlighting morphological characteristics of the lesion, tumor size, and even residual tumor [15]. Thus, a higher accuracy in evaluating NAC response has emerged with respect to other imaging techniques [16,17]. In the state-of-the-art, some studies focused on MRI-based radiomics have exploited tissue, peritumoral, or intratumoral features in combination with histopathological information to give an early prediction of pCR, i.e., prior to therapy or during the earliest stages of the therapy [18–22], with the final goal of optimizing treatment planning for each individual patient. However, these kinds of features, which are extracted from raw images, are handcrafted by experts in the field and could be influenced by human bias. More recently, methods based on deep learning, a branch of artificial intelligence (AI), have been designed to automatically extract relevant features from images, including MRIs, without human intervention. The task of “early” prediction of pCR to NAC has been achieved through such deep learning techniques exploiting algorithms, known as Convolutional Neural Networks (CNNs), which resemble the neuron functions of the human brain. Compared to handcrafted features, the exploitation of CNNs to extract features from images has proven to be more promising [23]. MRI examinations acquired prior to treatment (baseline MRI) or at initial stages of NAC, e.g., after the initial cycles, have been used as input data for CNNs [24–28]. As far as we know, the methods that have already proposed to predict pCR to NAC in breast cancer patients have been developed and tailored either for sagittal or axial MRI acquisition views. Hence, more generalizable methods, rather than protocol-specific approaches, need to be developed with the aim of being easily utilized in large multi-institutional studies where the acquisition protocols may differ across diverse institutions. In this work, an AI approach based on deep learning and specifically on a pre-trained CNN has been developed and separately applied on both baseline axial and sagittal pre-treatment MRI examinations to give an early prediction of

pCR in breast cancer patients undergoing NAC. More in detail, the developed model wants to give a prediction of the outcome of NAC in terms of pCR before the beginning of the therapy. The robustness of the approach at various MRI acquisition views, i.e., sagittal and axial views, in reference to a public DB and a private DB has been evaluated. Finally, this study is a first effort to pave the way to the design of an effective approach that could, on the one hand, be able to predict pCR to NAC “early on” in breast cancer patients and, on the other hand, represent a flexible predictive tool to be easily utilized in clinical practice across diverse institutions in accordance with their own imaging acquisition protocols.

## 2. Materials and Methods

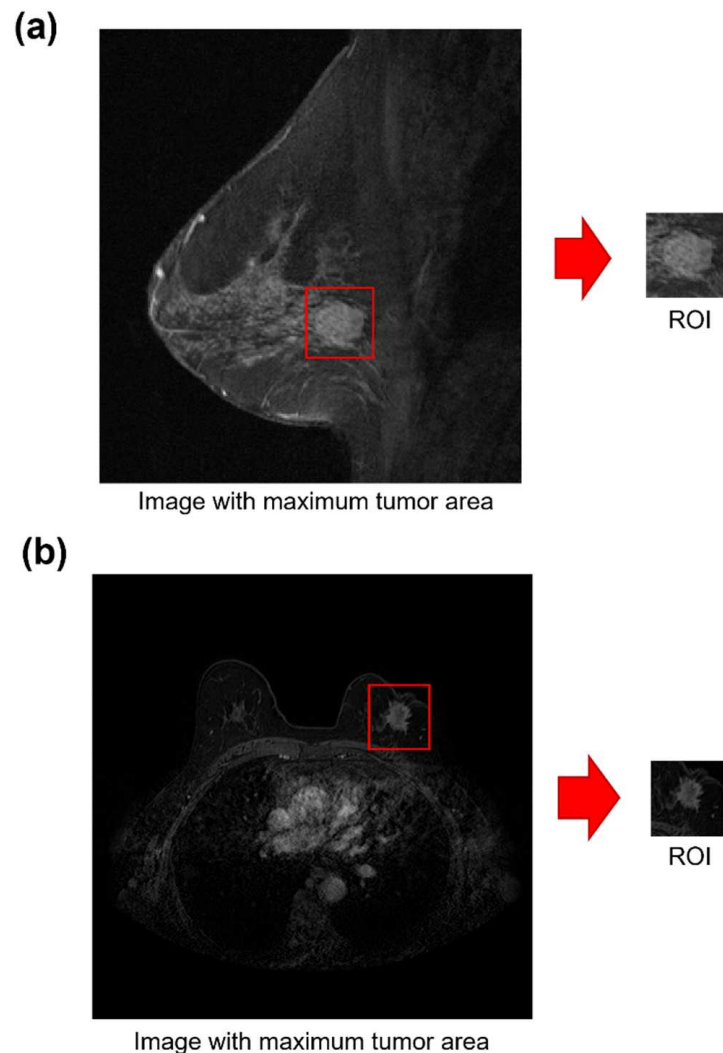
### 2.1. Data Collection

A binary classification task was developed for an early prediction of breast cancer patients who have, or have not, achieved pCR to NAC and whose classifications were indicated as pCR and non-pCR, respectively. The term pathological complete response indicates the absence of residual invasive disease or metastatic lymph nodes at the completion of the entire course of treatment. It is evaluated at the end of chemotherapy and after surgery. For the intended purpose, pre-treatment DCE-MRI examinations were analyzed. Specifically, we defined an AI framework and evaluated its performance to give an early prediction of pCR on two different case studies that dealt with DCE-MRI examinations of breast cancer patients undergoing NAC: a public DB containing sagittal DCE-MRIs and a private DB consisting of axial DCE-MRIs. The public DB, entitled Investigation of Serial Studies to Predict Your Therapeutic Response with Imaging and Molecular Analysis (I-SPY1 TRIAL) [6,29,30], is available online on The Cancer Imaging Archive (<https://wiki.cancerimagingarchive.net>; accessed on 25 November 2021) [31] and includes cases of 230 women with breast tumors of at least 3 cm in size. They were recruited between 2002 and 2006 and underwent NAC according to an anthracycline–cyclophosphamide (AC) regimen either on its own or followed by taxane. Sagittal MRI examinations at different timepoints were obtained. At each timepoint, three images were acquired using 1.5 T field-strength MR imaging systems: a single pre-contrast image and two images taken approximately 2 min and 7 min post contrast injection. Among all the subjects, a set of 151 patients (42 pCR; 109 non-pCR) was then considered, as they had undergone an MRI scan prior to treatment. The private DB consists of a set of 74 patients who were registered as having a first breast tumor diagnosis between 2018 and 2021 at Istituto Tumori “Giovanni Paolo II” in Bari (Italy) and received NAC. Among them, 36 patients were treated with AC followed by taxane; 7 patients underwent AC followed by taxane and trastuzumab; 2 patients received taxane and trastuzumab; 4 patients followed a regimen consisting of AC, pertuzumab, trastuzumab and taxane; 15 patients were treated with pertuzumab, trastuzumab and taxane; and 3 patients followed other treatment regimens. Axial MRI examinations at different timepoints were performed. At each timepoint, six images were acquired in the prone position with a dedicated seven-channel breast coil on a 1.5 Tesla PHILIPS scanner (Achieva, Philips Medical Systems, Amsterdam, The Netherlands): a single pre-contrast image and five images corresponding to approximately each minute post contrast injection, respectively. An MRI examination prior to treatment was acquired for all 74 patients (22 pCR; 52 non-pCR).

### 2.2. Data Pre-Processing

The baseline MRI examination at around the second minute post contrast injection was processed for both datasets. Since previous studies have underlined that the peritumoral region may benefit from a more accurate prediction of pCR in breast cancer patients undergoing NAC [20,27], a semi-automatic algorithm to define a Region Of Interest (ROI) also including the peritumoral zone was developed for all the images belonging to the two case studies: for each patient, an ROI around the center of the tumor mass and containing both intratumoral and peritumoral regions from the MRI scan with the largest tumor area was identified. All the ROIs were reviewed by our expert breast imaging radiologist with

over 20 years of experience. Examples of sagittal and axial acquisition from the public DB and private DB with their corresponding ROIs are shown in Figure 1a,b, respectively.



**Figure 1.** Examples of baseline MR image acquired according to (a) a sagittal view of a public DB patient and (b) an axial view of a private DB patient.

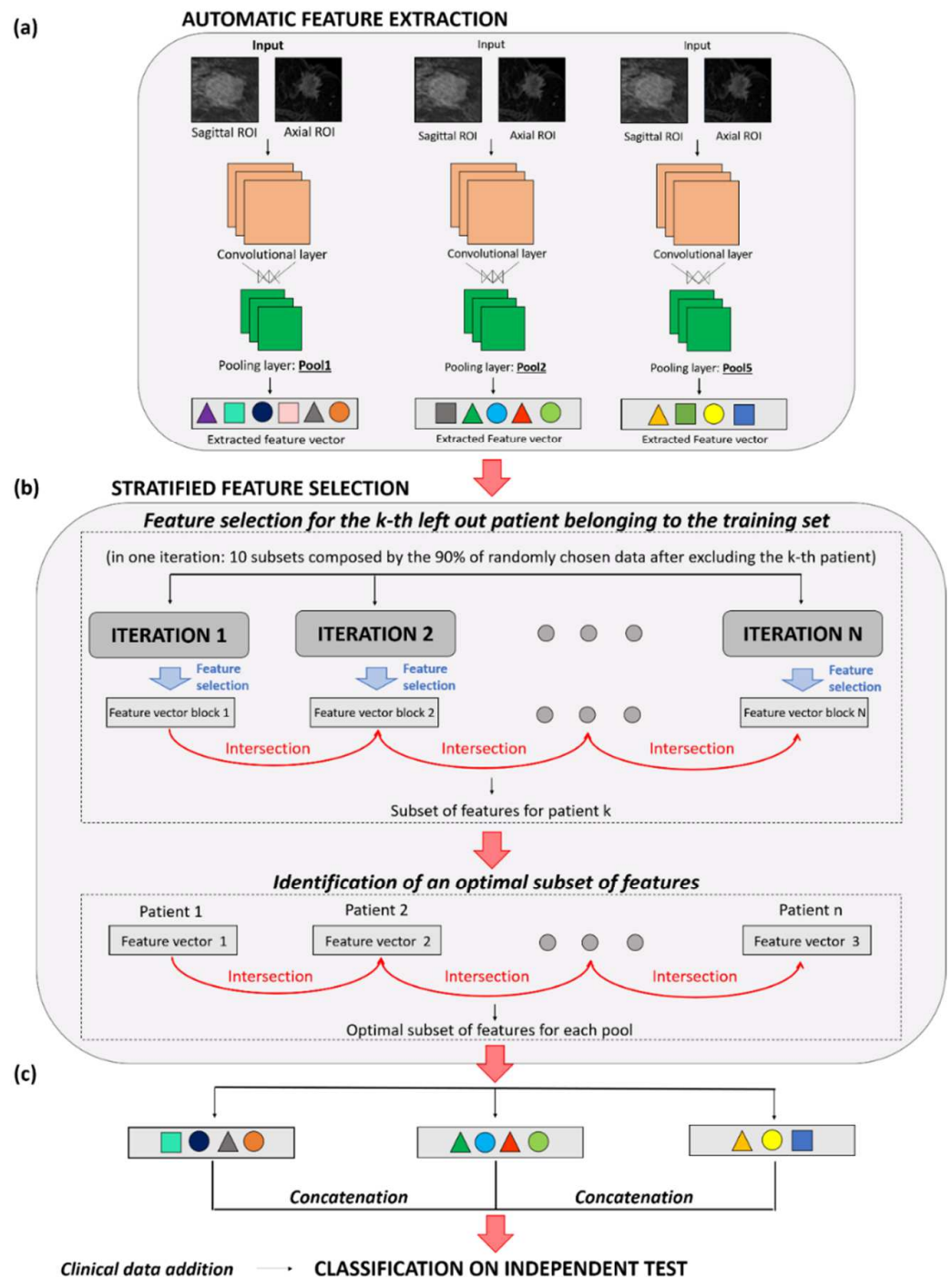
### 2.3. Statistical Analysis

Some clinical variables were included in the information available online on the public DB: age, ER, PgR, HER2 and tumor size (T), which is the largest diameter of the lesion evaluated from the baseline MRI examination. We were provided with two other variables by the authors of the public DB: Ki67 and grading. The same variables were considered for the private DB. The relationship between each clinical feature and the pCR value (0 for the patients belonging to the non-pCR classification and 1 for the patients belonging to the pCR classification) was evaluated by means of an overall statistical test on the datasets independently: the Wilcoxon–Mann–Whitney test [32] was used for continuous features, and the Spearman rank test [33] was used for features measured on an ordinal scale. A result was considered statistically significant when the  $p$ -value was less than 0.10.

### 2.4. An AI Framework to Predict pCR “Early On” from Baseline MRI Examinations

The main steps of the proposed method are briefly reported in the following and represented in Figure 2. The method was trialed on a set of patients (training set) and then validated on a set of patients (independent test set) identified on both the two databases separately. Specifically, the two datasets were divided into training and test sets (70% and

30% of patients, respectively) according to random stratified sampling. In this section, the main details of the method have been reported. Please refer to Appendix A for a more detailed description of the methods used. Features were automatically extracted from three different layers (called *pool1*, *pool2* and *pool5*) of a pre-trained Convolutional Neural Network (CNN), named AlexNET and firstly introduced by A. Krizhevsky and his colleagues [34] (Figure 2a). A stratified feature selection process was then developed to identify the most stable CNN features.



**Figure 2.** Workflow of the proposed AI framework for early pCR prediction. The approach consists of three main steps: (a) Automatic feature extraction through a pre-trained CNN; (b) Stratified feature selection; (c) Classification on the independent test. The method has been applied on sagittal and axial baseline MRIs separately.

Such features were identified in 70% of the overall data for the two databases (training sets), i.e., 106 patients (29 pCR; 77 non-pCR) from the public DB and 52 patients (15 pCR; 37 non-pCR) from the private DB. First, a set of optimal CNN-extracted features was obtained in correspondence with each of the three layers (Figure 2b). Next, a feature concatenation was conducted to merge the optimal features related to the three different layers: an optimal set of features, named F-merged, was constructed. With the aim of validating the proposed approach, the remaining 30% of patients on the two databases was used to define two independent tests, one for each of the case studies (Figure 2c). A total of 45 patients (13 pCR; 32 non-pCR) formed the independent test for the public DB, whereas the test for the private DB was composed of 22 patients (7 pCR; 15 non-pCR). An SVM classifier was built by exploiting the F-merged feature set alone or in combination with some clinical variables. The performances of these classifiers were compared with an SVM classifier using the clinical variables alone. The performance achieved by the designed classifiers was measured in terms of the Area Under the Curve (AUC), the Receiver Operating Characteristic (ROC) curve and other standard metrics, such as accuracy, sensitivity and specificity. 'Accuracy' evaluates the rate of correct classification between the groups of patients who have achieved pCR or not. 'Sensitivity' and 'specificity' measure the proportion of pCR and non-pCR subjects who were correctly identified. AUC, instead, indicates the ability of the classifier to correctly assign patients to the two classes (pCR and non-pCR) by assuming values ranging from 50% (meaning random guessing) to 100% (meaning perfect separability). Each patient with a classification score exceeding a threshold determined by the ratio of the number of patients belonging to the pCR-class over the overall number of patients comprising the dataset [35], i.e., 0.28 for the public DB and 0.30 for the private DB, was assigned to the pCR-class. All the steps of our analysis were performed by using the MATLAB R2019a (MathWorks, Inc., Natick, MA, USA) software.

### 3. Results

#### 3.1. Statistical Analysis Results

In Table 1, the clinical characteristics of patients belonging to the public DB and private DB are split into patients who have achieved pCR (pCR, 42 and 22 for the public DB and private DB, respectively) and patients who have not achieved pCR (non-pCR, 109 and 52 for the public DB and private DB, respectively). The rate of patients who achieved pCR (pCR) was very similar between the two case studies, i.e., they corresponded to 28% and 30% of the public DB and the private DB, respectively. The public DB contains the ER and PgR variables as categorical features. Specifically, they were binary features assuming a negative value (0) or a positive value (1). They were determined by immunohistochemistry (IHC) and were considered positive if the Allred score was  $\geq 3$  [36,37]. The grading variable was evaluated according to the SBR/Elston Classification and the Ki67 IHC staining was performed at the University of North Carolina by using the standard avidin–biotin complex technique. Four classes representing four different levels of Ki67 were recognized [36]: negative if Ki67 was equal to 0; low if Ki67 was less than 10%; intermediate if the variable was between 10% and 25%; and high if Ki67 was greater than 25%. The HER2 variable was evaluated according to IHC and/or fluorescent in situ hybridization assays. The T variable expressing the largest diameter of the lesion within the sagittal MRI examination was part of the clinical information available online [31].

In our DB, the variables ER (Clone EP1 DAKO), PgR (Clone PgR636) and Ki67 (Clone MIB1 DAKO) were evaluated in percentage values; the grading values were assessed according to Elston Classification, whereas the HER2 (polyclonal Rabbit Anti-Human c-erb 2 Oncoprotein) variable was measured according to the ASCO-CAP guidelines. For the sake of a fair comparison between the two case studies, the percentage values of ER and PgR being compared in our DB were converted into categorical binary variables (negative if ER and PgR were equal to 0; positive if ER and PgR assumed values greater than or equal to 1%). Similarly, the Ki67 value was converted in the same categories as the public DB. In

this case, the T variable represented the maximum diameter of the lesion evaluated and reported by our radiologist during the axial MRI scan acquisition.

**Table 1.** Patient characteristics.

	Public DB			Private DB	
	pCR	Non-pCR		pCR	Non-pCR
Overall	42 (28%)	109 (72%)	Overall	22 (30%)	52 (70%)
Age (years)			Age (years)		
Mean ± std	46.81 ± 8.59	49.06 ± 9.15	Mean ± std	51.55 ± 12.72	52.02 ± 12.39
T (mm)			T (mm)		
Mean ± std	75.86 ± 36.39	65.24 ± 27.66	Mean ± std	35.78 ± 20.95	36.05 ± 18.13
Grading			Grading		
G1	0 (0%)	8 (7.4%)	G1	1 (4.5%)	1 (1.9%)
G2	12 (28.6%)	60 (55.0%)	G2	1 (4.5%)	16 (30.8%)
G3	27 (64.3%)	41 (37.6%)	G3	20 (91.0%)	30 (57.7%)
NA	3 (7.1%)	0 (0%)	NA	0 (0%)	5 (9.6%)
ER			ER		
Negative	29 (69.0%)	38 (34.9%)	Negative	12 (54.5%)	13 (25.0%)
Positive	13 (31.0%)	71 (65.1%)	Positive	10 (45.5%)	38 (73.1%)
NA	0 (0%)	0 (0%)	NA	0 (0%)	1 (1.9%)
PgR			PgR		
Negative	34 (81.0%)	49 (45.0%)	Negative	17 (77.3%)	21 (40.4%)
Positive	8 (19.0%)	60 (55.0%)	Positive	5 (22.7%)	30 (57.7%)
NA	0 (0%)	0 (0%)	NA	0 (0%)	1 (1.9%)
Ki67			Ki67		
Negative	2 (4.8%)	4 (3.7%)	Negative	0 (0%)	0 (0%)
Low	2 (4.8%)	28 (25.7%)	Low	0 (0%)	2 (3.8%)
Intermediate	7 (16.7%)	32 (29.4%)	Intermediate	4 (18.2%)	15 (28.8%)
High	20 (47.6%)	34 (31.1%)	High	18 (81.8%)	34 (65.5%)
NA	11 (26.1%)	11 (10.1%)	NA	0 (0%)	1 (1.9%)
HER2			HER2		
Negative	24 (57.1%)	86 (78.9%)	Negative	10 (45.5%)	35 (67.3%)
Positive	17 (40.5%)	22 (20.1%)	Positive	12 (54.5%)	16 (30.8%)
NA	1 (2.4%)	1 (1.0%)	NA	0 (0%)	1 (1.9%)

In the brackets, percentage values are specified. The abbreviation NA indicates missing values.

Table 2 summarizes the *p*-values obtained by performing an association test between pCR and therapy and each clinical factor. No significant association was observed between the couples age–pCR and T–pCR for either of the two case studies (*p*-value > 0.10). Significant associations (*p* < 0.05) emerged between pCR and the histological variables ER, PgR and grading for both the case studies. Moreover, pCR and HER2 showed a significant association, although less significance was found in the private DB (*p*-value < 0.10). However, this was acceptable given the reduced sample size. It is worth noting that the Ki67 was closely associated with pCR only for the public DB (*p*-value < 0.05).

By considering all the patients from the private DB, who underwent different types of therapy pathways, no significant association between Ki67 and pCR emerged (*p*-value > 0.10). However, when the statistical test was performed only on patients of the private DB who underwent the AC + taxane scheme, a statistically significant association between Ki67 and pCR was identified (*p*-value < 0.05). Furthermore, since patients belonging to the private DB underwent different therapy schemes, a variable outlining the performed therapy (therapy type) was also considered, but it did not show a significant association with pCR (*p*-value > 0.10). In our further analysis, when SVM classifiers exploiting clinical variables were designed, the variables that resulted as significantly associated with pCR for at least one of the two case studies were included. They were ER, PgR, HER2, grading and Ki67.

**Table 2.** Statistical analysis on clinical features.

Variable	Type	DB	p-Value
Age	Continuous	public	0.1673
		private	0.8805
T	Continuous	public	0.2508
		private	0.8097
ER	Categorical (binary)	public	$1.2 \times 10^{-4}$
		private	0.0164
PgR	Categorical (binary)	public	$4.9 \times 10^{-5}$
		private	0.0046
HER2	Categorical (binary)	public	0.0087
		private	0.0617
Grading	Categorical	public	$1.4 \times 10^{-4}$
		private	0.0286
Ki67	Categorical	public	0.0116
		private	0.3494
		private (over patients AC + tax)	0.0995

The Wilcoxon–Mann–Whitney test was performed for continuous features, whereas Spearman rank test was used for categorical features. A result was considered statistically significant when the p-value was less than 0.10.

*3.2. Evaluation Performance Achieved by the AI Model on Sagittal and Axial Baseline MRIs*

The results achieved by the same AI approach on the training sets related to the two case studies are reported in Appendix A and represented in Table S1. A total of 29 and 28 features were selected as optimal features by applying all the steps of the AI method for both the public DB and the private DB. The proposed AI framework was then validated on two independent tests, one for each of the two case studies. Table 3 summarizes the results obtained from the independent tests by the models exploiting only clinical features, only F-merged features and clinical features in combination with the F-merged feature set.

**Table 3.** Summary of the performances achieved by the pCR prediction models in terms of AUC.

Set	Model	N. Features	AUC	Acc.	Sens.	Spec.
Public DB Independent test: 45 patients (13 pCR)	Clinical	5	58.2%	64.4%	53.6%	68.8%
	F-merged	29	75.0%	73.3%	<b>69.2%</b>	75.0%
	F-merged + clinical	34	<b>80.3%</b>	<b>84.4%</b>	<b>69.2%</b>	<b>90.6%</b>
Private DB Independent test: 22 patients (7 pCR)	Clinical	5	56.0%	59.1%	42.9%	66.7%
	F-merged	28	72.4%	<b>77.3%</b>	57.1%	<b>86.7%</b>
	F-merged + clinical	33	<b>78.0%</b>	<b>77.3%</b>	<b>71.4%</b>	80.0%

Accuracy (Acc.), Sensitivity (Sens.), and Specificity (Spec.) on the independent tests of the public DB and private DB. The number of features comprising each model is also reported. The best results achieved for each of the evaluation metrics are indicated in bold.

The model, with input of the clinical variables alone, reached an AUC value of 58.2% and 56.0% and an accuracy of 64.4% and 59.1% for the public DB and the private DB, respectively. In this case, the specificity values were greater than the sensitivity values, similar to the results on the training sets (see Appendix A and Table S1). The usage of the F-merged feature set allowed us to improve the accuracy of the results. Overall, the best results were obtained when the clinical features were added to the F-merged features: an AUC value of 80.3% and 78.0%, an accuracy of 84.4% and 77.3%, a sensitivity of 69.2% and 71.4% and a specificity of 90.6% and 80.0% were achieved on the public DB and private DB, respectively. Slightly higher figures for the public DB might be related to the larger values of tumor size (see Table 1). Moreover, the public DB is composed of a greater number of patients with more homogenous characteristics, such as the NAC scheme undergone.



However, more balanced results in terms of all the evaluation metrics were achieved on the private DB. As a final result, the AI method proved itself to be robust at interpreting various MRI acquisition views.

#### 4. Discussion

With the increased use of deep learning techniques in all areas of the biomedical field [38,39], several attempts to solve the early prediction of pCR to NAC in breast cancer patients have been proposed. Liu et al. [40] used the first post-contrast pre-treatment MRI examinations from 131 patients (40 pCR; 91 non-pCR) of the I-SPY1 TRIAL public database to design a CNN-based method to predict pCR. As a result, a mean AUC value of 72% was returned. Ravichandran et al. [26] trained a CNN that made use of pre-contrast and post-contrast pre-treatment MRI scans referred to 133 patients of the I-SPY1 TRIAL public database. The method was validated on 33 patients, reaching an AUC value of 70% and 77% when the post-contrast MRI examinations were considered alone or in conjunction with the pre-contrast MR images, respectively. In this study, we developed and applied an AI framework to two different case studies based on MRI examinations acquired according to varying orientations (sagittal and axial). To the best of our knowledge, all previously developed models have been tailored for a specific protocol (axial or sagittal acquisition view). No investigation into the robustness of AI methods in various case studies with reference to different MRI acquisition views has previously been carried out for the early prediction of pCR to NAC in breast cancer patients. This aspect can be crucial to prove the generalizability of the method on data provided by multiple cancer institutions. As demonstrated elsewhere [26,41], the addition of clinical variables could contribute to improving the performances achieved by using an AI approach. By combining the CNN-features with the clinical variables, the overall performances were stable at varying MRI acquisition views: an AUC value of 80.3% and 78.0% was achieved on the independent tests conducted for the public DB and the private DB, respectively. Among the clinical variables, ER, PgR and HER2 had a significant association with pCR for both case studies. The variable Ki67 was significantly associated with pCR only for patients undergoing the AC + taxane therapeutic scheme. We can observe how the mean T appears differently between the two datasets and across the two classes (pCR and non-pCR, see Table 1). This is because of the timelines of the two case studies: while several years ago NAC was only used for patients with specific requirements in terms of tumor diameter (e.g., greater than 3 cm), nowadays, NAC has become a standard therapy in clinical practice and is received by the majority of patients regardless of their tumor size, and especially if the tumor is categorized as Triple-Negative or HER2+ [42–44]. Despite the promising results, our study has some limitations. The cohort of patients used for our analysis was relatively small, especially in reference to the patients on the private DB that were referred to our Institute. Moreover, an expected drop of performances in passing from the training sets to the independent tests was observed: since there was a limited number of patients belonging to the training sets, a feature selection was more prone to overfitting. To overcome this limitation in a future extension of the study, the robustness and flexibility of our findings will be proven on a larger cohort of patients across multiple institutions and protocols. Manifold are the possible extensions of this work. Future works will be focused on a volumetric analysis by jointly involving all slices of the MRI examinations [45]. An improvement in the prediction performances could be obtained by developing an AI model which could integrate multimodal data, including pre-treatment clinical information joined with features extracted from several kinds of images, such as pre-treatment MRI examinations before and after injection, as well as diffusion weighted images. The fusion of the information from different sources, and especially from different kinds of images, through AI models has already been proven as promising [46]. Each datum of a different nature contains part of the description of the same objects of interest. By combining together all these data, hidden complex relationships between the different modalities can be recognized by AI models [47,48].

## 5. Conclusions

In conclusion, we proposed an AI framework based on deep learning that has revealed itself to be able to give an early prediction of NAC outcome in terms of pCR in breast cancer patients undergoing NAC. In clinical practice, this aspect is essential for medical figures to promote personalized tools with the aim of contributing to the optimal selection of treatment and therapeutic options. Despite several efforts having been made in the state-of-the-art to give an early prediction of pCR to NAC, there is a lack of generalizable methods, rather than approaches applied only on sagittal or axial MRI examinations. Hence, more generalizable methods need to be developed with the aim of being easily utilized in large multi-institutional studies where the acquisition protocols may differ across diverse institutions. In this work, we wanted to fill this gap. The proposed approach appeared robust at various acquisition views of the MRI examinations performed prior to treatment: both axial and sagittal. This work represents a first effort towards the implementation of a more generalizable approach that can overcome protocol-specific methods, so that it can be utilized in multi-center studies exploiting diverse imaging acquisition protocols.

**Supplementary Materials:** The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/jpm12060953/s1>, Table S1: Summary of the performances achieved by the pCR prediction models in terms of AUC, Accuracy (Acc.), Sensitivity (Sens.), and Specificity (Spec.) on the training sets of the public DB and private DB. The number of features comprising each model is also reported. The best results achieved for each of the evaluation metrics are indicated in bold.

**Author Contributions:** Conceptualization, M.C.C., R.M. and A.F.; methodology, M.C.C., R.M. and A.F.; software, M.C.C.; validation, M.C.C.; formal analysis, M.C.C., R.M., A.F., D.L.F. and A.L.; resources, V.L. and R.M.; data curation M.C.C., R.M. and A.F.; writing—original draft preparation, M.C.C., R.M., A.F. and S.B.; writing—review and editing, M.C.C., R.M., S.B., V.D., G.G., F.G., A.F., D.L.F., A.L., M.I.P., D.P., L.R., P.T., A.Z., V.L. and A.V.P.; supervision, R.M., A.F., V.L. and A.V.P. 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 according to the guidelines of the Declaration of Helsinki and approved by the Scientific Board of Istituto Tumori ‘Giovanni Paolo II’—Bari, Italy. The number of the Protocol approved by the Scientific Board of Istituto Tumori ‘Giovanni Paolo II’ (Bari, Italy) is 6629/21.

**Informed Consent Statement:** No consent or waiver was required for data taken from the public database as they were obtained already anonymized from the National Cancer Institute’s Cancer Imaging Archive. Written informed consent for participation was not required for collecting retrospective data from the private database in accordance with the institutional requirements.

**Data Availability Statement:** Data images refer to the public database that is part of The Cancer Imaging Archive (TCIA) at the following link: <https://wiki.cancerimagingarchive.net> (accessed on 25 November 2021). Data obtained from the private database are available on request from the corresponding author. These data are not publicly available because they are the propriety of I.R.C.C.S.—Istituto Tumori ‘Giovanni Paolo II’—Bari, Italy.

**Acknowledgments:** We thank David Newitt and Nola Hylton from the University of California, San Francisco, for providing us with some clinical variables related to the public DB. This work was supported by funding from the Italian Ministry of Health “Ricerca Finalizzata 2018”.

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

## Appendix A

### *Appendix A.1. Technical Details about the Proposed AI Framework*

The main steps of the proposed AI framework are summarized in the following.

### Appendix A.1.1. Feature Extraction

Artificial intelligence and in particular deep learning models based on Convolutional Neural Networks (CNNs) have been successfully adopted as classifiers to solve several tasks in medicine [41,47,48] which involve image analysis. However, they are also used to extract features from the raw images. Here, we exploited a pre-trained CNN, called AlexNET [34], to automatically extract features from ROIs related to baseline MRIs. ROIs were input directly to the CNN architecture, which was pre-trained on ImageNet, a dataset of over one million of natural images designed for general object recognition. In this study, the knowledge learned by the network during such a training phase was transferred on the MRI exams. These features were then utilized to discriminate “early on” if a patients achieved pCR or not (pCR vs. non-pCR). The usage of a pre-trained network was preferred here since both the public and private DBs were composed by a number of images that was not sufficient to train a customized network [28,47]. The used network consists of 25 layers in total, including 5 convolutional layers, 3 pooling layers and 3 fully connected layers [34], where convolutional layers represent features at different level of input representation and abstraction, whereas pooling layers make features invariant to truncation, occlusion and translation [11]. Starting layers capture low-level features, which are features related to the local structure of the image such as lines, edges and blobs. Deeper layers return more complex and abstract features related to global characteristics of the image. They are obtained by combining low-level features. In this work, features from the three pooling layers of the network (called *pool1*, *pool2* and *pool5*) were automatically extracted to utilize the information within the feature maps at different resolutions and abstractions (see Figure 2 in the main text). Each input image that was the ROI extracted from the pre-processing phase was then resized to  $227 \times 227$  pixels, since the pre-trained network required input data of this size. In details, the output of *pool1 layer* had dimensions of  $27 \times 27 \times 96$ . A single 69,984-length vector was finally obtained by flattening the output. The *pool2 layer* had an output with dimensions of  $13 \times 13 \times 256$  that was flattening to a single 43,264-length vector, whereas the *pool5 layer* returned an output of  $6 \times 6 \times 256$  that was transformed in a 9216-length vector.

### Appendix A.1.2. Stratified Feature Selection

A Stratified Feature Selection (SFS) was performed to finally choose a set of most stable features in correspondence of each of the three pooling layers used for feature extraction. Such a process was conducted on 70% of the overall data for the two databases separately (training sets). The remaining 30% of data for both datasets were employed as independent tests to validate the proposed model. The SFS consists of two main stages. As the first stage, a feature selection procedure was implemented according to a Leave-One-Out (LOO) cross-validation scheme to obtain a feature vector for each left-out patient of the training set (see the 2b in the main text). When a patient was left out, an iterative procedure based on stacking 10 subsets composed by the remaining 90% of the data, which were randomly selected, was performed. For each stacked subset of each iteration, important features were selected by applying two feature selection techniques via a cascade mechanism. A first well-known feature selection method, such as the non-parametric statistical Wilcoxon–Mann–Whitney test [32], evaluated the discrimination power with respect to the class (pCR or non-pCR) of each feature separately. The features that were resulted as statistically significant (with  $p < 0.001$  for the *pool1* and *pool2* layers and with  $p < 0.05$  for the *pool5* layer) became the input of a second feature selection algorithm, namely, Random Forest (RF), that chose a feature as important by comparing its importance with respect to all the features at disposal. In our simulations, the configuration of RF counted 100 trees. The features selected for each of the stacked subsets were unified in a single set. As shown in Figure 2 in the main text, at one iteration, one subset of features was associated. For each patient left out, the subsets of features obtained from the following iterations were intersected among each other thus finally obtaining a single feature set for that patient. The number of iterations ( $n$ ) was experimentally estimated as  $n = 20$ . As the second stage of the SFS

procedure, a single feature vector that was meaningful for all the patients of the training set was determined for each of the three pooling layers of the network, separately. Given a layer, such a feature vector was computed by intersecting all the subsets of features (one for each patient) that were outcome of the previous stage of the SFS procedure. These last feature sets were called as *F-pool1*, *F-pool2* and *F-pool5* in correspondence of *pool1*, *pool2* and *pool5* layers. Finally, the three feature sets contained the optimal features, i.e., the most stable features, that were less susceptible to variations in training samples. Several Support Vector Machine (SVM) classifiers were designed on these set of patients and a leave-one-out cross validation procedure on the training sets was implemented. A first classifier took in input the features belonging to *F-pool1*; a second classifier used the features of the *F-pool2* set; a third classifier employed the *F-pool5* as set of features. Another SVM classifier exploiting the F-merged set was designed. Finally, some clinical variables were added to the F-merged and exploited to build one further SVM classifier. The performances of these classifiers were compared with an SVM classifier using the clinical variables alone.

#### Appendix A.1.3. Classification

The power of the optimal selected features in discriminating “early on” patients who achieved pCR or not was assessed by separately training SVM classifiers on the sets of patients from the private DB and the public DB, which were used to identify the optimal subset of CNN features extracted for each layer. The classifiers were then validated on independent tests from the public and the private DB, respectively. A feature concatenation was conducted to merge the optimal features related to the three different layers: an optimal set of features, named F-merged, was constructed and then used to train another SVM classifier. Moreover, some clinical variables were evaluated alone or in addition to the F-merged. In correspondence, two diverse SVM classifiers were designed. Finally, clinical variables contain a few missing data. Such missing data were estimated by means of a proximity technique. For the training and test sets separately, the proximity technique allows us to replace the missing data of a patient with the respective data values referred to the patient without missing features with the closest HR and HER2 values: the Euclidean distance among the HR and HER2 values of the patients with missing data and completed data was computed.

#### Appendix A.2. Results on Training Sets

Here, we present the results achieved by the same AI approach on the training sets related to the public DB and the private DB, respectively. They are reported in Table S1. The model exploiting clinical variables achieved comparable results on the two case studies (see the rows with clinical model in Table S1). An AUC value of 51.0% and 59.8% and an accuracy of 58.5% and 55.8% were achieved on the public DB and the private DB, respectively. Overall, the specificity assumed a higher value than sensitivity in both cases. The models built by using the optimal CNN-features extracted by the sagittal and axial MRIs, referred to the public and private DBs, respectively, were evaluated. The number of features contained in the feature sets in correspondence of the three layers is reported in Table S1. Features from the three pooling layers (*pool1*, *pool2* and *pool5*) of the network were automatically extracted to utilize the information within the feature maps at different resolutions and abstractions. Starting layers capture low-level features, which are features related to the local structure of the image such as lines, edges and blobs. Deeper layers return more complex and abstract features related to global characteristics of the image that are obtained by combining low-level features. The highest number of features was obtained for the *pool2* layer (13 features) and the *pool1* layer (12 features) in the case of the public DB and private DB, respectively. The lower number of features was returned for the *pool5* layer with respect to both DBs (6 features for the public DB and 5 features for the private DB, respectively). On the sagittal MRIs of the public DB, the models exploiting the CNN-extracted features belonging to the sets *F-pool1* and *F-pool5* reached almost comparable performances in terms of AUC values (i.e., AUC values of 82.6% and 81.7% respectively).

Among the models employing the CNN-features extracted from the three layers separately, the highest specificity value (96.1%) was achieved by the model using the *F-pool2*. On the axial MRIs of the private DB, while the highest AUC value was achieved by using *F-pool1* (82.7%), the best accuracy was referred to the model exploiting *F-pool2* (84.6%). Merging the features belonging to *F-pool1*, *F-pool2* and *F-pool3* in a single set, i.e., F-merged, allowed it to globally outperform the previous models, especially in terms of AUC values: an AUC value of 86.7% and 86.1% was obtained on the public DB and private DB, respectively. Apart from the specificity value, the best performances were reached by combining the features belonging to the F-merged with the clinical features above mentioned in the previous subparagraph: an AUC value of 92.9% and 88.4%, an accuracy of 82.1% and 88.5% were achieved on the public DB and private DB, respectively.

## References

- Mieog, J.S.D.; Van Der Hage, J.A.; Van De Velde, C.J.H. Neoadjuvant chemotherapy for operable breast cancer. *Br. J. Surg.* **2007**, *94*, 1189–1200. [[CrossRef](#)] [[PubMed](#)]
- Cain, H.; Macpherson, I.R.; Beresford, M.; Pinder, S.E.; Pong, J.; Dixon, J.M. Neoadjuvant Therapy in Early Breast Cancer: Treatment Considerations and Common Debates in Practice. *Clin. Oncol.* **2017**, *29*, 642–652. [[CrossRef](#)] [[PubMed](#)]
- Minarikova, L.; Bogner, W.; Pinker, K.; Valkovič, L.; Zarić, O.; Bago-Horvath, Z.; Bartsch, R.; Helbich, T.H.; Trattning, S.; Gruber, S. Investigating the prediction value of multiparametric magnetic resonance imaging at 3 T in response to neoadjuvant chemotherapy in breast cancer. *Eur. Radiol.* **2017**, *27*, 1901–1911. [[CrossRef](#)] [[PubMed](#)]
- Nguyen, A.A.T.; Arasu, V.A.; Strand, F.; Li, W.; Onishi, N.; Gibbs, J.; Jones, E.F.; Joe, B.N.; Esserman, L.J.; Newitt, D.C.; et al. Comparison of segmentation methods in assessing background parenchymal enhancement as a biomarker for response to neoadjuvant therapy. *Tomography* **2020**, *6*, 101–110. [[CrossRef](#)]
- 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](#)]
- Hylton, N.M.; Blume, J.D.; Bernreuter, W.K.; Pisano, E.D.; Rosen, M.A.; Morris, E.A.; Weatherall, P.T.; Lehman, C.D.; Newstead, G.M.; Polin, S.; et al. Locally advanced breast cancer: MR imaging for prediction of response to neoadjuvant chemotherapy—Results from ACRIN 6657/I-SPY TRIAL. *Radiology* **2012**, *263*, 663–672. [[CrossRef](#)]
- Choi, W.J.; Kim, W.K.; Shin, H.J.; Cha, J.H.; Chae, E.Y.; Kim, H.H. Evaluation of the Tumor Response After Neoadjuvant Chemotherapy in Breast Cancer Patients: Correlation Between Dynamic Contrast-enhanced Magnetic Resonance Imaging and Pathologic Tumor Cellularity. *Clin. Breast Cancer* **2018**, *18*, e115–e121. [[CrossRef](#)]
- Li, X.; Abramson, R.G.; Arlinghaus, L.R.; Kang, H.; Chakravarthy, A.B.; Abramson, V.G.; Farley, J.; Mayer, I.A.; Kelley, M.C.; Meszoely, M.C.; et al. Multiparametric magnetic resonance imaging for predicting pathological response after the first cycle of neoadjuvant chemotherapy in breast cancer. *Investig. Radiol.* **2015**, *50*, 195–204. [[CrossRef](#)]
- Smith, I.C.; Heys, S.D.; Hutcheon, A.W.; Miller, I.D.; Payne, S.; Gilbert, F.J.; Ah-See, A.K.; Eremin, O.; Walker, L.G.; Sarkar, T.K.; et al. Neoadjuvant chemotherapy in breast cancer: Significantly enhanced response with docetaxel. *J. Clin. Oncol.* **2002**, *20*, 1456–1466. [[CrossRef](#)]
- Gonzalez-Angulo, A.M.; Morales-Vasquez, F.; Hortobagyi, G.N. Overview of resistance to systemic therapy in patients with breast cancer. In *Breast Cancer Chemosensitivity*; Advances in Experimental Medicine and Biology; Springer: Berlin/Heidelberg, Germany, 2007; Volume 608, pp. 1–22. [[CrossRef](#)]
- Wei, D.; Jahani, N.; Cohen, E.; Weinstein, S.; Hsieh, M.-K.; Pantalone, L.; Kontos, D. Fully automatic quantification of fibroglandular tissue and background parenchymal enhancement with accurate implementation for axial and sagittal breast MRI protocols. *Med. Phys.* **2021**, *48*, 238–252. [[CrossRef](#)]
- Fausto, A.; Fanizzi, A.; Volterrani, L.; Mazzei, F.G.; Calabrese, C.; Casella, D.; Marcasciano, M.; Massafra, R.; La Forgia, D.; Mazzei, M.A. Feasibility, image quality and clinical evaluation of contrast-enhanced breast mri performed in a supine position compared to the standard prone position. *Cancers* **2020**, *12*, 2364. [[CrossRef](#)] [[PubMed](#)]
- Margolies, L.R.; Salvatore, M.; Tam, K.; Yip, R.; Bertolini, A.; Henschke, C.I.; Yankelevitz, D.F. Breast mass assessment on chest CT: Axial, sagittal, coronal or maximal intensity projection? *Clin. Imaging* **2020**, *63*, 60–64. [[CrossRef](#)] [[PubMed](#)]
- Mann, R.M.; Cho, N.; Moy, L. Breast MRI: State of the art. *Radiology* **2019**, *292*, 520–536. [[CrossRef](#)] [[PubMed](#)]
- Fowler, A.M.; Mankoff, D.A.; Joe, B.N. Imaging neoadjuvant therapy Response in Breast Cancer. Online SA-CME STATE OF THE ART: Imaging Neoadjuvant Therapy Response in Breast Cancer Fowler et al. *Radiology* **2017**, *285*, 358–375. [[CrossRef](#)]
- Scheel, J.R.; Kim, E.; Partridge, S.C.; Lehman, C.D.; Rosen, M.A.; Bernreuter, W.; Pisano, E.D.; Marques, H.; Morris, E.A.; Weatherall, P.T.; et al. MRI, clinical examination, and mammography for preoperative assessment of residual disease and pathologic complete response after neoadjuvant chemotherapy for breast cancer: ACRIN 6657 trial. *Am. J. Roentgenol.* **2018**, *210*, 1376–1385. [[CrossRef](#)]

17. Park, J.S.; Moon, W.K.; Lyou, C.Y.; Cho, N.; Kang, K.W.; Chung, J.-K. The assessment of breast cancer response to neoadjuvant chemotherapy: Comparison of magnetic resonance imaging and 18F-fluorodeoxyglucose positron emission tomography. *Acta Radiol.* **2011**, *52*, 21–28. [[CrossRef](#)]
18. La Forgia, D.; Vestito, A.; Lasciarrea, M.; Comes, M.C.; Diotaiuti, S.; Giotta, F.; Latorre, A.; Lorusso, V.; Massafra, R.; Palmiotti, G.; et al. Response Predictivity to Neoadjuvant Therapies in Breast Cancer: A Qualitative Analysis of Background Parenchymal Enhancement in DCE-MRI. *J. Pers. Med.* **2021**, *11*, 256. [[CrossRef](#)]
19. Eun, N.L.; Kang, D.; Son, E.J.; Park, J.S.; Youk, J.H.; Kim, J.A.; Gweon, H.M. Texture analysis with 3.0-T MRI for association of response to neoadjuvant chemotherapy in breast cancer. *Radiology* **2020**, *294*, 31–41. [[CrossRef](#)]
20. Braman, N.M.; Etesami, M.; Prasanna, P.; Dubchuk, C.; Gilmore, H.; Tiwari, P.; Plecha, D.; Madabhushi, A. Intratumoral and peritumoral radiomics for the pretreatment prediction of pathological complete response to neoadjuvant chemotherapy based on breast DCE-MRI. *Breast Cancer Res.* **2017**, *19*, 57. [[CrossRef](#)]
21. Tahmassebi, A.; Wengert, G.J.; Helbich, T.H.; Bago-Horvath, Z.; Alaei, S.; Bartsch, R.; Dubsy, P.; Baltzer, P.; Clauser, P.; Kapetas, P.; et al. Impact of machine learning with multiparametric magnetic resonance imaging of the breast for early prediction of response to neoadjuvant chemotherapy and survival outcomes in breast cancer patients. *Investig. Radiol.* **2019**, *54*, 110–117. [[CrossRef](#)]
22. Drisis, S.; Metens, T.; Ignatiadis, M.; Stathopoulos, K.; Chao, S.L.; Lemort, M. Quantitative DCE-MRI for prediction of pathological complete response following neoadjuvant treatment for locally advanced breast cancer: The impact of breast cancer subtypes on the diagnostic accuracy. *Eur. Radiol.* **2016**, *26*, 1474–1484. [[CrossRef](#)]
23. Mencattini, A.; Di Giuseppe, D.; Comes, M.C.; Casti, P.; Corsi, F.; Bertani, F.R.; Ghibelli, L.; Businaro, L.; Di Natale, C.; Parrini, M.C.; et al. Discovering the hidden messages within cell trajectories using a deep learning approach for in vitro evaluation of cancer drug treatments. *Sci. Rep.* **2020**, *10*, 7653. [[CrossRef](#)] [[PubMed](#)]
24. Huynh, B.Q.; Antropova, N.; Giger, M.L. Comparison of breast DCE-MRI contrast time points for predicting response to neoadjuvant chemotherapy using deep convolutional neural network features with transfer learning. In *Medical Imaging 2017: Computer-Aided Diagnosis, Proceedings of the SPIE Medical Imaging, Orlando, FL, USA, 11–16 February 2017*; SPIE: Bellingham, WA, USA, 2017; Volume 10134, p. 101340U. [[CrossRef](#)]
25. Ha, R.; Chin, C.; Karcich, J.; Liu, M.Z.; Chang, P.; Mutasa, S.; Van Sant, E.P.; Wynn, R.T.; Connolly, E.; Jambawalikar, S. 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*, 693–701. [[CrossRef](#)]
26. Ravichandran, K.; Braman, N.; Janowczyk, A.; Madabhushi, A. A deep learning classifier for prediction of pathological complete response to neoadjuvant chemotherapy from baseline breast DCE-MRI. In *Medical Imaging 2018: Computer-Aided Diagnosis, Proceedings of the SPIE Medical Imaging, Houston, TX, USA, 10–15 February 2018*; SPIE: Bellingham, WA, USA, 2018; Volume 10575. [[CrossRef](#)]
27. El Adoui, M.; Drisis, S.; Benjelloun, M. Multi-input deep learning architecture for predicting breast tumor response to chemotherapy using quantitative MR images. *Int. J. Comput. Assist. Radiol. Surg.* **2020**, *15*, 1491–1500. [[CrossRef](#)]
28. Comes, M.C.; Fanizzi, A.; Bove, S.; Didonna, V.; Diotaiuti, S.; La Forgia, D.; Latorre, A.; Martinelli, E.; Mencattini, A.; Nardone, A.; et al. Early prediction of neoadjuvant chemotherapy response by exploiting a transfer learning approach on breast DCE-MRIs. *Sci. Rep.* **2021**, *11*, 14123. [[CrossRef](#)]
29. Newitt, D.; Hylton, N. Multi-center breast DCE-MRI data and segmentations from patients in the I-SPY 1/ACRIN 6657 trials. *Cancer Imaging Arch.* **2016**, *10*, 7. [[CrossRef](#)]
30. Hylton, N.M.; Gatsonis, C.A.; Rosen, M.A.; Lehman, C.D.; Newitt, D.C.; Partridge, S.C.; Bernreuter, W.K.; Pisano, E.D.; Morris, E.A.; Weatherall, P.T.; et al. Neoadjuvant chemotherapy for breast cancer: Functional tumor volume by MR imaging predicts recurrence-free survival-results from the ACRIN 6657/CALGB 150007 I-SPY 1 TRIAL. *Radiology* **2016**, *279*, 44–55. [[CrossRef](#)] [[PubMed](#)]
31. Clark, K.; Vendt, B.; Smith, K.; Freymann, J.; Kirby, J.; Koppel, P.; Moore, S.; Phillips, S.; Maffitt, D.; Pringle, D.; et al. The cancer imaging archive (TCIA): Maintaining and operating a public information repository. *J. Digit. Imaging* **2013**, *26*, 1045–1057. [[CrossRef](#)] [[PubMed](#)]
32. Mann, H.B.; Whitney, D.R. On a Test of Whether one of Two Random Variables is Stochastically Larger Larger than the other. *Ann. Math. Stat.* **1947**, *18*, 50–60. [[CrossRef](#)]
33. Lyerly, S.B. The average spearman rank correlation coefficient. *Psychometrika* **1952**, *17*, 421–428. [[CrossRef](#)]
34. Russakovsky, O.; Deng, J.; Su, H.; Satheesh, S.; Ma, S.; Huang, Z.; Karpathy, A.; Khosla, A.; Bernstein, M.; Berg, A.C.; et al. ImageNet Large Scale Visual Recognition Challenge. *Int. J. Comput. Vis.* **2015**, *115*, 211–252. [[CrossRef](#)]
35. Amoroso, N.; Errico, R.; Bruno, S.; Chincarini, A.; Garuccio, E.; Sensi, F.; Tangaro, S.; Tateo, A.; Bellotti, R. Hippocampal unified multi-atlas network (HUMAN): Protocol and scale validation of a novel segmentation tool. *Phys. Med. Biol.* **2015**, *60*, 8851–8867. [[CrossRef](#)] [[PubMed](#)]
36. Esserman, L.J.; Berry, D.A.; DeMichele, A.; Carey, L.; Davis, S.E.; Buxton, M.; Hudis, C.; Gray, J.W.; Perou, C.; Yau, C.; et al. Pathologic complete response predicts recurrence-free survival more effectively by cancer subset: Results from the I-SPY 1 TRIAL—CALGB 150007/150012, ACRIN 6657. *J. Clin. Oncol.* **2012**, *30*, 3242–3249. [[CrossRef](#)] [[PubMed](#)]
37. Schirosi, L.; Strippoli, S.; Gaudio, F.; Graziano, G.; Popescu, O.; Guida, M.; Simone, G.; Mangia, A. Is immunohistochemistry of BRAF V600E useful as a screening tool and during progression disease of melanoma patients? *BMC Cancer* **2016**, *16*, 905. [[CrossRef](#)]

38. Esteva, A.; Robicquet, A.; Ramsundar, B.; Kuleshov, V.; DePristo, M.; Chou, K.; Cui, C.; Corrado, G.; Thrun, S.; Dean, J.; et al. A guide to deep learning in healthcare. *Nat. Med.* **2019**, *25*, 24–29. [[CrossRef](#)]
39. Comes, M.C.; Filippi, J.; Mencattini, A.; Casti, P.; Cerrato, G.; Sauvat, A.; Vacchelli, E.; De Ninno, A.; Di Giuseppe, D.; D’Orazio, M.; et al. Multi-scale generative adversarial network for improved evaluation of cell–cell interactions observed in organ-on-chip experiments. *Neural Comput. Appl.* **2021**, *33*, 3671–3689. [[CrossRef](#)]
40. Liu, M.Z.; Mutasa, S.; Chang, P.; Siddiqueb, M.; Jambawalikara, S.; Ha, R. A novel CNN algorithm for pathological complete response prediction using an I-SPY TRIAL breast MRI database. *Magn. Reson. Imaging* **2020**, *73*, 148–151. [[CrossRef](#)]
41. Comes, M.C.; La Forgia, D.; Didonna, V.; Fanizzi, A.; Giotta, F.; Latorre, A.; Martinelli, E.; Mencattini, A.; Paradiso, A.V.; Tamborra, P.; et al. Early Prediction of Breast Cancer Recurrence for Patients Treated with Neoadjuvant Chemotherapy: A Transfer Learning Approach on DCE-MRIs. *Cancers* **2021**, *13*, 2298. [[CrossRef](#)]
42. Wein, L.; Luen, S.J.; Savas, P.; Salgado, R.; Loi, S. Checkpoint blockade in the treatment of breast cancer: Current status and future directions. *Br. J. Cancer* **2018**, *119*, 4–11. [[CrossRef](#)]
43. Bartsch, R.; Bergen, E.; Galid, A. Current concepts and future directions in neoadjuvant chemotherapy of breast cancer. *Memo-Mag. Eur. Med. Oncol.* **2018**, *11*, 199–203. [[CrossRef](#)]
44. Granzier, R.W.Y.; van Nijnatten, T.J.A.; Woodruff, H.C.; Smidt, M.L.; Lobbes, M.B.I. Exploring breast cancer response prediction to neoadjuvant systemic therapy using MRI-based radiomics: A systematic review. *Eur. J. Radiol.* **2019**, *121*, 108736. [[CrossRef](#)] [[PubMed](#)]
45. Peng, S.; Chen, L.; Tao, J.; Liu, J.; Zhu, W.; Liu, H.; Yang, F. Radiomics Analysis of Multi-Phase DCE-MRI in Predicting Tumor Response to Neoadjuvant Therapy in Breast Cancer. *Diagnostics* **2021**, *11*, 2086. [[CrossRef](#)] [[PubMed](#)]
46. Gao, J.; Li, P.; Chen, Z.; Zhang, J. A survey on deep learning for multimodal data fusion. *Neural Comput.* **2020**, *32*, 829–864. [[CrossRef](#)] [[PubMed](#)]
47. Bellotti, R.; Bagnasco, S.; Bottigli, U.; Castellano, M.; Cataldo, R.; Catanzariti, E.; Cerello, P.; Cheran, S.C.; De Carlo, F.; Delogu, P.; et al. The MAGIC-5 project: Medical applications on a grid infrastructure connection. *IEEE Nucl. Sci. Symp. Conf. Rec.* **2004**, *3*, 1902–1906. [[CrossRef](#)]
48. Bellotti, R.; De Carlo, F.; Massafra, R.; de Tommaso, M.; Sciruicchio, V. Topographic classification of EEG patterns in Huntington’s disease. *Neurol. Clin. Neurophysiol. NCN* **2004**, *2004*, 37. [[PubMed](#)]