

# Multi-Center Evaluation of Artificial Intelligent Imaging And Clinical Models For Predicting Neoadjuvant Chemotherapy Response In Breast Cancer

**Tan Hong Qi**

National Cancer Centre Singapore

**Ong Hiok Hian**

Nanyang Technological University

**Arjunan Muthu Kumaran**

National Cancer Centre Singapore

**Tira J. Tan**

National Cancer Centre Singapore

**Ryan Shea Tan Ying Cong**

National Cancer Centre Singapore

**Ghislaine Lee Su-Xin**

National Cancer Centre Singapore

**Elaine Lim Hsuen**

National Cancer Centre Singapore

**Raymond Ng Chee Hui**

National Cancer Centre Singapore

**Richard Yeo Ming Chert**

National Cancer Centre Singapore

**Faye Lynette Lim Wei Tching**

National Cancer Centre Singapore

**Zhang Zewen**

National Cancer Centre Singapore

**Christina Yang Shi Hui**

National Cancer Centre Singapore

**Wong Ru Xin**

National Cancer Centre Singapore

**Gideon Ooi Su Kai**

National Cancer Centre Singapore

**Lester Leong Chee Hao**

Loading [MathJax]/jax/output/CommonHTML/jax.js

**Tan Su Ming**

National Cancer Centre Singapore

**Madhukumar Preetha**

National Cancer Centre Singapore

**Sim Yirong**

National Cancer Centre Singapore

**Veronique Tan Kiak Mien**

National Cancer Centre Singapore

**Joe Yeong**

National Cancer Centre Singapore

**Wong Fuh Yong**

National Cancer Centre Singapore

**Cai Yiyu**

Nanyang Technological University

**Wen Long Nei** (✉ [nei.wen.long@singhealth.com.sg](mailto:nei.wen.long@singhealth.com.sg))

National Cancer Centre Singapore <https://orcid.org/0000-0001-6027-1652>

---

**Research article**

**Keywords:** Neoadjuvant Chemotherapy, Breast Cancer, Machine learning, Deep Learning, Artificial Intelligence, Radiomics

**Posted Date:** September 3rd, 2021

**DOI:** <https://doi.org/10.21203/rs.3.rs-838461/v1>

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

[Read Full License](#)

---

# Abstract

## Background:

Neoadjuvant chemotherapy (NAC) plays an important role in the management of locally advanced breast cancer. It allows for downstaging of tumours, potentially allowing for breast conservation. NAC also allows for in-vivo testing of the tumours' response to chemotherapy and provides important prognostic information. There are currently no clearly defined clinical models that incorporate imaging with clinical data to predict response to NAC. Thus, the aim of this work is to develop a predictive AI model based on routine CT imaging and clinical parameters to predict response to NAC.

## Methods:

The CT scans of 324 patients with NAC from multiple centers in Singapore were used in this study. Four different radiomics models were built for predicting pathological complete response (pCR): first two were based on textural features extracted from peri-tumoral and tumoral regions, the third model based on novel space-resolved radiomics which extract feature maps using voxel-based radiomics and the fourth model based on Deep Learning (DL). Clinical parameters were included to build a final prognostic model.

## Results:

The best performing models were based on space-resolved and deep learning approaches. Space-resolved radiomics improves the clinical AUCs of pCR prediction from 0.743 (0.650 to 0.831) to 0.775 (0.685 to 0.860) and our DL model improved it from 0.743 (0.650 to 0.831) to 0.772 (0.685 to 0.853). The tumoral radiomics model performs the worst with no improvement of the AUC from the clinical model. The peri-tumoral combined model gives moderate performance with an AUC of 0.765 (0.671 to 0.855).

## Conclusions:

Radiomics features extracted from diagnostic CT augments the predictive ability of pathological complete response when combined with clinical features. The novel space-resolved radiomics and deep learning radiomics approaches outperformed conventional radiomics techniques.

# Introduction

Neoadjuvant Chemotherapy (NAC) is currently standard of care for locally advanced breast cancer<sup>1</sup>. NAC allows for down-staging of bulky tumours, allowing previously inoperable tumours to become operable and allowing for Breast Conserving Surgery (BCS)<sup>2,3</sup> where mastectomy was only possible. In addition, it potentially spares axillary clearance in biopsy proven node positive disease that resolves clinically<sup>4</sup>. Relatively high pathological complete response (pCR) rates have been observed in HER2-positive and Triple negative breast cancer(TNBC)<sup>5</sup>. Response to NAC also reflects tumour chemosensitivity and

Loading [MathJax]/jax/output/CommonHTML/jax.js

prognosis with complete primary and nodal pathological response associated with improved event free as well as overall survival<sup>6</sup> and conversely, poor response or progression during NAC can result in delayed surgery, inoperability, greater risk of local progression and distant metastasis.

Aside from selection of breast cancer subtypes more sensitive to chemotherapy and/or targeted therapy, there are no good clinical models to predict NAC response. Use of functional imaging<sup>7</sup> and incorporation of biomarkers<sup>8</sup> have been studied to improve response predictions but validated and accurate clinical models are still lacking.

Radiomics, or the extraction of quantitative radiomic features from medical imaging coupled with machine learning<sup>9-11</sup> or deep learning<sup>12</sup>, may offer a solution. It has been used for other image-based classification problems such as diagnosing pulmonary tuberculosis from chest X-radiographs<sup>13</sup> and diabetic macular oedema from optical coherence tomography images<sup>14</sup>. In predicting response to NAC in BC, promising results have been reported using intra- and peri-tumoural textural analysis of multi parametric MRI (mpMRI) and CT images<sup>15-18</sup>. The authors thus aim to develop an AI model based on CT breast images to predict NAC response in BC.

## Patients And Methods

### Patients

The Centralised Institutional Review Board approved this retrospective study (IRB 2019/4219 Outcomes research in Breast Cancer Care). 361 patients treated with NAC from January 2014 to December 2017 were included in this study. 37 patients were excluded due to missing clinical information or tumours that were difficult to define on CT scans, leaving 324 patients. Of these, 173 patients had their scans performed in National Cancer Centre Singapore (NCCS) and Singapore General Hospital (SGH) while the remaining 151 patients had their scans done in Kangkar Kerbau Women and Children Hospital (KKWCH) and Changi General Hospital (CGH) in Singapore. The former and latter cohort was assigned as the training dataset and testing dataset, respectively. The binary endpoint of interest in this study is pathological complete response which is defined as achieving ypT0N0 after NAC.

All patients were staged according to American Joint Committee on Cancer 7th edition. Histological subtypes of cancers were approximated from the tumour Estrogen receptor (ER), Progesterone receptor (PR) and HER-2 receptor positivity: ER/PR<sup>+</sup> HER2<sup>-</sup> (ER and/or PR positive and HER-2 negative), ER/PR<sup>+</sup> HER2<sup>+</sup> (ER and/or PR positive and HER-2 positive), ER/PR<sup>-</sup> HER2<sup>+</sup> (ER and PR negative, and HER-2 positive) and ER/PR<sup>-</sup> HER2<sup>-</sup> (ER and PR negative, and HER-2 negative). ER and PR staining intensity for nuclear positivity of the cells were graded as 0, 1, 2, and 3 for none, mild, moderate, and strong intensity, respectively. HER-2 positivity is reported as per American Pathologists guideline recommendations.

### Analysis Pipelines

Four different radiomics algorithms were applied on the training datasets. Two of the algorithms rely on conventional radiomics approach applied to different segmentations – tumoral (entire Gross Tumour Volume, GTV) and peri-tumoral regions. This approach can only extract a single radiomics feature vector representative of the entire segmentation and the meaning of this vector can be nebulous when distinct heterogeneous regions exist in it. The third approach was developed by us to overcome this shortcoming. This novel technique makes use of voxel-based based radiomics which extracts radiomics feature from small kernel in the segmentation. The outputs are feature vectors defined for every voxel in the segmentations. Further processing is carried out to condense the high dimensional features into practical feature vector for modelling purpose. This method is known in the manuscript as *space-resolved radiomics* to distinguish it from the conventional radiomics. The last method used deep learning approaches on a fix-sized volume around the centroid of the GTV.

Feature reductions techniques were applied to the radiomics feature vectors extracted from the first three algorithms before being input into the LASSO algorithm<sup>19</sup> for model building. The first feature reduction technique involved removing correlated features with Spearman correlation greater than 0.80. The second feature reduction technique only applied to the first two algorithms and it removes non-robust features according to a modified method by A. Zwanenburg et al<sup>20</sup>. A ten-fold cross validation was used to determine the optimal lambda parameter in the LASSO model. The deep learning radiomics approach directly learnt the model and output the predicted probability from the input image and no features pre-processing are required. These constitute the *radiomics model* in this manuscript. *Clinical model* was constructed using stepping logistic regression with statistically significant covariates under univariate analysis. Lastly, *clinico-radiomics models* were obtained using similar stepping logistic regression method on the similar set of clinical covariates and an additional variable known as *radiomics score*. The radiomics score is calculated by applying a *logit* function to the predicted probability from the radiomics approach.

## Image Acquisition, Segmentations and Radiomics Feature Extraction

The 3D volumetric images in the entire dataset were acquired and reconstructed with 11 different CT scanners. The images with contrast were captured with two different CT scanners located in the centre's radiotherapy department. The first was the GE LightSpeed VCT and the second was Siemens SOMATOM Definition. The CT scanners from external hospitals were made up of primarily GE Revolution and Canon Aquilion model. The images were reconstructed with 90 to 140 kVp X-ray with variable slice thickness of 2.0 mm, 2.5 mm, 3.0 mm and 5.0 mm. The focal spot of the scanner ranges from 0.7 mm to 1.2 mm. The in-slice resolution was 512 by 512 for all images and the in-slice pixel widths are sub-millimeters.

The Gross Tumor Volume (GTV) was contoured by two experienced radiation oncologists (F. Y Wong and W. L Nei) for all the training and testing datasets without knowledge of the patient pathologic outcome. These segmentations were known as the *Intensity Mask* under the IBSI naming convention.

Radiomics features were extracted using pyradiomics v3<sup>21</sup>. It comprised of shape, first order and second order textural features (GLRLM, GLSZM, NGDTM and GLCM). The features are compliant with Image Biomarker Standardization Initiative (IBSI)<sup>22</sup>. CT image was first interpolated with 1 mm isotropic voxel before feature extraction<sup>23,24</sup>. The radiomics features were extracted from the original CT image and the filtered CT image. LoG (Laplacian of Gaussian) filter with sigma 1.0, 2.0 and 3.0 mm and wavelet filters were applied on the image. This gives a total of 1130 radiomics features per CTV and GTV segmentation. A constant bin width of 10 HU was used for textural calculation.

We designed a procedure to select a subset of features from the original 1130 radiomics features which are robust to CT scanner variation and inter-rater variation in GTV contouring. The contouring variability is simulated by performing morphological dilatation and erosion operations of up to 2.0 mm on the segmentation (this generated four additional structures per patient). The CT scanner variability is simulated by adding an independent Gaussian, Poisson and Uniform noise to the CT image. The mean of the Gaussian, Poisson and Uniform noise distributions is 20 HU and the standard deviation of the Gaussian and Uniform distribution is 20 HU as well. These parameters were estimated from a phantom study on the two CT machines in NCCS. The process of adding the noise to the original CT image was repeated five times for each distribution, resulting in fifteen different CT image for each patient. Overall, the morphological operations and noise addition generated 19 additional set radiomics features per patient. Details are shown in the Fig. 3. Intraclass Correlation Coefficients (ICC) with a threshold of 0.600 is used to select the robust features.

## Tumoral and Peri-tumoral Radiomics

Tumoral radiomics features are features extracted directly from the GTV contoured by the physician, and represents the textural, shape and intensity descriptors of the entire tumor core. The peri-tumoral regions are obtained by extracting an annulus ring segment from the boolean subtraction of GTV with 1 mm dilatation and 1 mm erosion. These features describe the gray level characteristic of the peripheral of the tumor.

## Space-resolved Radiomics

Voxel-based radiomics are first used to extract features heat map from the GTV segmentation. Voxel-based extraction involves defining a kernel size (5 voxels isotropically in this work) and calculate the radiomics features (first order and textural) for each kernel centered on every voxel within the GTV. The choice of the kernel size can be interpreted as the degree of locality the texture calculation is sensitive to; a small kernel size could result in a noisy texture map while a large kernel size will lose the local information and become increasingly smoother. A total of 88 first order, GLCM, GLRLM, GLSZM and NGDTM feature maps are calculated for each GTV.

It is impractical to use all the feature maps directly for modeling; thus, the second step involves condensing all the information in the 88 feature maps into a fixed size feature vector for each GTV. The following procedure is applied consistently to all the 88 feature maps. The feature values for all the

Loading [MathJax]/jax/output/CommonHTML/jax.js

voxels in the entire training dataset are combined together and are used to generate the bin edges values by constructing a histogram with 20 equal-sized bins. Feature values beyond the 2 and 98 percentiles are excluded while calculating the bin edges of the histogram to prevent the influence of outliers. The bin edges of the histogram are then used to bin all the feature values of all the voxels in each GTV and the percentage number of voxels in the 20 bins are used as the *feature vectors* to represent the particular feature map.

Space-resolved radiomics has another advantage over conventional radiomics approach in that it can show which are the voxels within the segmentations that are driving the prediction model. The significant features in the model are first identified and voxels within the segmentation with feature value lying within the bin can be highlighted.

## Deep Learning Radiomics

A fixed size box of dimensions 64 x 64 x 32 voxels centered at the centroid of the GTV is extracted and serves as the input image to the Deep Learning algorithm. Data augmentations consisting of a random translational shift of 1 to 5 voxels (uniform distribution) is applied to the box in all three directions, random 90 degree rotations about the z axis and 180 degree rotations about the x and y axes were applied to the box. MultiResUnet3D architecture<sup>25</sup> was selected for this work and the network was trained from scratch using Tensorflow framework. Adam optimizer was used with a learning rate of  $1.0 \times 10^{-4}$  with exponential decay. Several modifications were made to the open source architecture. Firstly, *He normal* weight initialization was used and a L2 weight regularization of  $1.0 \times 10^{-7}$  was applied to all Conv3D layer. Secondly, the last Conv3D was removed and replaced with a Flatten layer follow by a 128-units fully connected layer with reLu activation function and a 1-unit fully connected layer with sigmoid activation unit. A 0.50 and 0.20 dropout were applied to the first and second fully connected layer respectively. This architecture modification is necessary because the original architecture was meant for segmentation task with probability output per voxel while we are working on binary classification task with a single probability output. The 173 patients are divided into 120 training data and 53 validation data. The model with the highest validation AUC is applied to the testing data. The testing datasets are similar across all the different algorithms. Gradient-weighted class activation map (Grad-CAM)<sup>26</sup> are calculated at the encoded layer of the network to highlight the important regions in the image contributing to the pCR prediction.

## Assessing performance of models

The performance of the four algorithms were evaluated on the test datasets. For each algorithm, three different models comprising of clinical, radiomics and clinico-radiomics are generated for comparison. Two performance metrics are used to compare between algorithms and determine if the radiomics model value-add to the clinical model. The first performance metric is the AUC which is calculated from the ROC curve and serves as a measure of the discriminatory ability of the algorithm. The second performance metric is the calibration curve which measures how closely the predicted probability matches the observed one and Brier's score is calculated to quantify the degree of departure from logistic regression

model. The observed probabilities are partitioned into ten groups for the calculation of calibration curve. A lower Brier's score indicates better agreement between the predicted and observed probability.

## Statistical Analysis

All statistical tests were performed using R statistical software (version 3.4.2; R Foundation for Statistical Computing). A two-sided  $P$ -value  $< 0.05$  was considered significant.

## Results

### Patient Clinical Characteristics and analysis pipelines

A total of 361 patients met the inclusion criteria and were included in this study. 37 patients were excluded from the study, leaving 324. Of the 37 patients, 23 have missing clinical information while 13 patients had tumours that were difficult to define on CT scans (such as diffuse tumours or multi-focal tumours). The baseline characteristics of the patients in the training and testing datasets are shown in Table 1. The patient characteristics distributions were similar across the two datasets. The number of pCR events are 21.4% and 24.5% in the training and testing dataset respectively. The schematic of the training and testing dataset split is shown in Fig. 1A. Univariate analysis was also performed, and the  $P$ -values are shown in the table. HER2, ER, PR and histological subtypes were statistically significant under the analysis. Information on the chemotherapy drug usage in this study cohort is shown in **Supplementary Table 1**. The chemotherapy usage was evenly distributed across the two datasets with frequent usage of cyclophosphamide, taxanes and anthracyclines. No precise chemotherapy regimen information was known for 7 (0.9%) patients. More than 90% of patients received at least third generation chemotherapy, namely taxanes in addition to an anthracycline based regimen. Platinum-based chemotherapy was given to 52.5% of TN breast cancer patients. In Her-2 positive patients, 94.2% received trastuzumab. Of these, 24.5% also received pertuzumab.

Table 1  
Baseline characteristics of the patients.

	No. Of Patients	Training Set		Testing Set
		No. (%)	P-value	No. (%)
Total No. of Patients	324	173		151
<b>Clinical diagnosis</b>				
Age, years	324	53.8 (10.4)	0.971	55.6 (11.1)
cT Stage			0.206	
T1	4	1 (0.58)		3 (1.99)
T2	109	56 (32.4)		53 (35.1)
T3	97	57 (32.9)		40 (26.5)
T4	110	56 (32.4)		54 (35.8)
Missing	4	3 (1.73)		1 (0.66)
cN Stage			0.306	
N0	57	31 (17.9)		26 (17.2)
N1	296	108 (62.4)		88 (58.3)
N2	31	13 (7.51)		18 (11.9)
N3	36	18 (10.4)		18 (11.9)
Missing	4	3 (1.73)		1 (0.66)
HER2			< 0.001	
Negative	200	108 (62.4)		92 (60.9)
Positive	124	65 (37.6)		59 (39.1)
ER			< 0.001	
Negative	138	82 (47.4)		56 (37.1)
Positive	186	91 (52.6)		95 (62.9)
PR			< 0.001	
Negative	189	107 (61.8)		82 (54.3)
Positive	135	66 (38.2)		69 (45.7)

	No. Of Patients	Training Set	Testing Set
Histological Subtype			< 0.001
Luminal A	123	61 (35.3)	62 (41.1)
Luminal B	71	36 (20.8)	35 (23.2)
HER2 enriched	52	30 (17.3)	22 (14.6)
Basal	78	46 (26.6)	32 (21.2)
<b>Scanner</b>			
kVp			0.700
90 to 110	62	47 (27.2)	15 (9.93)
120 to 140	262	126 (72.8)	136 (90.1)
<b>pCR</b>			
No	250	136 (78.6)	114 (75.5)
Yes	74	37 (21.4)	37 (24.5)
All the data are categorical in nature and are shown as the number (percentage).			

**Abbreviations:** ER, Estrogen receptors; PR, Progesterone Receptor; Her2 +ve, human epidermal growth factor receptor 2.

The overview of the analysis pipeline is shown in Fig. 1B. Three of the four radiomics approach requires significant feature pre-processing (scaling and selection) before being input into the model building pipeline, while the deep learning approach directly output the predicted probability of pCR. The combined model is then compared primarily with the clinical model to determine if radiomics features augment the predictive power clinical model.

## Clinical Model

Four clinical covariates - HER2, ER, PR and Molecular subtypes are input into the stepping logistic equation and the resulting clinical model contains only two significant covariate – ER and HER2 with a coefficient of -1.37 (-1.79 to -0.95) and 1.31 (0.91 to 1.71) respectively.

## Image Acquisition, Segmentations and Radiomics Feature Extraction

Siemen Somatom Definition and GE LightSpeed VCT makes up 68.2 and 31.8 % of the CT machines in the training dataset (from our centre). Canon Aquilion and GE Revolution makes up 35.7 and 19.9% of the CT machine in the testing dataset. These two forms the largest bulk of the CT machines. The remaining

CT machines in the testing dataset comprise of Philips iCT, Philips Mx8000 Quad, Siemens Somatom Definition and Siemens Sensation Cardiac.

Examples of the segmentations of the tumoral and peri-tumoral regions are shown in the Fig. 2. The purple shaded regions in the images in the first column are the GTVs contoured by our radiation oncologists. The purple ring in the second column are the peri-tumoral regions extracted using morphological operation. A single feature heatmap obtained from voxel-based radiomics is shown in the third column for comparison to show that our method is inherently different from the conventional radiomics approach.

## Tumoral and Peri-tumoral Radiomics Model

The overview of the robust feature selection procedure is shown in Fig. 3. Noise addition to the CT image and morphological erosion and dilatation operations on the segmentations are shown in the figure. ICCs are calculated for 19 sets of radiomics features extracted from the GTVs (15 from random noise addition and 4 sets from morphological operations) and a robust feature is defined by having ICC value greater than 0.600. The plots of the right show the robust features as being yellow color coded while the non-robust features are blue color coded. First order, GLCM, GLRLM and GLDM robust features are shown in the figure as an example. In practice, the ICCs are calculated for all 1130 radiomics features. A total of 367 features are determined to be robust by this procedure. These robust features are used for modelling (after removing correlated features) in tumoral and peri-tumoral radiomics.

The significant radiomics features and its associated coefficients in the radiomics model are shown in Table 2. The final clinico-radiomics tumoral model consists of ER, HER2 and radiomics score with coefficients - 1.59 (-3.60 to 0.44), 0.98 (0.43 to 2.29) and 1.02 (0.31 to 3.35) respectively. Similarly, the peri-tumoral model have coefficients of -1.56 (-3.44 to 0.45), 0.94 (0.44 to 2.15) and 1.15 (0.29 to 4.00) for ER, HER2 and radiomics score respectively. All the coefficients are statistically significant with *P-value* < 0.05.

## Space-resolved Radiomics Model

The comparison between the conventional radiomics approach and our novel space-resolved radiomics is illustrated in Fig. 4. The main difference lies in the radiomics feature extraction process; conventional radiomics extract only a single feature vector for each segmentation (shown as  $\vec{f}$  in the left blue box) whereas our approach relies on voxel-based radiomics where the feature vector,  $\vec{f}$ , is extract for each kernel. The output of voxel-based radiomics are radiomics feature heatmap as shown in the second step of the spaced-resolved radiomics pipeline in Fig. 4. The last step of the space-resolved radiomics pipeline involves binning the feature values in each feature heatmap into histogram; all the histograms from the 88 feature heatmaps are concatenated together to form a single vector before being input into the model.

Table 2. Coefficients of LASSO for tumoral, peri-tumoral and space-resolved radiomics model

Loading [MathJax]/jax/output/CommonHTML/jax.js

Tumoral Radiomics			Peri-tumoral Radiomics		Space-resolved Radiomics	
	Feature	Coefficients	Feature	Coefficients	Feature	Coefficients
1	original_glcM_MCC	-0.08689029	original_glcM_Correlation	-0.1231956	original_firstorder_RootMeanSquared_B9	-0.08462663
2	wavelet_LLH_ngtdm_Busyness	-0.2418746	original_glcM_MCC	-0.02719496	original_gldm_DependenceVariance_B2	0.03582108
3	wavelet_HHL_firstorder_RootMeanSquared	0.1559246	wavelet_LLH_glszm_LargeAreaLowGrayLevelEmphasis	-0.1038756	original_gldm_DependenceVariance_B7	-0.08454196
4	wavelet_HHL_glszm_GrayLevelNonUniformityNormalized	-0.01081828	wavelet_HLH_firstorder_RootMeanSquared	0.111683	original_gldm_LargeDependenceLowGrayLevelEmphasis_B9	0.01070056
5			wavelet_HLH_ngtdm_Busyness	-0.04837576	original_glrIm_ShortRunLowGrayLevelEmphasis_B2	-0.1247833
6			wavelet_HHL_firstorder_RootMeanSquared	0.03228319	original_glrIm_ShortRunLowGrayLevelEmphasis_B8	-0.04928155
7			log.sigma.1.0.mm.3D_glcM_MaximumProbability	-0.04103622	original_glrIm_ShortRunLowGrayLevelEmphasis_B9	-0.09386563
8			log.sigma.2.0.mm.3D_ngtdm_Busyness	-0.06626755		
9			log.sigma.3.0.mm.3D_glszm_SizeZoneNonUniformityNormalized	-0.2150385		

The significant radiomics features and its associated coefficients in the radiomics model are shown in Table 2. The suffix represents the bin number of the histogram. ie. Original\_firstorder\_RootMeanSquared\_B9 means the ninth bin in the original\_firstorder\_RootMeanSquared histogram. The final clinico-radiomics space-resolved radiomics model consists of ER, HER2 and radiomics score with coefficients - 1.70 (-3.85 to 0.44), 1.13 (0.43 to 2.64) and 1.96 (0.75 to 2.62) respectively.

The highest coefficient features in the model are selected and the voxels represented by each feature is shown in Fig. 5. The left and right images show six different true pCR and true non-pCR respectively. The top, middle and bottom rows show voxels with original\_glrIm\_ShortRunLowGrayLevelEmphasis lying within the second, eighth and ninth bin, original\_gldm\_DependenceVariance lying within the second and seventh bins and original\_firstorder\_RootMeanSquared lying within the ninth bin respectively. The voxels driving the prediction for each feature are shown in purple.

## Deep Learning Radiomics Model

An overview of the deep learning radiomics pipeline is shown in Fig. 6A. A fixed size bounding volume around the centroid of the tumor is used as an input into the deep learning architecture. The model was trained for 120 epochs and the highest validation AUC model was used as the final deep learning radiomics model. The class activation maps of a true positive and true negative example calculated in the axial slice of the tumour centroid are shown in Fig. 6B. The map shows that the tumour contributes significantly to the explanation of pCR in the deep learning model. The highest achieved validation AUC is 0.746 and it happens at the fifth epoch.

## Assessing Performance of Models

The results of the ROC curve comparisons between the four algorithms and between clinical, radiomics and clinico-radiomics models are shown in Fig. 7 (for train data) and Fig. 8 (for test data). The blue,

yellow and green lines represent the clinical, radiomics and clinico-radiomics model respectively. A higher AUCs are obtained for all the combined models compared to the clinical models in the train data. The greatest increase is observed in the space-resolved model followed by the peri-tumoral model. Looking at the test data results in Fig. 8, only the peri-tumoral, space-resolved and deep learning radiomics approach are able to enhance the AUC of clinical model. Similar to the train data, the highest enhancement is achieved in space-resolved radiomics where the clinical model AUC increases from 0.743 (0.650 to 0.831) to 0.775 (0.685 to 0.860) by including the radiomics score. The deep learning model achieves similar magnitude enhancement of 0.772 (0.685 to 0.853). The ROC curves comparison of all the four clinico-radiomics models is shown in Fig. 9. Tumoral radiomics model clearly under-performs compare to the rest of the models.

ROC curves are also calculated for all the histological subtypes in the test data using the four radiomics models. The results are shown in **Supplementary Table 2**. The performances of the models are generally non-uniform across the subtypes. Most of the models are able to enhance the AUCs of the clinical models except for the tumoral radiomics model which results in a lower combined AUCs than clinical models in ER/PR<sup>+</sup>HER2<sup>-</sup>, ER/PR<sup>+</sup>/HER2<sup>+</sup> and ER/PR<sup>-</sup>HER2<sup>-</sup> subtypes. Space-resolved and deep learning radiomics models perform the best in the Basal subtypes (with highest combined AUCs) while the tumoral radiomics model performs the best in ER/PR<sup>-</sup>HER2<sup>+</sup> subtype.

The compilation of the calibration curve results of all the four models is shown in Fig. 10. The blue, yellow and green lines represent the clinical, radiomics and clinico-radiomics model respectively. The Brier score of the individual model is shown in the parenthesis in the legend. Similar to the ROC curve results, only peri-tumoral, space-resolved and deep learning models show improvement in the Brier's score from pure clinical model alone.

## Discussion

Our results suggest that the addition of radiomic features to clinical data can improve the prediction of pCR after NAC in BC. A key advantage of this approach is that it uses CT imaging which is commonly available and we have tested this in an external validation cohort. Apart from the tumoral radiomics model, the peri-tumoral, space-resolved and DL models show improvement on the clinical model. This suggests that heterogeneity within certain parts of the tumour, in particular, the peri-tumoral ring, may yield more information about the tumour than analyzing the whole tumour alone. This is consistent with other studies<sup>17,26</sup>. Braman et al<sup>17</sup> showed that a combination of intra and peritumoral radiomics was able to give a good prediction pCR in MRI images. Features from the 0- to 3-mm peritumoral region were significantly associated with the density of tumor-infiltrating lymphocytes. Similar conclusion on the importance of peritumoral region is found in a separate radiomics study conducted by Sun et al<sup>27</sup>. The author found that CT radiomic signatures extracted from 4mm peritumoral ring were correlated with CD8<sup>+</sup> T-cells infiltration and response to immunotherapy.

Our novel spatial radiomics method improves on the peri-tumoral ring method which extracts only a single feature vector for each segmentation (shown as  $\vec{f}$  in the left blue box in Fig. 4). Our approach relies on voxel-based radiomics where the feature vector is extracted for each kernel (as shown in the red box of Fig. 4). The improved discriminatory ability could be attributed to the fact that the proposed method extracts more information from the lesion and better characterizes the tumoral heterogeneity. Furthermore, this method is capable of outputting an equivalence of “class activation map”<sup>28</sup> which shows the region within the segment driving the prediction (**shown in** Fig. 5). In fact, from Fig. 5, one can observe the difference in patterns and regions driving the prediction between pCR and non pCR patients.

Currently, to the authors knowledge there are no DL models developed yet that can predict NAC response in BC. While traditional radiomics depend on thousands of user-defined features, DL radiomics is not constrained with this pre-determined set of features but uses a convolutional neural network to learn relevant radiomics features for its task. Our result shows the DL radiomics approach outperforms conventional radiomics approach (both peri-tumoral and tumoral), and interestingly, matched our space-resolved radiomics method in terms of AUC performance. Andre damant et al<sup>29</sup> has similarly applied this DL radiomics approach to predict outcomes in head and neck cancer and found that it outperformed the traditional radiomics prediction.

A common difficulty in the application of these radiomics model has been its generalizability. Different scanners and scanning parameters can give vastly different results and there is a potential of overfitting on the training data and generalizing to out-of-class distribution<sup>30,31</sup>. We mitigated the problem of overfitting by using two approaches. First, our model has been trained and tested on CT images retrieved from various machines and hospitals. Second, we included a robustness pipeline in our conventional radiomics approach to select our features that are stable to noise and contour perturbations. An advantage of using CT imaging is that they are usually done as part of the staging workup for breast cancer patients and our models could potentially be easily integrated into current workflows. CT scanning protocols and parameters are also relatively more uniform across centers than compared to MRI. Following the definition of radiomics quality score (RQS) defined by P. Lambin et al<sup>32</sup>, our assessment of this work has a RQS of 18 out of 36<sup>33</sup>.

Despite the encouraging results from our study, the main limitation of our study is the retrospective nature of the study and the relatively small cohort size. Further validation of our model in a larger cohort and in prospective studies is being planned.

## Conclusion

Peri-tumoral, space-resolved and DL CT radiomics features when combined with clinical features improve the prediction of pCR after NAC in BC. Conventional radiomics under-perform our proposed space-resolved radiomics and deep learning radiomics approaches. The new proposed radiomics technique is a

promising way to characterize tumor heterogeneity compared to conventional textural radiomics and could be used to develop better predictive and prognostic models.

## Declarations

### Ethics approval and consent to participate

The Centralised Institutional Review Board approved this retrospective study (IRB 2019/4219 Outcomes research in Breast Cancer Care). Waiver of consent was obtained.

### Consent for publication

Not applicable.

### Availability of data and materials

The datasets generated and/or analysed during the current study are not publicly available due data privacy and intellectual property issues but are available from the corresponding author on reasonable request.

### Competing interests

The authors declare that they have no competing interests.

### Funding

W.L.N is supported by the National Medical Research Council Fellowship (NMRC/ MOH-000166-00).

### Authors' contributions

T.H.Q., O.H.K. have led on the study, written the initial draft, and created the figures with the input of the senior authors W.F.Y., C.Y. and W.L.N. T.H.Q. and O.H.K. contributed equally as co-first authors. W.F.Y., C.Y. and W.L.N., contributed equally as senior authors. The following authors helped writing the manuscript, and approved the final version: T.H.Q., O.H.K., A.M.K., W.F.Y., C.Y. and W.L.N. The following authors provided data from the different centers, and approved the final version: T.T.J.Y., R.S.T.Y.C., G.L.S., E.L.H., R.N.C.H., R.Y.M.C., F.L.L.W.T., L.Y.G., Z.Z., C.Y.S.H., W.R.X., G.O.S.K, L.L.C.H., J.W.S.L., T.S.M., M.P., S.Y., V.T.K.M., J.Y., W.F.Y. The following authors analyzed the data and/or helped with the interpretation of data, helped writing the manuscript, and approved the final version: T.H.Q., O.H.K., A.M.K., W.F.Y., C.Y. and W.L.N.; W.F.Y., C.Y. and W.L.N., supervised the research, provided funding, and helped in writing the manuscript.

### Acknowledgement

Ong Hiok Kian would like to express his appreciation to The URECA program with Nanyang Technological University. Loading [MathJax]/jax/output/CommonHTML/jax.js would like to express our gratitude to Phyu-Nitar, Yeo Sook

Kwan and Siti Nurhidayah Binte Kamaludin for their help with the study. W.L.N is supported by the National Medical Research Council Fellowship (NMRC/ MOH-000166-00). We would like to acknowledge Joint Breast Cancer Registry (JBCR) and Artificial Intelligence, Innovation, Imaging and datascience laboratory (Ai3) members for their support.

JBCR members: Amit Jain; Benita Tan Kiat Tee; Bryan Ho Shihan; Caleb Ho ; Chay Wen Yee ; Chong Jun Hua ; Chua Eu Tiong ; Chua Hui Wen ; Ding Zee Pin ; Elaine Lim Hsuen ; Eliza Sin I-Lin ; Eric Andrew Finkelstein ; Ewe See Hooi ; Faye Lynette Lim Wei Tching ; Gail Chua Wan Ying ; Ghislaine Lee Su-Xin ; Grace Kusumawidjaja ; Gudi Mihir Ananta ; Hanis Mariyah Binte Mohd Ishak ; Jack Chan Junjie ; Janice Tan Ser Huey ; Jill Wong Su Lin ; Joanne Ngeow Yuen Yie ; Joe Yeong Poh Sheng ; Johan Chan ; Joshua Lim Sheng Hao ; Joycelyn Lee Jie Xin ; Julie Liana Bte Hamzah ; Kiley Loh Wei-Jen ; Koo Si-Lin ; Lee Jung Ah ; Lester Leong Chee Hao ; Lim Geok Hoon ; Lim Sheng An ; Lim Sue Zann ; Lim Swee Ho ; Lita Chew Sui Tjien ; Lloyd Tan Kuan Rui ; Mabel Wong ; Melvin Chew Ming Long ; Natalie Koh Si Ya ; Nei Wen Long ; Nelson Yit Ling Fung; Ng Choon Ta ; Ng Wee Loon ; Phyu Nitar ; Preetha Madhukumar ; Rachel Sim Hui Zhen ; Raymond Ng Chee Hui ; Rayson Lee Rui Sheng ; Rebecca Dent ; Richard Yeo Ming Chert ; Ryan Heng ; Ryan Shea Tan Ying Cong ; Sabrina Ngaserin Ng Hui Na ; Sharon Poh Shuxian ; Sheriff Quek Zhan Hong ; Sim Yirong ; Tan Hong Qi ; Tan Puay Hoon ; Tan Qing Ting ; Tan Si Ying ; Tan Su-Ming ; Tan Yongcheng ; Tira Tan Jing Ying ; Veronique Tan Kiak Mien ; Vivianne Shih Lee Chuen ; Wee Hwee Lin ; Wong Chow Yin ; Wong Fuh Yong ; Wong Ru Xin ; Yap Yoon Sim ; Yeo Khung Keong ; Yong Wei Sean ; Zhang Zewen ; Zubin Master

Ai3 Lab members: Wen Long Nei, Tan Hong Qi, Eric Pang Pei Ping, Bryan Ho Shihan, Jeffrey Tuan Kit Loong

## References

1. Sun Y, Liao M, He L, Zhu C. Comparison of breast-conserving surgery with mastectomy in locally advanced breast cancer after good response to neoadjuvant chemotherapy. *Med (United States)*. Published online 2017. doi:10.1097/MD.00000000000008367
2. Asselain B, Barlow W, Bartlett J, et al. Long-term outcomes for neoadjuvant versus adjuvant chemotherapy in early breast cancer: meta-analysis of individual patient data from ten randomised trials. *Lancet Oncol*. Published online 2018. doi:10.1016/S1470-2045(17)30777-5
3. Killelea BK, Yang VQ, Mougalian S, et al. Neoadjuvant chemotherapy for breast cancer increases the rate of breast conservation: Results from the national cancer database. *J Am Coll Surg*. Published online 2015. doi:10.1016/j.jamcollsurg.2015.02.011
4. King TA, Morrow M. Surgical issues in patients with breast cancer receiving neoadjuvant chemotherapy. *Nat Rev Clin Oncol*. Published online 2015. doi:10.1038/nrclinonc.2015.63

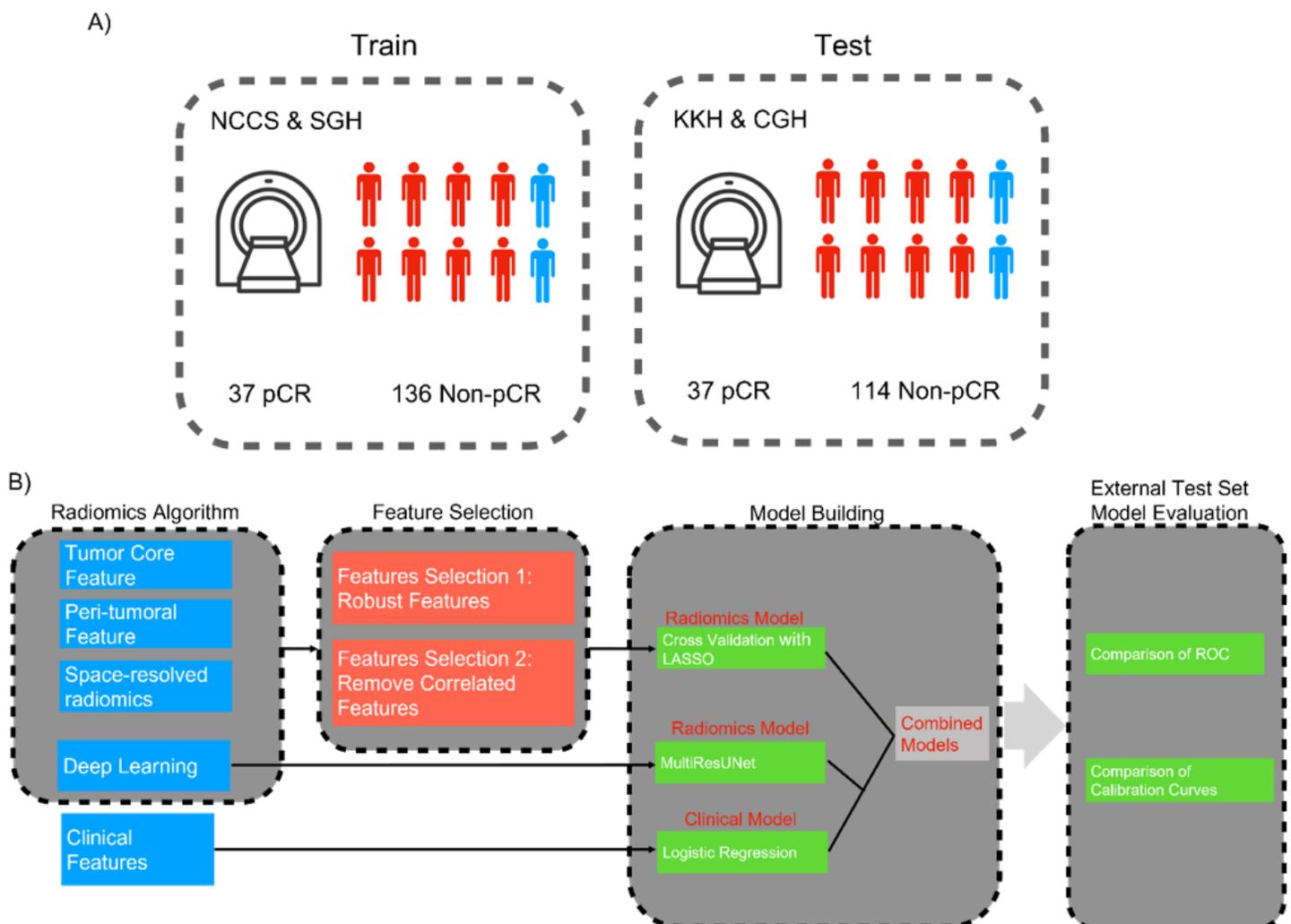
5. Houssami N, MacAskill P, Von Minckwitz G, Marinovich ML, Mamounas E. Meta-analysis of the association of breast cancer subtype and pathologic complete response to neoadjuvant chemotherapy. *Eur J Cancer*. Published online 2012. doi:10.1016/j.ejca.2012.05.023
6. Cortazar P, Zhang L, Untch M, et al. Pathological complete response and long-term clinical benefit in breast cancer: The CTNeoBC pooled analysis. *Lancet*. Published online 2014. doi:10.1016/S0140-6736(13)62422-8
7. Fowler AM, Mankoff DA, Joe BN. Imaging neoadjuvant therapy response in breast cancer. *Radiology*. Published online 2017. doi:10.1148/radiol.2017170180
8. Barker AD, Sigman CC, Kelloff GJ, Hylton NM, Berry DA, Esserman LJ. I-SPY 2: An adaptive breast cancer trial design in the setting of neoadjuvant chemotherapy. *Clin Pharmacol Ther*. Published online 2009. doi:10.1038/clpt.2009.68
9. Peeken JC, Bernhofer M, Wiestler B, et al. Radiomics in radiooncology – Challenging the medical physicist. *Phys Medica*. 2018;48:27-36. doi:10.1016/j.ejmp.2018.03.012
10. Gillies RJ, Kinahan PE, Hricak H. Radiomics: Images are more than pictures, they are data. *Radiology*. 2016;278(2):563-577. doi:10.1148/radiol.2015151169
11. Afshar P, Mohammadi A, Plataniotis KN, Oikonomou A, Benali H. From hand-crafted to deep learning-based cancer radiomics: Challenges and opportunities. *arXiv*. Published online 2018.
12. Lecun Y, Bengio Y, Hinton G. Deep learning. *Nature*. Published online 2015. doi:10.1038/nature14539
13. Ting DSW, Yi PH, Hui F. Clinical applicability of deep learning system in detecting tuberculosis with chest radiography. *Radiology*. Published online 2018. doi:10.1148/radiol.2017172407
14. Lee CS, Tying AJ, Deruyter NP, Wu Y, Rokem A, Lee AY. Deep-learning based, automated segmentation of macular edema in optical coherence tomography. *Biomed Opt Express*. Published online 2017. doi:10.1364/boe.8.003440
15. Tahmassebi A, Wengert GJ, Helbich TH, 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. *Invest Radiol*. Published online 2019. doi:10.1097/RLI.0000000000000518
16. Eun NL, Kang D, Son EJ, et al. Texture analysis with 3.0-T MRI for association of response to neoadjuvant chemotherapy in breast cancer. *Radiology*. Published online 2020. doi:10.1148/radiol.2019182718
17. Braman NM, Etesami M, Prasanna P, et al. Intratumoral and peritumoral radiomics for the complete response to neoadjuvant chemotherapy based on

breast DCE-MRI. *Breast Cancer Res*. Published online 2017. doi:10.1186/s13058-017-0846-1

18. Granzier RWY, van Nijnatten TJA, Woodruff HC, Smidt ML, Lobbes MBI. Exploring breast cancer response prediction to neoadjuvant systemic therapy using MRI-based radiomics: A systematic review. *Eur J Radiol*. Published online 2019. doi:10.1016/j.ejrad.2019.108736
19. Tibshirani R. Regression Shrinkage and Selection Via the Lasso. *J R Stat Soc Ser B*. Published online 1996. doi:10.1111/j.2517-6161.1996.tb02080.x
20. Zwanenburg A, Leger S, Agolli L, et al. Assessing robustness of radiomic features by image perturbation. *Sci Rep*. Published online 2019. doi:10.1038/s41598-018-36938-4
21. Van Griethuysen JJM, Fedorov A, Parmar C, et al. Computational radiomics system to decode the radiographic phenotype. *Cancer Res*. Published online 2017. doi:10.1158/0008-5472.CAN-17-0339
22. Zwanenburg A, Vallières M, Abdalah MA, et al. The image biomarker standardization initiative: Standardized quantitative radiomics for high-throughput image-based phenotyping. *Radiology*. Published online 2020. doi:10.1148/radiol.2020191145
23. Mackin D, Fave X, Zhang L, et al. Harmonizing the pixel size in retrospective computed tomography radiomics studies. *PLoS One*. 2017;12(9):e0178524. doi:10.1371/journal.pone.0178524
24. Shafiq-Ul-Hassan M, Latifi K, Zhang G, Ullah G, Gillies R, Moros E. Voxel size and gray level normalization of CT radiomic features in lung cancer. *Sci Rep*. 2018;8(1):10545. doi:10.1038/s41598-018-28895-9
25. Ibtehaz N, Rahman MS. MultiResUNet: Rethinking the U-Net architecture for multimodal biomedical image segmentation. *Neural Networks*. Published online 2020. doi:10.1016/j.neunet.2019.08.025
26. Braman N, Prasanna P, Whitney J, et al. Association of Peritumoral Radiomics With Tumor Biology and Pathologic Response to Preoperative Targeted Therapy for HER2 (ERBB2)-Positive Breast Cancer. *JAMA Netw open*. Published online 2019. doi:10.1001/jamanetworkopen.2019.2561
27. Sun R, Limkin EJ, Vakalopoulou M, et al. A radiomics approach to assess tumour-infiltrating CD8 cells and response to anti-PD-1 or anti-PD-L1 immunotherapy: an imaging biomarker, retrospective multicohort study. *Lancet Oncol*. Published online 2018. doi:10.1016/S1470-2045(18)30413-3
28. Selvaraju RR, Cogswell M, Das A, Vedantam R, Parikh D, Batra D. Grad-CAM: Visual Explanations from Deep Networks via Gradient-Based Localization. *Int J Comput Vis*. Published online 2020. doi:10.1007/s11263-019-01228-7
29. Diamant A, Chatterjee A, Vallières M, Shenouda G, Seuntjens J. Deep learning in head & neck cancer outcome prediction. *Sci Rep*. Published online 2019. doi:10.1038/s41598-019-39206-1

30. Reuzé S, Orhac F, Chargari C, et al. Prediction of cervical cancer recurrence using textural features extracted from 18F-FDG PET images acquired with different scanners. *Oncotarget*. Published online 2017. doi:10.18632/oncotarget.17856
31. Reuzé S, Schernberg A, Orhac F, et al. Radiomics in Nuclear Medicine Applied to Radiation Therapy: Methods, Pitfalls, and Challenges. *Int J Radiat Oncol Biol Phys*. Published online 2018. doi:10.1016/j.ijrobp.2018.05.022
32. Lambin P, Leijenaar RTH, Deist TM, et al. Radiomics: The bridge between medical imaging and personalized medicine. *Nat Rev Clin Oncol*. Published online 2017. doi:10.1038/nrclinonc.2017.141
33. Sanduleanu S, Woodruff HC, de Jong EEC, et al. Tracking tumor biology with radiomics: A systematic review utilizing a radiomics quality score. *Radiother Oncol*. Published online 2018. doi:10.1016/j.radonc.2018.03.033

## Figures



The patient cohort and the analysis pipeline A) This figure shows the patient cohort in the train and test datasets. The pCR event rate is about similar in both datasets. (NCCS: National Cancer Center Singapore; SGH: Singapore General Hospital; KKH: Kandang Kerbau Hospital ; CGH: Changi General Hospital) B) This shows the entire analysis pipeline used for building the machine learning and deep learning model. Four different radiomics algorithm comprising of textural-based features and deep learning features are used for building the radiomics model.

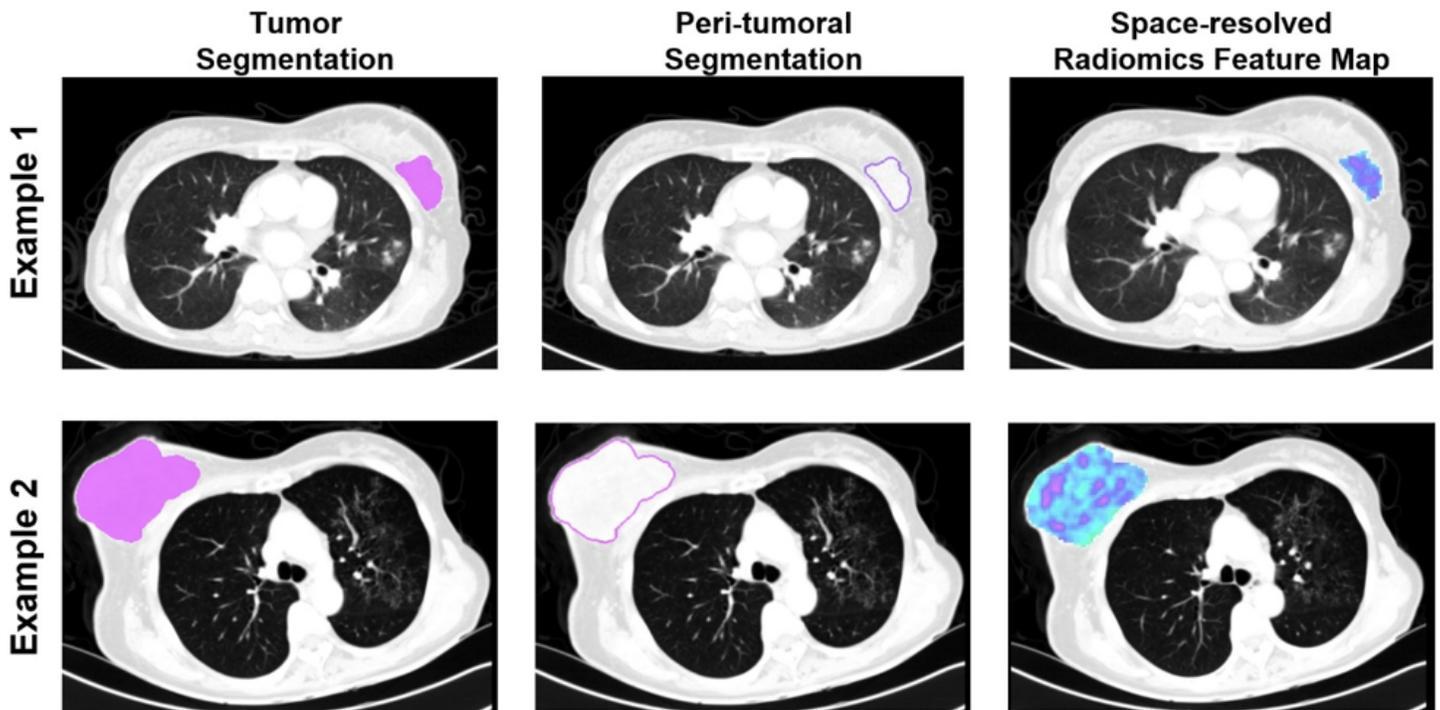
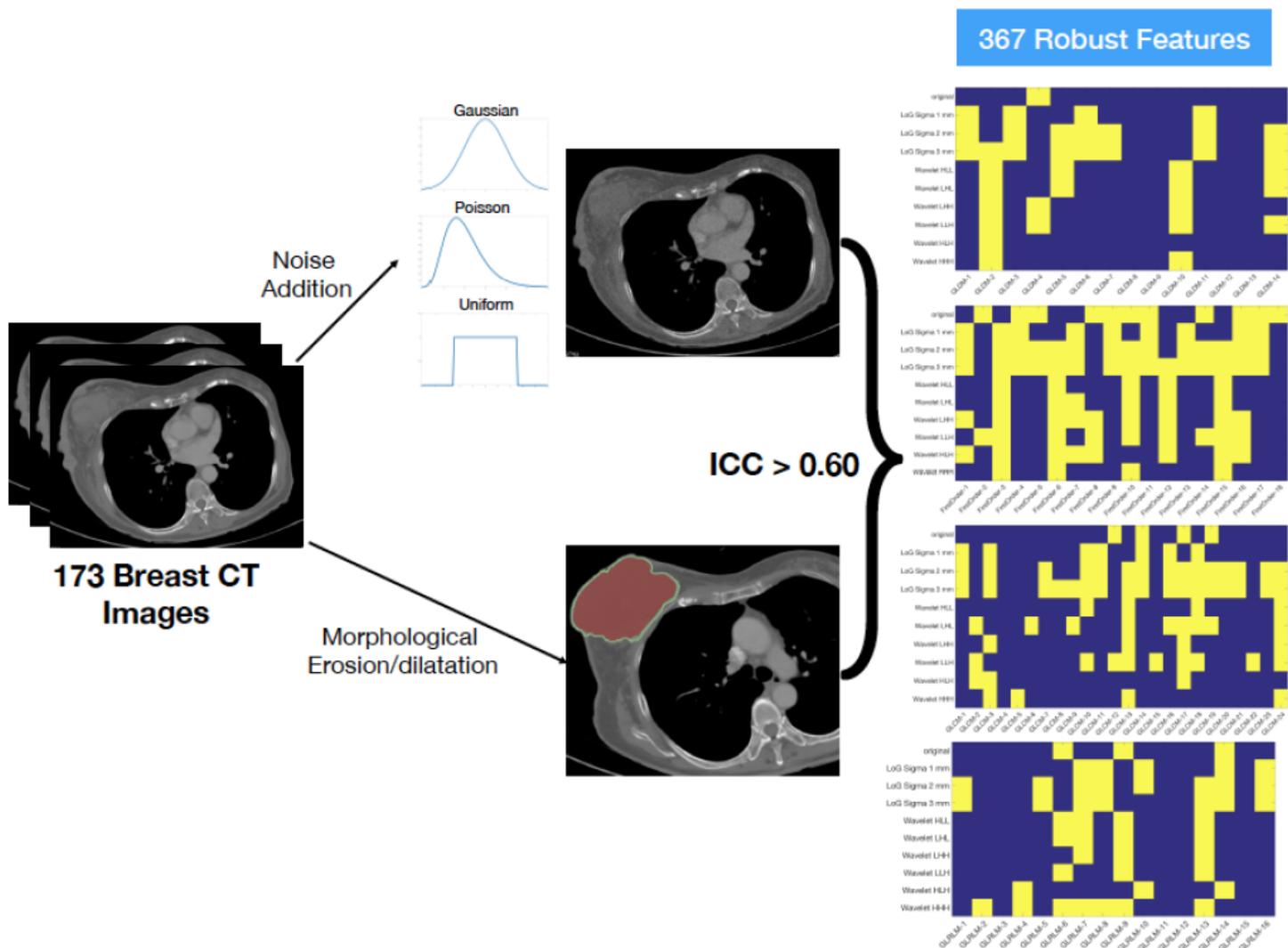


Figure 2

Examples of segmentations in tumoral and peri-tumoral radiomics and the texture map obtained from voxel-based radiomics. This figure shows two examples of tumoral and per-tumoral segmentations. The peri-tumoral segmentations are obtained from Boolean operation on the original tumoral segmentations. Examples of feature maps obtained from voxel-based radiomics are shown in the last column.



**Figure 3**

Pipeline for extracting robust radiomics features. This figure shows the pipeline for extracting robust radiomics features from the original set of features. Radiomics features are extracted from new CT sets (with Gaussian, Poisson and Uniform noise added) and new dilated and eroded segments to simulate scanner noise and contour variation between physicians. ICC (Intraclass correlation coefficient) threshold of 0.60 was used to identify the robust features. The yellow regions in last column shows the robust features.

## Comparison of Conventional vs Space-resolved Radiomics pipelines

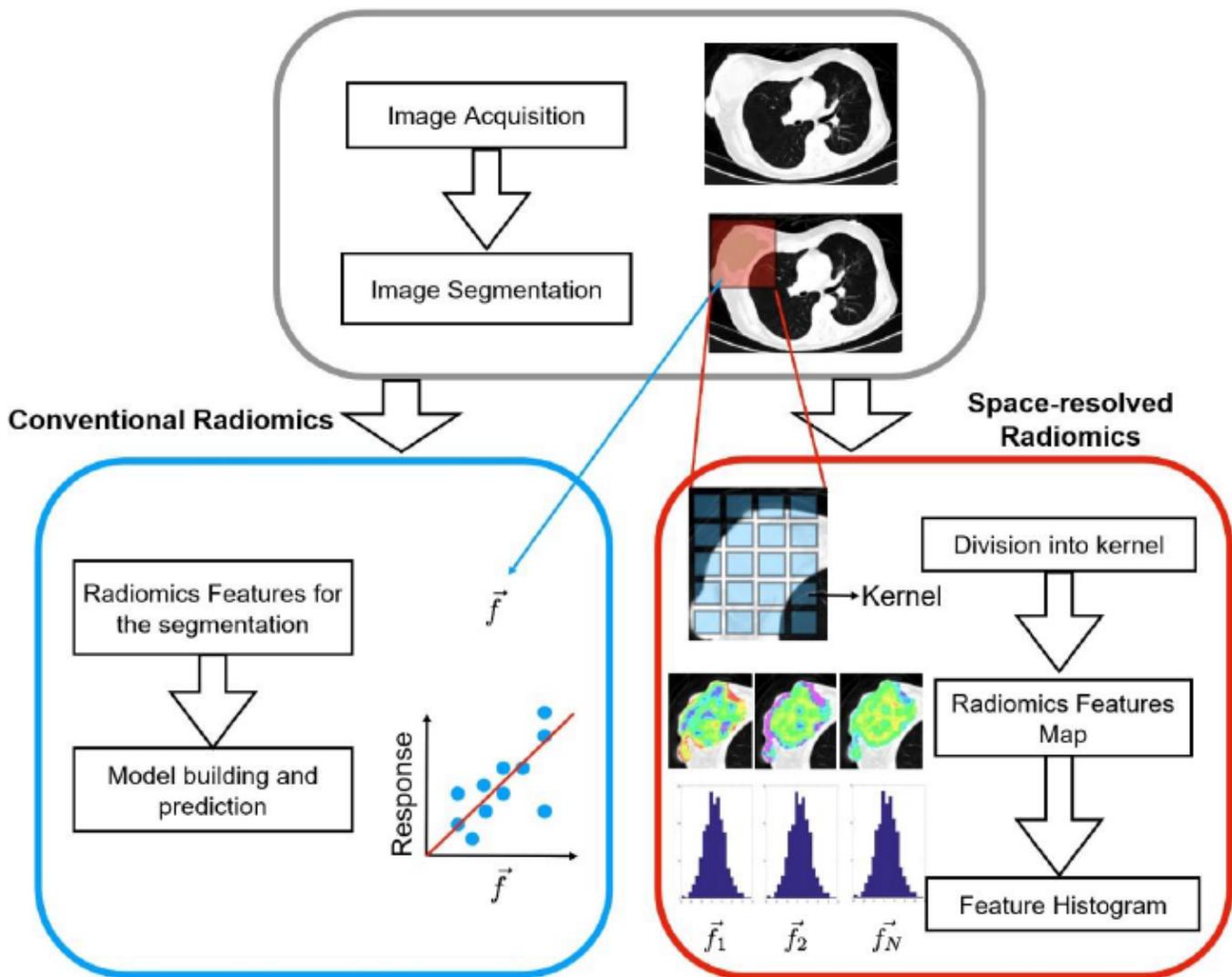
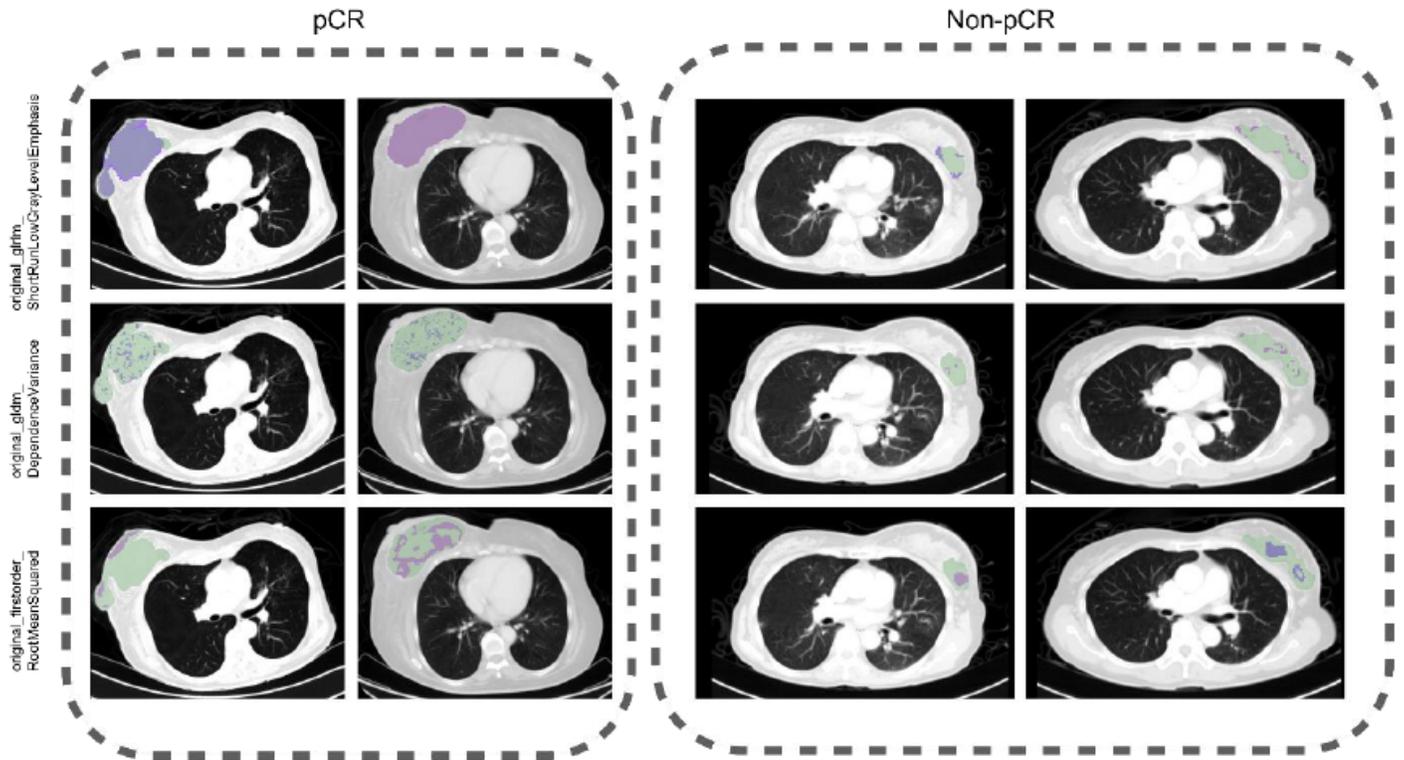


Figure 4

Comparison of conventional and space-resolved radiomics pipelines. This figure compares the conventional and space-resolved radiomics pipeline. The left blue box shows that a single radiomics features is extracted from each segment while the right red box shows that radiomics feature maps are extracted from the space-resolved radiomics pipeline. A histogram is generated from each radiomics feature map to characterize the tumor heterogeneity.



**Figure 5**

Regions driving the prediction in space-resolved radiomics. This figure shows the regions within the lesion driving the radiomics model prediction. The significant regions are shown in purple. Two true positive and true negative pCR and Non-pCR are selected in this work (column-wise). Three most significant features map from the LASSO are shown in rows in the figure.

# Deep Learning Architecture

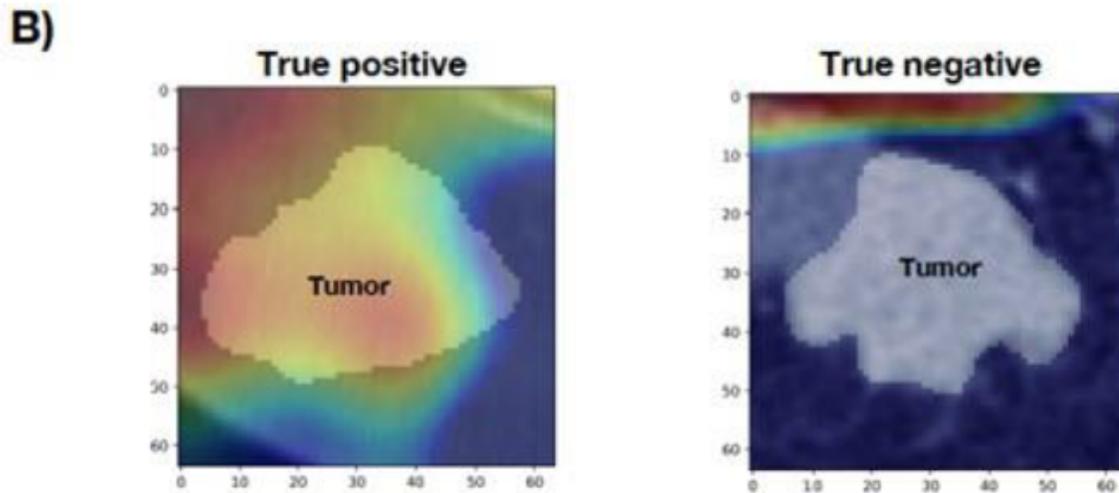
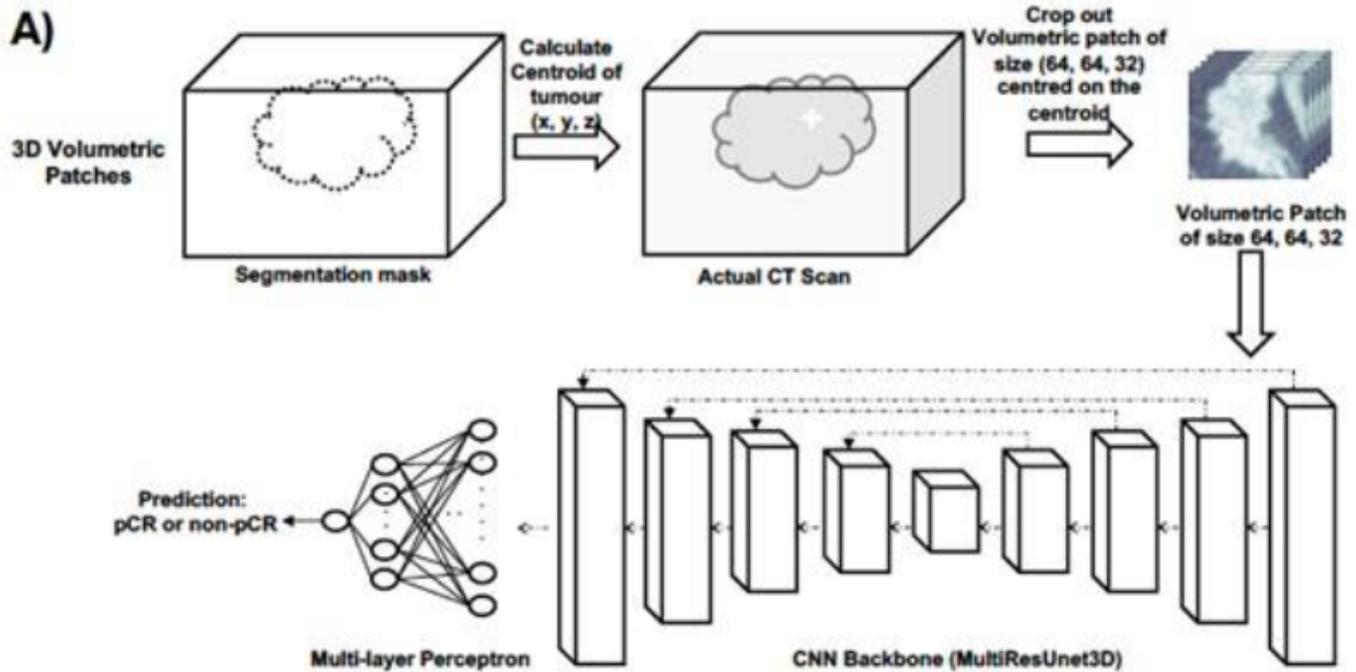
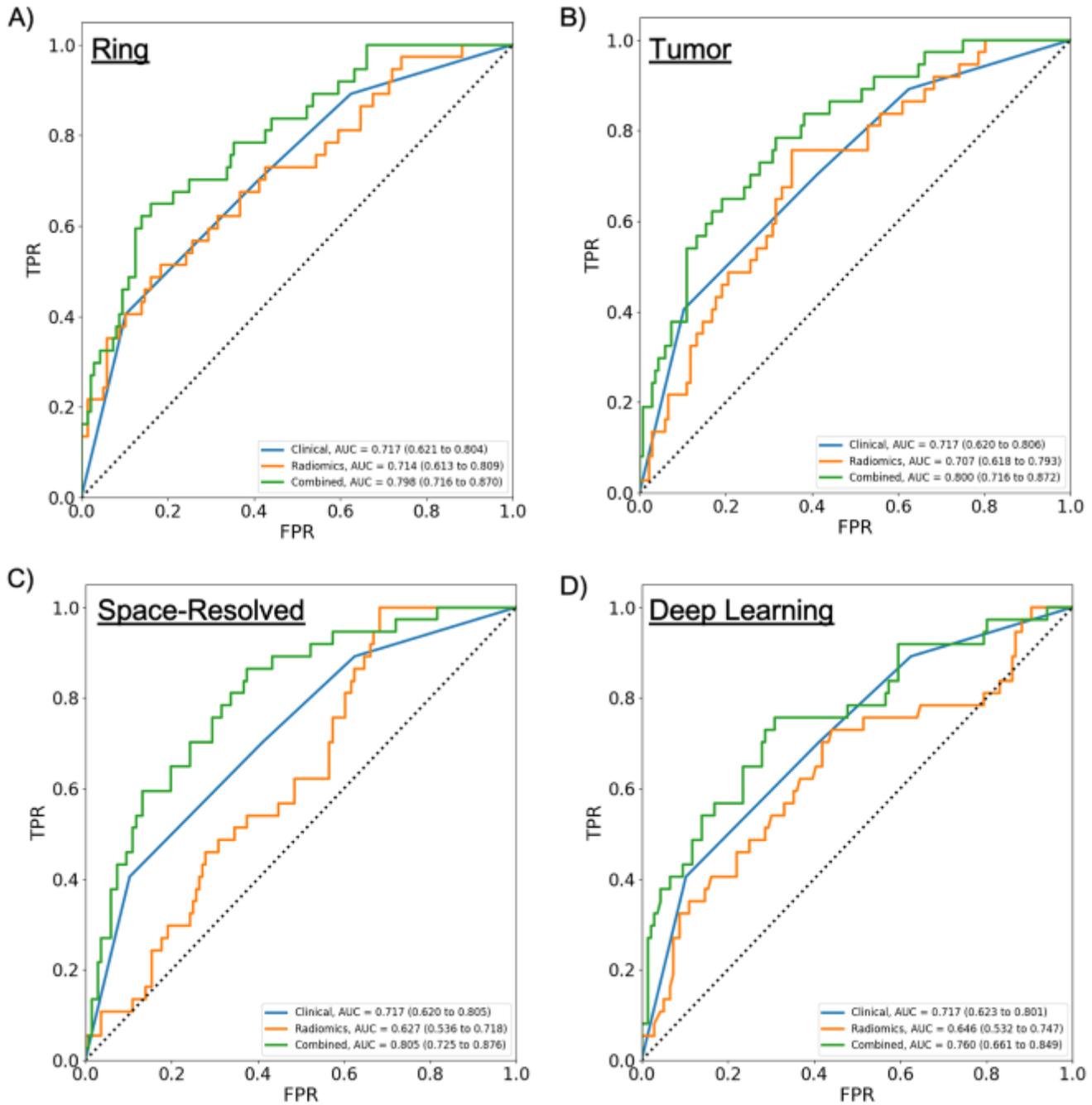


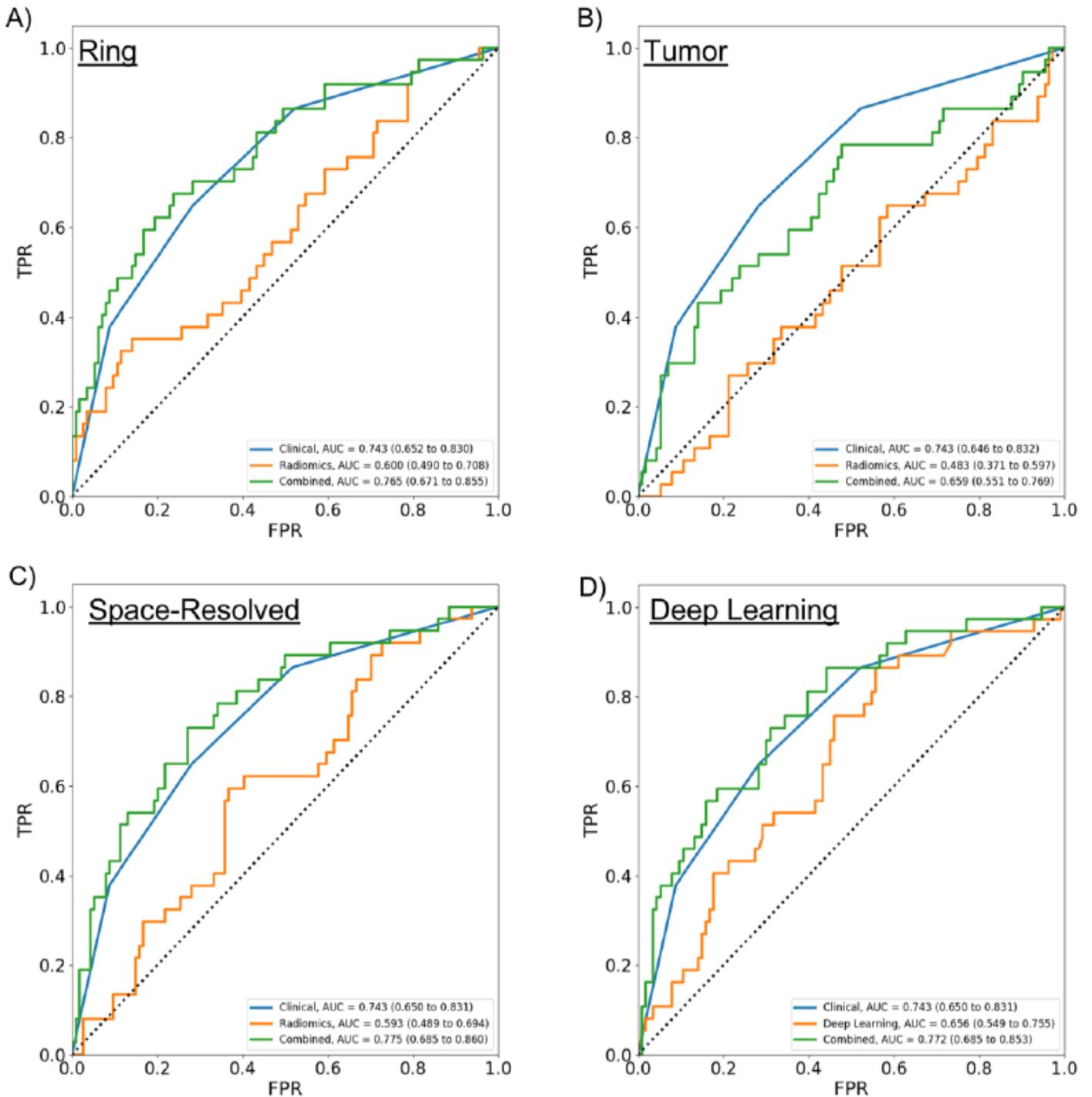
Figure 6

Deep Learning pipeline and class activation maps A) Image preprocessing and the deep learning pipeline. A MultiResUnet3D model with an additional Densenet connected to a sigmoid activation function is used for predicting the probability of pCR. B) Class activation maps calculated in the center axial slice of the tumor of a true positive and true negative examples in the test set.



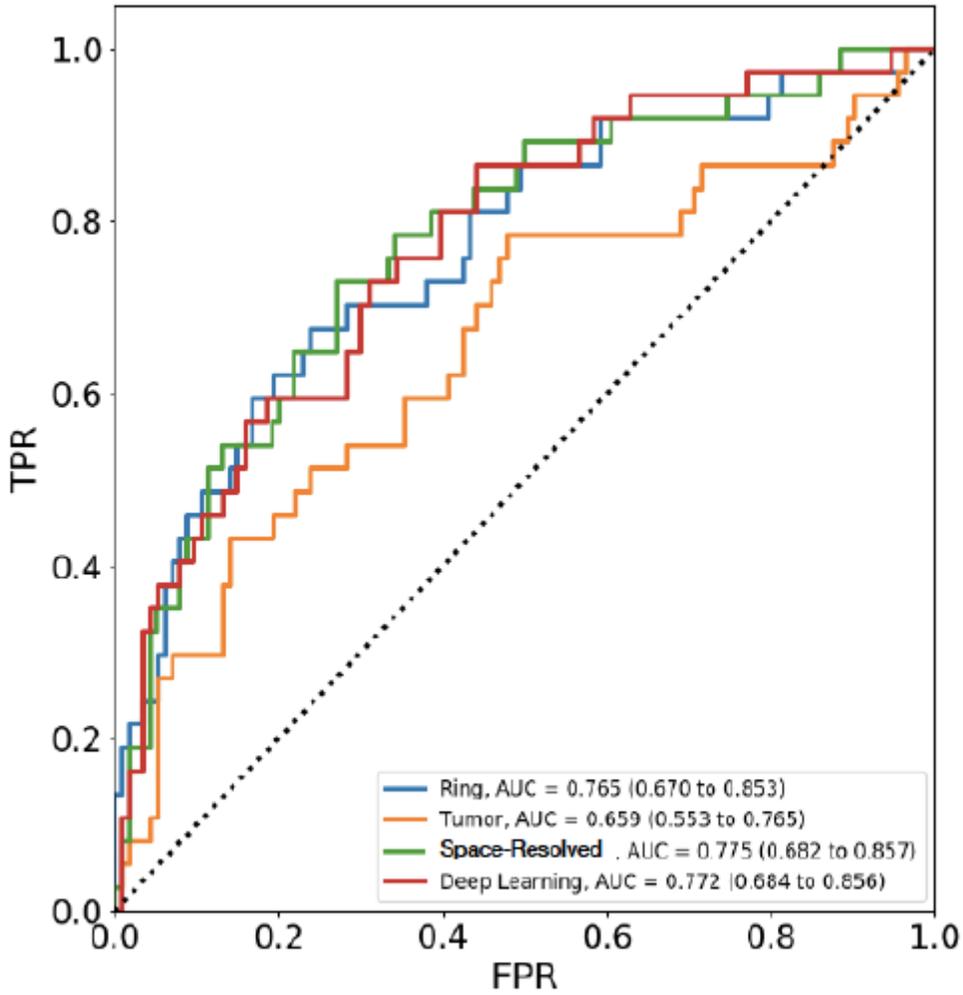
**Figure 7**

Comparison of ROC curves of the 4 different algorithms using clinical, radiomics and combined of the training data Figure A), B), C) and D) show the ROC curves of the clinical (in blue), radiomics (in yellow) and combined (in green) models for peri-tumoral, tumoral, space-resolved and Deep learning algorithms respectively. The 95% CI of the AUC are shown in parenthesis.



**Figure 8**

Comparison of ROC curves of the 4 different algorithms using clinical, radiomics and combined of the test data Figure A), B), C) and D) show the ROC curves of the clinical (in blue), radiomics (in yellow) and combined (in green) models for peri-tumoral, tumoral, space-resolved and Deep learning algorithms respectively. The 95% CI of the AUC are shown in parenthesis.



**Figure 9**

Comparison of ROC curves of the four different algorithms using combined radiomics and clinical model. This figure compares the final ROC curve of the combined model generated from peri-tumoral (blue), tumoral (yellow), space-resolved (green) and Deep learning radiomics algorithms (red). The 95% CI of the AUC are shown in parenthesis.

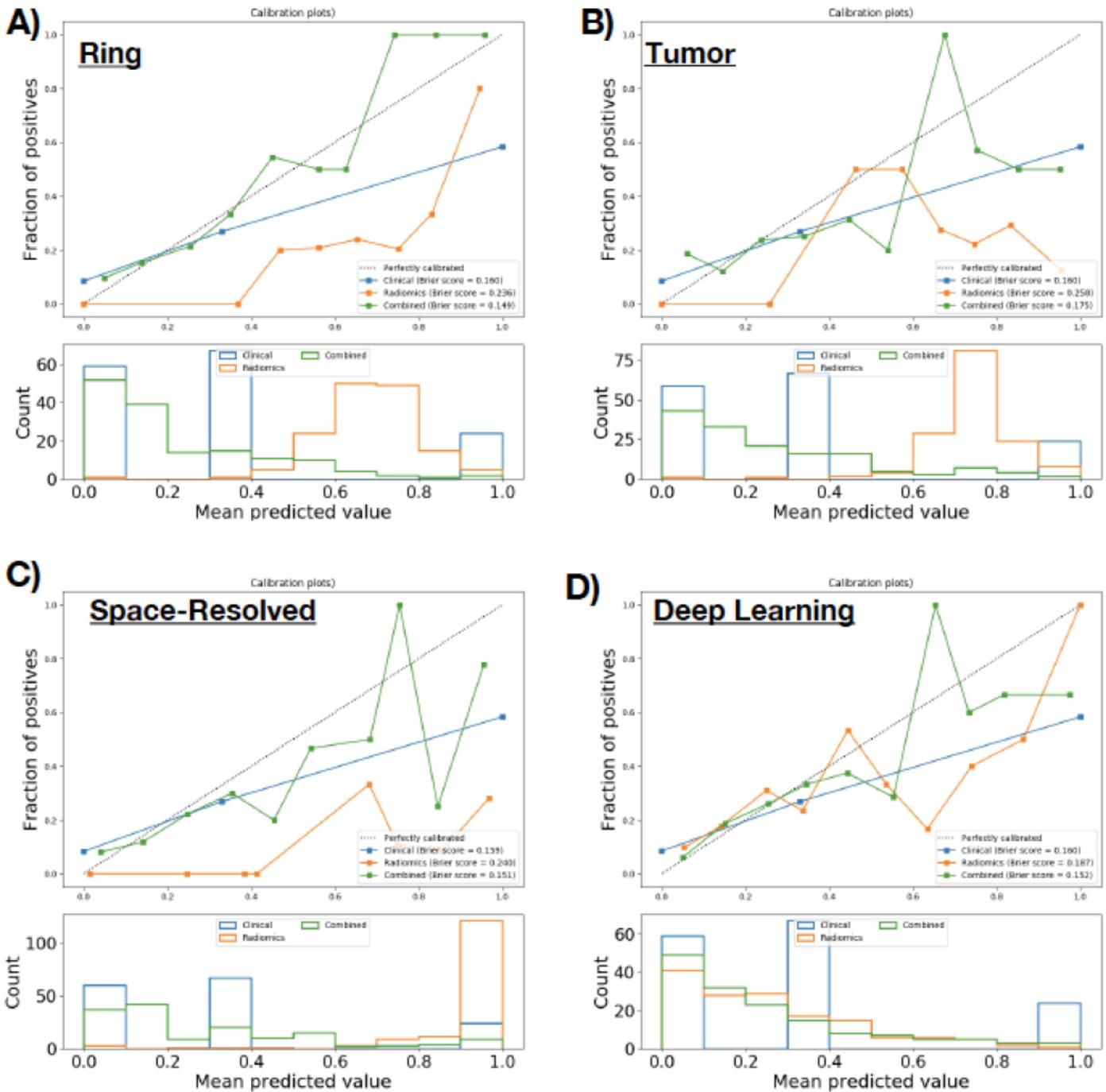


Figure 10

Comparison of calibration curves of the four different algorithms Figure A), B), C) and D) show the calibration curves of the clinical (in blue), radiomics (in yellow) and combined (in green) models for peri-tumoral, tumoral, space-resolved and Deep learning algorithms respectively. 10 groups are used for generating the calibration curves and the histogram of the predicted probability are shown in the figure as well.

Supplementary File

Loading [MathJax]/jax/output/CommonHTML/jax.js

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

- [SUPPLEMENTARY.docx](#)