**Supporting Information I: MRI Data Acquisition**

The patients which we retrospectively recruited from the four centers all underwent MRI scan using 1.5T or 3.0T scanners with double-breast coils. The image acquisition protocol included fat-suppressed T2WI, T1+C with fat suppression and DWI images. Imaging parameters of each sequence and the No. of patients scanned by different scanners were showed in Table S2. Following was the detailed description of the data acquisition procedure at different hospitals.

For the Guangdong General Hospital cohort, all MRI scans were performed either on a 1.5 T or 3.0 T scanners both from Philips (1.5 T Achieva and 3.0 T Ingenia, Philips Healthcare, Best, The Netherlands). At the scanning, an axial fat-suppressed T2WI sequence and axial DWI images were obtained using two b values (0 and 1,000 s/mm2) were acquired before contrast medium administration. An initial fat-saturated T1WI pre-contrast scan was collected before T1+C images scanning, and then T1+C images were acquired as six post contrast scans at intervals of 60 seconds following the intravenous injection of gadolinium contrast agent. A gadolinium-based agent (Magnevist; Bayer Healthcare, Berlin, Germany) was injected using an MR imaging compatible power injector at a rate of 2 ml/s and at a dose of 0.2 ml/kg of body weight, followed by a 20-ml saline flush with high-pressure injector.

For the Henan Cancer Hospital cohort, all MRI scans were performed on two 3.0 T scanners, one from Siemens (Skyra; Siemens, Erlangen, Germany) and the other from GE (Signa HDxt, GE Healthcare, USA).

For data from the Siemens scanner, an axial fat-suppressed T2WI sequence and axial DWI images were obtained using two b values (0 and 1,000 s/mm2) were acquired before contrast medium administration. An initial fat-saturated T1WI pre-contrast scan was collected before T1+C images scanning, and then T1+C images were acquired as 43 post contrast scans with no intervals following the intravenous injection of gadolinium contrast agent. A gadolinium-based agent (Gadopentetate; Consun, Guangzhou, China) was injected using an MR imaging compatible power injector at a rate of 2.5 ml/s and at a dose of 0.2 ml/kg of body weight, followed by an equal volume of normal saline solution to flush the tube with high-pressure injector.

For data from the GE scanner, an axial fat-suppressed T2WI sequence and axial DWI images were obtained using two b values (0 and 1,000 s/mm2) were acquired before contrast medium administration. An initial fat-saturated T1WI pre-contrast scan was collected before T1+C images scanning, and then T1+C images were acquired as seven post contrast scans at intervals of 50 seconds following the intravenous injection of gadolinium contrast agent. A gadolinium-based agent (Gadopentetate; Consun, Guangzhou, China) was injected using an MR imaging compatible power injector at a rate of 2.5 ml/s and at a dose of 0.2 ml/kg of body weight, followed by an equal volume of normal saline solution to flush the tube with high-pressure injector.

For the Yunnan Cancer Hospital cohort, all MRI scans were performed on two scanners with the same procedure. At the scanning, an axial fat-suppressed T2WI sequence and axial DWI images were obtained using two b values (0 and 1,000 s/mm2) were acquired before contrast medium administration. An initial fat-saturated T1WI pre-contrast scan was collected before T1+C images scanning, and then T1+C images were acquired as seven post contrast scans at intervals of 60 seconds following the intravenous injection of gadolinium contrast agent. A gadolinium-based agent (Omniscan; GE Healthcare, Ireland) was injected using an MR imaging compatible power injector at a rate of 2.5 ml/s and at a dose of 0.2 ml/kg of body weight, followed by a 20-ml saline flush with high-pressure injector.

For the cohort from Cancer Hospital Chinese Academy of Medical Sciences, all MRI scans were performed on one 3.0 T scanner (Signa HDxt, GE Healthcare, USA) with the same procedure. At the scanning, an axial fat-suppressed T2WI sequence and axial DWI images were obtained using two b values (0 and 1,000 s/mm2) were acquired before contrast medium administration. An initial fat-saturated T1WI pre-contrast scan was collected before T1+C images scanning, and then T1+C images were acquired as nine post contrast scans at intervals of 50 seconds following the intravenous injection of gadolinium contrast agent. A gadolinium-based agent (Magnevist; Bayer Healthcare, Berlin, Germany) was injected using an MR imaging compatible power injector at a rate of 2 ml/s and at a dose of 0.2 ml/kg of body weight, followed by a 20-ml saline flush with high-pressure injector.

**Supporting Information II: Radiomic Feature Extraction**

Since the intensity values of MR images distribute widely, we used z-score normalization to make the image intensities have the properties of a standard normal distribution with and , where is the mean value of the images, and is the standard deviation. The normalized values (also called z scores) of the image intensities (*x*) were calculated as follows:

After z-score normalization of image pixel intensities, a total of 4650 quantitative imaging features including 8 shape and size based features, 17 first order statistical features, 90 textural features and 4535 wavelet features (4280 features of Gabor-bank wavelet filtered images and 155 features of Law’s filtered images), were extracted respectively for T2 images, ADC maps and DCE MR images using corresponding ROIs.

**(1) Shape and size based features**

In this group of features, we included descriptors of the three-dimensional shape and size of the tumor region. Let in the following definitions V denote the volume and A the surface area of the volume of interest. We determined the following shape and size based features:

1. **Compactness 1**=
2. **Compactness 2**=
3. **Maximum 3d diameter**: The maximum three-dimensional tumor diameter is measured as the largest pairwise Euclidean distance, between voxels on the surface of the tumor volume.
4. **Spherical disproportion** = 
5. **Sphericity** =
6. **Surface area**: The surface area is calculated by triangulation (i.e. dividing the surface into connected triangles) and is defined as:



Where N is the total number of triangles covering the surface and a, b and c are edge vectors of the triangles.

1. **Surface to volume ratio** =
2. **Volume**: The volume (V) of the tumor is determined by counting the number of pixels in the tumor region and multiplying this value by the voxel size.

**(2) First order statistical features**

The following 17 statistical features were extracted.

Let **X** be the three dimensional image matrix with *N* voxels of the ROI and P be the first order histogram distribution with *Ng* discrete intensity levels.

1. **IntensityMax:** The maximum intensity value of **X.**
2. **IntensityMin:** The minimum intensity value of **X**.
3. **Median:** The median intensity value of **X**.
4. **IntensityStd:**



1. **Mean:**



1. **Variance:**



1. **Skewness:**



1. **Kurtosis:**



1. **Range:**

The range of intensity values of **X**.

1. **Mean absolute deviation:**

The mean of the absolute deviations of all voxel intensities around the mean intensity value

1. **Energy:**



1. **Entropy:**



1. **Entropy\_p:**



1. **Root mean square:**



1. **Uniformity:**



1. **Uniformity\_p:**



1. **Mass:**

The sum intensity value of .

**(3) Textural features**

Second order statistic texture features, and higher order statistic texture features were extracted. Forty-four second order statistic texture features could be calculated from the Gray Level Co-occurrence Matrix (GLCM). Forty-six high order statistic texture features were calculated from the Gray Level Size Zone Matrix (GLSZM), Gray Level Run Length Matrix (GLRLM), and Neighborhood Gray Tone Difference Matrix (NGTDM). All of the GLCM, GLSZM, GLRLM, and NGTDM based texture feature were calculated using a 2D analysis and then averaged for all slices within the three-dimensional tumor volume.

*Gray-Level Co-Occurrence Matrix based features (GLCM)*

GLCM based features were second-order statistical texture features, which are defined as a matrix *M* (*i, j; δ, θ*) to indicate the relative frequency with intensity values of pixels (*i* and *j*) at the distance of *δ* in direction *θ*.

Let:

*M*(*i, j*) be the co-occurrence matrix for an arbitrary *δ* and *θ*, set *δ=1 and θ=0 and 45*

*Ng* be the number of discrete intensity levels in the images, set as 25,

*μ* be the mean of *M*(*i, j*),

 be the marginal row probabilities,

 be the marginal column probabilities, and *uy ,μx,* be the mean of *mx* .and *my*

,

,

,

.

.

1. **Energy:**



1. **Contrast:**



1. **Entropy:**



1. **Homogeneity 1:**



1. **Homogeneity 2:**



1. **Correlation:**



1. **Variance:**



1. **Sum Average:**



1. **Sum Entropy:**



1. **Dissimilarity:**



1. **Inverse Difference Moment:**



1. **Autocorrelation:**



1. **Cluster Prominence**



1. **Cluster Shade**



1. **Cluster Tendency**



1. **Difference Entropy**



1. **Maximum Probability:**



1. **Sum variance**



1. **Informational measure of correlation 1 (IMC1):**



1. **Informational measure of correlation 2 (IMC2):**



1. **Inverse Difference Moment Normalized (IDMN):**



1. **Inverse Difference Normalized (IDN):**



*Gray Level Run Length Matrix based features (GLRLM)*

GLRLM based features were high-order statistical texture feature, which were defined as *P*(*i, j; θ*) to indicate the number of times j and gray level i appear consecutively in the direction *θ*.

Let:

*P*(*i, j; θ*) be the run-length matrix *P* for a direction *θ*, set *θ=0 and 45*

*Ng* be the number of discrete intensity values,

*Nr* be the number of different run lengths, and

*Np* be the number of voxels in the ROI.

1. **Short Run Emphasis (SRE):**



1. **Long Run Emphasis (LRE):**



1. **Gray-Level Nonuniformity (GLN):**



1. **Run-Length Nonuniformity (RLN):**



1. **Run Percentage (RP):**



1. **Low Gray-Level Run Emphasis (LGRE):**



1. **High Gray-Level Run Emphasis (HGRE):**



1. **Short Run Low Gray-Level Emphasis (SRLGE):**



1. **Short Run High Gray-Level Emphasis (SRHGE):**



1. **Long Run Low Gray-Level Emphasis (LRLGE):**



1. **Long Run High Gray-Level Emphasis (LRHGE):**



1. **Mean:**



1. **Entropy:**



1. **Energy:**



*Gray Level Size Zone Matrix based features (GLSZM)*

GLSZM based features were high-order statistical texture features, which were defined as *P*(*i, j*) to indicate the areas of size j and gray level i.

Let:

*P*(*i, j*) be the size zone of matrix *P*,

*Ng* be the number of discrete intensity values,

*Nr* be the number of different areas sizes,

*Np* be the number of voxels in the ROI.

1. **Small Zone Emphasis (SZE):**



1. **Large Zone Emphasis (LZE):**



1. **Gray-Level Nonuniformity (GLN):**



1. **Zone-Size Nonuniformity (ZSN):**



1. **Zone Percentage (ZP):**



1. **Low Gray-Level Zone Emphasis (LGZE):**



1. **High Gray-Level Zone Emphasis (HGZE):**



1. **Small Zone Low Gray-Level Emphasis (SZLGE):**



1. **Small Zone High Gray-Level Emphasis (SZHGE):**



1. **Large Zone Low Gray-Level Emphasis (LZLGE):**



1. **Large Zone High Gray-Level Emphasis (LZHGE):**



1. **Gray-Level Variance (GLV):**



1. **Zone-Size Variance (ZSV):**



*Neighborhood Gray Tone Difference Matrix based features (NGTDM)*

NGTDM based features were high-order statistical texture features, which were defined as *S(i)* to indicate the sum of the absolute value between gray intensity level i and it’s neighbors’ average intensity.

Let:

*S(i)* be the sum of absolute value between gray intensity level i and its neighbors’ average intensity,

*C(i)* be the number of voxels with the gray intensity level I,

*Ng* be the number of discrete intensity values.

1. **Coarseness:**



1. **Contrast:**



1. **Busyness:**



1. **Complexity:**



1. **Strength:**



**(4) Wavelet features: first order statistical and texture features of a wavelet filtered image.**

A total of 4535 wavelet based features were extracted for each sequence.

With the Gabor wavelet, we obtained five different wave lengths and eight different orientations filtered images. These wavelet-based features were computed on the filtered images. The original image was filtered by a two-dimensional Gabor filter defined as:



Here, five wave lengths and eight orientations were used. After filtering, 40 filtered images were generated. For each image, the first order statistical and texture features were computed. Finally, 4280 Gabor wavelet based features were extracted.

With the Law’s filter, we obtained fifteen filtered images. The filtered images were defined as follows:

L5L5,E5L5,E5E5,S5L5,S5E5,S5S5,W5L5,W5E5,W5S5,W5W5,R5L5,R5S5,R5W5,R5R5

Here, it meant as follows:



After filtering, 15 filtered images were generated. For each image, the 17 first order statistic features were computed. Finally, 255 Law’s wavelet based features were extracted.

**Table S1**. MR scanning parameters for the patients

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Hospital** | **Scanner** | **Patients No.** | **Sequence** | **TR/TE**  **(ms)** | **FOV**  **(mm)** | **Matrix** | **Slice Thickness (mm)** | **Slice Gap**  **(mm)** | **Slices** | **Flip Angle** | **Acquisition**  **Time (min)** | **Scans** |
| **Guangdong General Hospital** | Philips 1.5T  (Achieva) | 98 | T2WI | 3400/90 | 260×320 | 348×299 | 3 | 0.3 | 44 | 120° | 4min4s |  |
| DWI | 2000/103 | 320×320 | 160×160 | 5 | 1 | 32 | 90° | 1min36s |  |
| T1+C | 5.4/2.4 | 300×320 | 300×320 | 1 | 0 | 300 | 15° | 7min2s | 6 |
| Philips 3.0T  (Ingenia) | 30 | T2WI | 4495/70 | 280×340 | 332×377 | 3 | 0 | 52 | 90° | 3min53s |  |
| DWI | 7011/67 | 320×340 | 148×153 | 4 | 1 | 30 | 90° | 1min24s |  |
| T1+C | 4.8/2.1 | 280×340 | 280×339 | 1 | 0 | 300 | 12° | 6min59s | 6 |
| **Henan Cancer Hospital** | Siemens 3.0T  (Skyra) | 91 | T2WI | 3600/54 | 340×319 | 384×384 | 4 | 0.4 | 28 | 120° | 3min16s |  |
| DWI | 6700/93 | 340×153 | 200×200 | 4 | 0.4 | 28 | 180° | 2min57s |  |
| T1+C | 5.65/2.46 | 360×360 | 384×384 | 2.5 | 0.5 | 60 | 15° | 5min07s | 43 |
| GE 3.0T  (Signa HDx) | 8 | T2WI | 5950/85 | 320×320 | 320×192 | 4 | 0.5 | 28 | 90° | 2min38s |  |
| DWI | 8000/86.8 | 320×320 | 128×128 | 4.5 | 0.5 | 28 | 90° | 1min52s |  |
| T1+C | 6.5/2.1 | 380×342 | 256×256 | 3 | 0 | 54 | 10° | 6min14s | 7 |
| **Yunnan Cancer Hospital** | Siemens 1.5T  (Avanto) | 100 | T2WI | 5600/56 | 340×340 | 320×313 | 4 | 0.8 | 34 | 142° | 2min55s |  |
| DWI | 4900/84 | 185×340 | 220×220 | 4 | 0.8 | 24 | 90° | 1min43s |  |
| T1+C | 4.43/1.5 | 340×340 | 448×336 | 1.7 | 0.34 | 120 | 10° | 8min22s | 7 |
| Philip 3.0T  (Ingenia) | 7 | T2WI | 4451/70 | 280×339 | 352×406 | 3 | 0 | 50 | 90° | 3min7s |  |
| DWI | 3070/62 | 280×342 | 108×129 | 3 | 0.6 | 40 | 90° | 2min21s |  |
| T1+C | 4.3/2.1 | 280×339 | 280×337 | 2 | -1 | 150 | 12° | 8min33s | 7 |
| **Cancer Hospital Chinese Academy of Medical Sciences** | GE 3.0T  (Signa HDx) | 80 | T2WI | 3800/80 | 340×340 | 384×224 | 5 | 1 | 50 | 90° | 3min7s |  |
| DWI | 6030/68 | 360×360 | 108×129 | 5 | 1 | 40 | 90° | 2min21s |  |
| T1+C | 5/1.9 | 240×240 | 288×192 | 3.8 | 0 | 150 | 10° | 8min25s | 9 |

*Abbreviations*: FOV, field of view; TR, repetition time; TE, echo time; T2WI, T2 weighted imaging; DWI, diffusion weighted imaging; CE, contrast enhancement.

**Table S2.** Details of NAC Regimens in each cohort

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Guangdong General Hospital | Henan Cancer Hospital | Yunnan Cancer Hospital | Cancer Hospital Chinese Academy of Medical Sciences |
| Taxane-based | 113 | 31 | 21 | 28 |
| Anthracycline-based |  | 51 | 10 | 2 |
| Anthracycline- and Taxane-based | 15 | 17 | 76 | 50 |

**Table S3**. Characteristics of Patients in the Primary and Validation Cohorts

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Characteristics | Primary Cohort  (N = 128) | Validation Cohort 1  (N = 99) | Validation Cohort 2  (N = 107) | Validation Cohort 3  (N = 83) | P |
| **Age, mean ± SD, years** | 47.99 ± 9.051 | 46.06 ± 8.468 | 45.98 ± 8.778 | 44.52±11.495 | 0.078 |
| **Stage** **(%)** | | | | | 0.001\* |
| I | 13(10.15) | 1(1.01) | 0(0.00) | 2(2.50) |  |
| II | 94(73.44) | 45(45.45) | 78(72.90) | 25(31.25) |  |
| III | 21(16.41) | 53(53.54) | 29(27.10) | 53(66.25) |  |
| **ER Status (%)** | | | | | 0.001\* |
| Positive | 81(63.28) | 72(72.73) | 84(78.50) | 42(52.50) |  |
| Negative | 47(36.72) | 27(27.28) | 23(21.50) | 38(47.50) |  |
| **PR Status (%)** | | | | | 0.001\* |
| Positive | 66(51.56) | 66(66.67) | 80(74.77) | 43(53.75) |  |
| Negative | 62(48.44) | 33(33.33) | 27(25.23) | 37(46.25) |  |
| **HER2 Status (%)** | | | | | 0.001\* |
| Positive | 57(44.53) | 21(21.21) | 10(9.35) | 69(86.25) |  |
| Negative | 71(55.47) | 78(78.79) | 97(90.65) | 11(13.75) |  |
| **Ki-67 Status (%)** | | | | | 0.001\* |
| Positive | 105(82.03) | 90(90.91) | 63(58.88) | 12(15.00) |  |
| Negative | 23(17.97) | 9(9.09) | 44(41.12) | 68(85.00) |  |
| **Cancer Subtype (%)** |  |  |  |  | 0.001\* |
| HR+ and HER2- | 51(39.84) | 62(62.63) | 82(76.63) | 43(53.75) |  |
| HER2+ | 57(44.53) | 21(21.21) | 10(9.35) | 12(15.00) |  |
| Triple-negative | 20(15.63) | 16(16.16) | 15(14.02) | 25(31.25) |  |
| **Response to NAC (%)** | | | | | 0.001\* |
| *pCR* | 56(43.75) | 16(16.17) | 28(26.17) | 12(15.00) |  |
| Non-*pCR* | 72(56.25) | 83(83.84) | 79(73.83) | 68(85.00) |  |

*Note*: The ER and PR threshold value for level was ≤1%, and the threshold value for Ki-67 was ≤20%.

*Abbreviations*: SD, standard deviation; ER, Estrogen receptor; PR, Progesterone receptor; HER2, Human epidermal growth factor receptor 2; NAC, neoadjuvant chemotherapy; *pCR*, pathological complete response.

*\* P < 0.05*

**Table S4.** Radiomic Features selected for the construction of radiomics signature based on multi-parametric MRI

|  |  |  |  |
| --- | --- | --- | --- |
| Features | Sequence | Normalized Values  (mean ± standard deviation) | |
| *pCR* cohort | *Non-pCR* cohort |
| Gabor1\_GLRLM45\_LRHGLE | T1+C | -0.3071±0.4952 | -0.4883±0.4884 |
| Laws1\_FOS\_mean | ADC | 0.1615±0.4546 | 0.0547±0.4260 |
| Laws2\_FOS \_mass | ADC | -0.6978±0.3423 | 0.0210±0.4517 |
| Gabor17\_ GLRLM45\_LRHGLE | ADC | -0.0748±0.4629 | -0.4112±0.3414 |
| Gabor22\_FOS\_minimum | ADC | -0.2831±0.4169 | -0.7503±0.3190 |
| Gabor10\_GLCM\_IMC1 | T2WI | -0.2397±0.4345 | 0.1645±0.4899 |
| Gabor17\_FOS\_maximum | T2WI | 0.0836±0.4448 | -0.0918±0.4358 |
| Gabor18\_FOS\_root\_mean\_square | T2WI | -0.0001±0.5057 | -0.3031±0.3683 |

*Abbreviations*: ADC, apparent diffusion coefficient; T2WI, T2-weighted images; T1+C, contrast enhanced T1-weighted imaging; GLCM, gray-level co-occurrence matrix; GLRLM, gray level run length matrix; FOS, first order statistics; Laws and Gabor is wavelet filter.

**Table S5.** AUCs of radiomics signatures in primary and validation cohorts

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Model | Primary Cohort | p value | Validation Cohort 1 | Validation Cohort 2 | Validation Cohort 3 |
| RS (95% CI) | 0.79 (0.71-0.87) |  | 0.70 (0.58-0.82) | 0.68 (0.56-0.79) | 0.79 (0.65-0.93) |
| T1+C (95% CI) | 0.64 (0.54-0.73) | 0.002\* | 0.53 (0.38-0.68) | 0.59 (0.46-0.71) | 0.52 (0.33-0.71) |
| T2WI (95% CI) | 0.69 (0.60-0.78) | 0.042\* | 0.64 (0.52-0.77) | 0.55 (0.43-0.67) | 0.56 (0.37-0.75) |
| ADC (95% CI) | 0.69 (0.59-0.79) | 0.053\* | 0.54 (0.40-0.68) | 0.62 (0.49-0.75) | 0.63 (0.43-0.84) |

*Abbreviations*: AUC, area under receiver operating characteristic curve; *pCR*, pathological complete response; RS, radiomics signature; T1+C, contrast enhanced T1-weighted imaging; T2WI, T2-weighted images; ADC, apparent diffusion coefficient.

p value refers to Delong test for the differences of AUCs between different metrics

\* p < 0.05, with significant or marginally significant differences for AUCs of radiomics signature based on T1+C, T2WI and ADC compared with that of Radiomics Signature based on multi-parametric MRI

**Table S6.** AUCs of radiomics signature based on multi-parametric MRI with separation by breast cancer subtypes

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Subtype | Primary Cohort | Validation Cohort 1 | Validation Cohort 2 | Validation Cohort 3 |
| HR+, HER2- (95% CI) | 0.81 (0.69-0.93) | 0.78 (0.56-0.99) | 0.71 (0.59-0.85) | 0.87 (0.66-1) |
| HER2+ (95% CI) | 0.70 (0.56-0.85) | 0.79 (0.59-0.99) | 0.58 (0.13-1) | 0.62 (0.25-0.99) |
| Triple Negative (95% CI) | 0.96 (0.89-1) | 0.79 (0.55-1) | 0.82 (0.59-1) | 0.84 (0.69-1) |

*Abbreviations*: AUC, area under receiver operating characteristic curve; HR, hormone receptor; HER2, human epidermal growth factor receptor 2.

**Table S7.** Related Factors for pCR prediction in Breast cancer

|  |  |  |  |
| --- | --- | --- | --- |
| Intercept and Variables | *β* | Odds Ratio (95% CI) | P |
| Intercept | -2.312 |  | <0.001\* |
| PR status | -1.626 | 5.086 (2.009-12.873) | <0.001\* |
| HER2 status | 1.362 | 3.905 (1.585-9.618) | 0.003\* |
| Radiomic Signature | 4.959 | 7.569 (3.229-17.741) | <0.001\* |

*Note*: *β* is the regression coefficient. \**P ＜ 0.05.*

*Abbreviations*: PR, progesterone receptor; HER2, human epidermal growth factor receptor 2.

**Table S8.** AUCs of RMM, radiomics signature, and clinical model for pCR prediction in primary and validation cohorts

|  |  |  |  |
| --- | --- | --- | --- |
|  | RMM (95% CI) | RS (95% CI) | Clinical model (95% CI) |
| Primary Cohort | 0.86 (0.80-0.92) | 0.79 (0.71-0.87) | 0.77 (0.69-0.86) |
| p value |  | 0.019\* | 0.007\* |
| Validation Cohort 1 | 0.79 (0.69-0.90) | 0.70 (0.58-0.82) | 0.76 (0.64-0.89) |
| p value |  | 0.044\* | 0.046\* |
| Validation Cohort 2 | 0.71 (0.61-0.81) | 0.68 (0.56-0.79) | 0.60 (0.49-0.71) |
| p value |  | 0.412 | 0.004\* |
| Validation Cohort 3 | 0.80 (0.67-0.91) | 0.79 (0.65-0.93) | 0.79 (0.68-0.89) |
| p value |  | 0.613 | 0.386 |

*Abbreviations*: AUC, area under receiver operating characteristic curve; *pCR*, pathological complete response; RMM, radiomics of multiparametric MRI; RS, radiomics signature.

p value refers to Delong test for the differences of AUCs between different metrics in different cohorts

\* p < 0.05, with significant differences for AUCs of RS and Clinical model compared with that of RMM