Appendix 1

Methods

Image acquisition

Scanning parameters were divided into 2 parts based on different voxel sizes both in phantom and clinical data. The first test scanning parameters were as follows: matrix, 512×512; field of view (FOV), 350 mm; slice thickness, 2 mm; tube voltage, 120 kV; tube current, auto mA; pitch, 0.813; collimator width, 16×0.575 mm;; rotation time, 0.5 s; lung window settings (width/level), 1,200/–600 Hounsfield units (HU); and mediastinal window settings (width/level), 350/40 HU. After repositioning, the retest data of the phantom was obtained by repeated scanning. The second test scanning parameters were as follows: matrix, 1,024×1,024; FOV, 180 mm; slice thickness, 1 mm; The remaining scan parameters were set the same as those of the first scanning.

Radiomic features extraction

The uAI Research Portal software was used to extract radiomic features, including first-order statistics, shape, and texture features. Texture features included gray-level co-occurrence matrix (GLCM), gray-level size zone matrix (GLSZM), gray-level run length matrix (GLRLM), gray-level dependence matrix (GLDM), and neighborhood gray tone difference matrix (NGTDM). The features extraction was filtered by 14 parameters including boxmean, additivegaussiannoise, binomialblurimage, curvatureflow, boxsigmaimage, normalize, laplaciansharpening, discretegaussian, mean, specklenoise, recursivegaussian, shotnoise, laplacian of gaussian, and wavelet.

The inclusion/exclusion criteria of validation data

A clinical validation dataset with 512×512 matrix containing 186 lung nodules (pathologically confirmed LUAD) with a mean diameter of less than 2 cm was used in this study from July 1, 2018 to June 30, 2020 at Beilun Second People's Hospital (Zhejiang, China) and Shanghai Changzheng Hospital (Shanghai China). Patients that met any 1 of the following criteria was excluded from the study: absence of CT examination within 1 month before surgery; absence of consecutive CT images with 1.5 mm thickness or less; with maximum diameter more than 2 cm; complications with other tumors or pulmonary disease (such as obstructive pneumonia); with severe respiratory motion artifacts; with preoperative treatment (such as neoadjuvant therapy).

https://dx.doi.org/10.21037/qims-22-992

Table S1 Parameters used for CT images both in phantom and clinical data

Scanning parameters	Phantom data		Clinical data	
	512×512	1,024×1,024	512×512	1,024×1,024
FOV (mm)	350	180	318–395	149–232
Thickness (mm)	2	1	2	1
Pitch	0.891	0.891	0.891	0.891
Rotation time (s)	0.5	0.5	0.5	0.5
Tube voltage (kV)	120	120	120	120
Tube current (mA)	Auto	Auto	Auto	Auto
Voxel size (mm^3)	0.68×0.68×2.00	0.18×0.18×1.00	0.62×0.62×2.00–0.77×0.77×2.00	0.15×0.15×1.00–0.23×0.23×1.00

CT, computed tomography; FOV, field of view.

Table S2 The clinical and CT images characteristics of participants with LUAD in clinical validation

Characteristics	Total (n=186)	IAC group (n=88)	MIA-AIS group (n=98)	P value
Age (years)	57.96±11.12	60.77±10.47	55.43±11.08	0.001
Gender (F/M)	125/61	52/36	73/25	0.019
Nodule type				0.000
pGGN	100 (55.76)	26 (29.55)	74 (75.51)	
mGGN	72 (38.71)	49 (55.68)	23 (23.47)	
Solid nodule	14 (7.53)	13 (14.77)	1 (1.02)	

Data are presented as mean ± standard deviation, n, or n (%). CT, computed tomography; LUAD, lung adenocarcinoma; IAC, invasive adenocarcinoma; MIA, minimally invasive adenocarcinoma; AIS, adenocarcinoma in situ; F, female; M, male; pGGN, pure ground glass nodule; mGGN, mixed ground glass nodule.

https://dx.doi.org/10.21037/qims-22-992

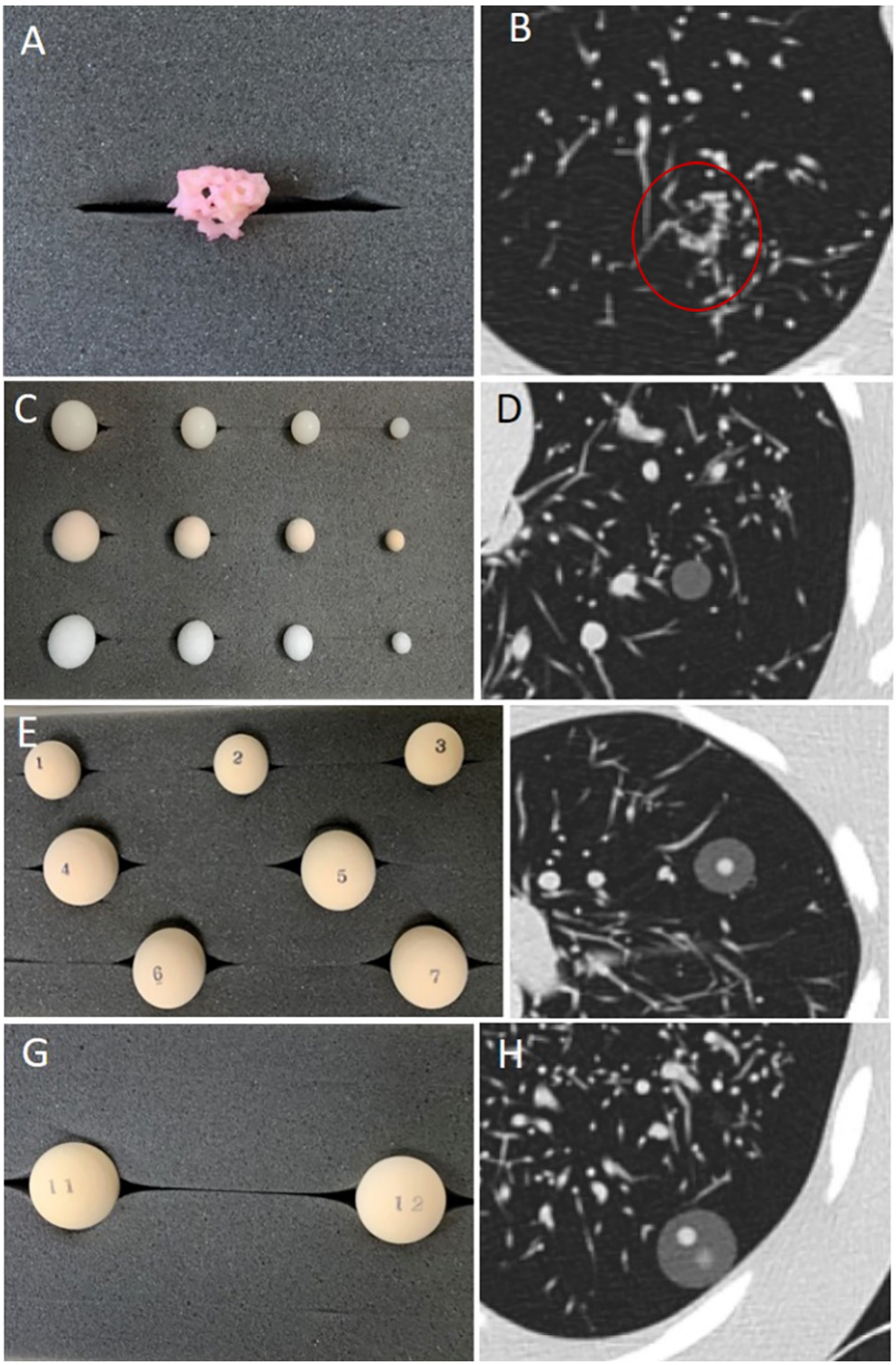

Figure S1 CT scanning of the artificial nodules embedded in the artificial vascular bundle in the chest phantom and. (A,B) Irregular nodule with 15 mm in diameter (red circle); (C,D) pure nodules, including 4 pure solid nodules and 8 pGGNs; (E,F) concentric mGGNs of different diameters (15 and 20 mm) with different spherical solid components (3, 5, 7, 9 mm); (G,H) double eccentric mGGN of 20 mm in diameter with different spherical solid components (3, 5, 7 mm). CT, computed tomography; pGGN, pure ground glass nodule; mGGN, mixed ground glass nodule.

https://dx.doi.org/10.21037/qims-22-992

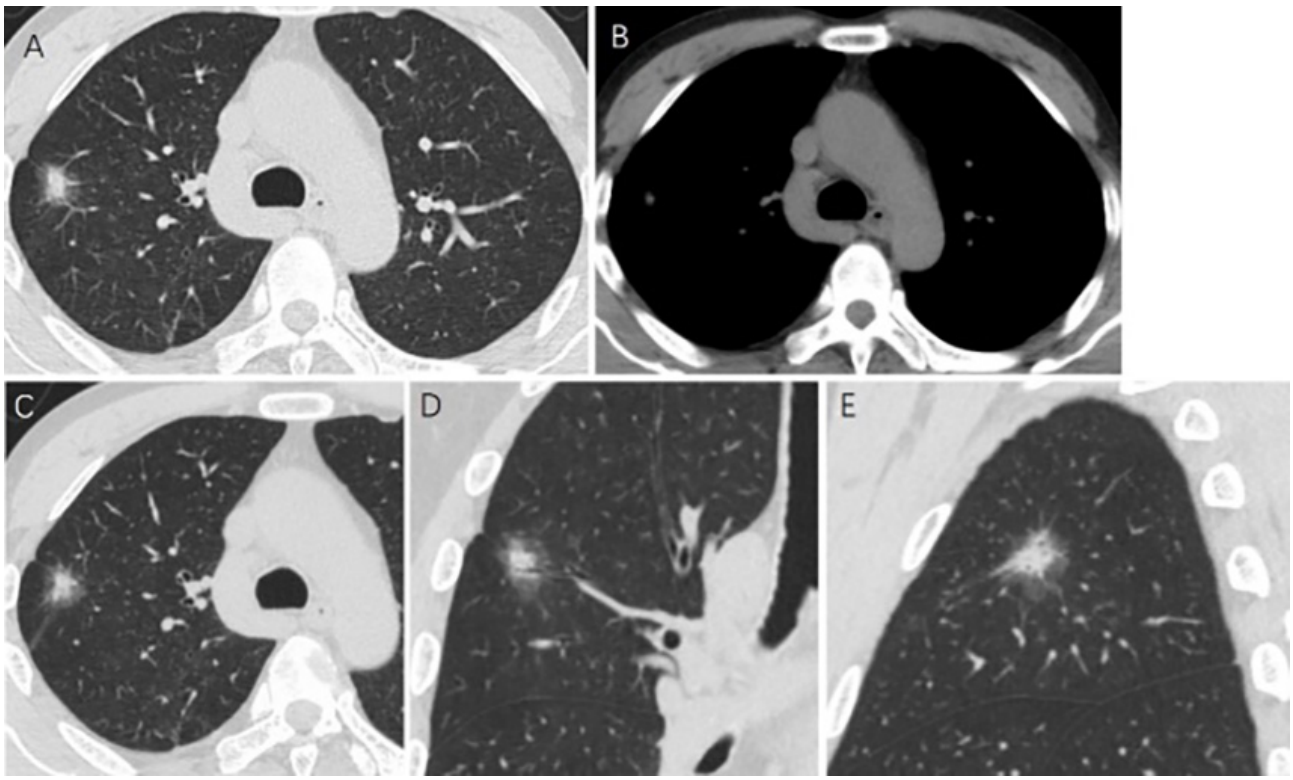

Figure S2 CT scanning of a 52-year-old man with pulmonary adenocarcinoma. (A,B) Images of lung and mediastinal window on conventional CT scanning with 512×512 matrix, the voxel size was 0.68×0.68×2.00 mm^3; (C-E) the transverse, coronal, and sagittal images on CT target scanning with 1,024×1,024 matrix, respectively, and the voxel size was 0.18×0.18×1.00 mm^3. CT, computed tomography.

https://dx.doi.org/10.21037/qims-22-992

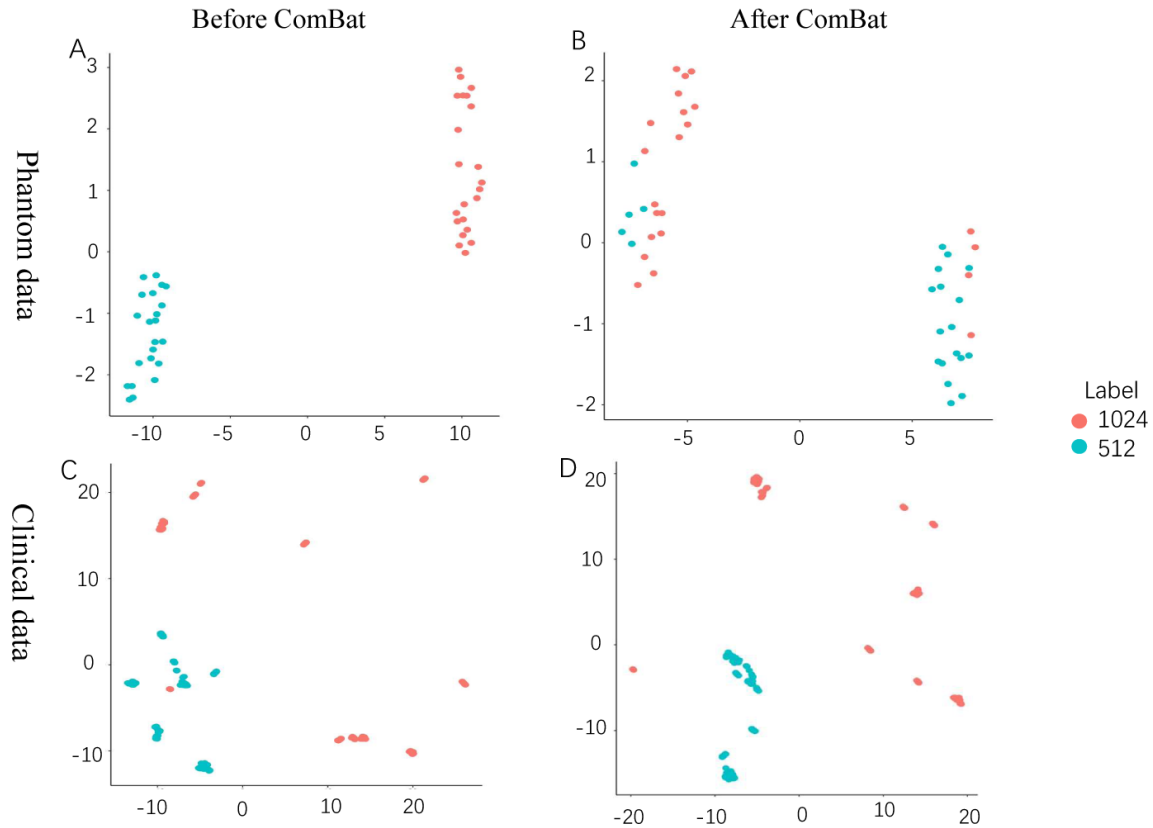

Figure S3 The UMAP by ComBat compensation method both in phantom and clinical application. (A,B) Batch effect decreased from 46.39% to 14.90% in phantom application; (C,D) batch effect decreased from 25.76% to 6.15% in clinical application. ComBat, combatting batch effect; UMAP, manifold approximation and projection.

https://dx.doi.org/10.21037/qims-22-992

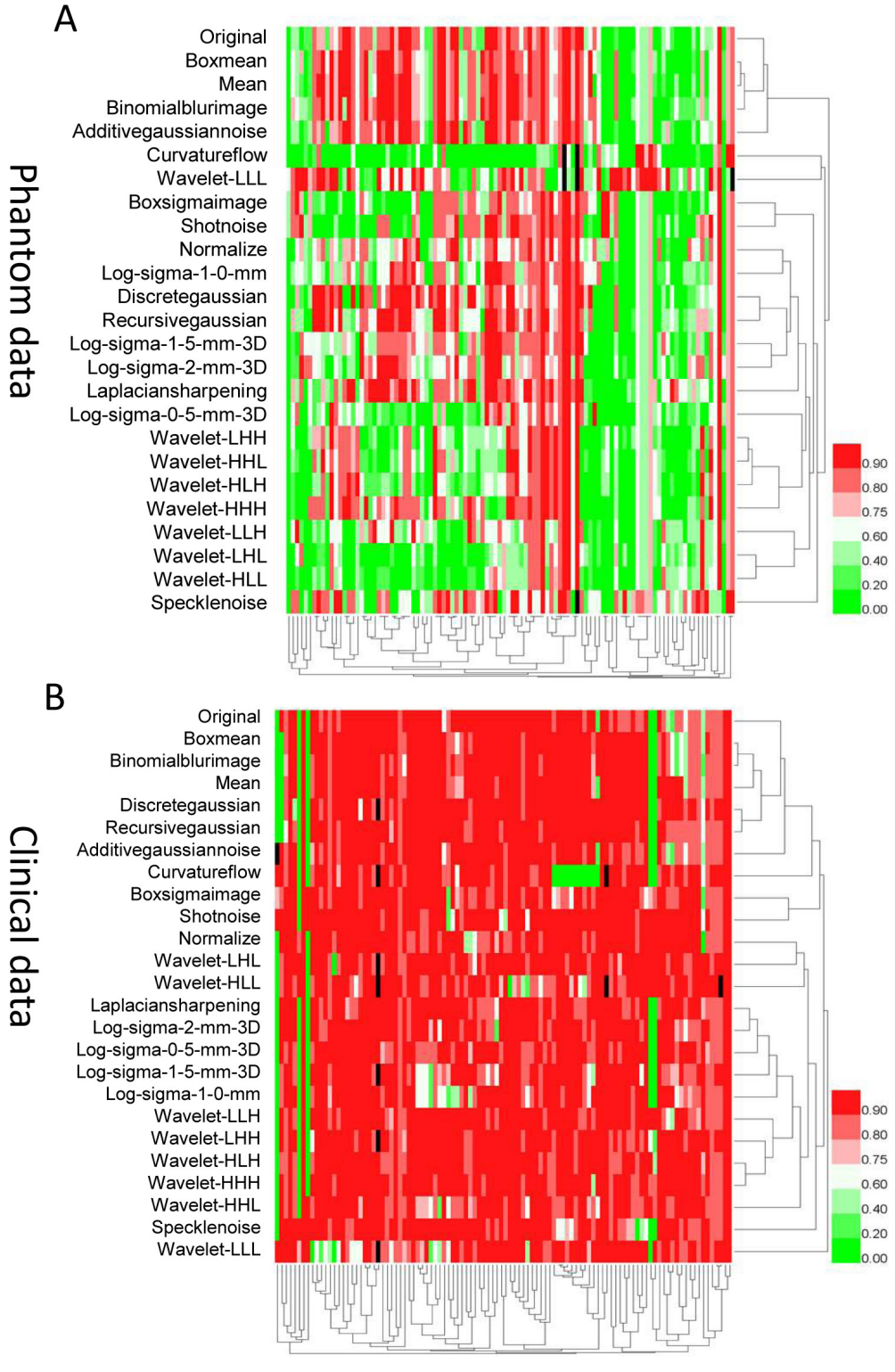

Figure S4 Heatmap of hierarchical clustering analysis of phantom data with ComBat correction. The excellent features were increased after ComBat correction and the poor features were reduced both in phantom and clinical data. (A) Phantom data of 1,085 (41.73%) excellent features, 693 (26.65%) moderate features, and 822 (31.62%) poor features. The hierarchical clustering was stopped arbitrarily at 15 groups of features; (B) clinical data of 2,365 (90.96%) excellent features, 106 (4.08%) moderate features, and 129 (4.92%) poor features. The hierarchical clustering was stopped arbitrarily at 15 groups of features. The color map of the clustogram ranges from 0 to 1, values close to 1 have a red shade and values close to 0 have a green shade. The groups with red shade mean relatively high feature values. Black shade: empty value. ComBat, combatting batch effect.

https://dx.doi.org/10.21037/qims-22-992

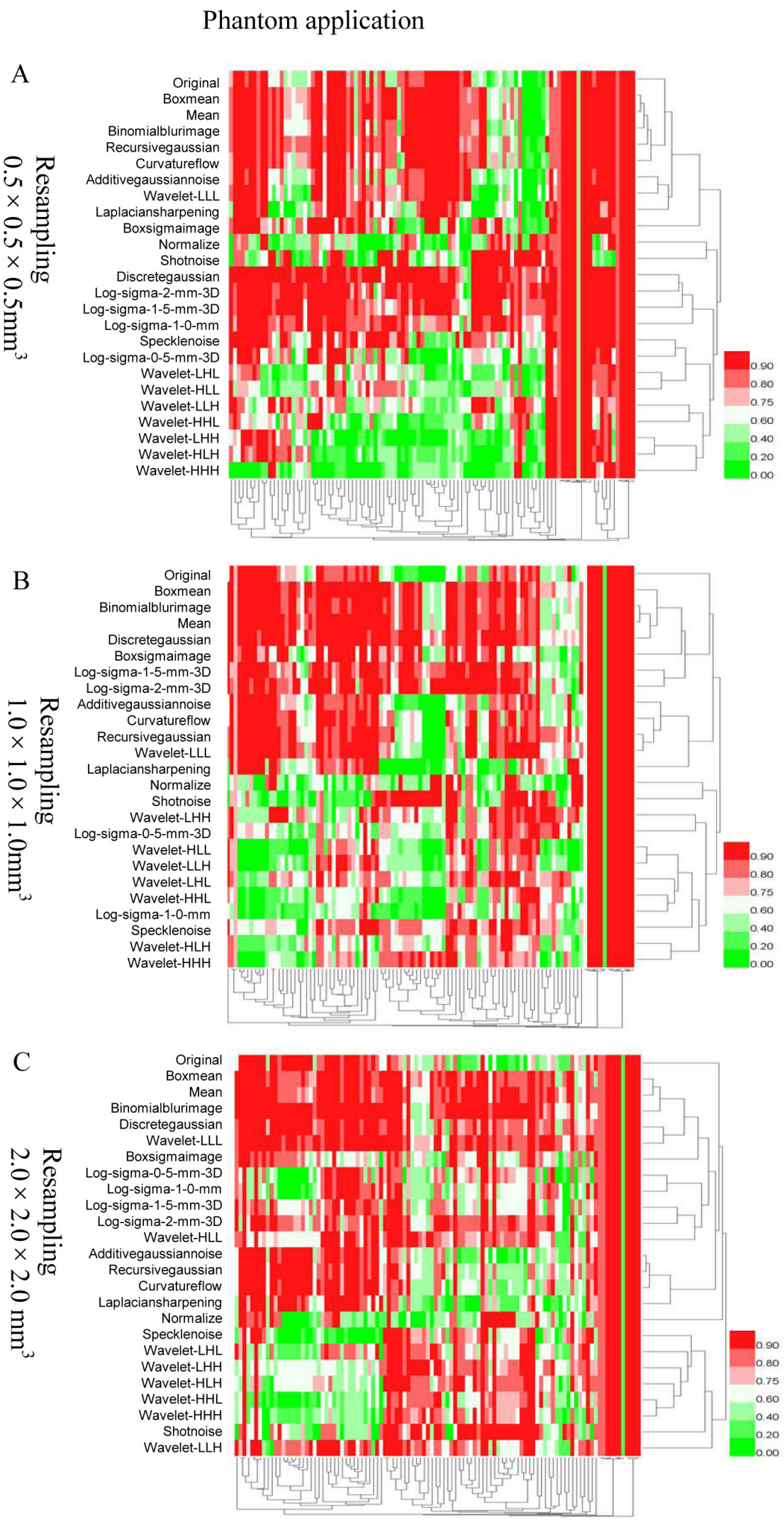

Figure S5 Heatmap of hierarchical clustering analysis of phantom data with image resampling. The excellent features were increased after image resampling and the poor features were reduced. (A) There were 1,655 (63.65%) excellent features, 502 (19.31%) moderate features, and 443 (17.04%) poor features after image resampling 0.5×0.5×0.5 mm^3; (B) there were 1,558 (59.92%) excellent features, 661 (25.43%) moderate features, and 381 (14.65%) poor features after image resampling 1.0×1.0×1.0 mm^3; (C) there were 1,655 (63.65%) excellent features, 654 (25.16%) moderate features, and 291 (11.19%) poor features after image resampling 2.0×2.0×2.0 mm^3. The gradual color changes from green to red in heatmap represent the steady increase in feature values from 0 to 1. H, high; L, low; 3D, three-dimensional.

https://dx.doi.org/10.21037/qims-22-992

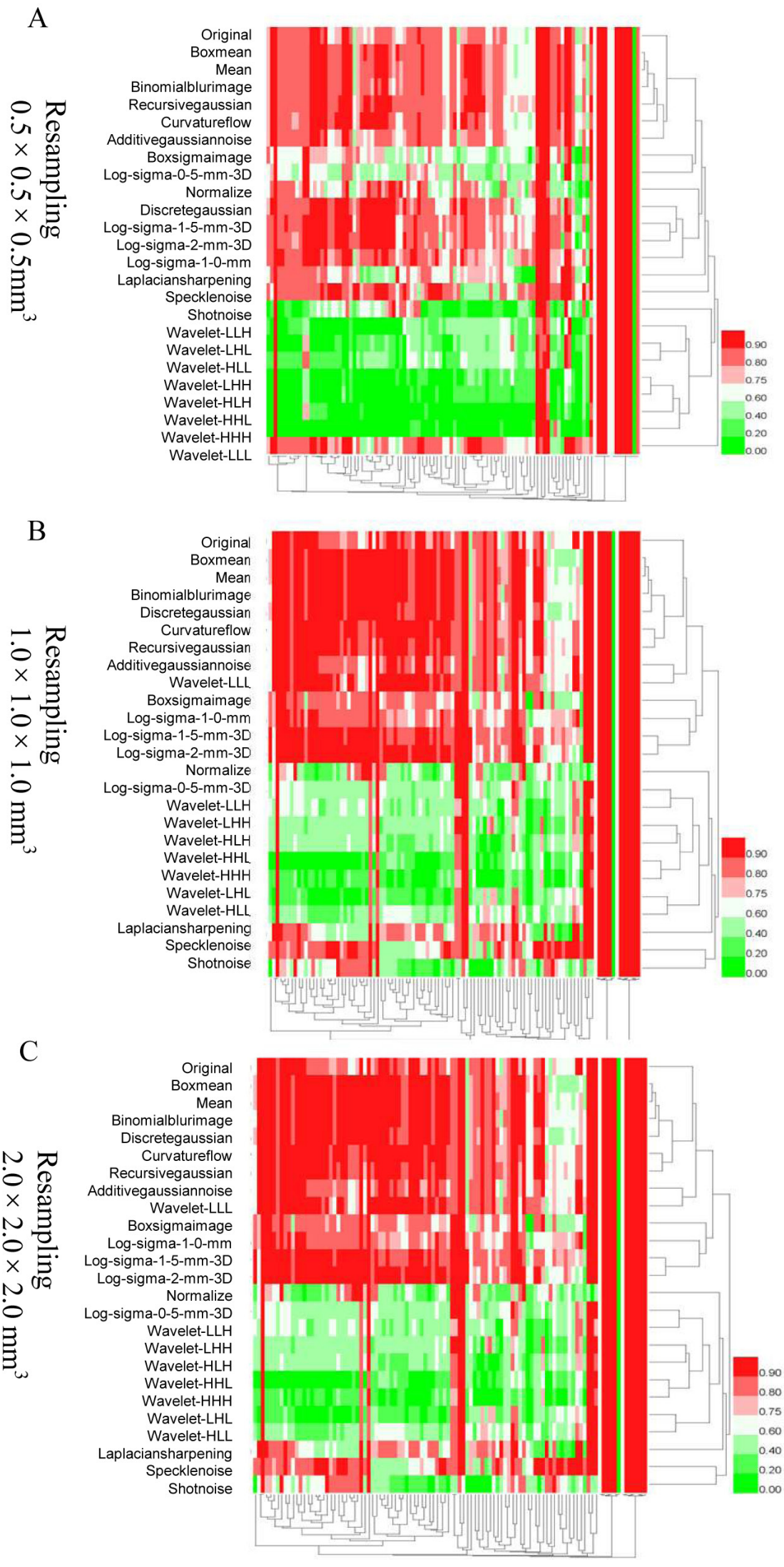

Figure S6 Heatmap of hierarchical clustering analysis of clinical data with image resampling. The excellent features were increased after image resampling, whereas poor features were reduced. (A) There were 1,394 (53.62%) excellent features, 637 (24.50%) moderate features, and 569 (21.88%) poor features after image resampling 0.5×0.5×0.5 mm^3; (B) there were 1,519 (58.42%) excellent features, 745 (28.66%) moderate features, and 336 (12.92%) poor features after image resampling 1.0×1.0×1.0 mm^3; (C) there were 1,251 (48.12%) excellent features, 1,116 (42.92%) moderate features, and 233 (8.96%) poor features after image resampling 2.0×2.0×2.0 mm^3. The gradual color changes from green to red in heatmap represent the steady increase in feature values from 0 to 1. 3D, three-dimensional; H, high; L, low.

https://dx.doi.org/10.21037/qims-22-992

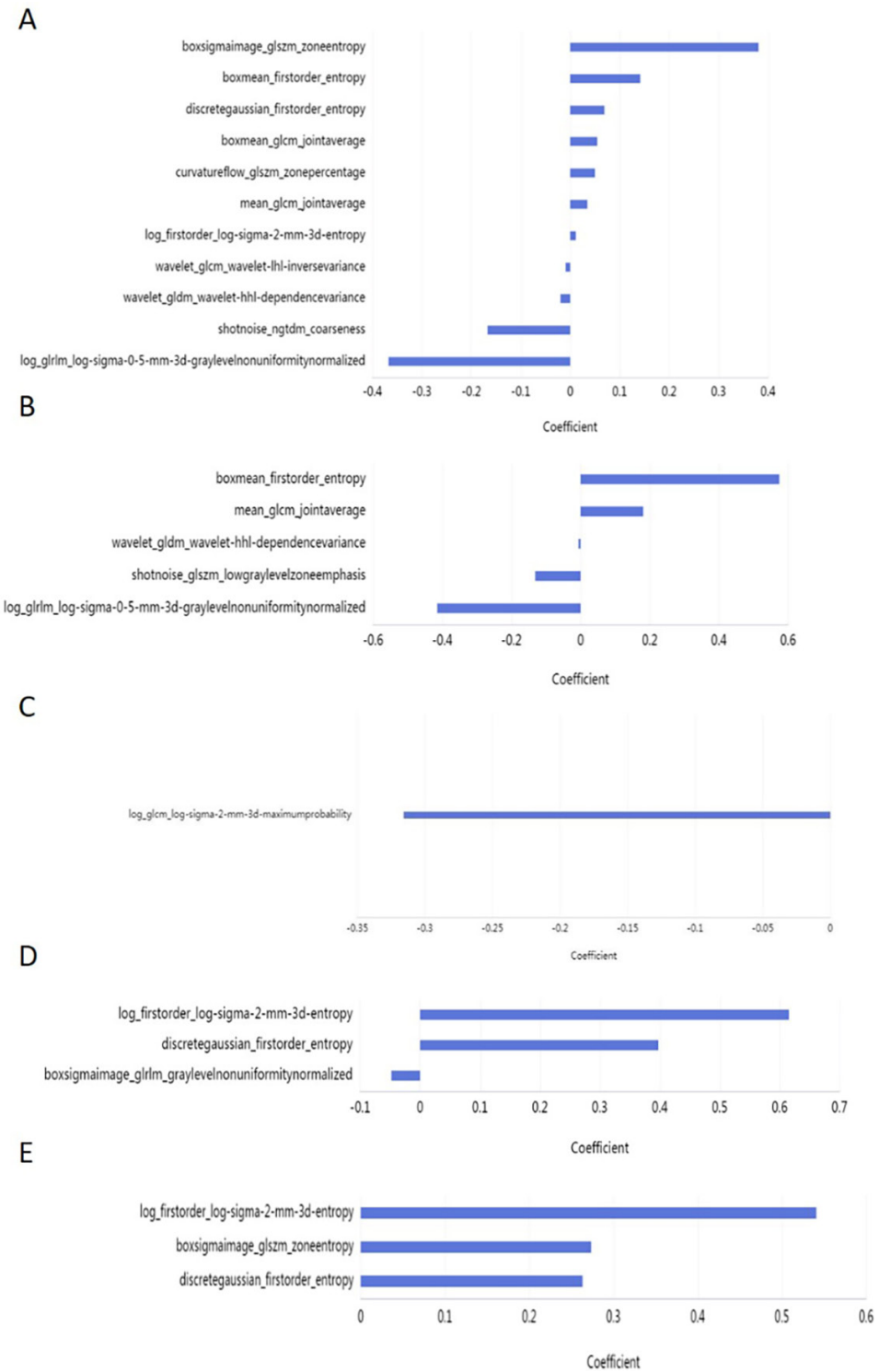

Figure S7 The selected radiomics features with non-zero coefficients. (A) The 11 features of model before optimization; (B) the 5 features of model with ComBat correction; (C-E) the features of models with image resampling correction of 0.5×0.5×0.5, 1.0×1.0×1.0, and 2.0×2.0×2.0 mm^3, respectively. 3D, three-dimensional; ComBat, combatting batch effect.

https://dx.doi.org/10.21037/qims-22-992