Appendix 1

Methods

Radiomics feature extraction

Feature extraction followed the Image Biomarker Standardization Initiative (IBSI) guideline (35) in this study.

To ensure data validity and accuracy, 2 radiologists (readers 1 and 2, with 6 and 12 years of chest diagnostic experience, respectively) independently performed manual segmentation of computed tomography (CT) images using ITK-SNAP software, and were blinded to the clinical and histological data. A senior radiologist with 20 years of experience confirmed the segmentation. Reader 1 (G.J.) segmented all training cases, and reader 2 (L.S.) segmented all validation cases. Reader 3 (W.F.) with 20 years of experience confirmed the segmentation when the 2 radiologists were uncertain. Regions of interest (ROI) were manually delineated on the CT lung window (width, 1500 HU; level, -500 HU), then the segmented regions delineated on each slice were merged to generate a volume of interest (VOI).

To assess the reproducibility and robustness of feature extraction, 1 month later, 40 patients in the training set were randomly selected and re-segmented by Reader 1 and Reader 2 to construct a re-segmentation set, and 40 patients were randomly selected from each CT scanner to construct different sets of CT scanners, which were used to calculate the intra-class/inter-class correlation coefficient (ICC), respectively.

The ICCs were calculated to assess the agreement of features extracted separately by 2 radiologists and different CT scanners, and all values were >0.75, reflecting good agreement.

In total, 1,727 radiomics features were extracted from each VOI of the CT images. All specific calculation formulas could be easily obtained in the open-source software package PyRadiomics 3.0.1 or previous studies (36). Here, we only listed several categories that these features could be divided into. Details of radiomics features were as follows:

(I) 16 shape features,
(II) 324 first order features,
(III) 1,387 texture features,
  (i) 418 gray-level co-occurrence matrices (GLCM) features,
  (ii) 304 gray-level run-length matrix (GLRLM) features,
  (iii) 304 gray-level size zone matrix (GLSZM) features,
  (iv) 95 neighboring gray tone difference matrix (NGTDM) features,
  (v) 266 gray-level dependence matrix (GLDM) features.

First order features and texture features were extracted from original pictures as well as 8 filters, including Wavelet filter, Laplacian of Gaussian (LoG) filter, Local Binary Pattern (LBP) 3D filter, Square filter, Square Root filter, Logarithm filter, Gradient filter, and Exponential filter. The shape features were extracted from original pictures.

Visualization of the deep learning model

Grad-CAM (23) uses the gradient of network back-propagation to calculate the weight of each channel of the feature map to obtain the heat-map. The weight calculation formula for Grad-CAM is as follows:

$$\alpha_k^c = \frac{1}{Z}\sum_i\sum_j \frac{\partial y^c}{\partial A_{ij}^k} \qquad [1]$$

Where α_k^c represents the weight, c represents the category, k represents the feature map, Z represents the size of the feature map (i.e., length × width), y^c is the logits corresponding to the category (before the softmax), A^k represents the feature map of the convolution output, and i and j represent the abscissa and ordinate of the feature map, respectively.

After obtaining the weights, the channel linear weights of the feature map were fused to obtain the heat-map. Grad-CAM adds an *ReLU* operation to the fused heat-map, reserving only the area with a positive effect on category c. The Grad-CAM fusion formula is as follows:

$$L_{\text{Grad-CAM}}^c = ReLU\left(\sum_k \alpha_k^c A^k\right) \qquad [2]$$

References

35. Zwanenburg A, Vallières M, Abdalah MA, Aerts HJWL, Andrearczyk V, Apte A, et al. The Image Biomarker Standardization Initiative: Standardized Quantitative Radiomics for High-Throughput Image-based Phenotyping. Radiology 2020;295:328-38.

https://dx.doi.org/10.21037/qims-24-33

36. van Griethuysen JJM, Fedorov A, Parmar C, Hosny A, Aucoin N, Narayan V, Beets-Tan RGH, Fillion-Robin JC, Pieper S, Aerts HJWL. Computational Radiomics System to Decode the Radiographic Phenotype. Cancer Res 2017;77:e104-7.

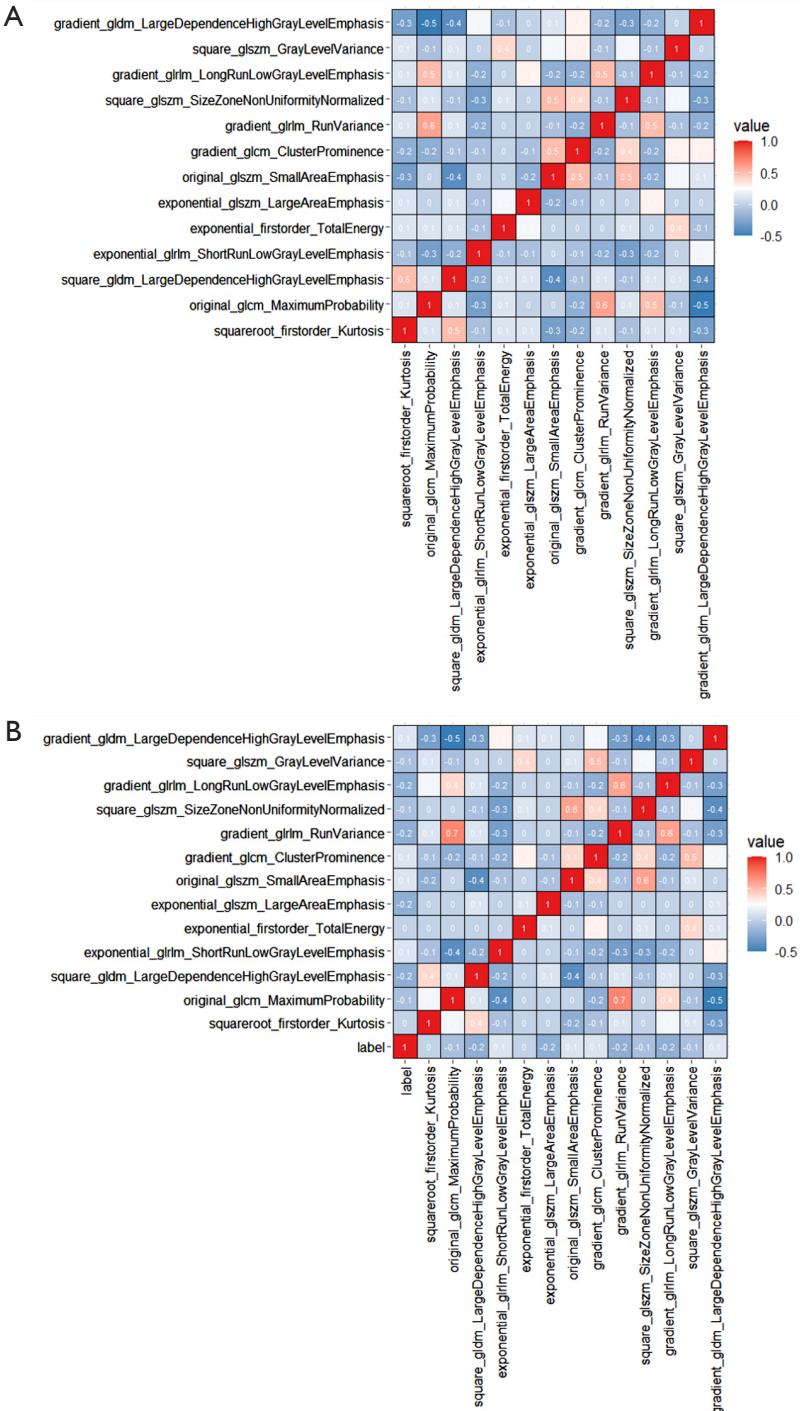

Figure S1 Related heat maps in training (A) and test (B) sets. Correlation analysis showed that the absolute value of the correlation of each feature between *EGFR* mutation group and wild-type group was less than 0.75. *EGFR*, epidermal growth factore receptor.

https://dx.doi.org/10.21037/qims-24-33

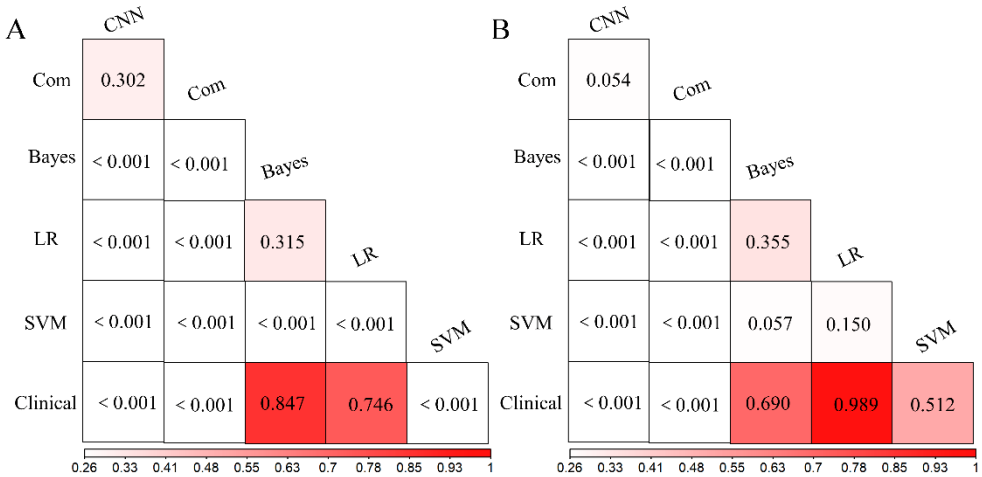

Figure S2 Delong test was used to compare the performance differences of different prediction models in predicting *EGFR* mutation status. (A) Training set; (B) test set. CNN, convolutional neural network; Com, comprehensive model; Bayes, naïve Bayes; LR, logistic regression; SVM, support vector machine; *EGFR*, epidermal growth factore receptor.

https://dx.doi.org/10.21037/qims-24-33