Supplementary

Appendix 1

Methods 1

PET/CT image acquisition parameters

Zhongda Hospital: A GE Discovery LS PET/CT scanner was used for the study. Scanning parameters: tube voltage: 140 kV, tube current: 80 mA, layer thickness: 3.75 mm, layer spacing: 5 mm. All patients were fasting for >6 h before the examination and had fasting blood glucose ≤11.10 mmol/L prior to intravenous ^{18}F-FDG administration at a dose of 5.55 MBq/kg body mass. Whole-body images were acquired between 50 minutes and 1 hour, with the CT localization scan performed first, followed by the PET scan. The PET scan covered the area from the top of the head to the upper part of the femur, with a total of 6–8 beds, and the acquisition time was 3 minutes per bed. The CT data were used for attenuation correction. The corrected PET images were reconstructed using the ordered subset maximum expectation method (OSEM).

Affiliated Drum Tower Hospital: Philips GEMINI GXL PET/CT scanner was used for the study. Scanning parameters: tube voltage: 120 kV, tube current: 120 mA, layer thickness: 5 mm, layer spacing: 5 mm. All patients were fasting for >6 h before the examination and had fasting blood glucose ≤10.0 mmol/L prior to intravenous ^{18}F-FDG administration at a dose of 5.18 MBq/kg body mass. Whole-body images were acquired between 45 minutes and 1 hour, with the CT localization scan performed first, followed by the PET scan. The PET scan covered the area from the top of the head to the middle and lower part of the femur, with a total of 9-10 beds, and the acquisition time was 2 minutes per bed. The CT data were used for attenuation correction. The corrected PET images were reconstructed using the ordered subset maximum expectation method (OSEM).

Shengjing Hospital tube voltage: GE Discovery Elite PET/CT scanner was used for the study. Scanning parameters: tube voltage: 140 kV, tube current: 180–240 mA, layer thickness: 3.75 mm, layer spacing: 5 mm. All patients were fasting for >6 h before the examination and had fasting blood glucose ≤11.10 mmol/L prior to intravenous 18F-FDG administration at a dose of 3.7–5.5 MBq/kg body mass. Whole-body images were acquired between 50 minutes and 1 hour, with the CT localization scan performed first, followed by the PET scan. The PET scan covered the area from the top of the head to the middle part of the femur, with a total of 6–7 beds, and the acquisition time was 1.5 minutes per bed. The CT data were used for attenuation correction. The corrected PET images were reconstructed using the ordered subset maximum expectation method (OSEM).

Tumour segmentation for radiomics

The first nuclear medicine physician, without knowledge of the clinicopathological information, manually drew the lymph node boundaries layer by layer on the axial images of the mediastinal window by importing the thin-layer CT images into 3D Slicer (version 5.0.3) to generate a three-dimensional volume of interest (3D-VOI). Whole-body PET and CT images were then imported into LIFEx software (version v7.3.6) using the fixed-threshold method with 40% of SUVmax as the threshold. The software automatically performed volumetric segmentation of lymph nodes in transverse, sagittal, and coronal planes. Images were resampled and normalized prior to feature extraction to mitigate the effects of different machines and machine scanning parameters, such as slice thickness and slice spacing, on CT images. Images were resampled 1×1×1 to minimize the effects of these variations. One month later, 50 lymph nodes were randomly selected by the primary nuclear medicine physician to be outlined again. The objective was to assess the similarity of the extracted radiomics features by calculating the Intraclass Correlation Coefficient (ICC).

Radiomics signature development

We extracted a total of 2632 features from PET and CT images of seven types: First-order statistics, Gray Level Cooccurrence Matrix, Gray Level Run Length Matrix, Gray Level Size Zone Matrix, Neighboring Gray Tone Difference Matrix, Gray Level Dependence Matrix, Shape-based features. We set the bin width to 25 for CT images and 0.1 for

https://dx.doi.org/10.21037/qims-24-100

PET images. For feature extraction, we used both primary and matched images. The derived images are then subjected to feature extraction using logarithmic, exponential, gradient, square, square root, and wavelet transform filters. Whereas eight decompositions obtain the wavelet transform image after wavelet filtering. Eight combinations are obtained by applying a high pass filter (H) or low pass filter (L) to the 3D image: HHH, HHL, HLL, HLH, HLL, LHL, LHH, LLL.

First-order statistics (N=18)

- Interquartile Range
- Skewness
- Uniformity
- Median
- Energy
- Robust Mean Absolute Deviation
- Mean Absolute Deviation
- Total Energy
- Maximum
- Root Mean Squared
- 90Percentile
- Minimum
- Entropy
- Range
- Variance
- 10Percentile
- Kurtosis
- Mean

High-order statistics

(I) Gray Level Cooccurrence Matrix (N=24)

- Joint Average
- Joint Entropy
- Cluster Shade
- Maximum Probability
- Idmn
- Joint Energy
- Contrast
- Difference Entropy
- Inverse Variance
- Difference Variance
- Idn
- Idm
- Correlation
- Autocorrelation
- Sum Entropy
- Sum of Squares
- Cluster Prominence
- Imc2
- Imc1
- Difference Average
- Id

https://dx.doi.org/10.21037/qims-24-100

- Cluster Tendency
- Sum Average
- MCC

(II) Gray Level Size Zone Matrix (N=16)

- Gray Level Variance
- Zone Variance
- Gray Level Non-Uniformity Normalized
- Size Zone Non-Uniformity Normalized
- Size Zone Non-Uniformity
- Gray Level Non-Uniformity
- Large Area Emphasis
- Small Area High Gray Level Emphasis
- Zone Percentage
- Large Area Low Gray Level Emphasis
- Large Area High Gray Level Emphasis
- High Gray Level Zone Emphasis
- Small Area Emphasis
- Low Gray Level Zone Emphasis
- Zone Entropy
- Small Area Low Gray Level Emphasis

(III) Gray Level Run Length Matrix (N=16)

- Short Run Low Gray Level Emphasis
- Gray Level Variance
- Low Gray Level Run Emphasis
- Gray Level Non-Uniformity Normalized
- Run Variance
- Gray Level Non-Uniformity
- Long Run Emphasis
- Short Run High Gray Level Emphasis
- Run Length Non-Uniformity
- Short Run Emphasis
- Long Run High Gray Level Emphasis
- Run Percentage
- Long Run Low Gray Level Emphasis
- Run Entropy
- High Gray Level Run Emphasis
- Run Length Non-Uniformity Normalized

(IV) Neighboring Gray Tone Difference Matrix (N=5)

- Coarseness
- Complexity
- Strength
- Contrast
- Busyness

(V) Gray Level Dependence Matrix (N=14)

- Gray Level Variance
- High Gray Level Emphasis
- Dependence Entropy
- Dependence Non-Uniformity

https://dx.doi.org/10.21037/qims-24-100

- Gray Level Non-Uniformity
- Small Dependence Emphasis
- Small Dependence High Gray Level Emphasis
- Dependence Non-Uniformity Normalized
- Large Dependence Emphasis
- Large Dependence Low Gray Level Emphasis
- Dependence Variance
- Large Dependence High Gray Level Emphasis
- Small Dependence Low Gray Level Emphasis
- Low Gray Level Emphasis

Shape-based features (N=14)

- Maximum 2D Diameter Slice
- Sphericity
- Minor Axis
- Compactness1
- Elongation
- Surface Volume Ratio
- Volume
- Maximum 3D Diameter
- Major Axis
- Surface Area
- Flatness
- Least Axis
- Maximum 2D Diameter Column
- Maximum 2D Diameter Row

Methods 2

The architecture of the Resnet18

The Resnet18 deep learning model (22,46) is an end-to-end deep convolutional neural network (DCNN) architecture consisting of 18 layers. The architecture comprises an input layer, convolutional layer, batch normalization layer, ReLU activation function layer, and maximum pooling layer. The input data size is 3×224×224. The Resnet18 network structure contains a significant amount of residual blocks. Resnet18 contrasts the collaborative learning method used in AlexNet, VGG, and other classical network structures. It employs two layers of local convolutional blocks, succeeded by a residual block layer, and ultimately a fully connected layer for image classification. To minimize overfitting of the model, we included a dropout layer and finally calculated the probability of lymph node metastasis risk in a linear layer using a softmax function.

Radiomics Signature Selection and Modelling

First, we selected histological features (12) with ICC above 0.75, resulting in 1655 stable features in total. After standardizing the characteristics of the training and test sets, we harmonized the characteristics of the three centers using the COMBAT coordination approach. A machine learning (47) algorithm called Minimum Redundancy Maximum Relevance (mRMR) (12,48) filters the harmonized features to eliminate redundancies. To reduce overfitting, the filtered characteristics underwent a 10-fold cross-validated Least Absolute Shrinkage and Selection Operator (LASSO) logistic regression analysis in the training group. The analysis screened for features with correlation coefficients that could discern between mediastinal lymph nodes' benign and malignant nature that were not zero. Finally, 17 features, including PET and CT images, were selected. Of

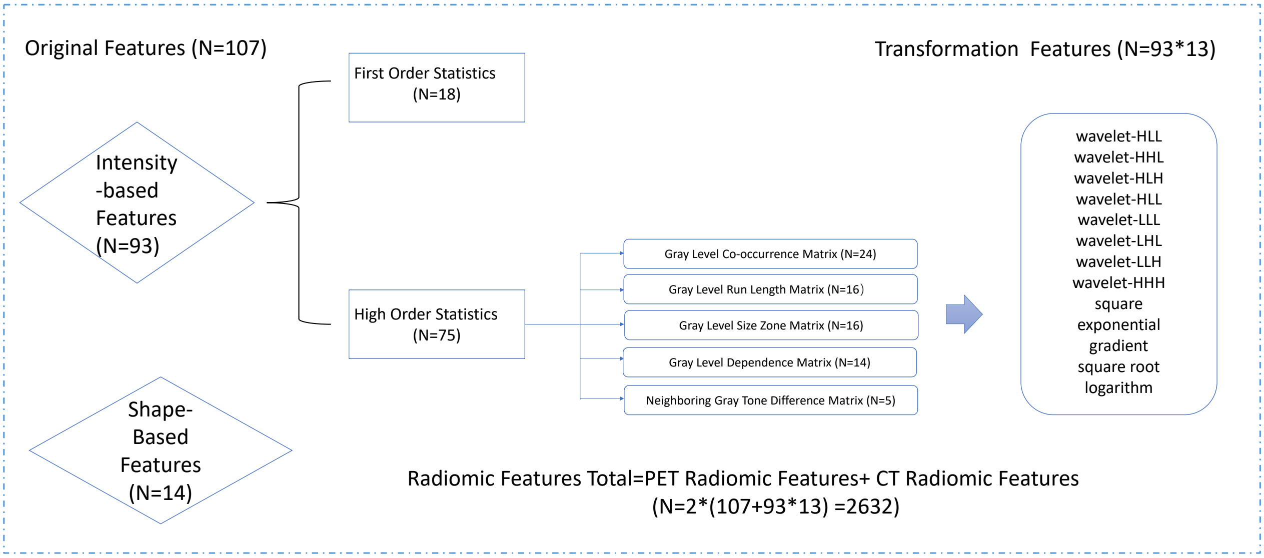
Original Features (N=107)
Transformation Features (N=93*13)
Intensity -based Features (N=93)
First Order Statistics (N=18)
High Order Statistics (N=75)
Gray Level Co-occurrence Matrix (N=24)
Gray Level Run Length Matrix (N=16)
Gray Level Size Zone Matrix (N=16)
Gray Level Dependence Matrix (N=14)
Neighboring Gray Tone Difference Matrix (N=5)
wavelet-HLL
wavelet-HHL
wavelet-HLH
wavelet-HLL
wavelet-LLL
wavelet-LHL
wavelet-LLH
wavelet-HHH
square
exponential
gradient
square root
logarithm
Shape-Based Features (N=14)
Radiomic Features Total=PET Radiomic Features+ CT Radiomic Features
(N=2*(107+93*13) =2632)

Figure S1 Details of extracted radiomics features.

Table S1 The pairwise correlation evaluation of PET/CT clinical parameters using Spearman correlation coefficient

Parameters	SAD	SUVmax	SUVavg	SUVpeak	MTV	TLG	Calcification
SAD	1.000	0.516	0.517	0.555	0.506	0.614	0.035
SUVmax	0.516	1.000	0.974	0.852	0.461	0.783	–0.004
SUVavg	0.517	0.974	1.000	0.8333	0.430	0.771	0.003
SUVpeak	0.555	0.852	0.833	1.000	0.568	0.785	0.041
MTV	0.506	0.461	0.430	0.568	1.000	0.898	0.024
TLG	0.614	0.783	0.771	0.785	0.898	1.000	0.026
Calcification	0.035	–0.004	0.003	0.041	0.024	0.026	1.000

PET/CT, positron emission tomography/computed tomography; SAD, short-axis diameter; SUVmax, maximal standard uptake value; SUVavg, mean standardized uptake value; SUVmin, minimal standard uptake value; SUVpeak, peak standardized uptake value; MTV, metabolic tumor volume; TLG, total lesion glycolysis.

these, 6 features were identified from PET images, and 11 were identified from CT images. The radiomic score (Rad-Score) for each lymph node was calculated by the linear combination of the filtered features after individual weighting by their coefficients. The Rad-Score formula is as follows:

Rad-Score= 0.26630411
+0.02757353×PET.wavelet.HLH_glrlm_LongRunLowGrayLevelEmphasis
+0.34770776×CT.exponential_firstorder_Skewness
-0.02168469×CT.wavelet.LHL_gldm_DependenceVariance
-0.21287140×CT.wavelet.LHL_firstorder_InterquartileRange
+0.03656187×PET.wavelet.HHL_glcm_MCC
+0.07056995×CT.original_gldm_LargeDependenceHighGrayLevelEmphasis
-0.22968557×CT.gradient_firstorder_InterquartileRange
+0.02477114×PET.wavelet.HHH_glcm_SumEntropy

Table S2 Model performance of 5 machine learning algorithms in training and validation sets

Model	AUC (95% CI)	Sensitivity	Specificity
Training set			
GBDT-LR	92.17% (89.04–95.30%)	87.3%	89.8%
LR	86.19% (82.26–90.12%)	72.6%	85.0%
NB	80.54% (75.82–85.25%)	85.4%	65.9%
DT	100% (100.00–100.00%)	100.0%	100.0%
SVM	99.24% (98.70–99.77%)	93.0%	97.6%
Internal test set			
GBDT-LR	84.55% (76.15–92.96%)	84.6%	73.8%
LR	80.83% (71.25–90.41%)	84.6%	71.4%
NB	78.69% (68.50–88.89%)	69.0%	84.6%
DT	70.7% (60.88–80.52%)	79.5%	61.9%
SVM	80.28% (70.56–90.00%)	71.8%	81.0%
External test set			
GBDT-LR	84.57% (76.98–92.17%)	84.9%	74.5%
LR	78.06% (68.98–87.14%)	73.6%	76.5%
NB	68.96% (58.73–79.19%)	73.6%	58.8%
DT	59.25% (51.87–66.63%)	28.3%	90.2%
SVM	71.59% (61.69–81.49%)	75.5%	62.7%

AUC, area under the curve; CI, confidence interval; GBDT-LR, Gradient Boosting Decision Tree-Logistic Regression; LR, logistic regression; NB, Naive Bayes model; DT, decision tree; SVM, support vector machine.

+0.09788875×PET.wavelet.HLH_glszm_HighGrayLevelZoneEmphasis
-0.08838097×PET.diagnostics_Image.interpolated_Minimum
-0.20688460×CT.wavelet.LLH_gldm_DependenceVariance
-0.27071977×CT.wavelet.LHL_gldm_DependenceEntropy
+0.09883175×CT.wavelet.LHL_firstorder_Median
-0.17084967×CT.wavelet.LHL_firstorder_Skewness
+0.18283536×CT.wavelet.HLL_glcm_MCC
+0.30407387×CT.original_shape_Flatness
+0.44781302×PET.wavelet.HHL_glszm_HighGrayLevelZoneEmphasis

To enable the comparison of the classification performance of multiple machine learning algorithms in identifying benign and malignant mediastinal lymph nodes, a variety of additional machine learning models were constructed, including Support Vector Machines (SVMs), Logistic Regression, Decision Trees (DTs), Naive Bayes (NB), and Gradient Boosting Tree-Logistic Regression (GBDT-LR). The GBDT-LR algorithm divides the training set into two equally sized parts. The data from training set 1 trains the GBDT model, while data from training set 2 prevents overfitting by being fed into the trained GBDT model. The discrete vectors of the feature combinations are then obtained and fed into the logistic regression model with the original features for training. Finally, the internal validation set and the external validation set are provided in the model to obtain the prediction results. As shown in *Table S2*.

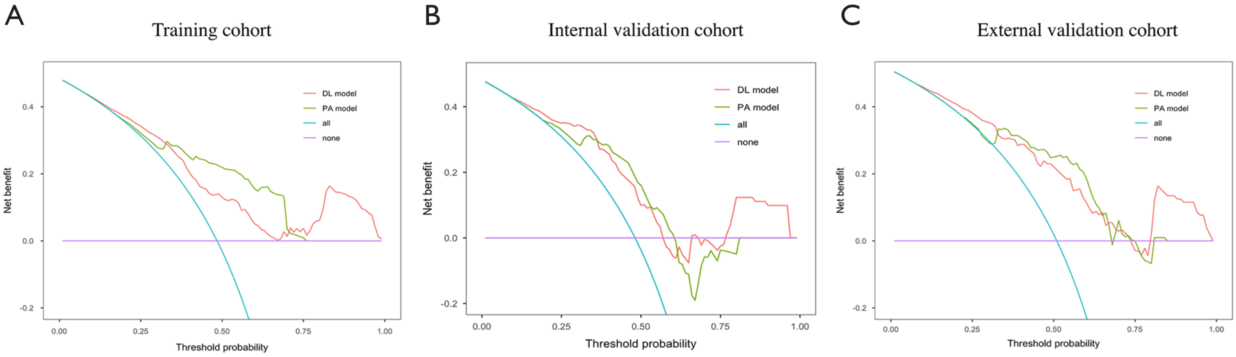

Figure S2 Decision curve analyses. Decision curve analyses of DL model, PA model in training cohort (A), internal validation cohort (B), and external validation cohort (C). DL, deep learning; PA, parametric.

References

46. Tran KA, Kondrashova O, Bradley A, Williams ED, Pearson JV, Waddell N. Deep learning in cancer diagnosis, prognosis and treatment selection. Genome Med 2021;13:152.
47. Yoo J, Cheon M, Park YJ, Hyun SH, Zo JI, Um SW, Won HH, Lee KH, Kim BT, Choi JY. Machine learning-based diagnostic method of pre-therapeutic (18)F-FDG PET/CT for evaluating mediastinal lymph nodes in non-small cell lung cancer. Eur Radiol 2021;31:4184-94.
48. Nambu A, Kato S, Sato Y, Okuwaki H, Nishikawa K, Saito A, Matsumoto K, Ichikawa T, Araki T. Relationship between maximum standardized uptake value (SUVmax) of lung cancer and lymph node metastasis on FDG-PET. Ann Nucl Med 2009;23:269-75.