Appendix 1 Materials and methods

Radiomics predictive model

Radiomics feature extraction

To mitigate the effects of varying acquisition parameters from different CT devices on the stability of radiomics features, we performed normalization of the original DICOM images and segmentation results based on pixel spacing and slice thickness. The CT images were then standardized using z-score normalization to achieve a normal distribution of image intensities.

We extracted a total of 1,454 radiomics features from each Region of Interest (ROI), which included: First-order features, Shape-based features, Gray Level Co-occurrence Matrix (GLCM) features, Gray Level Dependence Matrix (GLDM) features, Gray Level Run Length Matrix (GLRLM) features, Gray Level Size Zone Matrix (GLSZM) features, Neighbouring Gray Tone Difference Matrix (NGTDM) features. Shape-based features were extracted exclusively from the original images, while the remaining six sets of features were derived from both the original and processed images. The tasks of normalizing the images and ROIs, as well as extracting the radiomics features, were conducted using PyCharm (version 2019.1.3). For a detailed description of each radiomic feature, please refer to PyRadiomics.

Feature selection

We combined the extracted radiomics features with feature-vectorized clinical data. Each feature underwent a significance test to identify those with high predictive power, retaining only features with a P value less than 0.05. Next, we conducted pairwise comparisons of all retained features. If two features had a Pearson correlation coefficient greater than 0.85, the feature with the smaller P value from the significance test was selected for further processing. Following this, we applied the Least Absolute Shrinkage and Selection Operator (LASSO) with 5-fold cross-validation to identify features with non-zero coefficients from the entire set of extracted radiomics features. In this study, the optimal λ value was determined by the lowest binomial deviation, allowing us to retain a relatively small and refined set of features for the subsequent model fitting in the training cohort.

Model building

After selecting the radiomics features, we built a radiomics model using the training cohort and applied a five-fold cross-validation technique to ensure robustness. This model was then tested on an independent testing cohort. To construct the final prediction model, we employed a support vector machine (SVM) classifier.

Deep learning predictive model

Data preprocessing

To address variations in pixel spacing and slice thickness, we linearly interpolated the CT images into 3D isotropic images with a voxel spacing of 0.6×0.6×0.6 mm^3. Extensive data augmentation was performed on all initially generated image patches containing the nodules, each sized 85×85×85. This augmentation included random rotations from 0 to 360 degrees, random zooming in and out, and random cropping and flipping. The processed patches were then resized to 64×64 64 for training the algorithm.

DL model

The DL model consisted 3 branches: global feature branch, medium feature branch and local feature branch.

Discriminative filter learning (DFL): this module is employed in both the medium feature branch and the local feature

https://dx.doi.org/10.21037/tlcr-24-1039

branch. It aims to detect discriminative local features for fine-grained classification, achieving multi-scale learning to enhance the recognition ability of image patches.

Global feature branch: this branch seeks to extract global contextual information from the entire input, constructed using a backbone model. We use a 3D Dense-Net as the backbone model for FGP-NET. Compared to the original 2D network architecture, as shown in the figure, all convolutional and pooling layers are modified to 3D. Additionally, the first convolutional layer is replaced with a 3×3×3 convolutional layer, and the pooling layer in the third Dense Block is removed, resulting in only three down-sampling operations through the backbone.

Medium feature branch: this branch is designed to utilize features at a medium resolution. It starts from the feature map after the third Dense Block, completing two down-sampling operations. The feature map is then input into the DFL module to extract local features with a volume of 4×4×4 pixels. Consequently, we apply this to our model to detect the most discriminative parts of the nodules, such as spiculation and lobulation.

Local feature branch: this branch focuses on capturing local details of the given nodule. It has two inputs: one is the feature map from the DFL module of the medium feature branch, and the other is the feature map from the global feature branch after the second Dense Block, which has undergone one down-sampling operation. These two inputs are concatenated, followed by an additional Dense Block. Finally, the DFL module is used on top of these to extract local features with a volume of 2×2×2 pixels.

Calculation of the loss

The loss of the entire network is calculated as follows.

$$Loss = Loss_{global} + 0.5 \times Loss_{medium} + 0.5 \times Loss_{local}$$

Where L_global is the loss function constructed from the global feature branch, L_medium is the loss function constructed from the medium feature branch, L_local is the loss function constructed from the local feature branch. All functions are in the form of cross-entropy loss.

https://dx.doi.org/10.21037/tlcr-24-1039