Supplementary

## Appendix 1

Univariate and multivariate analysis

| Variables | Univariate analysis | | | | | Multivariate analysis | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | β | S.E | Z | P | OR (95% CI) | β | S.E | Z | P | OR (95% CI) |
| Spiculation | | | | | | | | | | |
| 0 | | | | | 1.00 (Reference) | | | | | 1.00 (Reference) |
| 1 | 2.00 | 0.28 | 7.04 | <0.001 | 7.36 (4.22–12.84) | 1.08 | 0.36 | 3.05 | 0.002 | 2.96 (1.47–5.94) |
| Lobulation | | | | | | | | | | |
| 0 | | | | | 1.00 (Reference) | | | | | 1.00 (Reference) |
| 1 | 0.52 | 0.29 | 1.77 | 0.076 | 1.68 (0.95–2.98) | −0.76 | 0.42 | −1.82 | 0.068 | 0.47 (0.21–1.06) |
| Vacuole sign | | | | | | | | | | |
| 0 | | | | | 1.00 (Reference) | | | | | 1.00 (Reference) |
| 1 | 0.72 | 0.27 | 2.69 | 0.007 | 2.06 (1.22–3.50) | 0.75 | 0.35 | 2.16 | 0.031 | 2.13 (1.07–4.21) |
| Pleural Indentation | | | | | | | | | | |
| 0 | | | | | 1.00 (Reference) | | | | | 1.00 (Reference) |
| 1 | 1.14 | 0.22 | 5.24 | <0.001 | 3.11 (2.04–4.76) | −0.06 | 0.30 | −0.20 | 0.840 | 0.94 (0.52–1.69) |
| Pericystic emphysema | | | | | | | | | | |
| 0 | | | | | 1.00 (Reference) | | | | | 1.00 (Reference) |
| 1 | 1.04 | 0.50 | 2.09 | 0.037 | 2.82 (1.07–7.44) | 0.01 | 0.83 | 0.01 | 0.993 | 1.01 (0.20–5.11) |
| Tumor-lung interface | | | | | | | | | | |
| 0 | | | | | 1.00 (Reference) | | | | | 1.00 (Reference) |
| 1 | −0.65 | 0.45 | −1.46 | 0.143 | 0.52 (0.22–1.25) | −0.60 | 0.55 | −1.09 | 0.275 | 0.55 (0.18–1.62) |
| Gender | | | | | | | | | | |
| 0 | | | | | 1.00 (Reference) | | | | | 1.00 (Reference) |
| 1 | 0.76 | 0.21 | 3.57 | <0.001 | 2.14 (1.41–3.24) | 0.53 | 0.28 | 1.90 | 0.058 | 1.70 (0.98–2.96) |
| Density | | | | | | | | | | |
| 1 | | | | | 1.00 (Reference) | | | | | 1.00 (Reference) |
| 2 | −2.14 | 0.29 | −7.45 | <0.001 | 0.12 (0.07–0.21) | −1.81 | 0.32 | −5.58 | <0.001 | 0.16 (0.09–0.31) |
| 3 | −3.73 | 0.42 | −8.89 | <0.001 | 0.02 (0.01–0.05) | −2.98 | 0.51 | −5.84 | <0.001 | 0.05 (0.02–0.14) |
| Air bronchogram | | | | | | | | | | |
| 0 | | | | | 1.00 (Reference) | | | | | 1.00 (Reference) |
| 1 | 1.01 | 0.21 | 4.72 | <0.001 | 2.76 (1.81–4.20) | 0.47 | 0.30 | 1.59 | 0.112 | 1.60 (0.90–2.86) |
| Age | 0.06 | 0.01 | 5.42 | <0.001 | 1.06 (1.04–1.08) | 0.03 | 0.01 | 2.29 | 0.022 | 1.03 (1.01–1.06) |
| Maximum diameter | 0.09 | 0.01 | 7.06 | <0.001 | 1.10 (1.07–1.13) | 0.03 | 0.02 | 1.91 | 0.056 | 1.03 (1.00–1.06) |

OR, odds ratio; CI, confidence interval.


https://dx.doi.org/10.21037/tlcr-24-646

## Appendix 2

### *3D DCNN model*

Four 3D DCNN architectures were employed in the DCNN models to assess the utilization and integration of various data modalities. These architectures encompassed 3D ResNet18 (*Figure S1*), 3D ResNet50 (*Figure S2*), 3D DenseNet121 (*Figure S3*), and 3D DenseNet201 (*Figure S4*). During the optimization of each model, a batch size of 16 was maintained for training purposes. The models' parameters were iteratively updated using the Adam optimizer with exponential decay. The training process was terminated after 150 epochs. Additionally, the dropout layer parameter within the network was 0.5.

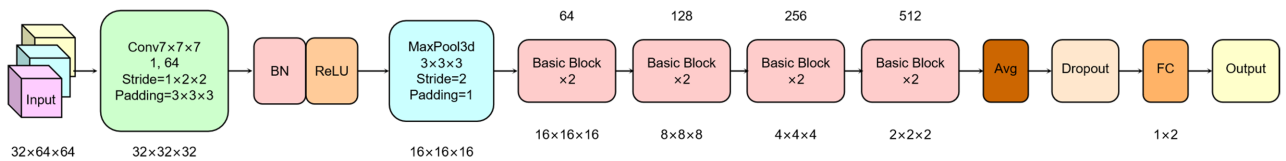


**Figure S1** Model structure of 3D ResNet18.

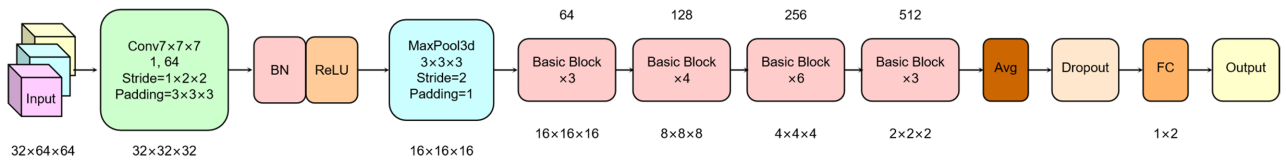


**Figure S2** Model structure of 3D ResNet50.

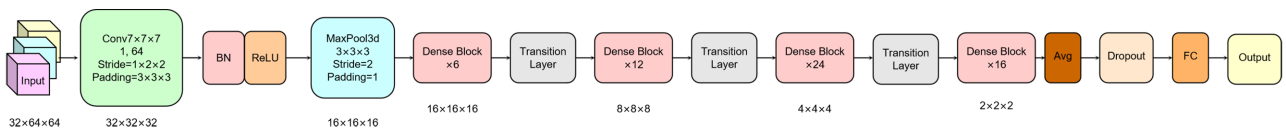


**Figure S3** Model structure of 3D DenseNet121.

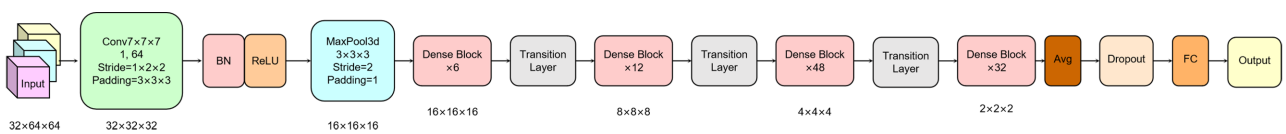


**Figure S4** Model structure of 3D DenseNet201.

### *2.5D DCNN Model*

Training 3D networks with limited medical imaging data is prone to overfitting, while training 2D networks lacks innovation and may lose their three-dimensional information. Here, the 2.5D network specifically refers to a way of representing the three-dimensional information of the dataset in the form of three channels, training a 2D network, and avoiding overfitting to some extent (31).


https://dx.doi.org/10.21037/tlcr-24-646

The model adopts the 2D DenseNet121 network (*Figure S5*). During the optimization process, this study trains with a batch size of 16, and uses the Adam optimizer to update the model parameters by reducing the learning rate by 0.8 times when the loss function does not decrease for 5 consecutive epochs. The training was stopped after 150 epochs.

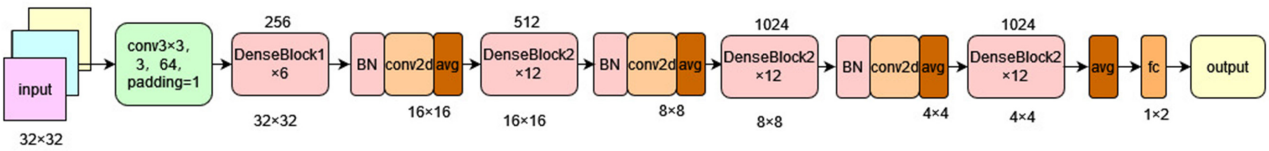


**Figure S5** Model structure of 2.5D DenseNet.

### *Integrating Clinical Features into Deep Learning Models*

The research aims to incorporate clinical features into deep learning models to achieve more accurate classification results. However, integrating clinical features into the 3DResNet basicblock is complex and requires significant effort. Therefore, the experiment chooses to integrate clinical features in the feature space. Specifically, there are two methods: concatenation and bias (32).

In the concatenation method (*Figure S6*), clinical features are concatenated with the feature vectors of deep learning after the dropout layer. The concatenated features are then input into the fully connected layer for classification. Clinical feature data values include variables (such as age, maximum diameter, etc.) and Boolean values (gender, cavitation sign, etc.), with significant differences in distribution. Directly inputting these features into the network without passing through the batch normalization (bn) layer can easily affect the classification results, leading to predictions that are all 0 or all 1.

Another method is to transform the clinical features into the same size as the deep learning feature vector through a fully connected layer, and then add the clinical features to the deep learning features in the form of bias by addition, and finally output the model classification results through subsequent dropout layers, fully connected layers, etc. (*Figure S7*).

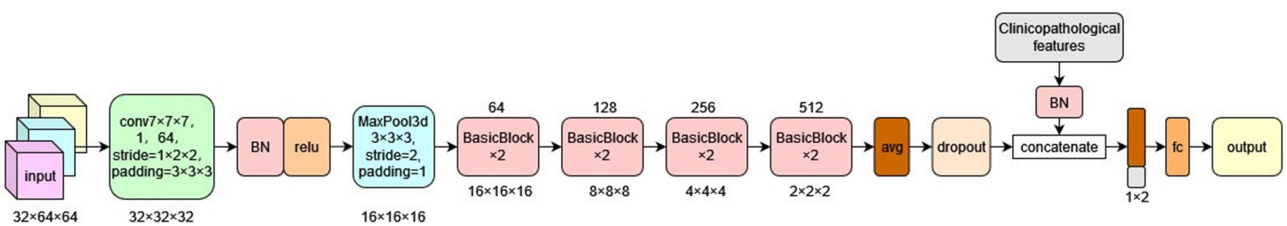


**Figure S6** Concatenation method.

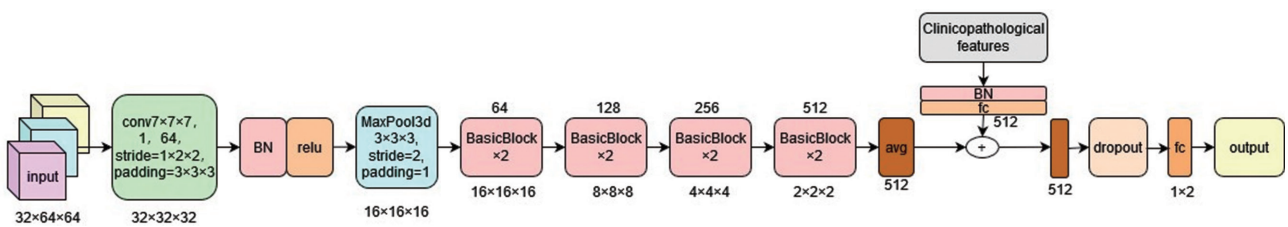


**Figure S7** Bias method.


https://dx.doi.org/10.21037/tlcr-24-646

### *Comparison of deep learning and machine learning in predicting STAS*

The ROC curves of the models besides those with highest performance for both deep learning and machine learning are presented in *Figure S8*.

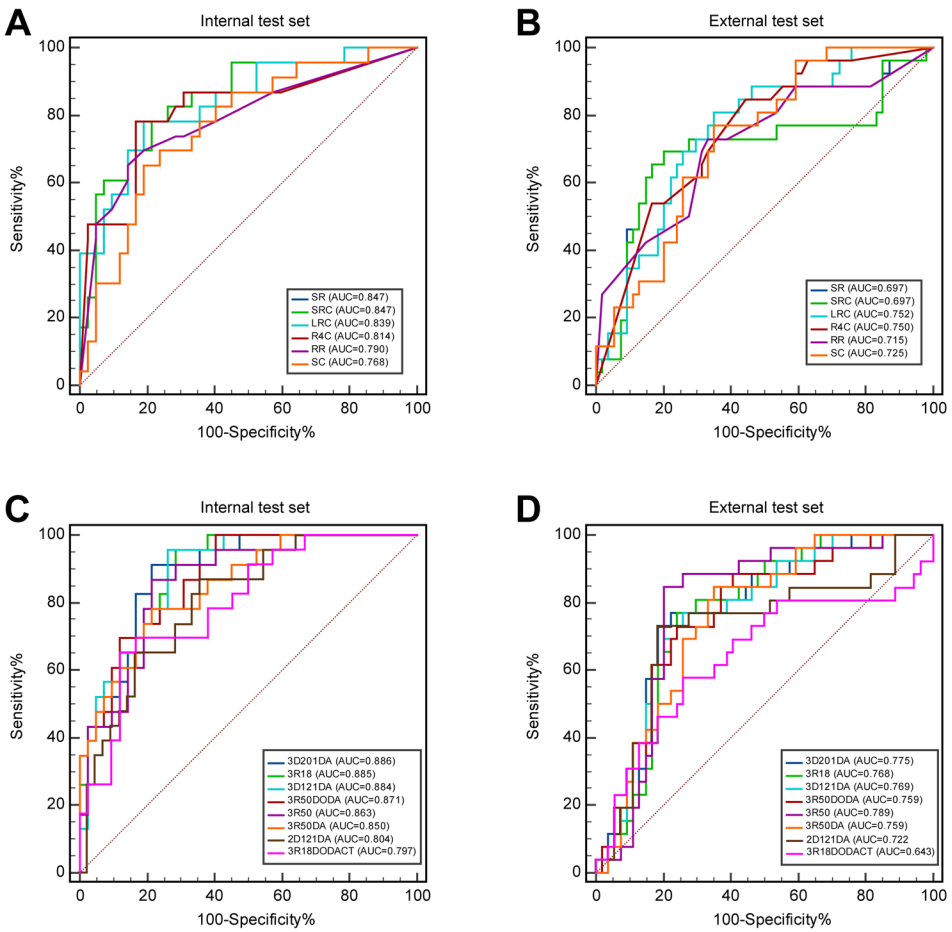


**Figure S8** Performances for spread through air spaces (STAS) prediction. The receiver operating characteristic (ROC) curves of high-performance machine learning models in internal test set (A), external test set (B), and high-performance deep learning models in internal test set (C), and external test set (D). AUC indicates area under the curve.


https://dx.doi.org/10.21037/tlcr-24-646

### *Comparison of AI performance with that of human readers*

To further evaluate the performance of our algorithm, we conducted a comprehensive comparison with human readers. Two radiologists with 6 years of experience were invited to evaluate 80 chest CT scan randomly selected from test set, and we compared the results with the gold standard. *Figure S9* illustrates the confusion matrix, Cohen's kappa and accuracy analysis for the two readers, 3R18DA (Model 1), 3R18DODA4CC (Model 2), 3R18DODA4CB (Model 3) in relation to the gold standard. High agreement values of κ=0.4844 and κ=0.4196 were found in Model 1 and Reader 2 respectively, which demonstrates a moderate agreement between our model and gold standard (*Figure S9B*). However, Model 2, Model 3 and Reader 1 showed only weak agreement values with the gold standard with κ=0.3833, 0.3957 and 0.4195, respectively. The accuracy analysis showed similar trends as depicted in *Figure S8C*. The best model demonstrated substantial discriminatory accuracy of 0.7625 and Cohen's kappa values of 0.4844, superior to both the two readers. The other two models also established comparable results to Reader 2 and superior results to Reader 1. The results indicate that AI-based algorithm has the potential to serve as a reliable tool for STAS diagnosis.

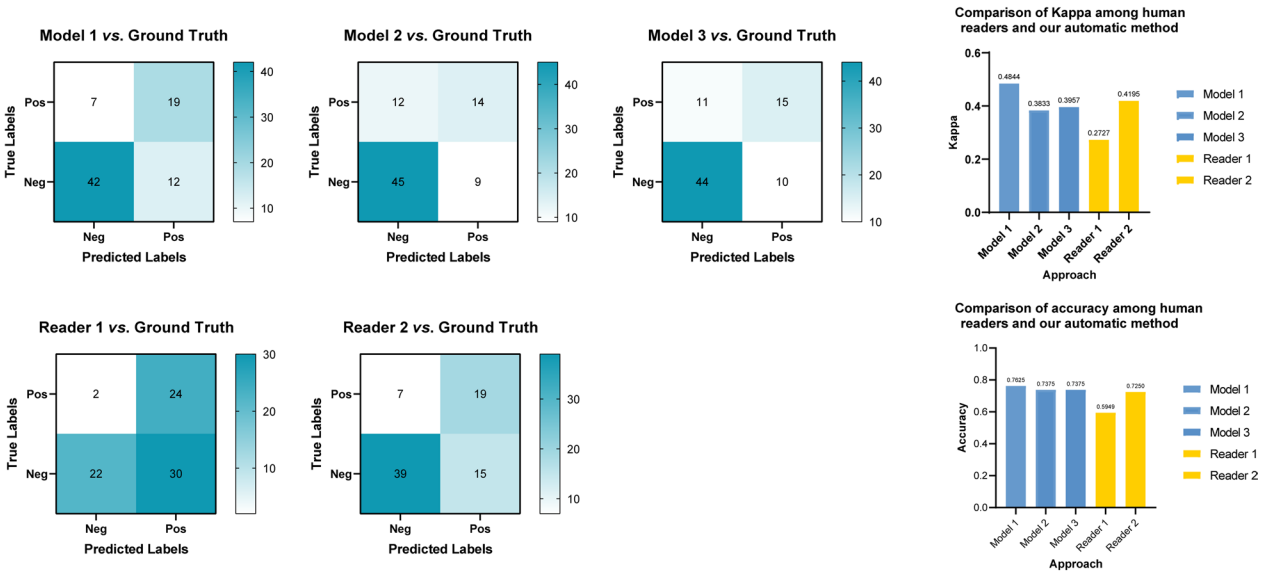


**Figure S9** Confusion matrix of the human experts' labels and Kappa visualization to compare the performance of human experts and our model. (A) Confusion matrix of the human experts' labels, illustrating the distribution of the human labeling. (B) Visualization of the kappa to compare the performance of human experts and our model. (C) Visualization of the accuracy to compare the performance of human experts and our model.


https://dx.doi.org/10.21037/tlcr-24-646

### *Abbreviations*

All the abbreviations and counterparts of models are provided in *Table S1*.

**Table S1** Abbreviations and counterparts of different models

| | Abbreviation | Counterpart |
|---|---|---|
| 1 | LRC | Logistic Regression + Radiomics + Clinical data |
| 2 | LR | Logistic Regression + Radiomics |
| 3 | LC | Logistic Regression + Clinical data |
| 4 | L4C | Logistic Regression + Four Clinical data |
| 5 | RRC | Random Forest + Radiomics + Clinical data |
| 6 | RR | Random Forest + Radiomics |
| 7 | RC | Random Forest + Clinical data |
| 8 | R4C | Random Forest + Four Clinical data |
| 9 | SRC | SVM + Radiomics + Clinical data |
| 10 | SR | SVM + Radiomics |
| 11 | SC | SVM + Clinical data |
| 12 | S4C | SVM + Four Clinical data |
| 13 | 2D121DA | 2.5d densenet121 + Data augmentation |
| 14 | 3D121DA | 3D Densenet121 + Data augmentation |
| 15 | 3D201DA | 3D Densenet201 + Data augmentation |
| 16 | 3R18DODA4CB | 3D Resnet18 + dropout 0.5 + Data augmentation + Four Clinical data (bias) |
| 17 | 3R18DODA4CC | 3D Resnet18 + dropout 0.5 + Data augmentation + Four Clinical data (concatenate) |
| 18 | 3R18DODA11CB | 3D Resnet18 + dropout 0.5 + Data augmentation + 11 Clinical data (bias) |
| 19 | 3R18DODA11CC | 3D Resnet18 + dropout 0.5 + Data augmentation + 11 Clinical data (concatenate) |
| 20 | 3R18DODA | 3D Resnet18 + dropout 0.5 + Data augmentation |
| 21 | 3R18 | 3D resnet18 |
| 22 | 3R18DA | 3D Resnet18 + Data augmentation |
| 23 | 3R50DODA | 3D Resnet50 + dropout 0.5 + Data augmentation |
| 24 | 3R50 | 3D resnet50 |
| 25 | 3R50DA | 3D Resnet50 + Data augmentation |

## References


31. Kim H, Lee D, Cho WS, et al. CT-based deep learning model to differentiate invasive pulmonary adenocarcinomas appearing as subsolid nodules among surgical candidates: comparison of the diagnostic performance with a size-based logistic model and radiologists. Eur Radiol 2020;30:3295-305.
32. Dumoulin V, Perez E, Schucher N, et al. Feature-wise transformations. Distill 2018. doi: 10.23915/distill.00011.