



A CT Radiomics Model for Classifying Pulmonary Metastases and Predicting Their Primary Origin

Ruichen Cui, MD; Xiaohu Hao, MD; Xiaolong Zhang, MD; Jiahan Cheng, MD; Qiang Pu, MD*

West China Hospital, Sichuan University, Chengdu, Sichuan, China | *Correspondence: puqjiang100@163.com



BACKGROUND & OBJECTIVE

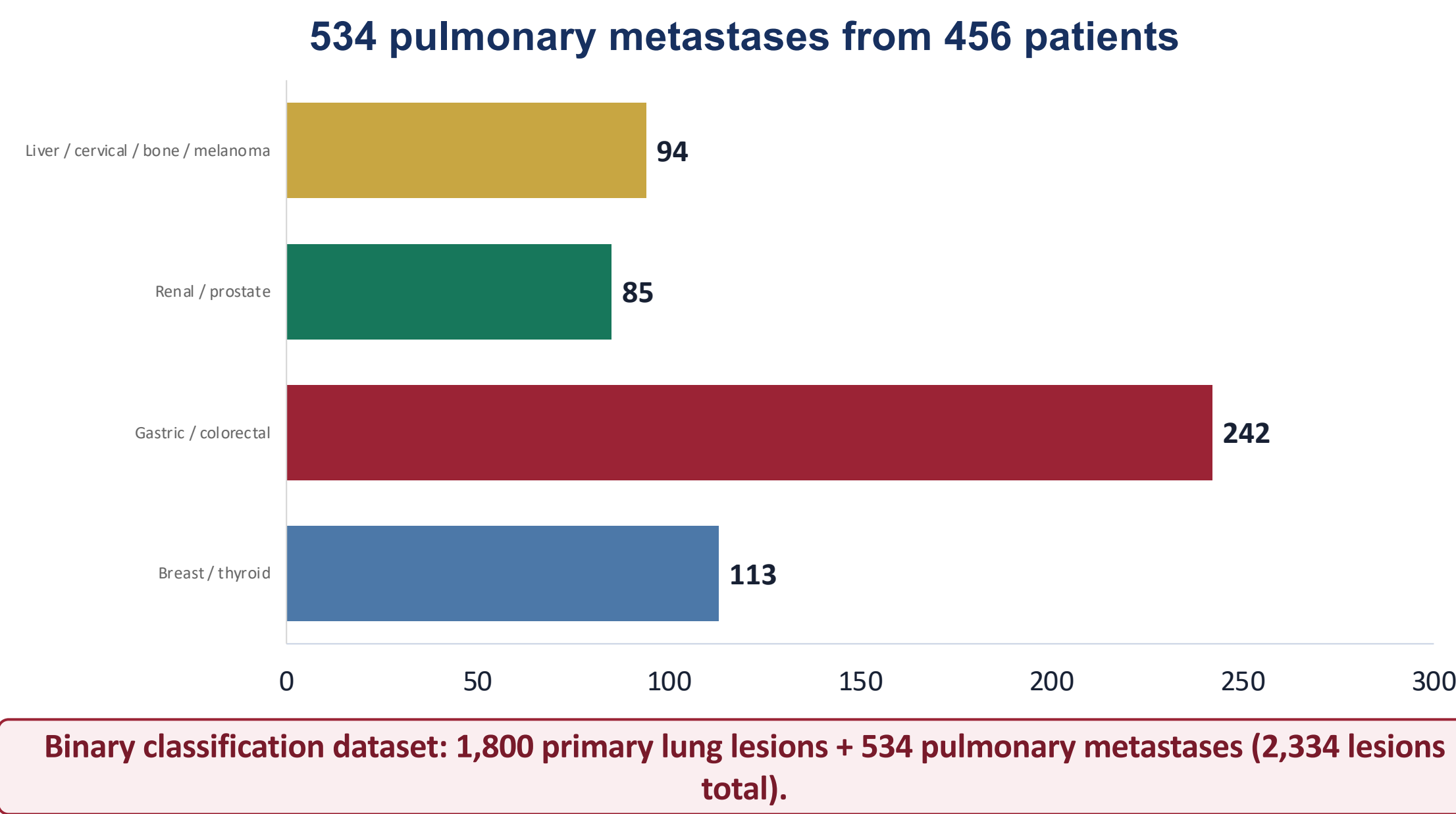
Pulmonary metastases can mimic primary lung malignancy on CT, and their biological behavior depends on the primary tumor. Existing radiomics studies often focus on a single cancer type, limiting use in patients with unknown or multiple primaries.

OBJECTIVE

Develop a generalizable two-stage CT radiomics framework to: (1) distinguish pulmonary metastases from primary lung lesions, and (2) predict the most likely primary-origin group.

RETROSPECTIVE COHORT

West China Hospital, 2013–2023 | Pathology-confirmed disease | Age 18–85 years | Complete clinical and pre-diagnostic chest CT data

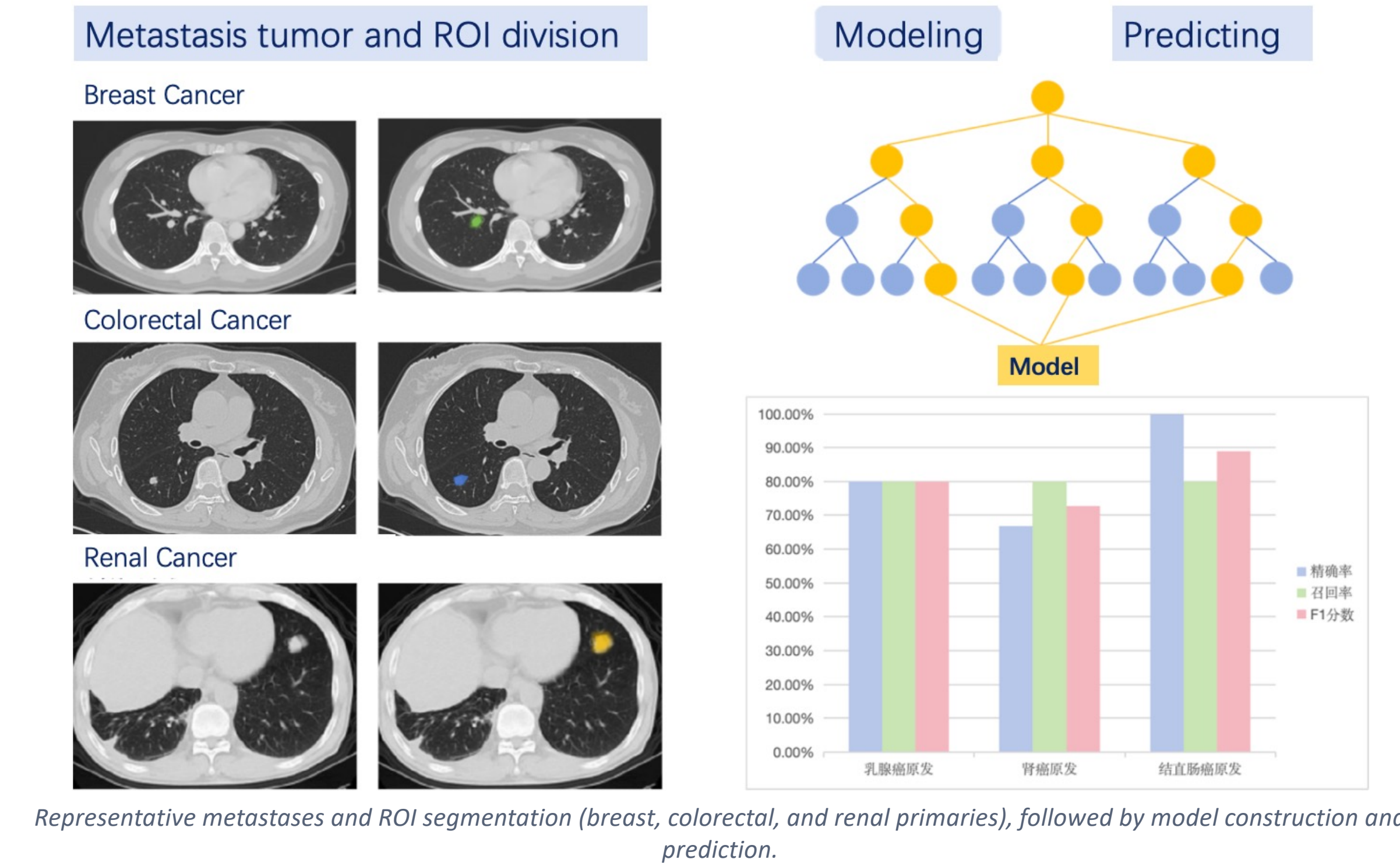


MODEL DEVELOPMENT

- Segment the pulmonary lesion ROI on chest CT.
- Extract quantitative radiomic features.
- Select features using LASSO.
- Stage 1: classify primary lung lesion vs pulmonary metastasis.
- Stage 2: predict pooled primary-origin group using random forest; optimize tree number, maximum depth, and minimum leaf size by grid search.
- Evaluate discrimination and classification using AUC, accuracy, precision, recall/sensitivity, specificity, and F1 score; obtain AUC confidence intervals by 1,000 bootstrap repetitions.

Planned next step: independent external validation across 9 centers.

TWO-STAGE RADIOMICS WORKFLOW



STAGE 1 · METASTASIS vs PRIMARY LUNG LESION

The binary classifier showed strong discrimination on the held-out test set.

0.979

AUC ± 0.013

0.937

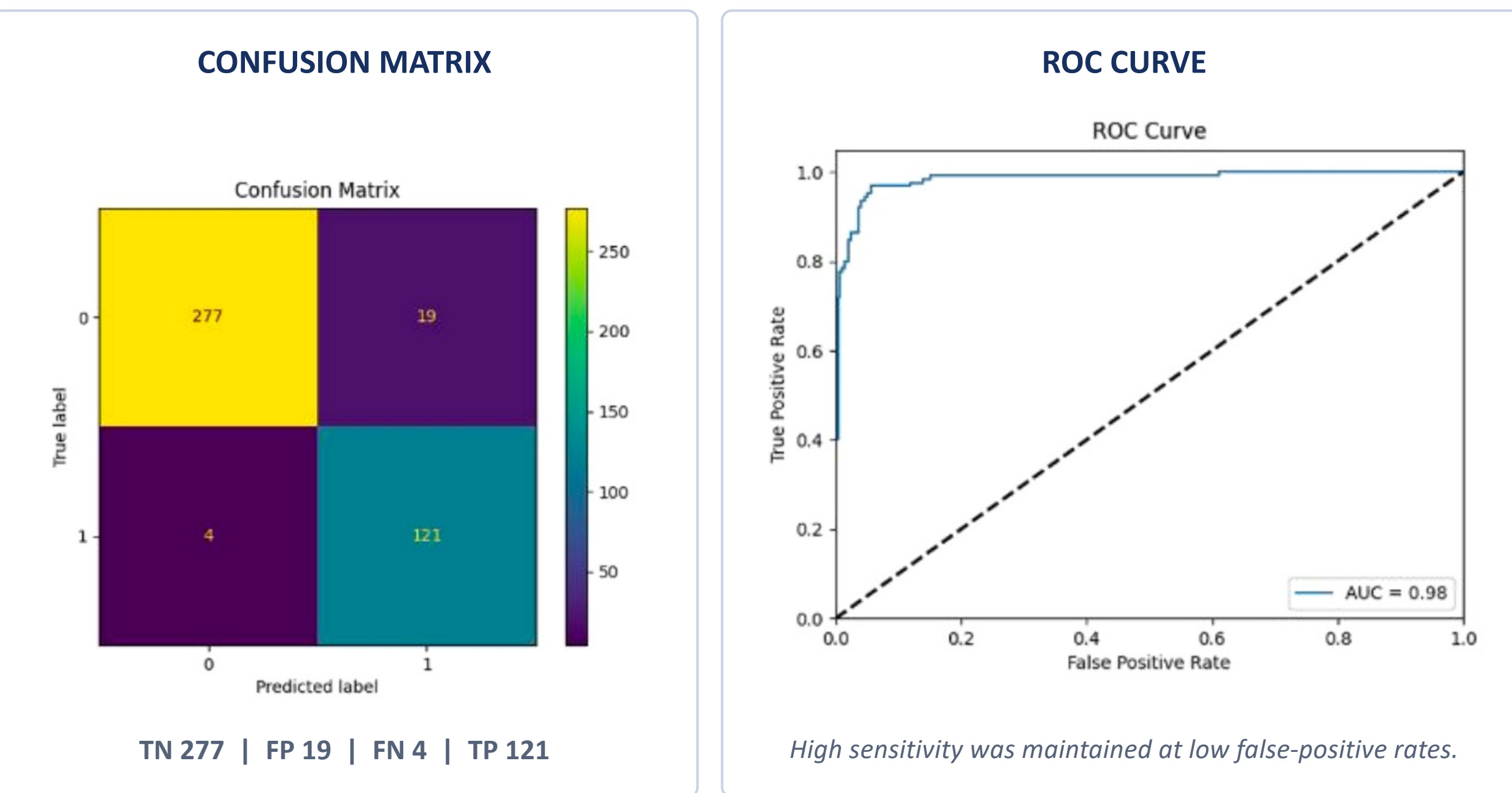
ACCURACY

0.891

SENSITIVITY

0.956

SPECIFICITY



INTERPRETATION

The model identified pulmonary metastases with few false negatives (4 cases) while retaining high specificity. These preliminary results support a sequential diagnostic strategy: first identify metastatic lesions, then estimate their origin.

Potential clinical value: prioritize the diagnostic work-up when the primary tumor is unknown, when multiple prior primaries exist, or when biopsy is difficult.

STAGE 2 · PRIMARY-ORIGIN GROUP PREDICTION

Four clinically pragmatic origin groups were modeled from 534 pulmonary metastases. Because the groups were imbalanced, five loss/sampling strategies were compared.

Metric	Cross-entropy	Oversampling	Undersampling	Weighted CE	Focal loss
Accuracy	0.878 ± 0.013	0.842 ± 0.028	0.890 ± 0.035	0.830 ± 0.021	0.827 ± 0.013
F1	0.810 ± 0.024	0.872 ± 0.029	0.864 ± 0.019	0.869 ± 0.013	0.838 ± 0.045
AUC	0.745 ± 0.025	0.735 ± 0.015	0.730 ± 0.013	0.808 ± 0.018	0.739 ± 0.020

Undersampling produced the highest accuracy (0.890), whereas weighted cross-entropy achieved the highest AUC (0.808) and sensitivity (0.857), suggesting a more balanced operating point for heterogeneous origins.

DISCUSSION & CONCLUSION

KEY MESSAGE

A two-stage CT radiomics framework can first distinguish pulmonary metastases from primary lung lesions and then provide a preliminary estimate of the primary-origin group.

CLINICAL IMPLICATION

The approach may support targeted diagnostic work-up in patients without a known primary or with multiple malignancies.

LIMITATIONS

Retrospective single-center development; heterogeneous pooled origin groups; class imbalance; lesion-level analyses may not fully account for within-patient correlation; prospective multicenter validation is required.

NEXT STEPS

External validation in 9 centers, patient-level split verification, calibration analysis, and multimodal fusion with clinical history and tumor markers.

REFERENCES & ACKNOWLEDGEMENTS

- Huang M, et al. Eur J Radiol. 2023;160:110671.
- Shang H, et al. Acad Radiol. 2023;30:40–46.
- Sollini M, et al. Eur J Nucl Med Mol Imaging. 2018. PMID: 29623375.

Acknowledgement: We thank the multidisciplinary thoracic oncology, radiology, pathology, and data-science teams at West China Hospital.