

Clinical and Serum Biomarkers Associated with Invasive Status in Early-Stage Lung Adenocarcinoma

R. Ling¹, W. Di², S. Su², Z. Yang¹, F-F. Yin¹

¹Duke Kunshan University, Kunshan, China; ²Kunshan Hospital of Traditional Chinese Medicine, Kunshan, China

Correspondence: rui.ling@dukekunshan.edu.cn

INTRODUCTION

Early-stage lung adenocarcinoma represents a heterogeneous spectrum ranging from indolent lesions to invasive tumors. Accurate assessment of invasive status is important for treatment planning and risk stratification. However, preoperative identification of invasive potential remains challenging, as conventional radiological features and biomarkers may not fully capture tumor heterogeneity. This study aimed to evaluate clinical, radiological, and serum biomarker differences between invasive and non-invasive early-stage lung adenocarcinoma and explore the potential value of machine learning-based integration of these features for risk stratification.

METHODS

Study Population

This retrospective study included 109 patients with early-stage lung adenocarcinoma who underwent surgical resection at Kunshan Traditional Chinese Medicine Hospital between March 2023 and January 2026. Patients were classified into invasive (n=42) and non-invasive (n=67) groups based on pathological evaluation.

Data Collection and Statistical Analysis

Clinical characteristics, radiological features, and serum biomarkers were collected from medical records. Continuous variables were compared using the Mann–Whitney U test, while categorical variables were analyzed using chi-square or Fisher's exact tests.

Exploratory Machine Learning Analysis

An exploratory machine learning analysis was performed using clinical, hematological, and serum biomarker features. Patients with complete feature data (n=96) were included. Models were evaluated using a stratified training/test split (76/20) with 5-fold cross-validation repeated three times.

DISCUSSION AND CONCLUSION

Key Findings

Early-stage lung adenocarcinoma with invasive status demonstrated distinct clinical, radiological, and serum biomarker profiles. Older age, larger tumor diameter, mixed/solid nodule components, and elevated CEA and CYFRA21-1 levels were associated with invasive disease.

Clinical Implication

Routinely available clinical and laboratory parameters may provide additional information for preoperative assessment of invasive potential. Exploratory machine learning analysis suggested that integrating multiple clinical and biomarker features may improve invasive status stratification.

Limitations

This study was conducted in a single-center cohort with a limited sample size.

The machine learning findings are exploratory and require validation in larger independent cohorts.

Take Away Message

Integrating routinely available clinical and serum biomarker features may provide additional support for identifying invasive early-stage lung adenocarcinoma.

Future Direction for Research

Future studies should validate the findings in larger multicenter cohorts and investigate whether combined clinical, biomarker, imaging, and molecular features can further improve invasive status prediction.

REFERENCES & ACKNOWLEDGEMENTS

Reference

Sucony, L., Rassi, D. M., Barker, A. P., McCaughan, F. M., & Rintoul, R. C. (2021). Adenocarcinoma spectrum lesions of the lung: Detection, pathology and treatment strategies. Cancer treatment reviews, 99, 102237. <https://doi.org/10.1016/j.ctrv.2021.102237>
Travis, W. D., Asamura, H., Bankier, A. A., Bessley, M. B., Detterbeck, F., Flieder, D. B., Goo, J. M., MacMahon, H., Naidich, D., Nicholson, A. G., Powell, C. A., Prokop, M., Rami-Porta, R., Rusch, V., van Schil, P., Yatabe, Y., & International Association for the Study of Lung Cancer Staging and Prognostic Factors Committee and Advisory Board Members (2016). The IASLC Lung Cancer Staging Project: Proposals for Coding T Categories for Subsolid Nodules and Assessment of Tumor Size in Part-Solid Tumors in the Forthcoming Eighth Edition of the TNM Classification of Lung Cancer. Journal of thoracic oncology : official publication of the International Association for the Study of Lung Cancer, 11(8), 1204–1223. <https://doi.org/10.1016/j.jtho.2016.03.025>

Acknowledgement

We thank **Zhuohang Liu** (Duke Kunshan University) for his support and guidance in conducting the exploratory machine learning analysis.

AI Tool Used: ChatGPT (OpenAI)

Description of AI Contribution: ChatGPT was used as a supportive tool for language refinement, poster organization, and improvement of scientific communication. AI was not used for data analysis, statistical interpretation, generation of results, or scientific decision-making. All study design, analyses, results, and conclusions were independently developed and verified by the authors.

RESULTS

1. Clinical and Serum Biomarkers Associated with Invasive Status

Patients in the invasive group were significantly older than those in the non-invasive group (median age: 66.5 [IQR 60.0–72.0] vs. 57.0 [47.5–64.0] years, $p < 0.001$) and had larger tumor diameter (20 vs. 9 mm, $p < 0.001$).

Serum biomarker analysis demonstrated higher levels of CEA (2.90 [1.60–4.07] vs. 1.65 [1.27–2.62] ng/mL, $p = 0.002$) and CYFRA21-1 (2.12 [1.70–3.13] vs. 1.60 [1.40–2.46] ng/mL, $p = 0.013$) in invasive tumors. Invasive lesions were more frequently associated with mixed or solid components, whereas non-invasive lesions predominantly showed ground-glass appearance ($p < 0.001$). The invasive group also had a higher proportion of male patients ($p < 0.001$).

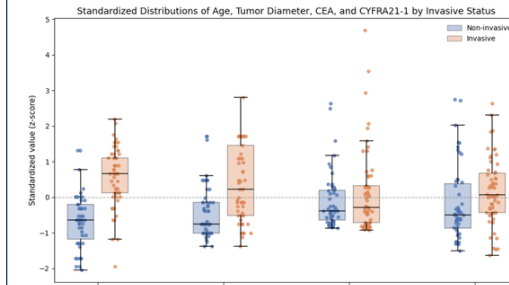


Figure 1. Standardized distributions of age, tumor diameter, CEA, and CYFRA21-1 by invasive status. Values were z-score normalized. Boxplots show median and IQR with individual data points. The invasive group shows a consistent shift toward higher values. Red boxes represent the invasive group, and blue boxes represent the non-invasive group.

Results at a glance

Clinical & Biomarker Findings

Older age, larger tumor diameter, and higher CEA and CYFRA21-1 were associated with invasive status.

Exploratory ML

Integrated features showed moderate discrimination, with a held-out test AUC of 0.768.

Overall Interpretation

Routine clinical and laboratory features may support invasive status assessment and future risk stratification.

2. Exploratory Machine Learning Analysis

RF-SVM showed the highest cross-validation ROC-AUC, although differences across models were modest. Logistic Regression achieved an AUC of 0.768 on the held-out test set, with higher specificity (77.8%) than sensitivity (36.4%). Permutation analysis identified WBC count, hemoglobin, and CEA among the strongest contributors to model discrimination.



Figure 2. Repeated 5-fold cross-validation ROC-AUC (mean ± SD) across models; dashed line indicates chance level (AUC = 0.5).

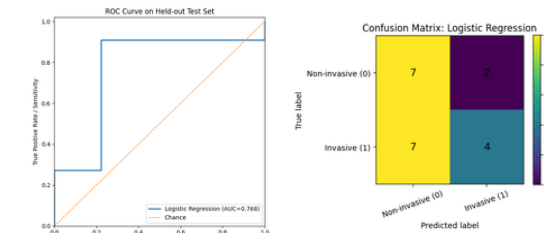


Figure 3. ROC curve of Logistic Regression on the held-out test set (n=20).

Figure 4. Confusion matrix of Logistic Regression on the held-out test set (n=20).

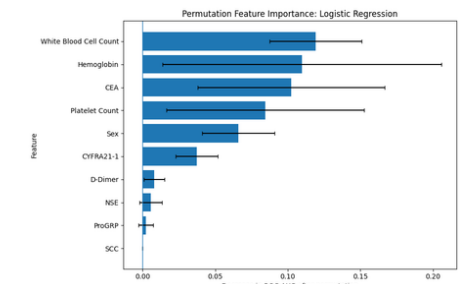


Figure 5. Permutation feature importance for Logistic Regression; larger ROC-AUC decreases indicate greater contribution to model discrimination.