Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Association of clinicopathological features with spread through air spaces in non-small cell lung cancer: a retrospective machine-learning study

Nikita Laptev

nikitalaptev77@gmail.com

Siberian State Medical University

Marina Zavyalova

Siberian State Medical University

Dmitry Pismenny

Siberian State Medical University

Alexander Zavyalov

Siberian State Medical University

Dmitry Loos

Siberian State Medical University

Tatiana Lubina

Siberian State Medical University

Donir Garmaev

Siberian State Medical University

Sergey Miller

Siberian State Medical University

Vladimir Perelmuter

Siberian State Medical University

Research Article

Keywords: spread through air spaces, non-small cell lung cancer, machine learning, internal validation, pathology, AutoGluon, repeated cross-validation

Posted Date: September 8th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-10784971/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Background: Spread through air spaces (STAS) is associated with adverse outcomes in non-small cell lung cancer (NSCLC). We investigated the association between the combined set of available clinicopathological characteristics and STAS status and assessed the ability of a multivariable model to discriminate STAS-positive from STAS-negative cases under internal validation.

Methods: The single-center retrospective cohort comprised 230 patients with NSCLC who underwent surgical treatment between 2022 and 2025: 81 with STAS and 149 without STAS. An AutoGluon TabularPredictor model was trained using 19 clinicopathological predictors. Internal evaluation used five repetitions of 5-fold outer cross-validation (25 outer test folds); within each outer training set, 4-fold cross-validation was used to select the resampling strategy. Labels from the outer test folds were not used for configuration selection. ROC-AUC and average precision (AP) were calculated from mean patient-level out-of-fold (OOF) probabilities, whereas F1 and other threshold-dependent metrics were calculated using majority voting at a prespecified threshold of 0.5. Ninety-five percent confidence intervals were estimated using 1,000 nonparametric patient-level bootstrap samples from the aggregated 230-patient OOF table.

Results: For the 19-feature model, ROC-AUC was 0.685 (95% CI 0.611–0.751), AP was 0.523 (95% CI 0.422–0.641), F1 was 0.548 (95% CI 0.435–0.621), and the Brier score was 0.211. At the 0.5 threshold, sensitivity was 0.531, specificity 0.779, PPV 0.566, and NPV 0.753 (TP = 43, FN = 38, FP = 33, TN = 116); 38 of 81 STAS-positive cases were classified as negative. After Holm correction, none of the six paired comparisons between the full and reduced feature configurations reached statistical significance.

Conclusions: The available clinicopathological data contain a moderate internal discriminatory signal for STAS; however, a substantial proportion of predictors become available only after resection, and probability calibration shows noticeable deviations from ideal calibration. The resulting configuration should therefore be considered an exploratory model of association with STAS rather than a preoperative clinical tool. Further studies should assess the incremental value of truly preoperative clinical, imaging, and molecular features, followed by independent external validation.

Association of clinicopathological features with spread through air spaces in non-small cell lung cancer: a retrospective machine-learning study

Nikita V. Laptev[1*], Marina V. Zavyalova[1,2], Dmitry S. Pismenny[1,2], Alexander V. Zavyalov[1], Dmitry M. Loos[1,2], Tatiana A. Lubina[1], Donir B. Garmaev[1], Sergey V. Miller[1,2], Vladimir M. Perelmuter[1,2]

[1*]Siberian State Medical University, Ministry of Health of the Russian Federation, 2 Moskovsky Trakt, Tomsk, 634050, Russia.
[2]Cancer Research Institute, Tomsk National Research Medical Center, Russian Academy of Sciences, 5 Kooperativny Lane, Tomsk, 634009, Russia.

*Corresponding author(s). E-mail(s): nikitalaptev77@gmail.com;

Keywords: spread through air spaces, non-small cell lung cancer, machine learning, internal validation, pathology, AutoGluon, repeated cross-validation

1 Introduction

Selection of the extent of surgical resection in non-small cell lung cancer requires accurate assessment of factors associated with recurrence risk. One such factor is spread through air spaces (STAS), an invasion pattern in which tumor cells are identified within alveolar spaces beyond the edge of the main tumor. The association of STAS with recurrence and adverse outcomes has been demonstrated in lung adenocarcinoma [1–6] and squamous cell carcinoma [7–9].

A key clinical problem is that the final assessment of STAS is usually made on resection specimens, after the decision regarding the extent of surgery has already been made. The inclusion of STAS in the 2015 World Health Organization classification as an additional histologic descriptor of lung adenocarcinoma [10] emphasized

its relevance to contemporary thoracic pathology. However, recognition of this pattern requires expert morphologic assessment and remains vulnerable to interobserver variability even among thoracic pathologists [11]. This diagnostic gap provides a rationale for cautious investigation of features associated with STAS and of its potential preoperative assessment.

Previous studies have evaluated computed tomography features [12] and radiomic characteristics [13] as potential sources of information about STAS. Imaging data were not used in the present study; the analysis was limited to available clinical, morphologic, and molecular characteristics. The aim was to assess the association of their combined information with STAS status and the ability of the resulting multivariable model to discriminate STAS-positive from STAS-negative cases under internal validation.

Automated machine-learning methods, including AutoGluon TabularPredictor [14], enable the analysis of structured clinical data while maintaining separation between model tuning and independent internal evaluation. Using AutoGluon TabularPredictor, we assessed the discriminative performance of a 19-feature clinicopathological configuration and the uncertainty of the resulting estimates, and we examined the stability of permutation importance for individual features across repeated internal validation. Comparisons with 15- and 10-feature configurations were secondary exploratory analyses.

The findings may serve as a clinicopathological reference for subsequent multimodal studies of STAS. A promising direction is to assess the incremental information provided by imaging and molecular features, including mass-spectrometric profiles, beyond the baseline clinicopathological configuration.

2 Methods

2.1 Study design and cohort

We conducted a single-center retrospective study of patients who underwent surgical treatment for non-small cell lung cancer at the Cancer Research Institute, Tomsk National Research Medical Center, Russian Academy of Sciences (Tomsk, Russia), between 2022 and 2025. The analytical cohort included 230 patients with morphologically verified adenocarcinoma ($n = 137$) or squamous cell carcinoma ($n = 93$) and a defined STAS status. In the cohort, 192 patients underwent lobectomy, 20 underwent atypical resection, and 18 underwent pneumonectomy; no recorded neoadjuvant therapy was present in any patient. The outcome was the presence of STAS in the resection specimen. No external validation was performed; all reported performance estimates refer to internal validation within this single cohort.

Of the 19 analyzed variables, only age was documented as unequivocally preoperative; 11 features were postoperative pathologic characteristics, while the acquisition timing of seven variables could not be unambiguously reconstructed. Accordingly, the analysis was designed to investigate the association of clinicopathological characteristics with STAS status and should not be interpreted as a preoperative clinical prediction tool.

2.2 Outcome and predictors

STAS was determined by morphologic examination of resection specimens in accordance with the 2015 World Health Organization classification [10], as the presence of tumor cells or tumor cell clusters within alveolar spaces beyond the edge of the main tumor. Histologic slides were reviewed by four pathologists. When assessments differed, the final STAS status was established by consensus. The analytical cohort comprised 230 observations: 81 STAS-positive (35.2%) and 149 STAS-negative (64.8%).

The primary configuration included 19 clinicopathological predictors. In the source table, features were represented by numeric values or numeric codes; by data type, however, the set comprised continuous, binary, and ordinal/categorical variables. Categorical and ordinally coded variables were processed separately from continuous variables: missing values in categorical variables were imputed with the most frequent value, whereas missing values in continuous variables were imputed using K-nearest neighbors. Imputation parameters were fitted using only the training portion of each split and subsequently applied to the corresponding test portion. When SMOTENC was used, categorical variables were passed to the algorithm as discrete features. The acquisition timing of each variable is documented in Table 1. The overall proportion of missing values in the feature matrix was 25.5%. The origin of the reduced feature sets could not be confirmed as fold-local within a unified feature-selection procedure, and the exact composition of local reduced feature sets could not be reconstructed from the retained artifacts.

2.3 Training and evaluation scheme

Training and internal evaluation were performed using repeated nested cross-validation. The outer level consisted of 5-fold cross-validation repeated five times with different fixed random seeds, yielding 25 outer test folds in total. Within each outer cycle, 4-fold inner cross-validation was performed on the corresponding training portion to select the class-imbalance correction strategy. An AutoGluon TabularPredictor pipeline [14] was used for each outer fold. The class-imbalance correction strategy was selected exclusively from the inner training data, with a fixed seed, among RandomOverSampler, ADASYN [15], Borderline-SMOTE, SMOTENC [16, 17], and no resampling. Resampling was applied only to training subsets. The outer test fold was never resampled, and its labels were not used to select the resampling strategy, features, or threshold. The threshold of 0.5 was prespecified and used only for threshold-dependent summaries.

2.4 Out-of-fold aggregation and statistical analysis

For each repetition of the outer cross-validation, out-of-fold (OOF) predictions were obtained, after which repeated predictions from the five outer repetitions were aggregated at the patient level. Each participant could receive up to five predictions. ROC-AUC and AP were calculated using the mean patient-level probability; F1, sensitivity, specificity, PPV, NPV, and the confusion matrix were calculated using majority

voting at the fixed threshold of 0.5. Ninety-five percent confidence intervals for aggregated metrics were estimated using a nonparametric patient-level bootstrap (1,000 repetitions) applied to the aggregated OOF table ($n = 230$). The resampling unit was the patient. Bootstrapping was performed after aggregation of repeated OOF predictions and therefore quantifies uncertainty in the aggregated metrics but does not capture the full variability associated with repeated model training. Calibration of the aggregated OOF probabilities for the primary 19-feature configuration was additionally assessed using a calibration curve and the Brier score.

The primary null hypothesis was that ROC-AUC equaled 0.5. This was tested with a two-sided patient-level permutation test using fixed aggregated probabilities and 10,000 permutations. AP was tested as a secondary metric against the observed positive-class prevalence of 0.352 (81/230); this analysis was post hoc.

Six paired comparisons were performed: 19 versus 15 features and 19 versus 10 features for ROC-AUC, AP, and F1. Each comparison used 10,000 valid paired bootstrap repetitions (no rejected repetitions) and a two-sided paired randomization test with 10,000 randomizations. Holm correction was applied across the six hypotheses. The aggregation rule for repeated predictions was not preserved in the historical artifacts; the present analysis therefore uses the explicitly stated aggregation rule. Point estimates from the 25 outer folds are dependent descriptive quantities because participants contribute to multiple outer folds. Reporting followed the TRIPOD+AI recommendations [18].

3 Results

The primary measure of internal discrimination was ROC-AUC; results of secondary comparisons with the reduced feature sets are reported below.

3.1 Cohort, internal discrimination, and threshold-dependent performance

Cohort and demographic characteristics are presented in Table 2.

Age, sex, vascular invasion, tumor location, and histologic type were available for $N = 230$. Tumor size, neural invasion, and pleural invasion were available for $n = 142$; the distribution of available observations by STAS status is given in the caption to Table 2.

For the primary 19-feature configuration, ROC-AUC was 0.685 (95% CI 0.611–0.751), indicating moderate ability to discriminate STAS-positive from STAS-negative cases under internal validation. AP was 0.523 (95% CI 0.422–0.641). The Brier score for the primary configuration was 0.211. The calibration curve deviated from the line of ideal calibration, with the most pronounced deviations in some ranges of high predicted probabilities; therefore, the resulting probabilities should be interpreted as exploratory estimates rather than clinically calibrated individual probabilities of STAS (Supplementary Fig. S6). At the threshold of 0.5, the confusion matrix contained TP = 43, FN = 38, FP = 33, and TN = 116; sensitivity was 0.531, specificity 0.779, PPV 0.566, and NPV 0.753. These threshold-dependent measures describe the error profile at the prespecified threshold, whereas ROC-AUC remains the primary

threshold-independent result. Performance estimates for all three configurations are shown in Table 3.

ROC and precision–recall curves for the three feature-set sizes are shown in Supplementary Fig. S1.

Threshold-dependent summaries for the three configurations at the fixed threshold of 0.5 are provided in Supplementary Table S1. The confusion matrix for the 19-feature configuration is shown in Supplementary Fig. S2; its key feature is the presence of 38 false-negative predictions among 81 STAS-positive cases.

3.2 Permutation tests and paired comparisons of feature sets

The primary permutation test yielded a ROC-AUC of 0.685061 versus a null value of 0.5 (difference 0.185061; two-sided $p = 0.0001$; 10,000 permutations; seed 20260811). The secondary AP test yielded an AP of 0.522875 versus 0.352174 (difference 0.170701; $p = 0.0002$; 10,000 permutations; seed 20260812) (Supplementary Fig. S3).

The 19-feature configuration yielded slightly higher point estimates; however, after Holm correction, none of the paired comparisons reached statistical significance, and all confidence intervals included zero. Thus, the data did not demonstrate a convincing advantage of one configuration over another, but neither do they establish interchangeability (Supplementary Table S2).

3.3 Outer-fold variability and permutation importance of features

Supplementary Fig. S4 shows point estimates of ROC-AUC and AP for the 25 outer folds. These values are dependent descriptive quantities (see Methods); statistical inference is based on the aggregated patient-level OOF estimates.

Permutation importance was used to describe which variables were most consistently used by the models across outer folds. Tumor grade, pN, vascular invasion, solid and micropapillary patterns, tumor size, and age occupied the most stable positions. Supplementary Table S3 and Fig. 1 summarize stability across all 19 predictors, while Supplementary Fig. S5 shows their ranking. Histologic type had a mean importance of 0.034, a mean rank of 9.4, a positive-importance frequency of 1.00, and availability in 25/25 folds; ROS1 was available in 24/25 folds, had a positive-importance frequency of 0.417 and mean importance of 0.003, and did not enter the top 10. These results represent post hoc attribution for the fitted models and do not establish causal relationships between predictors and STAS.

4 Discussion

4.1 Main findings

This study evaluated the internal discriminative performance of a 19-feature clinicopathological configuration using repeated nested cross-validation with a deterministic prediction pipeline. ROC-AUC was 0.685 (95% CI 0.611–0.751), and AP was 0.523 (95% CI 0.422–0.641). The permutation test showed that ROC-AUC differed from the null value of 0.5, with a difference of 0.185061 ($p = 0.0001$).

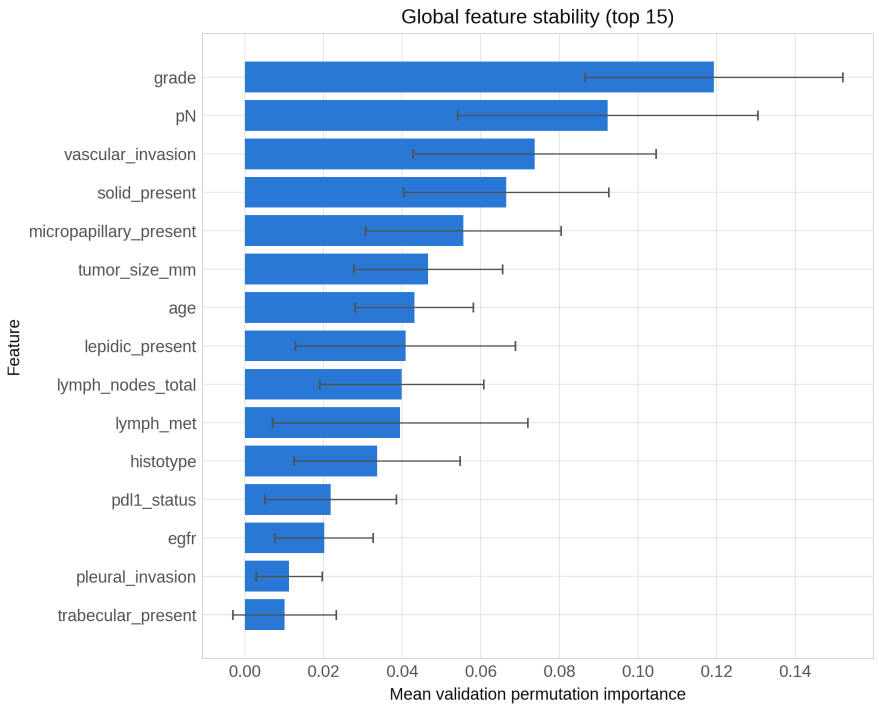


Figure 1 Mean permutation importance of the 15 features with the highest mean importance across 25 outer test folds. Horizontal error bars reflect variability of estimates across folds; numerical values for mean importance, mean rank, and the frequency of positive importance are provided in Supplementary Table S3.

Taken together, the results indicate a statistically detectable but moderate discriminatory signal contained in the combined set of evaluated clinicopathological features. The 19-feature configuration produced the highest point estimates among the three variants; however, after Holm correction, paired comparisons did not confirm a statistically significant advantage. These findings are best regarded as a basis for refining research hypotheses rather than as a ready-to-use clinical rule.

4.2 Clinical interpretation

At the fixed threshold of 0.5, the model failed to identify 38 of 81 STAS-positive cases. Together with the fact that a substantial proportion of the analyzed features become available only after surgical resection, this error profile precludes direct use of the resulting configuration as a preoperative clinical tool.

At the same time, the study provides a reproducible research framework comprising a deterministic pipeline, fixed random seeds, an explicit rule for aggregation of repeated predictions, and complete reporting of outer-fold results. These properties make the model suitable as a baseline for subsequent comparisons with extended predictor sets and candidate biomarkers.

4.3 Relationship to previous studies and future directions

STAS was recognized as an invasion pattern in the 2015 WHO classification [10]. Numerous studies have associated STAS with adverse outcomes and clinicopathological characteristics [1–3, 5–9], and a systematic review and meta-analysis confirmed its adverse prognostic significance in NSCLC [4]. CT features and radiomic models have also been investigated as potential sources of information for preoperative assessment of STAS [12, 13]. The present analysis complements this line of research by evaluating the association between routine clinicopathological characteristics and STAS status before incorporating imaging and additional molecular data sources.

The resulting clinicopathological configuration may serve as a baseline for subsequent multimodal studies. The next step should be to assess the incremental information provided by preoperative imaging and candidate molecular features, including mass-spectrometric profiles, in an independent cohort using a prespecified analysis protocol.

4.4 Limitations

This was a single-center retrospective study with an overall missing-data rate of 25.5% in the feature matrix; missing values were handled automatically within the pipeline. The complete predictor contract includes postoperative variables and variables with uncertain acquisition timing, and the model therefore cannot be regarded as a preoperative tool. Threshold-dependent metrics—F1, sensitivity, specificity, PPV, NPV, and the confusion matrix—describe model performance at the fixed threshold of 0.5. Test-fold labels were not used for threshold or feature selection; however, the rule used to aggregate repeated predictions was not preserved in the historical artifacts and is explicitly defined in the present analysis. Bootstrapping was performed on aggregated OOF predictions and does not account for additional uncertainty associated with complete retraining of the prediction pipeline. The Brier score and calibration curve indicate that the aggregated probabilities should not be interpreted as clinically calibrated individual risk estimates.

The AP test and comparisons of reduced feature sets were secondary post hoc analyses; the origin of the reduced feature sets could not be confirmed as fold-local in the retained artifacts. ADASYN operates in a numerical feature space and does not directly account for the categorical nature of numerically encoded variables; its use in some inner cycles should therefore be considered an additional methodological limitation. Permutation importance does not establish causal relationships. Feature interactions, prognostic performance in subgroups, and clinical utility were not evaluated.

5 Conclusions

The available clinical and morphologic characteristics contain a measurable but moderate internal discriminatory signal for STAS. The 19-feature configuration discriminated STAS-positive from STAS-negative cases better than chance; however, its error profile, calibration, and predictor composition do not support translating this

result into preoperative decision-making. The principal contribution of this study is a quantitative assessment of the association between a combined set of clinicopathological features and STAS status, together with a transparent methodological basis for the next stage of investigation.

Further studies should assess the incremental value of truly preoperative clinical, imaging, and molecular features, followed by independent external validation of the complete analytical pipeline.

Abbreviations

STAS	Spread through air spaces
NSCLC	Non-small cell lung cancer
ROC AUC	Area under the receiver operating characteristic curve
AP	Average precision
F1	Harmonic mean of precision and recall
PPV	Positive predictive value
NPV	Negative predictive value
OOF	Out-of-fold
SMOTENC	Synthetic Minority Oversampling Technique for Nominal and Continuous features
ADASYN	Adaptive Synthetic Sampling
CI	Confidence interval
AutoML	Automated machine learning

Ethics approval and consent to participate. The study was conducted in accordance with the principles of the Declaration of Helsinki and was approved by the independent Ethics Committee of Siberian State Medical University (2 Moskovsky Trakt, Tomsk 634050, Russia), protocol No. 10198 dated 22 December 2025.

Consent for publication. Not applicable. The manuscript contains no individually identifiable patient data.

Availability of data and materials. The data used in this study are deposited in Zenodo and are available at https://doi.org/10.5281/zenodo.22038864. The current version of the analysis code and software pipeline is available in the public GitHub repository at https://github.com/Nikita75699/stas-nsclc-autogluon. The code repository is under active development; the version corresponding to the final manuscript will be archived as a fixed release before publication.

Competing interests. The authors declare that they have no competing interests.

Funding. This study was supported by the Russian Science Foundation (RSF), grant No. 26-15-00029, "Development of technologies for the preoperative diagnosis of spread through air spaces in patients with non-small cell lung cancer using high-throughput chromatography–mass spectrometry."

Authors' contributions. N.V.L. — machine-learning methodology, software pipeline development, formal data analysis, visualization, and preparation of the original manuscript draft. M.V.Z. — study conceptualization, methodology of the

pathology component, morphologic assessment and data interpretation, supervision, and critical revision of the manuscript. D.S.P. — morphologic assessment and data verification, curation of morphologic data, interpretation of results, and critical revision of the manuscript. A.V.Z. — collection and systematization of clinicopathological data, preparation of the source dataset, and participation in manuscript revision. D.M.L. — morphologic assessment of surgical specimens, verification of morphologic features, interpretation of results, and critical revision of the manuscript. T.A.L. — collection and curation of clinicopathological data, preparation of study materials, and manuscript revision. D.B.G. — collection and curation of clinicopathological data, preparation of study materials, and manuscript revision. S.V.M. — establishment and clinical characterization of the surgical cohort, clinical interpretation of results, and critical revision of the manuscript. V.M.P. — study conceptualization, methodology of the pathology component, morphologic assessment, supervision, interpretation of results, and critical revision of the manuscript. All authors read and approved the final manuscript.

Acknowledgements.

Authors' information. **Marina Viktorovna Zavyalova**, MD, Dr Sci, Professor, Head of the Department of Pathological Anatomy and Researcher at the Central Research Laboratory, Siberian State Medical University; Leading Researcher, Department of General and Molecular Pathology, Cancer Research Institute, Tomsk National Research Medical Center, Russian Academy of Sciences. ORCID: 0000-0001-9429-9813.

Nikita Vitalievich Laptev, PhD in Engineering, Researcher, Department of Medical Data and Bioinformatics, Siberian State Medical University. ORCID: 0000-0003-0709-9974.

Dmitry Sergeevich Pismenny, MD, PhD, Associate Professor, Department of Pathological Anatomy, and Researcher, Central Research Laboratory, Siberian State Medical University; Clinical Laboratory Diagnostics Physician, Department of General and Molecular Pathology, Cancer Research Institute, Tomsk National Research Medical Center, Russian Academy of Sciences. ORCID: 0000-0001-8973-8439.

Alexander Vasilievich Zavyalov, medical student and Laboratory Assistant, Department of Pathological Anatomy and Central Research Laboratory, Siberian State Medical University. ORCID: 0009-0009-0266-6707.

Dmitry Maksimovich Loos, MD, PhD, Researcher, Central Research Laboratory, Siberian State Medical University; Pathologist, Department of General and Molecular Pathology, Cancer Research Institute, Tomsk National Research Medical Center, Russian Academy of Sciences. ORCID: 0000-0002-4098-6076.

Tatiana Andreevna Lubina, postgraduate researcher, Department of Pathological Anatomy, Siberian State Medical University. ORCID: 0000-0002-2264-1074.

Donir Bairovich Garmaev, postgraduate researcher, Department of Pathological Anatomy, Siberian State Medical University. ORCID: 0009-0003-3924-4045.

Sergey Viktorovich Miller, MD, Dr Sci, Researcher, Central Research Laboratory, Siberian State Medical University; Head of the Thoracic Department, Cancer Research Institute, Tomsk National Research Medical Center, Russian Academy of Sciences. ORCID: 0000-0002-5365-9840.

Vladimir Mikhailovich Perelmuter, MD, Dr Sci, Professor, Department of Pathological Anatomy, Siberian State Medical University; Chief Researcher, Department of General and Molecular Pathology, Cancer Research Institute, Tomsk National Research Medical Center, Russian Academy of Sciences. ORCID: 0000-0002-7633-9620.

References

[1] Kadota K, Nitadori Ji, Sima CS, Ujiie H, Rizk NP, Jones DR, et al. Tumor Spread through Air Spaces is an Important Pattern of Invasion and Impacts the Frequency and Location of Recurrences after Limited Resection for Small Stage I Lung Adenocarcinomas. Journal of Thoracic Oncology. 2015;10(5):806–814. https://doi.org/10.1097/JTO.0000000000000486.

[2] Warth A, Muley T, Kossakowski CA, Goeppert B, Schirmacher P, Dienemann H, et al. Prognostic Impact of Intra-alveolar Tumor Spread in Pulmonary Adenocarcinoma. American Journal of Surgical Pathology. 2015;39(6):793–801. https://doi.org/10.1097/PAS.0000000000000409.

[3] Shiono S, Yanagawa N. Spread through air spaces is a predictive factor of recurrence and a prognostic factor in stage I lung adenocarcinoma. Interactive CardioVascular and Thoracic Surgery. 2016;23(4):567–572. https://doi.org/10.1093/icvts/ivw211.

[4] Chen D, Mao Y, Wen J, She Y, Zhu E, Zhu F, et al. Tumor Spread Through Air Spaces in Non-Small Cell Lung Cancer: A Systematic Review and Meta-Analysis. The Annals of Thoracic Surgery. 2019;108(3):945–954. https://doi.org/10.1016/j.athoracsur.2019.02.045.

[5] Toyokawa G, Yamada Y, Tagawa T, Kozuma Y, Matsubara T, Haratake N, et al. Significance of Spread Through Air Spaces in Resected Pathological Stage I Lung Adenocarcinoma. The Annals of Thoracic Surgery. 2018;105(6):1655–1663. https://doi.org/10.1016/j.athoracsur.2018.01.037.

[6] Dai C, Xie H, Su H, She Y, Zhu E, Fan Z, et al. Tumor spread through air spaces affects the recurrence and overall survival in patients with lung adenocarcinoma >2 to 3 cm. Journal of Thoracic Oncology. 2017;12(7):1052–1060. https://doi.org/10.1016/j.jtho.2017.03.020.

[7] Lu S, Tan KS, Kadota K, Eguchi T, Bains S, Rekhtman N, et al. Spread through Air Spaces (STAS) Is an Independent Predictor of Recurrence and Lung Cancer-Specific Death in Squamous Cell Carcinoma. Journal of Thoracic Oncology. 2017;12(2):223–234. https://doi.org/10.1016/j.jtho.2016.09.129.

[8] Kadota K, Kushida Y, Katsuki N, Ishikawa R, Ibuki E, Motoyama M, et al. Tumor Spread Through Air Spaces Is an Independent Predictor of Recurrence-Free Survival in Patients With Resected Lung Squamous Cell Carcinoma. American

Journal of Surgical Pathology. 2017;41(8):1077–1086. https://doi.org/10.1097/PAS.0000000000000872.

[9] Yanagawa N, Shiono S, Endo M, Ogata Sy. Tumor spread through air spaces is a useful predictor of recurrence and prognosis in stage I lung squamous cell carcinoma, but not in stage II and III. Lung Cancer. 2018;120:14–21. https://doi.org/10.1016/j.lungcan.2018.03.018.

[10] Travis WD, Brambilla E, Nicholson AG, Yatabe Y, Austin JHM, Beasley MB, et al. The 2015 World Health Organization Classification of Lung Tumors: Impact of Genetic, Clinical and Radiologic Advances Since the 2004 Classification. Journal of Thoracic Oncology. 2015;10(9):1243–1260. https://doi.org/10.1097/JTO.0000000000000630.

[11] Garlin-Politis M, Saqi A, Mino-Kenudson M. Spread Through Air Spaces: Interresponder Agreement and Comparison Between Pulmonary and General Pathologists. Modern Pathology. 2024;37(11):100596. https://doi.org/10.1016/j.modpat.2024.100596.

[12] Yin Q, Wang H, Cui H, Wang W, Yang G, Qie P, et al. Meta-analysis of association between CT-based features and tumor spread through air spaces in lung adenocarcinoma. Journal of Cardiothoracic Surgery. 2020;15(1). https://doi.org/10.1186/s13019-020-01287-9.

[13] Onozato Y, Nakajima T, Yokota H, Morimoto J, Nishiyama A, Toyoda T, et al. Radiomics is feasible for prediction of spread through air spaces in patients with nonsmall cell lung cancer. Scientific Reports. 2021;11(1). https://doi.org/10.1038/s41598-021-93002-4.

[14] Erickson N, Mueller J, Shirkov A, Zhang H, Larroy P, Li M, et al.: AutoGluon-Tabular: Robust and Accurate AutoML for Structured Data. arXiv:2003.06505. Available from: https://arxiv.org/abs/2003.06505.

[15] He H, Bai Y, Garcia EA, Li S. ADASYN: Adaptive Synthetic Sampling Approach for Imbalanced Learning. In: 2008 IEEE International Joint Conference on Neural Networks (IEEE World Congress on Computational Intelligence); 2008. p. 1322–1328.

[16] Chawla NV, Bowyer KW, Hall LO, Kegelmeyer WP. SMOTE: Synthetic Minority Over-sampling Technique. Journal of Artificial Intelligence Research. 2002;16:321–357. https://doi.org/10.1613/jair.953.

[17] Fernandez A, Garcia S, Herrera F, Chawla NV. SMOTE for Learning from Imbalanced Data: Progress and Challenges, Marking the 15-year Anniversary. Journal of Artificial Intelligence Research. 2018;61:863–905. https://doi.org/10.1613/jair.1.11192.

[18] Collins GS, Moons KGM, Dhiman P, Riley RD, Beam AL, Van Calster B, et al. TRIPOD+AI statement: updated guidance for reporting clinical prediction models that use regression or machine learning methods. BMJ. 2024;p. e078378. https://doi.org/10.1136/bmj-2023-078378.

Table 1 Predictor contract for the 19-feature configuration. Acquisition timing was defined according to the audited availability manifest; "uncertain" indicates that the available provenance did not allow the variable to be classified unequivocally as preoperative. Missingness is reported among 230 observations.

Feature	Description	Acquisition timing	Missing, n (%)
pN	Pathologic N stage	Postoperative pathology	0 (0.0)
Lymphatic metastasis	Indicator of lymphatic metastasis	Postoperative pathology	0 (0.0)
Tumor size	Tumor size, mm	Uncertain	88 (38.3)
EGFR	EGFR mutation status	Uncertain	88 (38.3)
KRAS	KRAS mutation status	Uncertain	88 (38.3)
ALK	ALK status	Uncertain	88 (38.3)
ROS1	ROS1 status	Uncertain	88 (38.3)
PD-L1 status	PD-L1 status	Uncertain	0 (0.0)
Age	Age, years	Preoperative	0 (0.0)
Number of lymph nodes	Total number of lymph nodes	Postoperative pathology	88 (38.3)
Tumor grade	Degree of tumor differentiation	Postoperative pathology	0 (0.0)
Solid pattern	Presence of a solid pattern	Postoperative pathology	91 (39.6)
Trabecular pattern	Presence of a trabecular pattern	Postoperative pathology	131 (57.0)
Micropapillary pattern	Presence of a micropapillary pattern	Postoperative pathology	93 (40.4)
Lepidic pattern	Presence of a lepidic pattern	Postoperative pathology	94 (40.9)
Vascular invasion	Presence of vascular invasion	Postoperative pathology	0 (0.0)
Neural invasion	Presence of neural invasion	Postoperative pathology	88 (38.3)
Pleural invasion	Presence of pleural invasion	Postoperative pathology	88 (38.3)
Histologic type	Adenocarcinoma or squamous cell carcinoma	Uncertain	0 (0.0)

Table 2 Study cohort and clinicopathological characteristics according to STAS status. Tumor size, neural invasion, and pleural invasion were available for 142 observations (92 STAS-negative and 50 STAS-positive); percentages for these variables were calculated among available observations within each group.

Characteristic	Value/category	All patients, $N = 230$	STAS-negative, $n = 149$	STAS-positive, $n = 81$
Age, years	Median [IQR]	63 [56–69]	63 [56–69]	63 [56–69]
	Mean ± SD	62.1 ± 8.9	62.0 ± 9.0	62.3 ± 8.7
Tumor size, mm	Median [IQR]	35 [20–45]	35 [20–45]	31 [19.3–45]
	Mean ± SD	34.8 ± 19.4	35.2 ± 20.3	34.0 ± 17.9
Sex	Male	158 (68.7%)	104 (69.8%)	54 (66.7%)
	Female	72 (31.3%)	45 (30.2%)	27 (33.3%)
Vascular invasion	Present	45 (19.6%)	18 (12.1%)	27 (33.3%)
	Absent	185 (80.4%)	131 (87.9%)	54 (66.7%)
Neural invasion	Present	8 (5.6%)	3 (3.3%)	5 (10.0%)
	Absent	134 (94.4%)	89 (96.7%)	45 (90.0%)
Pleural invasion	Present	14 (9.9%)	8 (8.7%)	6 (12.0%)
	Absent	128 (90.1%)	84 (91.3%)	44 (88.0%)
Tumor location	Central	115 (50.0%)	73 (49.0%)	42 (51.9%)
	Peripheral	115 (50.0%)	76 (51.0%)	39 (48.1%)
Histologic type	Adenocarcinoma	137 (59.6%)	82 (55.0%)	55 (67.9%)
	Squamous cell carcinoma	93 (40.4%)	67 (45.0%)	26 (32.1%)

Table 3 Internal discrimination for the 19-feature configuration and secondary exploratory comparisons with the 15- and 10-feature configurations

Feature set	ROC-AUC (95% CI)	AP (95% CI)	F1 (95% CI)
19 features	0.685 (0.611–0.751)	0.523 (0.422–0.641)	0.548 (0.435–0.621)
15 features	0.670 (0.594–0.737)	0.481 (0.383–0.597)	0.540 (0.354–0.557)
10 features	0.667 (0.593–0.734)	0.488 (0.389–0.599)	0.510 (0.380–0.575)

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- supplementary.pdf