

Development and Validation of a Nomogram and Machine Learning Models for Preoperative Prediction of IASLC Grade 3 Lung Adenocarcinoma

Chia Liu^{1,2,3} · Herdiantri Sufriyana³ · Chien-Sheng Huang^{1,2,*} · Emily Chia-Yu Su^{3,*}

¹Division of Thoracic Surgery, Taipei Veterans General Hospital, Taipei, Taiwan · ²School of Medicine, College of Medicine, National Yang Ming Chiao Tung University, Taipei, Taiwan · ³Institute of Biomedical Informatics, College of Medicine, National Yang Ming Chiao Tung University, Taipei, Taiwan

*Corresponding authors: huangcs@vghtpe.gov.tw; emilysu@nycu.edu.tw · Underlined = presenting author

BACKGROUND

The IASLC grading system identifies **grade 3 lung adenocarcinoma** as a high-risk subgroup. Grade is assigned on the resection specimen and is therefore unavailable at the time the operation is planned.

OBJECTIVE

To develop and temporally validate a **preoperative model for IASLC grade 3**, and to examine whether the predicted classification identifies patients in whom the adequacy of nodal evaluation is associated with survival.

COHORT

Cohort	n	Grade 3
Derivation 2006–2017	1,018	385 (37.8%)
Temporal validation 2018–2020	202	123 (60.9%)

METHODS

Inclusion. Curative-intent resection for solitary non-small cell lung cancer.

Exclusion. Non-adenocarcinoma on final pathology; microscopically or macroscopically positive margin (R1/R2); no preoperative FDG-PET/CT.

Model. Multivariable logistic regression; consolidation-to-tumor ratio (CTR) and SUVmax modeled with restricted cubic splines. Five machine-learning algorithms trained in parallel with five-fold cross-validation.

Nodal evaluation. Evaluation was defined as adequate when at least three N1 nodes from three N1 stations and three N2 nodes from three N2 stations, including the subcarinal station, were examined, which constitutes a complete resection under the IASLC definition (R0); otherwise the resection was uncertain [R(un)].

Endpoints. Overall, cancer-specific, and recurrence-free survival by Kaplan–Meier, with multivariable Cox regression for adjusted comparisons.

Cohort characteristics by IASLC grade

P1.162 · 3 / 8

DERIVATION 2006–2017 (n = 1,018)				TEMPORAL VALIDATION 2018–2020 (n = 202)			
Derivation cohort	Grade 1–2 (n=633)	Grade 3 (n=385)	p	Temporal validation	Grade 1–2 (n=79)	Grade 3 (n=123)	p
Male sex	38.1%	49.1%	<0.001	Male sex	32.9%	43.9%	0.119
Smoking history	23.9%	40.5%	<0.001	Smoking history	17.7%	33.3%	0.015
Elevated CEA	9.5%	27.3%	<0.001	Elevated CEA	13.9%	26.8%	0.030
CTR, median	0.54	0.76	<0.001	CTR, median	0.54	0.78	<0.001
Maximal SUV, median	1.9	6.1	<0.001	Maximal SUV, median	2.6	6.8	<0.001
Central location	11.4%	29.4%	<0.001	Central location	32.9%	43.1%	0.148
Lobectomy	81.7%	86.5%	0.045	Lobectomy	69.6%	91.1%	<0.001
Nodal metastasis (N1–2)	4.7%	21.8%	<0.001	Nodal metastasis (N1–2)	2.7%	31.1%	<0.001
Angiolymphatic invasion	9.6%	44.7%	<0.001	Angiolymphatic invasion	7.6%	58.7%	<0.001

Grade 3 tumors were consistently more often seen in men and smokers, with higher CEA, higher CTR and higher maximal SUV, more central tumours, and far more nodal metastasis and angiolymphatic invasion. Grade 3 prevalence rose from **37.8%** in derivation to **60.9%** in the validation era.

MODEL DEVELOPMENT		
Model	Variables	AUC
Random forest	10	0.856
Ridge regression	10	0.841
Support vector machine	10	0.839
Decision tree	10	0.835
Naïve Bayes	10	0.790
Nomogram	4	0.839

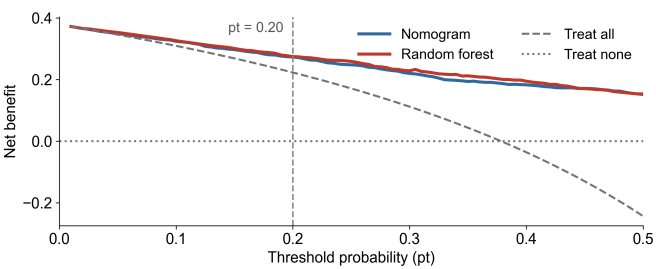
Derivation cohort (n = 1,018). Random forest gave the highest area under the curve (0.856); the nomogram reached 0.839 with only four variables. Hyperparameters were selected by five-fold cross-validation.

DECISION CURVE ANALYSIS

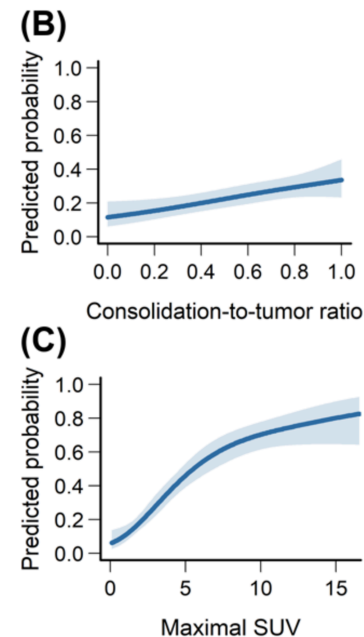
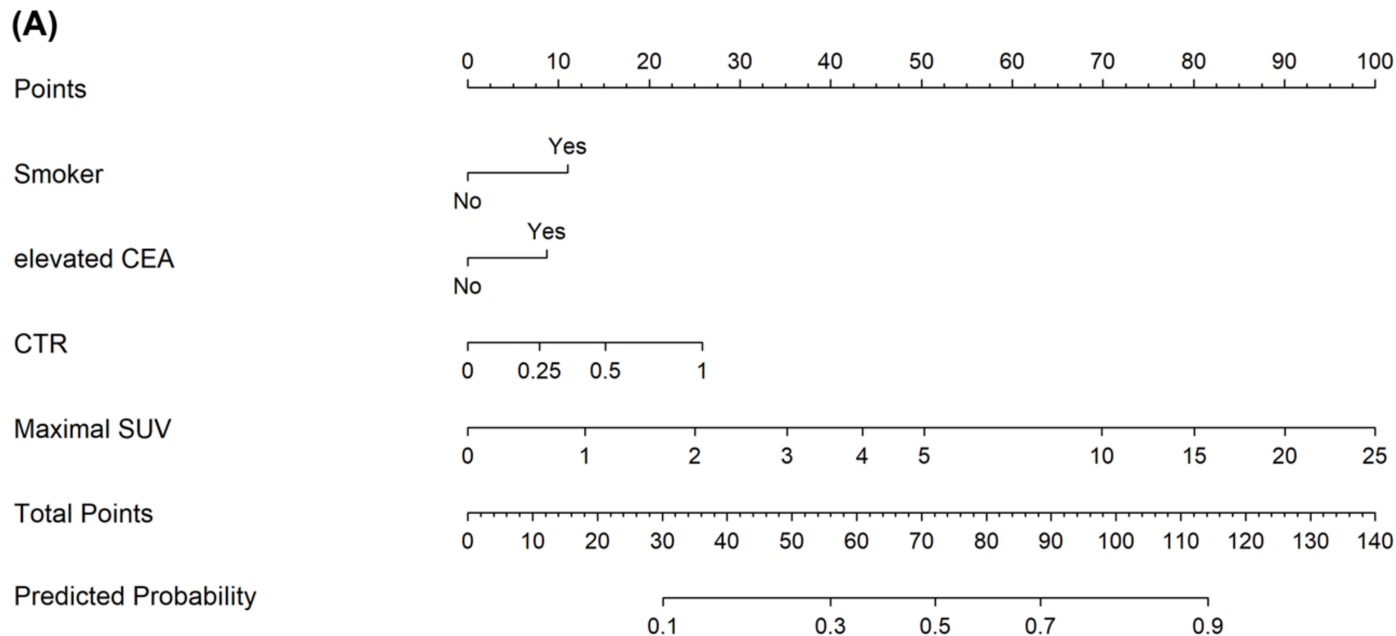
Derivation cohort. At a threshold of 0.20 the net benefit was **0.273** for the nomogram and **0.274** for the random forest, against 0.223 for treating all. The four-variable nomogram is therefore no less useful in decision terms than the ten-variable model.

NOMOGRAM PREDICTORS		
Predictor	OR (95% CI)	p
Smoking history	1.70 (1.22–2.36)	0.002
Elevated CEA	1.59 (1.05–2.41)	0.029
CTR	5.69 (2.92–11.07)	<0.001
Maximal SUV	1.34 (1.26–1.42)	<0.001

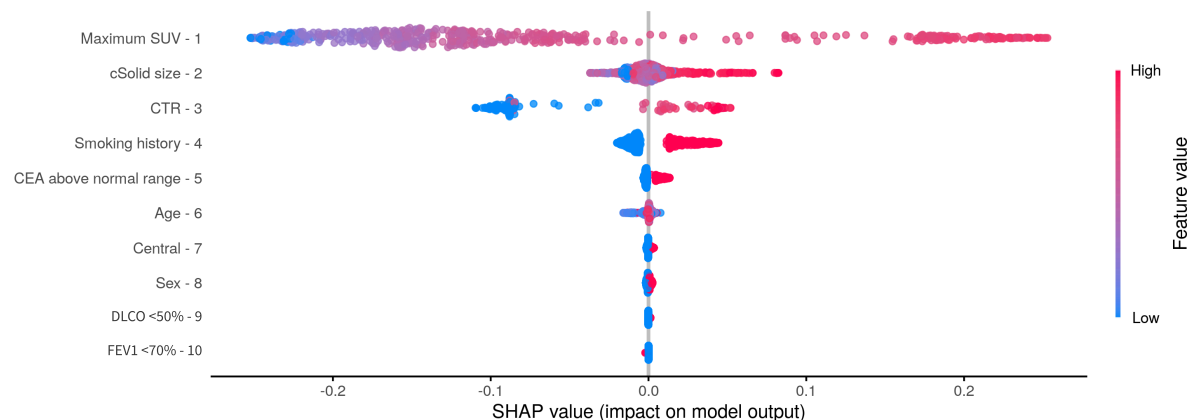
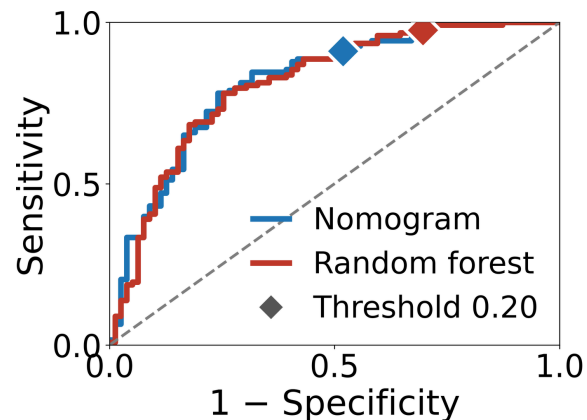
Linear-term odds ratios, shown for interpretability; the deployed model uses restricted cubic splines.



PREOPERATIVE NOMOGRAM AND SPLINE EFFECTS



(A) Four routinely available preoperative variables: smoking history, elevated CEA, CTR and PET SUVmax; total points map to the predicted probability of IASLC grade 3. **(B, C)** Partial effect of the two spline terms with 95% bands: the effect of SUVmax rises steeply to about 7 and then flattens, so a constant per-unit odds ratio overstates it at the top of the range.



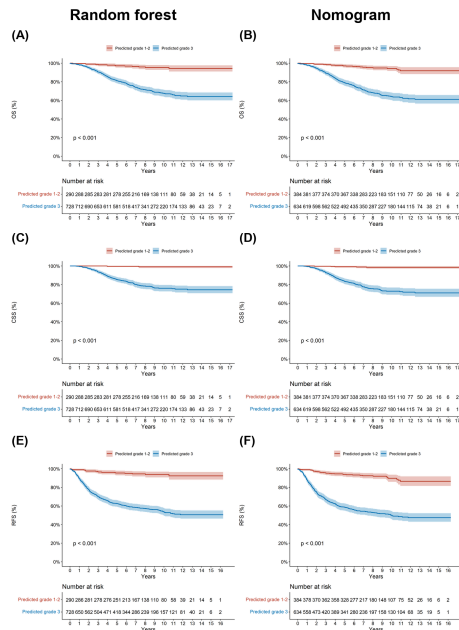
Left: temporal validation, with the threshold of 0.20 marked; the two curves are almost superimposed. **Right:** SHAP summary for the random forest in the derivation cohort — all four nomogram variables rank among its five most influential, and the remaining six contribute little.

TEMPORAL VALIDATION AT A THRESHOLD OF 0.20 (n = 202)

Model	AUC	Sensitivity	Specificity	PPV	NPV	Calibration slope
Random forest	0.811	97.1%	30.2%	67.9%	87.6%	0.97
Nomogram	0.815	91.1%	48.1%	73.2%	77.6%	0.90

Both models under-predicted risk as grade 3 prevalence rose from 37.8% to 60.9%; calibration intercept 0.79 for the random forest and 0.70 for the nomogram.

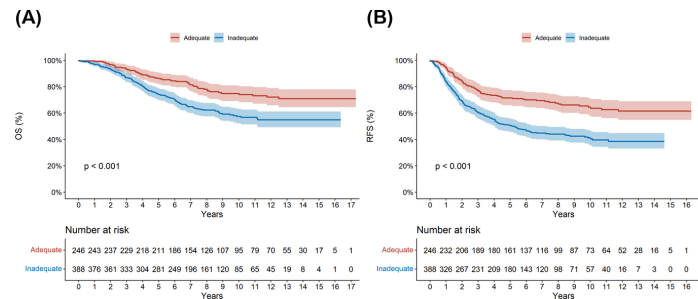
SURVIVAL BY PREDICTED CLASSIFICATION



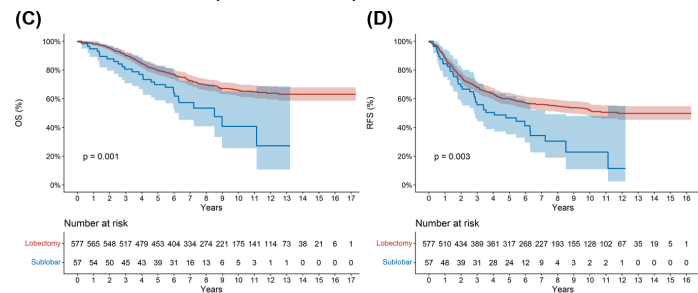
Derivation cohort ($n = 1,018$) at a threshold of 0.20; left column random forest, right column nomogram. (A, B) Ten-year OS 67.7% versus 95.4% and 64.3% versus 94.2%; (C, D) CSS 76.0% versus 99.1% and 72.9% versus 98.5%; (E, F) RFS 53.9% versus 93.8% and 50.2% versus 89.7% (all $p < 0.001$). Both models were developed on this cohort, so neither set of curves is out-of-sample.

SURGICAL IMPLICATIONS IN PREDICTED GRADE 3

Adequacy of nodal evaluation ($n = 246$ vs 388)



Extent of resection ($n = 577$ vs 57)



Nomogram-predicted grade 3 ($n = 634$). (A, B) Adequate (IASLC R0) versus inadequate [R(un)] nodal evaluation: ten-year OS 74.8% versus 57.4% ($p < 0.001$) and RFS 64.4% versus 41.1% ($p < 0.001$). (C, D) Lobectomy versus sublobar resection: OS 66.0% versus 40.8% ($p = 0.001$), RFS 52.2% versus 22.9% ($p = 0.003$). All comparisons are unadjusted.

NODAL EVALUATION VERSUS EXTENT OF RESECTION

Endpoint	Adequate nodal evaluation	Lobectomy vs sublobar
OS	0.58 (0.43–0.80)*	0.90 (0.56–1.45)
RFS	0.51 (0.39–0.66)*	0.73 (0.49–1.09)

Nomogram-predicted grade 3 (n = 634). Adjusted hazard ratios (95% CI); * $p < 0.01$. Adjusted for age, CCI, sex, FEV₁, DLCO, clinical stage and extent of resection.

CLINICAL APPLICATION

Predicted probability ≥ 0.20 . Systematic nodal evaluation should be prioritized: at least three N1 and three N2 stations, including the subcarinal station.

Below 0.20. A low prediction lowers the risk without removing it: 36 of 384 were still grade 3.

CONCLUSIONS

- 1 IASLC grade 3 was predicted accurately from routine preoperative data. In the derivation cohort the area under the curve was **0.839** for the four-variable nomogram and **0.790** to **0.856** across five machine-learning algorithms; in temporal validation the nomogram reached **0.815** and the random forest **0.811**.
- 2 Among patients classified as predicted grade 3, **adequate nodal evaluation** was independently associated with better survival for all three endpoints. Sublobar resection carried worse unadjusted survival, but **this did not persist after adjustment**, and the two comparisons are confounded.
- 3 Either model could identify, before operation, patients in whom **systematic nodal evaluation** is most likely to be relevant. **Prospective evaluation is required** before either is applied to operative decision-making.

Limitations: single center, retrospective design, and preoperative PET/CT was required for inclusion. The association between nodal evaluation and survival is observational and may be influenced by unmeasured differences between groups. This cohort overlaps two previous reports from our institution that addressed different questions (Eur J Surg Oncol 2024;50:108540 · Eur J Cardiothorac Surg 2025, doi:10.1093/ejcts/ezaf083).

AI use disclosure: Generative artificial intelligence was used to assist with language editing and figure layout for this presentation. The authors reviewed and verified all content and take full responsibility for its accuracy and integrity. The machine learning models reported here are the subject of the research and were developed by the authors.