Using integrated radiomic and pathomic-based models to predict progression-free survival in early-stage lung adenocarcinoma

Tengyue Zhang[1], Anil Yadav*[2,3], Ruiwen Ding*[2,3], Sean Johnson[3], Denise Aberle[2,3], Ashley Prosper[3], and William Hsu[2,3]

[1]BIG Summer Program, Institute for Quantitative and Computational Biosciences, UCLA, Los Angeles, CA, USA
[2]Department of Bioengineering, UCLA, Los Angeles, CA, USA
[3]Medical & Imaging Informatics, Department of Radiological Sciences, UCLA, Los Angeles, CA, USA
*Equal contributions

Background

- Non-small cell lung cancer (NSCLC) patients have a relatively high rate of recurrence after surgery, reflecting a need to identify patients with a high risk of recurrence and offer personalized adjuvant therapies.
- Images from radiology and pathology provide valuable information for the NSCLC recurrence prediction task.
- How can the combination of radiomic and pathomic features extracted from pre-surgery computed tomography (CT) scans and hematoxylin and eosin (H&E)-stained whole slide images (WSIs) effectively predict progression-free survival in early-stage lung adenocarcinoma (LUAD) patients?
- Progression-free survival is defined as the time from surgical resection to disease recurrence or lung cancer death.

Data Description

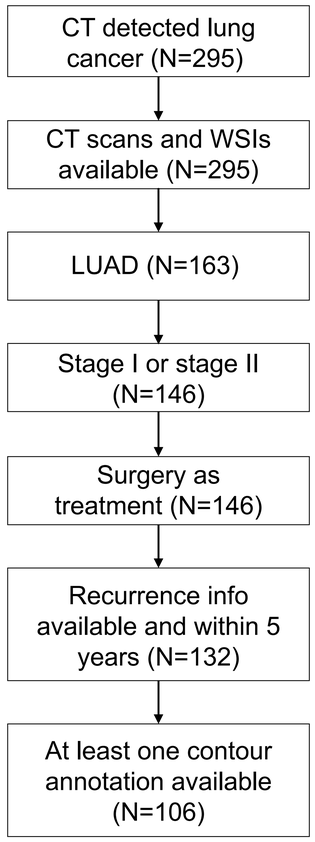

Figure 1. Dataset inclusion criteria for patient selection.

	Our cohort
# patients	106
Mean age	63.7
# stage I	89 (84.0%)
# Female	52 (49.1%)
# 5-year recurrence or lung-cancer death	26 (24.5%)
# 5-year recurrence and lung-cancer death	17 (16.0%)
Median time to even in days for non-censored	774 (157 – 1637)
Median follow-up in days for censored	1378.5 (54 – 2136)

Table 1. Clinical information and outcome of patients in our cohort selected from NLST [1].

Methods

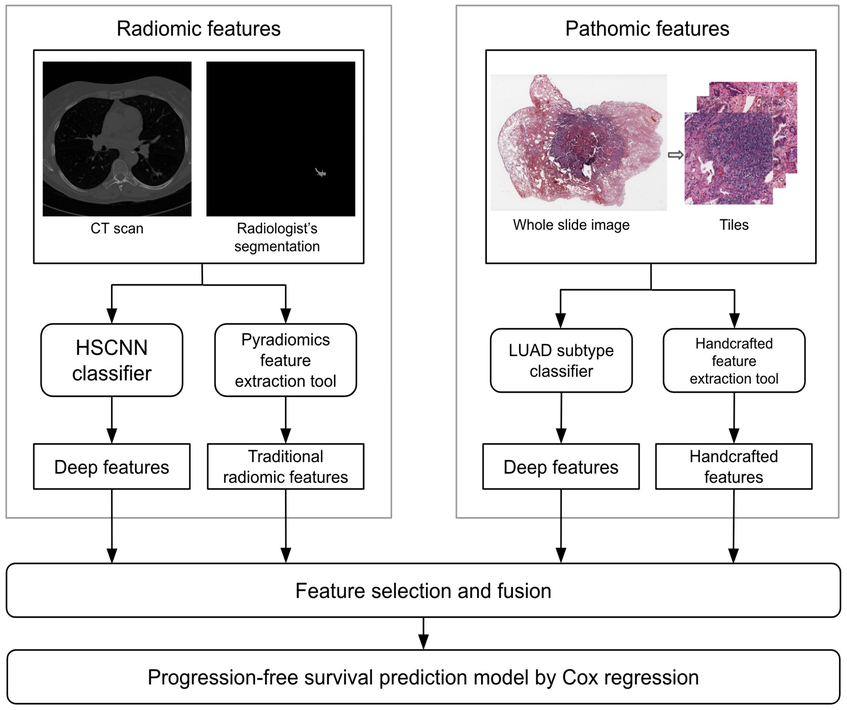

Figure 2. Flowchart illustrating the proposed method of using integrated radiomic- and pathomic-based models to predict progression-free survival in early-stage LUAD.

Radiomic Feature Extraction

- 256 deep features were extracted using the HSCNN model [2].
- 200 Pyradiomics [3] features were extracted from intratumoral and peritumoral regions.
  - Intratumoral region: the radiologist-outlined nodule region.
  - Peritumoral region: dilating the nodule contour to a 20-mm peritumoral radius and subtracting the nodule contour from the dilated contour to generate a ring-shaped region.
- Nodule-level features were aggregated into patient-level features.

Pathomic Feature Extraction

- 512 deep features were extracted from each tile using a LUAD histologic subtype classifier (unpublished).
- 27 hand-crafted features were extracted from each tile.
  - The hand-crafted features describe immune cell density and spatial colocalization with tumor cells within each tile [4,5].
- Tile-level features were aggregated into patient-level features.

Survival Analysis

- Features were selected via the Maximum Relevance Minimum Redundancy (mRMR) algorithm individually from each modality.
- Cox proportional hazards regression model with ridge regularization was used as the survival analysis tool.
- Cross-validation: 5-fold cross-validation with 10 repetitions, 60% training, 20% validation, and 20% held-out testing patient cases.

Results

	Training C-index	Validation C-index	Test C-index
(R, P)	0.8914 ± 0.0262	0.6311 ± 0.1279	**0.6336 ± 0.1302**
(R, P, C)	**0.8924 ± 0.0258**	0.6299 ± 0.1276	0.6273 ± 0.1236
(R)	0.7845 ± 0.0342	0.5829 ± 0.1416	0.6117 ± 0.1429
(P)	0.8520 ± 0.0382	**0.6338 ± 0.1413**	0.5838 ± 0.1494
(C)	0.6271 ± 0.0528	0.4985 ± 0.1456	0.4770 ± 0.1614

Table 2. Mean and standard deviation of concordance index (C-index) using repeated 5-fold cross validation. (R, P) is the fused model with radiomic and pathomic features. (R, P, C) is the fused modal with radiomic, pathomic, and clinical features. (R) is the radiomic-only model. (P) is the pathomic-only model. (C) is the clinical-only model.

Modality	Feature type	Number selected
Radiomic	Deep feature	1.86
	Pyradiomics feature	5.14
Pathomic	Deep feature	6.14
	Hand-crafted feature	0.86

Table 3. Summary of the average number of features selected in each type throughout repeated cross validation.

Conclusion

- The combined radiomic-pathomic model provides the most promising results.
- Pyradiomics radiomic and deep pathomic features were often selected as the most informative.
- This multi-modal approach underscores the value of combining multimodal data for prognostication and presents promising results toward informing treatment strategies in lung cancer care.
- In the future, other intermediate fusion techniques, such as canonical correlation analysis, can be explored, and datasets from other institutions will be added to further improve and validate the current model.

References

[1] National Lung Screening Trial Research Team, "The national lung screening trial: overview and study design," *Radiology* **258**, 243–253 (Jan. 2011).
[2] Shen, S. et al., "An interpretable deep hierarchical semantic convolutional neural network for lung nodule malignancy classification," *Expert Syst. Appl.* **128**, 84–95 (Aug. 2019).
[3] van Griethuysen, J. J. M. et al., "Computational radiomics system to decode the radiographic phenotype," *Cancer Res.* **77**, e104–e107 (Nov. 2017).
[4] Shaban, M. et al., "A novel digital score for abundance of tumour infiltrating lymphocytes predicts disease free survival in oral squamous cell carcinoma," *Sci. Rep.* **9**, 13341 (Sept. 2019).
[5] Ding, R. et al., "Image analysis reveals molecularly distinct patterns of TILs in NSCLC associated with treatment outcome," *NPJ Precis. Oncol.* **6**, 33 (June 2022).