#23875 Uncertainty-Aware Multiscale Treatment Response Prediction for Personalized Radiotherapy using Conformal Prediction in Advanced Non-Small Cell Lung Cancer (NSCLC)

Faisal Yaseen[1], Daniel S. Hippe[2], Sunan Cui[1], Jie Fu[1], Shouyi Wang[3], Chunyan Duan[4], John H. Gennari[1], Stephen R. Bowen[1,2]

1. University of Washington, Seattle. 2. Fred Hutchinson Cancer center. 3. University of Texas at Arlington. 4. Tongji University, China

2026 JULY 19–22 VANCOUVER, BC
JOINT AAPM | COMP MEETING

ABSTRACT

Objective: Multiscale treatment response prediction in advanced NSCLC enables personalized dose painting, yet prediction point estimates alone do not convey the uncertainty required for safe adaptive dose escalation. We developed a multiscale conformal prediction framework that provides predictions and intervals for mid treatment FDG-PET response at both voxel and lesion levels.

Methods: Data were utilized from two prospective clinical trials: FLARE-RT (NCT02773238, n=25 locally-advanced NSCLC) and BRIGHT (NCT04151940, n=19 metastatic NSCLC). Patients underwent FDG-PET/CT at baseline and week 3 of therapy (chemoradiation in FLARE-RT and chemoimmunotherapy in BRIGHT). A hybrid generalized least squares model was used to predict tumor voxel standardized uptake value (SUV) at mid-treatment from baseline patient-level and voxel-level covariates, building on a previously developed model (Voxel-Forecast) with re-optimized Stable variogram modeling for spatial correlation. We compared the forecast performance (cross-validated RMSE) using patient-level features only versus multiscale features. To quantify uncertainty, we applied nested leave-one-out jackknife conformal prediction framework (targeting 80% coverage, alpha = 0.2) with spatially adaptive weighting and multiscale calibration sets.

Results: Incorporating voxel-level features with patient-level features achieved significantly lower RMSE (0.22 vs. 0.34 [FLARE-RT] and 0.44 vs. 0.64 [BRIGHT], $p<0.05$) than patient-level only. Voxel-level coverage of the conformal prediction intervals was 79.0% ± 22.6% in FLARE-RT and 80.5% ± 15.0% in BRIGHT, close to the target level of 80%, with mean interval widths of 0.7± 0.3 and 1.0 ± 0.2, respectively. Lesion/patient-level coverage was lower: 71.4% ± 25.1% in FLARE-RT and 68.8% ± 17.5% in BRIGHT, with narrower mean widths of 0.6 ± 0.2 and 0.9 ± 0.3.

Conclusion: Incorporating patient- and voxel-level information improves response prediction and uncertainty-aware voxel-level forecasting to enable personalized adaptive dose painting. The proposed multiscale conformal prediction approach provides rigorous, distribution-free safety intervals for mid treatment FDG-PET response, supporting risk-aware dose escalation beyond point-based predictions.

CONTACT: fyaseen@uw.edu

INTRODUCTION

- Advanced NSCLC therapy (chemoradiation for locally-advanced disease; chemo-immunotherapy ± localized radiotherapy for metastatic disease) yields limited, toxicity-constrained benefit.
- Spatially-adaptive dose painting guided by mid-treatment FDG-PET response could improve the therapeutic ratio, but requires voxel-level forecasts that are both accurate and honest about uncertainty.
- Standard conformal prediction (CP) [1] gives distribution-free but spatially uniform, fixed-width intervals; we pair a re-optimized hybrid GLS/OLS Voxel-Forecast [2] with Residual-Variance CP (VarCP) that adapts interval width to local difficulty while preserving finite-sample coverage.

METHODS AND MATERIALS

Cohorts & imaging. FLARE-RT (n=25 locally-advanced, chemoradiation; 26,980 voxels) and BRIGHT (n=19 metastatic, chemo-ICI; 11,100 voxels). FDG-PET/CT at baseline and mid-treatment (week 3); response = mid/baseline SUV ratio.

Voxel-Forecast model. Hybrid GLS/OLS jointly fits patient-level coefficients (OLS) and voxel-level spatially-correlated residuals (GLS), using a two-stage Stable variogram selected from 5 families (16 configurations).

Conformal prediction. Nested leave-one-out calibration in log-space. StdCP: global residual quantile → fixed-width intervals. VarCP: residuals normalized by a learned Random-Forest scale model (200 trees, depth 6) → adaptive intervals.

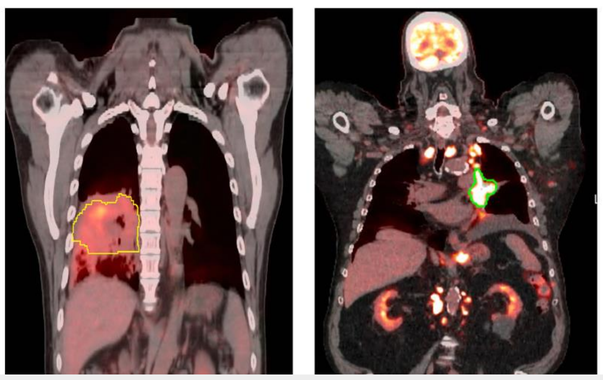

Fig 1. *FDG-PET/CT of locally-advanced (left) vs. metastatic (right) NSCLC; analyses restricted to the primary lung tumor*

Evaluation. α = 0.10 / 0.15 / 0.20 (90/85/80% intervals) at voxel and lesion (median-aggregated) levels. Metrics: coverage, mean interval width, per-patient paired Wilcoxon signed-rank tests.

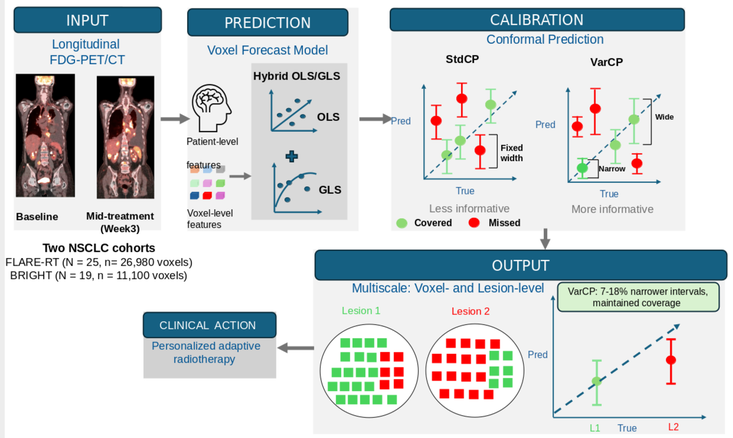

Fig 2. *Multiscale VarCP workflow: longitudinal FDG-PET/CT → hybrid OLS/GLS Voxel-Forecast → conformal calibration (StdCP fixed-width vs. VarCP adaptive) → voxel- & lesion-level prediction intervals → personalized adaptive radiotherapy.*

RESULTS

- **Cohort characteristics.** Comparable demographics across FLARE-RT (n=25) and BRIGHT (n=19): median age 62 vs 65 y, ~60% female, mostly former/never smokers.
- **Tumor burden & response.** BRIGHT tumors were larger and more heterogeneous (median MTV 101 vs 27 cm^3; intra-tumoral CV 0.50 vs 0.32); response ratio 0.78 vs 0.62, with 79% vs 92% of tumors responding (mid/pre <1.0).
- **Multiscale features improve accuracy.** Voxel + patient features beat patient-only in both cohorts (Fig 2). The Stable variogram was significantly superior to Matérn ($p<0.001$).

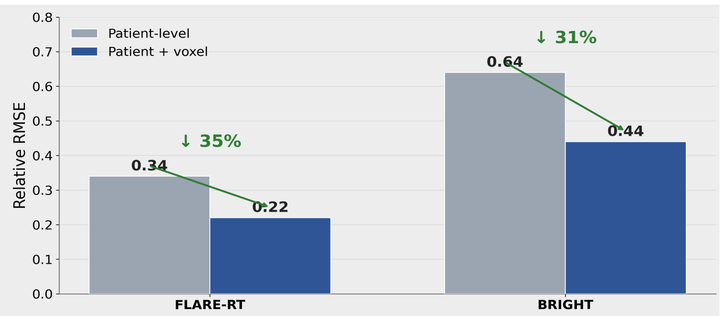

Fig 3. *Multiscale Adding Voxel-level to patient-level features lowers relative RMSE by 35% (FLARE-RT) and 31% (BRIGHT).*

VarCP narrows intervals. Across both cohorts and all α, VarCP gave narrower voxel intervals than StdCP: 7–18% (FLARE-RT) and 4–8% (BRIGHT) width reduction (Fig 4), significant per-patient (paired Wilcoxon).

Coverage holds near nominal. Empirical voxel coverage held near the (1−α) target for both methods (≤0.003 apart), so VarCP tightens intervals without losing validity (Fig 5).

RESULTS

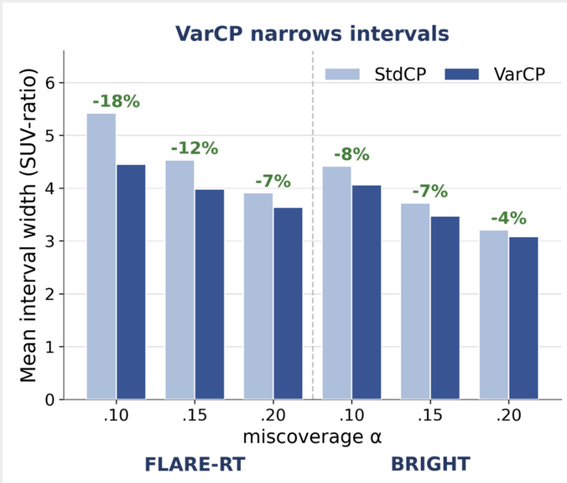

Fig 4. VarCP produces narrower voxel-level intervals than StdCP across all α, with 4–18% mean width reduction (per-patient paired Wilcoxon).

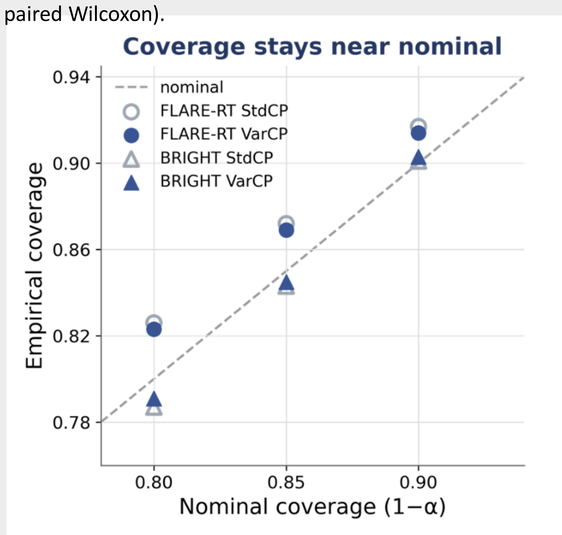

Fig 5. Empirical voxel-level coverage tracks the nominal (1−α) target for both StdCP and VarCP across cohorts and all three α levels.

DISCUSSION & CONCLUSION

- **Uncertainty-aware framework.** Coupling a Stable-variogram Voxel-Forecast with VarCP yields heteroscedastic, distribution-free intervals that preserve coverage; voxel-level gains propagate to median-aggregated lesion intervals.
- **Adaptive & interpretable.** VarCP tightens intervals where prediction is easy and widens them in peripheral, high-uptake regions, letting clinicians pair lesion-level robustness with voxel-level safety maps for risk-aware dose painting.
- **Limitations & next steps.** Small single-set cohorts (n=25/19), no external validation, and no clinician tool yet; next: larger cohorts, metastatic lesions, other tumors, and a prospective interface.

REFERENCES

1. Bowen SR, Hippe DS, Chaovalitwongse WA, Duan C, Thammasorn P, Liu X, Miyaoka RS, Vesselle HJ, Sinahan PE, Rengan R, Zeng J. Voxel Forecast for Precision Oncology: Predicting Spatially Variant and Multiscale Cancer Therapy Response on Longitudinal Quantitative Molecular Imaging. Clin Cancer Res. 2019 Aug 15;25(16):5027-5037. doi: 10.1158/1078-0432.CCR-18-3908. Epub 2019 May 29. PMID: 31142507; PMCID: PMC6697586.
2. A gentle introduction to conformal prediction, https://doi.org/10.1561/2200000101}