SIIM CAIMI26 Conference on Artificial Intelligence in Medical Imaging
OCT 26-27 | University of Pennsylvania, Philadelphia, PA

Class-conditional Conformal Prediction for Three-class WHO 2021 Glioma Molecular Subtyping: Development and External Validation of an MRI Radiomics Model

Ramkumar Rajabathar Babu Jai Shanker, MS, University of Chicago; Jonathan Neilio, MD;
Daniel Ginat, MD, MS

Introduction/Background
Purpose: To evaluate whether class-conditional conformal prediction provides usable uncertainty for three-class WHO 2021 glioma subtyping under domain shift, where single-label classification rarely identifies oligodendroglioma

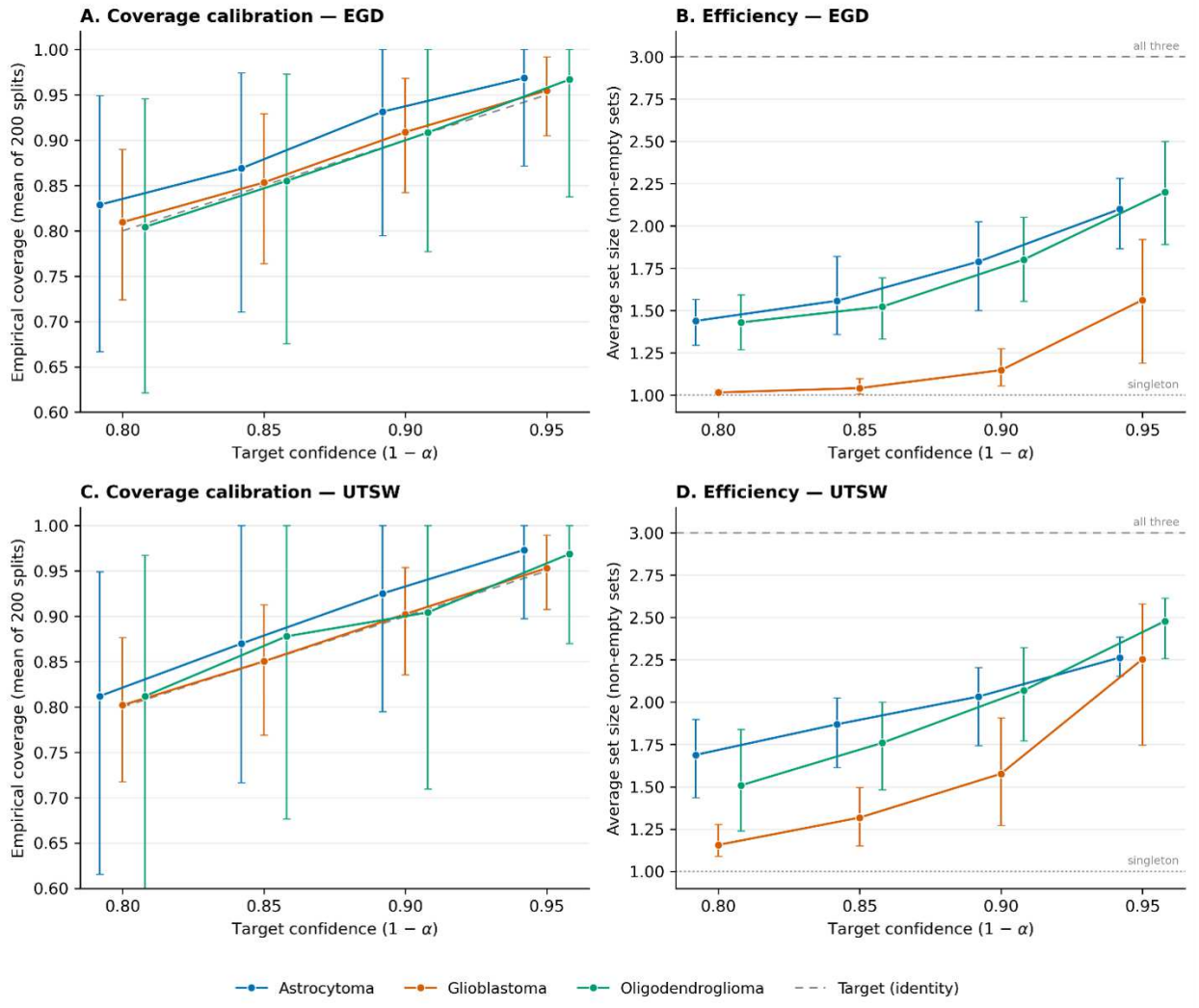
Methods/Intervention
Materials and Methods: In this retrospective study, 1544 patients with preoperative multiparametric MRI and molecular labels came from four datasets. A random forest trained on whole-tumor radiomic features in a development cohort (The Cancer Genome Atlas and UCSF-PDGM; n = 611) was applied to two external cohorts (Erasmus Glioma Database [EGD], n = 404; UTSW-Glioma, n = 529). Argmax, a class-weighted rule, and class-conditional (Mondrian) split-conformal prediction were compared, with per-cohort thresholds across 200 splits (80%-95% confidence).

Results/Outcome
Results: Discrimination held externally (AUC, 0.87-0.96 [EGD]; 0.82-0.91 [UTSW-Glioma]). Argmax identified oligodendroglioma in only 7 of 73 EGD (sensitivity, 0.10) and 4 of 63 UTSW-Glioma cases (0.06), and a class-weighted rule shifted the failure to another class. Conformal prediction held per-class coverage near target at a class-asymmetric cost: at the 90% target, glioblastoma sets were near-singleton whereas oligodendroglioma sets were predominantly two-class, and all-three-class sets became common at 95% (50% in UTSW-Glioma). Overall, 53% of cases resolved to a single-label set (94% accuracy), the rest deferred for molecular testing. Transporting calibration without a local split lowered coverage by up to 0.14.

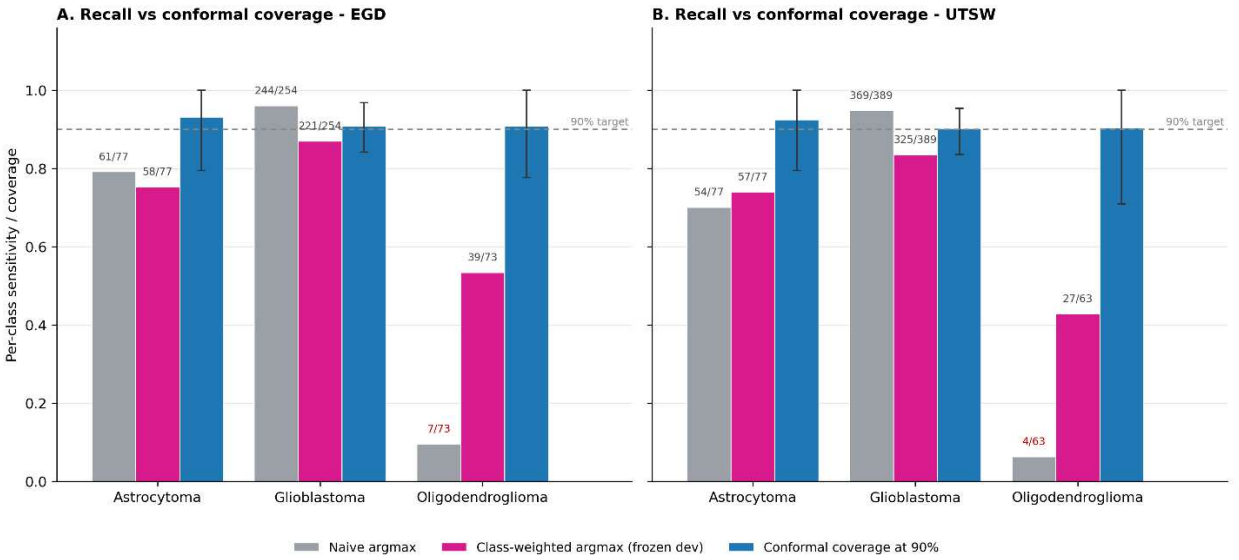
Conclusion
Conclusion: Conformal prediction did not improve accuracy but turned silent oligodendroglioma misclassification into larger prediction sets. The efficiency cost was class-asymmetric, larger in the more dissimilar cohort, and dependent on local calibration data; it is best used as a triage signal rather than a standalone classifier.

Statement of Impact
A single-label prediction forces one answer even when the classes cannot be separated, which externally produced near-silent oligodendroglioma misclassification (sensitivity 0.10 and 0.06). A prediction set instead reports every class that cannot be ruled out, turning the same case into a multi-label set that flags for molecular testing rather than a confident wrong label. The tradeoff is explicit: 53% resolve to a single label at

94% accuracy, the rest deferred. The cost is class-asymmetric and calibration must be local.



Empirical per-class coverage on the external test sets across target confidence levels (80%, 85%, 90%, 95%).



Comparison of argmax, class-weighted argmax, and class-conditional conformal prediction on the EGD and UTSW-Glioma external test sets.

Keywords

Uncertainty quantification; Conformal prediction; Glioma molecular subtyping; Radiomics