SIIM CMIMI24 Oct 21-22 | Boston, MA
Conference On Machine Intelligence In Medical Imaging

Cross-Vendor Reproducibility of Radiomics-based Machine Learning Models for Computer-aided Diagnosis

Jatin Chaudhary, University of Turku; Ivan K. Jambor; Hannu Aronen, MD, MSc; Otto Ettala, MD; Jani Saunavaara, PhD; Peter Bostrom, MD, PhD; Jukka Heikkonen, DTech; Rajeev Kanth, DSc; Harri Merisaari, PhD

Introduction/Background

Prostate cancer (PCa) remains the most common cancer among men in the western world and the second leading cause of cancer-related deaths. The integration of machine learning (ML) into Magnetic Resonance Imaging (MRI) holds promise for enhancing the accuracy and efficiency of PCa diagnostics. This study aims to evaluate the reproducibility and performance of ML models using radiomic features derived from Pyradiomics and MRCradiomics across different MRI scanners.

Methods/Intervention

We utilized imaging data from 637 men with clinical suspicion of PCa, enrolled in various clinical trials. The data included axial T2-weighted images scanned using Siemens MAGNETOM Verio 3T and Philips Ingenia 3T MRI devices. Radiomic features were extracted using Pyradiomics and MRCradiomics packages, yielding a total of 2693 features. Feature selection was conducted using the Maximum Relevance Minimum Redundancy (MRMR) method, reducing the set to 14 highly predictive features. We trained and evaluated Support Vector Machine (SVM) and Random Forest models on training, validation, and test datasets, assessing their performance using Area Under the Curve (AUC) metrics.

Results/Outcome

The SVM model achieved an AUC of 0.74 on the Multi-Improd dataset using combined Pyradiomics and MRCradiomics features, but the AUC dropped to 0.35 on the Philips test set. The Random Forest model showed similar trends with AUCs of 0.73 on the Multi-Improd set and 0.60 on the Philips set. Models trained exclusively on Pyradiomics features demonstrated higher robustness, with the Random Forest achieving an AUC of 0.78 on the Philips set. In contrast, models using only MRCradiomics features had varied outcomes, highlighting the challenge of reproducibility across different scanners.

Conclusion

This study underscores the significant impact of scanner-induced variability on the performance of ML models in PCa diagnostics. While combining Pyradiomics and MRCradiomics features enhances predictive performance, rigorous cross-platform validation is crucial to ensure model reliability.

Statement of Impact

Our findings emphasize the need for standardized imaging protocols and comprehensive validation frameworks to bridge the gap between innovative AI applications and practical, patient-centric healthcare solutions. This research advances the field of PCa imaging and promotes the broader adoption of AI in medical diagnostics, ultimately aiming to improve diagnostic accuracy and patient outcomes.

Keywords
Inter-Vendor Reproducibility; Radiomics; Diagnostic tools; Model Reproducibility; Machine Learning; Prostate Cancer