Radiation Oncology

Outcome Prediction for Head and Neck Cancer Adaptive Radiation Therapy Using Pre- and During-Treatment Imaging

Jing Wang, Ph.D.
Associate Professor
Director of Data Analytics and Informatics
Advanced Imaging and Informatics for Radiation Therapy (AIRT) lab and Medical Artificial Intelligence and Automation (MAIA) lab
Department of Radiation Oncology

Jing.Wang@utsouthwestern.edu

UT Southwestern Medical Center

Adaptive Radiation Therapy

- Adaptation to anatomical change
  - Increased therapeutic dose for non-responders, or reduced dose for early responders

- Risk adaptation for treatment (de-)intensification
  - Reduce or eliminate dose to low-risk targets
  - Increased therapeutic dose for non-responders, or reduced dose for early responders
  - Additional systemic therapy for patients at high-risk for distant metastasis

Biomarkers for risk stratification in head and neck cancer

- Human papillomaviruses (HPV) type 16 associated oropharyngeal cancers
  - Markedly improved survival

- Imaging-based markers
  - Pre-treatment PET has prognostic values
  - SUV of the primary tumor was associated with disease-free survival (DFS), OS and local control
  - Often based on a single measurement, e.g., SUVmax or SUVmean


AI or Machine Learning

- Analyze/model complex data
  - Integrate information from different sources
    - Imaging [radiological/pathological]
    - Clinical
    - Biology
    - Complex patterns
      - Texture of images

Radiomics-based Modeling

- Explosion of radiomics studies over last decade
  - Imaging-based predictive models
  - Extraction and analysis of large amount of features from medical images
  - Building predictive models from extracted imaging features, often in combination with other features such as clinical characteristics

Head and Neck Outcome Prediction

- Toxicity

- Treatment failure: distant metastasis, local regional failure

- Survival
How to choose a classifier?

- SVM
- Logistic Regression
- Decision Tree
- Discriminant Analysis
- K-Nearest Neighbors
- Naïve Bayesian
- Random forest
- CNN...

Model performance strongly depends on data: different runs on different training, validation, test splittings may result different preferred classifiers.

Multi-Classifier Multi-Objective and Multi-Modality (mCOM)

- Explicitly considers both sensitivity and specificity, critical for imbalanced dataset.
- Instead of choosing a specific classifier, we aim to maximally utilize information extracted by different classifiers.
- Lead to more robust prediction results.

Reliable fusion

- Fusing information extracted from individual classifier/source by combining the output scores with both weight and reliability.

- 277 patients from 4 institutions, a public H&N dataset downloaded from TCIA
- 40 experienced locoregional recurrence
- Median follow-up: 43 months
- Median time to locoregional recurrence: 18 months
- Model trained on data from two institutions while tested on other two institutions.

Performance of models built with different classifiers and features from different modalities

<table>
<thead>
<tr>
<th>Model</th>
<th>Classifier</th>
<th>Sensitivity</th>
<th>Specificity</th>
<th>Accuracy</th>
<th>AUC</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>SVM</td>
<td>0.96 ± 0.04</td>
<td>0.90 ± 0.02</td>
<td>0.86 ± 0.06</td>
<td>0.99 ± 0.02</td>
<td>0.01</td>
<td></td>
</tr>
<tr>
<td>LR</td>
<td>0.95 ± 0.03</td>
<td>0.93 ± 0.01</td>
<td>0.90 ± 0.04</td>
<td>0.98 ± 0.02</td>
<td>0.01</td>
<td></td>
</tr>
<tr>
<td>KNN</td>
<td>0.94 ± 0.05</td>
<td>0.92 ± 0.03</td>
<td>0.89 ± 0.05</td>
<td>0.97 ± 0.03</td>
<td>0.07</td>
<td></td>
</tr>
<tr>
<td>DT</td>
<td>0.93 ± 0.06</td>
<td>0.91 ± 0.04</td>
<td>0.88 ± 0.06</td>
<td>0.96 ± 0.03</td>
<td>0.08</td>
<td></td>
</tr>
<tr>
<td>RF</td>
<td>0.92 ± 0.07</td>
<td>0.90 ± 0.05</td>
<td>0.87 ± 0.07</td>
<td>0.95 ± 0.04</td>
<td>0.08</td>
<td></td>
</tr>
<tr>
<td>CNN</td>
<td>0.94 ± 0.06</td>
<td>0.91 ± 0.04</td>
<td>0.90 ± 0.06</td>
<td>0.97 ± 0.04</td>
<td>0.07</td>
<td></td>
</tr>
<tr>
<td>mCOM</td>
<td>0.96 ± 0.04</td>
<td>0.90 ± 0.02</td>
<td>0.86 ± 0.06</td>
<td>0.99 ± 0.02</td>
<td>0.01</td>
<td></td>
</tr>
</tbody>
</table>

Performance of models built with multiple classifiers using features from different modalities

<table>
<thead>
<tr>
<th>Modality</th>
<th>Classifier</th>
<th>Sensitivity</th>
<th>Specificity</th>
<th>Accuracy</th>
<th>AUC</th>
<th>PRD</th>
</tr>
</thead>
<tbody>
<tr>
<td>CT</td>
<td>MC</td>
<td>0.65 ± 0.10</td>
<td>0.64 ± 0.06</td>
<td>0.64 ± 0.05</td>
<td>0.69</td>
<td>0.67</td>
</tr>
<tr>
<td>PET</td>
<td>MC</td>
<td>0.64 ± 0.06</td>
<td>0.63 ± 0.05</td>
<td>0.63 ± 0.04</td>
<td>0.68</td>
<td>0.69</td>
</tr>
<tr>
<td>CT + PET</td>
<td>MC</td>
<td>0.62 ± 0.05</td>
<td>0.61 ± 0.04</td>
<td>0.61 ± 0.03</td>
<td>0.67</td>
<td>0.69</td>
</tr>
<tr>
<td>CT + PET + Chem</td>
<td>MC</td>
<td>0.60 ± 0.04</td>
<td>0.59 ± 0.03</td>
<td>0.59 ± 0.02</td>
<td>0.70</td>
<td>0.71</td>
</tr>
</tbody>
</table>


During-treatment Imaging

- 75 HNSCC treated by primary RT (+ chemotherapy) with curative intent and received FDG PET–CT before (prePET) and during third week of RT (iPET).
- A reduction of more than 50% in the node total lesion glycolysis (TLG) was the best biomarker for locoregional and regional failure-free survival (FFS), disease-free survival (DFS) and overall survival (OS)

Radiation Oncology

During-treatment CBCT

- Daily/weekly CBCT is routinely used for patient setup or adaptive therapy
- Change of CBCT-based radiomics (delta-CBCT-radiomics) could reflect the therapy included response.
- Adding delta-CBCT-radiomic may improve the performance of models based on baseline imaging/clinical characteristics.
- Cohort: 1:2 case-control cohort of patients with HNSCC treated at UTSW with definitive radiotherapy +/- chemotherapy. 90 patients (30 cases) were included with:
  - 89 primary GTVs (23 primaries with LF)
  - 209 nodal GTVs (15 nodes with LF)

Reproducible CBCT features

- Repeated CBCTs with the same fraction

Radiation Oncology

Intra-treatment FMISO PET

- In 19 patients with human papillomavirus-related oropharyngeal cancers, pre- and intra-treatment dynamic fluoro-18-labeled fluoromisonidazole positron emission tomography (PET) was used to assess tumor hypoxia.
- Patients without hypoxia at baseline or intra-treatment received 30 Gy; patients with persistent hypoxia received 70 Gy.
- Fifteen of 19 patients were deescalated to 30 Gy. Of these 15 patients, 11 had a pathologic complete response.
- Two-year locoregional control and overall survival were 94.4% (95% confidence interval = 84.4% to 100%) and 94.7% (95% confidence interval = 85.2% to 100%), respectively. No acute grade 3 radiation-related toxicities were observed.

Intra-treatment CBCT

- Daily/weekly CBCT is routinely used for patient setup or adaptive therapy.
- Change of CBCT-based radiomics (delta-CBCT-radiomics) could reflect the therapy included response.
- Adding delta-CBCT-radiomic may improve the performance of models based on baseline imaging/clinical characteristics.
- Cohort: 1:2 case-control cohort of patients with HNSCC treated at UTSW with definitive radiotherapy +/- chemotherapy. 90 patients (30 cases) were included with:
  - 89 primary GTVs (23 primaries with LF)
  - 209 nodal GTVs (15 nodes with LF)
Local failure prediction for primary structures

<table>
<thead>
<tr>
<th>Model</th>
<th>AUC</th>
<th>95% CI (lower, upper)</th>
<th>$P$ value (vs random)</th>
<th>$P$ value (vs fused ensemble)</th>
<th>Max TLD (rad)</th>
<th>Predicted score threshold at 90% of DPC (%)</th>
<th>Sensitivity (%)</th>
<th>Specificity (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fixed ensemble</td>
<td>0.671</td>
<td>0.674 (0.667, 0.676)</td>
<td>0.000</td>
<td>0.000</td>
<td>58.3</td>
<td>10.3</td>
<td>78.5</td>
<td>30.2</td>
</tr>
<tr>
<td>Combined feature ensemble</td>
<td>0.663</td>
<td>0.668 (0.664, 0.672)</td>
<td>0.000</td>
<td>0.000</td>
<td>58.3</td>
<td>10.3</td>
<td>78.5</td>
<td>30.2</td>
</tr>
<tr>
<td>Clinical Only ensemble</td>
<td>0.709</td>
<td>0.710 (0.708, 0.710)</td>
<td>0.000</td>
<td>0.000</td>
<td>58.3</td>
<td>10.3</td>
<td>78.5</td>
<td>30.2</td>
</tr>
<tr>
<td>Radiomic ensemble</td>
<td>0.715</td>
<td>0.716 (0.715, 0.716)</td>
<td>0.000</td>
<td>0.000</td>
<td>58.3</td>
<td>10.3</td>
<td>78.5</td>
<td>30.2</td>
</tr>
<tr>
<td>CT1 only ensemble</td>
<td>0.687</td>
<td>0.690 (0.683, 0.694)</td>
<td>0.000</td>
<td>0.000</td>
<td>58.3</td>
<td>10.3</td>
<td>78.5</td>
<td>30.2</td>
</tr>
<tr>
<td>Delta only ensemble</td>
<td>0.686</td>
<td>0.687 (0.682, 0.691)</td>
<td>0.000</td>
<td>0.000</td>
<td>58.3</td>
<td>10.3</td>
<td>78.5</td>
<td>30.2</td>
</tr>
</tbody>
</table>

H Morgan et al., Quantitative Imaging in Medicine and Surgery, 2021

Local failure prediction for nodal structures

<table>
<thead>
<tr>
<th>Model</th>
<th>AUC</th>
<th>95% CI (lower, upper)</th>
<th>$P$ value (vs random)</th>
<th>$P$ value (vs fused ensemble)</th>
<th>Max TLD (rad)</th>
<th>Predicted score threshold at 90% of DPC (%)</th>
<th>Sensitivity (%)</th>
<th>Specificity (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fixed ensemble</td>
<td>0.913</td>
<td>0.913 (0.910, 0.913)</td>
<td>0.000</td>
<td>0.000</td>
<td>58.3</td>
<td>10.3</td>
<td>78.5</td>
<td>30.2</td>
</tr>
<tr>
<td>Combined feature ensemble</td>
<td>0.893</td>
<td>0.893 (0.890, 0.893)</td>
<td>0.000</td>
<td>0.000</td>
<td>58.3</td>
<td>10.3</td>
<td>78.5</td>
<td>30.2</td>
</tr>
<tr>
<td>Clinical Only ensemble</td>
<td>0.868</td>
<td>0.869 (0.865, 0.872)</td>
<td>0.000</td>
<td>0.000</td>
<td>58.3</td>
<td>10.3</td>
<td>78.5</td>
<td>30.2</td>
</tr>
<tr>
<td>Radiomic ensemble</td>
<td>0.863</td>
<td>0.863 (0.860, 0.865)</td>
<td>0.000</td>
<td>0.000</td>
<td>58.3</td>
<td>10.3</td>
<td>78.5</td>
<td>30.2</td>
</tr>
<tr>
<td>CT1 only ensemble</td>
<td>0.864</td>
<td>0.865 (0.860, 0.866)</td>
<td>0.000</td>
<td>0.000</td>
<td>58.3</td>
<td>10.3</td>
<td>78.5</td>
<td>30.2</td>
</tr>
<tr>
<td>Delta only ensemble</td>
<td>0.857</td>
<td>0.857 (0.852, 0.863)</td>
<td>0.000</td>
<td>0.000</td>
<td>58.3</td>
<td>10.3</td>
<td>78.5</td>
<td>30.2</td>
</tr>
</tbody>
</table>

H Morgan et al., Quantitative Imaging in Medicine and Surgery, 2021

Prediction uncertainty?

- Patients are often limited
  - Model may not provide reliable predictions to all the testing samples, especially for those whose characteristics vary significantly from the training dataset distribution
  - Epistemic uncertainty
    - Can be estimated by anomaly scores
  - Inherent noise of input data
    - Aleatoric Uncertainty
    - Can be estimated by using test-time augmentation (TTA)

Locoregional recurrence prediction in HNC by learning with rejection option

- Kai Wang et al., under revision, Medical Physics

Personalized treatment target identification

- Involved nodal radiation therapy for head and neck cancer (HNC) patients
  - Majority of disease sites treated with RT no longer receive elective/prophylactic radiotherapy to clinically-negative areas
  - Despite our ability to tailor the radiotherapy volume and dose to specific areas, IMRT still targets the same lymph node regions as conventional 2D radiotherapy in HNC
  - The toxicity of associated with RT is very high, especially for patients receiving chemoradiation therapy, where acute and late toxicity rates of grade 3 or higher are 80% and 25%-60%, respectively

Cervical Lymph Node Malignancy Prediction

- There is often uncertainty about the malignant potential of small and less FDG avid lymph nodes (LN’s) in head and neck cancer.
- Malignant LN identification strongly depends on the physicians’ experience.
- AI-based clinical decision support tool for physicians to identify malignant LNs more consistently.

Normal
Suspicious
Involved
Combination of MO-Radiomics and CNN

Model deployment for a phase II trial

- INRT-AIR: A Prospective Phase II Study of Involved Nodal Radiation Therapy Using Artificial Intelligence-Based Radiomics for Head and Neck Squamous Cell Carcinoma (PI: David Sher).
- https://clinicaltrials.gov/ct2/show/NCT03953976
- Eliminating elective neck irradiation and strictly treating involved and suspicious lymph nodes

Normal tissue dose sparing with INRT-AIR
Comparison of Dose bath of 30 Gy and 10 Gy

Preliminary results of INRT-AIR trial

- With a median follow-up of surviving patients of 19.6 months, there were no solitary regional recurrences.
- The mean composite MDADI scores at 6 and 12 months were 90.7 and 89.8, respectively and 94.9 and 94.6 at 6 and 12 months with a baseline MDADI score > 75.
- These outcomes are much higher than a cohort of patients treated with standard IMRT with elective neck irradiation from a prospective cohort at Royal Marsden, where mean MDADI composite score 12 months after treatment completion was 72.

Currently employed in another prospective phase II trial (PI: David Sher)
- A Prospective Study of Daily Adaptive Radiotherapy to Better Organ-at-Risk Doses in Head and Neck Cancer (DARTBOARD)
https://clinicaltrials.gov/ct2/show/NCT04883281

Uncertainty quantification

- Model performance measured on the test data stratified by the median aleatoric uncertainty obtained from the incorrect predictions within the validation cohort.

Uncertainty quantification

- Model performance measured on the test data stratified by the median epistemic uncertainty obtained from the incorrect predictions within the validation cohort.

Anatomical change prediction

- Identify which patients can potentially benefit from adaptive RT
- Facilitate clinical workflow management

GTvP regression. AUC=0.75
GTvN regression. AUC=0.73
### Predicting Radiotherapy Induced Anatomic Change for Head and Neck Cancer Patients using Vision Transformer

Kai Wang et al., PO-GfPV-M-301

![Vision Transformer Diagram](image)

\[
\text{Loss} = L_{\text{identity}} + \|\text{Deform}(\phi)\|_{1} + L_{\text{MSE}}(\phi) + \mu_{\text{diff}}(\phi)
\]

<table>
<thead>
<tr>
<th>Volumetric Difference (cm³)</th>
<th>ASD (mm)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Planning CT 0.0106 (0.0107, 0.0133)</td>
<td>3.04 (2.94, 3.76)</td>
</tr>
<tr>
<td>Week 1 CBCT 0.0079 (0.0079, 0.010)</td>
<td>1.40 (1.40, 1.66)</td>
</tr>
<tr>
<td>Predicted week-3 CBCT 0.0068 (0.0067, 0.0088)</td>
<td>1.28 (1.26, 1.58)</td>
</tr>
</tbody>
</table>

**Summary**

- Intra-treatment imaging may capture therapeutic induced change
  - Response adaptive therapy
- AI-based imaging analysis could aid in adaptive radiation treatment strategy
  - Risk adaptive treatment management
  - Personalized treatment target

**Acknowledgements**

AIRT Lab (PI: Jing Wang/You Zhang)

- Kai Wang M.S.
- Mike Dohopolski M.D.
- Liyuan Chen Ph.D.
- Zhiguo Zhou Ph.D.
- Howard Morgan M.D.

NIH: R01 EB027898, R01EB020366, R01CA251792

CPRIT: RP130562, RP130109, RP160661

ACS: RSG-13-326-01-CCE

[Visit MAIA Laboratory](https://www.utsouthwestern.edu/labs/maia)