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@utsw.edu

UTSouthwestern 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

Economopoulou, Panagiota, et al. "Diagnostic tumor markers in head and neck squamous cell carcinoma (HNSCC) in the clinical setting." *Frontiers in oncology* 9 (2019): 827.

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
  - Carbonara et al. Investigation of Radiation-Induced Toxicity in Head and Neck Cancer Patients through Radiomics and Machine Learning: A Systematic Review, J Oncol. 2021
- Treatment failure: distant metastasis, local regional failure
  - Vallieres et al. Radiomics strategies for risk assessment of tumour failure in head-and-neck cancer. Sci Rep. 2017
- Survival
  - Starke, Sebastian, et al. "A hybrid radiomics approach to modeling progression-free survival in head and neck cancers." 3D Head and Neck Tumor Segmentation in PET/CT Challenge. Springer, Cham, 2021.

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 set 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.**

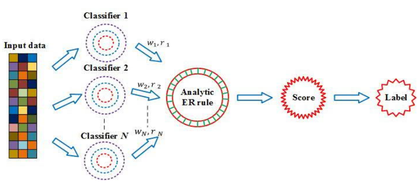
R. Wang, …, and **J. Wang**, Locoregional Recurrence Prediction in Head & Neck Cancer Based on Multi-modality and Multi-view Feature Expansion, *PMB*, 2022

Z. Zhou,…, **J. Wang**, Multifaceted radiomics for distant metastasis prediction in head & neck cancer, *PMB*, 2020

L. Chen, …, **J. Wang**, Combining Many-objective Radiomics and 3-dimensional Convolutional Neural Network through Evidential Reasoning to Predict Lymph Node Metastasis in Head and Neck Cancer, *PMB*, 2019 (Used in two prospective Phase II clinical trials)

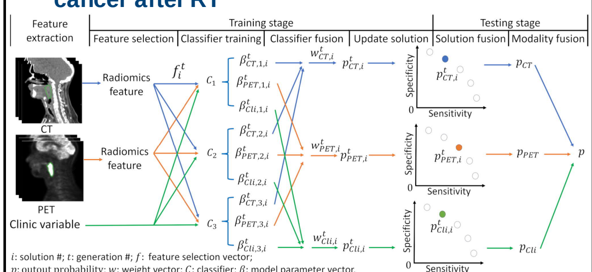
K. Wang, …, **J. Wang,** A multi-objective radiomics model for the prediction of locoregional recurrence in head and neck squamous cell cancers, *Medical Physics*, 2020

Reliable fusion



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

mCOM (multi-classifier, objective, modality) for locoregional failure prediction in H&N cancer after RT



(K. Wang, …, J. Wang, A multi-objective radiomics model for the prediction of locoregional recurrence in head and neck squamous cell cancers, Medical Physics, 2020)

- 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.

Vallieres M, et al. Radiomics strategies for risk assessment of tumour failure in head-and-neck cancer. Sci Rep, 2017.

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

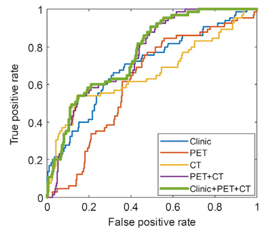
Modality	Classifier	Sensitivity	Specificity	Accuracy	AUC	*P*-value
Clinic	SVM	0.60 ± 0.20	0.67 ± 0.22	0.66 ± 0.16	0.66 ± 0.06	0.51
	LR	0.65 ± 0.07	0.63 ± 0.06	0.63 ± 0.04	0.68 ± 0.02	0.98
	DA	0.62 ± 0.08	0.63 ± 0.05	0.63 ± 0.04	0.67 ± 0.01	0.52
	MC	0.63 ± 0.06	0.65 ± 0.06	0.64 ± 0.06	0.68 ± 0.02	–
CT	SVM	0.52 ± 0.06	0.82 ± 0.08	0.78 ± 0.06	0.66 ± 0.01	0.89
	LR	0.52 ± 0.03	0.84 ± 0.01	0.79 ± 0.01	0.67 ± 0.02	0.59
	DA	0.54 ± 0.00	0.81 ± 0.02	0.77 ± 0.02	0.67 ± 0.01	0.87
	MC	0.54 ± 0.00	0.84 ± 0.02	0.80 ± 0.01	0.69 ± 0.02	–
PET	SVM	0.62 ± 0.09	0.49 ± 0.04	0.51 ± 0.02	0.52 ± 0.01	<0.01
	LR	0.59 ± 0.09	0.50 ± 0.04	0.51 ± 0.03	0.53 ± 0.01	0.02
	DA	0.58 ± 0.14	0.61 ± 0.03	0.61 ± 0.02	0.59 ± 0.03	0.26
	MC	0.62 ± 0.12	0.61 ± 0.03	0.61 ± 0.01	0.62 ± 0.03	–

DA, discriminant analysis, LR, logistic regression; MC, multi-classifier model; SVM, The classifiers comprise support vector machine.

(K. Wang, …, J. Wang, A multi-objective radiomics model for the prediction of locoregional recurrence in head and neck squamous cell cancers, *Med. Phys.* 2020)

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

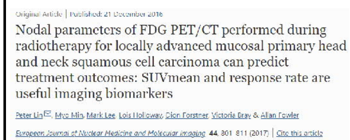
Modality	Classifier	Sensitivity	Specificity	Accuracy	AUC	P-value
Clinic	MC	0.63 ± 0.06	0.65 ± 0.06	0.64 ± 0.06	0.68 ± 0.02	0.02
CT	MC	0.54 ± 0.00	0.84 ± 0.02	0.80 ± 0.01	0.69 ± 0.02	<0.01
PET	MC	0.62 ± 0.12	0.61 ± 0.03	0.61 ± 0.01	0.62 ± 0.03	<0.01
CT + PET	MC	0.54 ± 0.00	0.85 ± 0.01	0.80 ± 0.01	0.76 ± 0.00	0.02
CT + PET+Clinic	MC	0.54 ± 0.00	0.84 ± 0.02	0.80 ± 0.02	0.77 ± 0.00	–



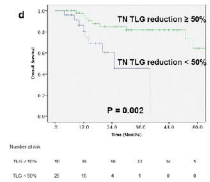
AUC=0.69 in Vallieres M, et al. Radiomics strategies for risk assessment of tumour failure in head-and-neck cancer. Sci Rep, 2017.

UTSouthwestern Medical Center

During-treatment Imaging

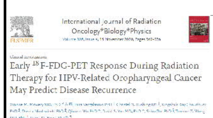

Nodal parameters of FDG PET/CT performed during radiotherapy for locally advanced mucosal primary head and neck squamous cell carcinoma can predict treatment outcomes: SUVmean and response rate are useful imaging biomarkers

European Journal of Nuclear Medicine and Molecular Imaging 44, 801–811 (2017)



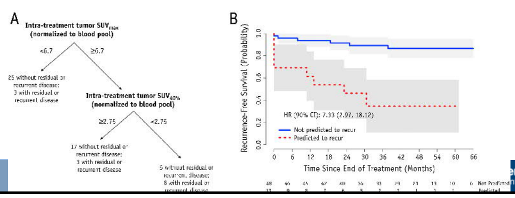
- 75 HNSCC treated by primary RT (± chemotherapy) with curative intent and received FDG PET–CT before (prePET) and during third week of RT (iPET).
- A reduction of more than 50 % in the node total lesion glycolysis (TLG) was the best biomarker for locoregional and regional failure-free survival (FFS),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

Radiation Oncology | AIRT and MAIA Labs | UTSouthwestern Medical Center


International Journal of Radiation Oncology*Biology*Physics

Early [18]F-FDG-PET Response During Radiation Therapy for HPV-Related Oropharyngeal Cancer May Predict Disease Recurrence

- 62 patients undergoing definitive (chemo)radiation for HPV-related/p16-positive OPC, prospectively enrolled. 18F-fluorodeoxyglucose PET/computed tomography scans performed at simulation and after 2 weeks at a dose of ∼20 Gy.
- Patients with intratreatment SUVmax (normalized to blood pool SUVmean) <6.7 or SUVmax (normalized to blood pool SUVmean) ≥6.7 with intratreatment SUV40% ≥2.75 as less likely to recur.



Intra-treatment FMISO PET

- In 19 patients with human papillomavirus–related oropharyngeal cancers, pre- and intratreatment dynamic fluorine-18-labeled fluoromisonidazole positron emission tomography (PET) was used to assess tumor hypoxia.
- Patients without hypoxia at baseline or intratreatment 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.

Nadeem Riaz, Eric Sherman, Xin Pei,..., Nancy Lee, Precision Radiotherapy: Reduction in Radiation for Oropharyngeal Cancer in the 30 ROC Trial, JNCI: Journal of the National Cancer Institute, Volume 113, Issue 6, June 2021, Pages 742–751

Radiation Oncology | AIRT and MAIA Labs | UTSouthwestern Medical Center

Intra-treatment CBCT

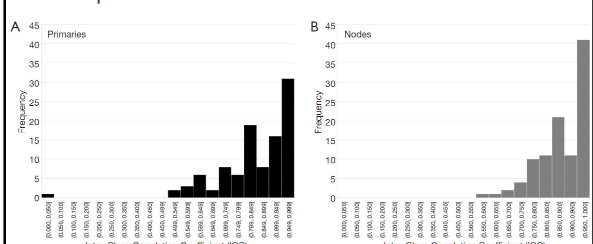
- Daily/weekly CBCT is routinely used for patient setup or adaptive therapy
- Change of CBCT-based radiomics (delta-CBCT-radiomics) could reflect the therapy included response
- Adding delta-CBCT-radiomic may improve the performance of models based on baseline imaging/clinical characteristics

- 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)

Radiation Oncology | AIRT and MAIA Labs | UTSouthwestern Medical Center

Reproducible CBCT features

- Repeated CBCTs with the same fraction



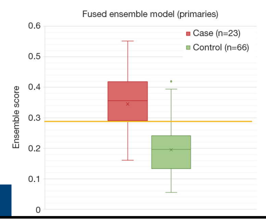
H Morgan, ..., J. Wang, “Exploratory ensemble interpretable model for predicting local failure in head and neck cancer: the additive benefit of CT and intra-treatment cone-beam computed tomography features”, Quantitative Imaging in Medicine and Surgery, 2021

Radiation Oncology | AIRT and MAIA Labs | UTSouthwestern Medical Center

Local failure prediction for primary structures

Model	AUC	95% CI (lower)	95% CI (upper)	P value (*vs.* random chance)†	P value (*vs.* fused ensemble)‡	Max Youden J statistic	Predicted score threshold at max J statistic	Sensitivity (%)	Specificity (%)
Fused ensemble	0.871	0.788	0.954	*0.000*	NA	0.692	0.290	78.3	90.9
Combined feature ensemble	0.853	0.771	0.935	*0.000*	0.494	0.565	0.161	91.3	56.5
Clinical Only ensemble	0.788	0.680	0.895	*0.000*	0.134	0.469	0.252	69.6	77.3
Radiomic ensemble (CT1 + Delta)	0.770	0.655	0.885	*0.000*	*0.017*	0.491	0.237	87.0	49.1
CT1 only ensemble	0.687	0.561	0.813	*0.004*	*0.004*	0.340	0.291	52.2	81.8
Delta only ensemble	0.696	0.571	0.822	*0.002*	*0.013*	0.345	0.208	73.9	60.6

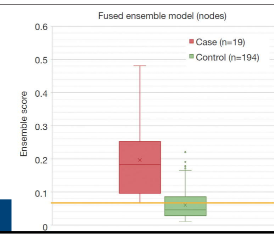
H Morgan et al, Quantitative Imaging in Medicine and Surgery, 2021



Local failure prediction for nodal structures

Model	AUC	95% CI (lower)	95% CI (upper)	P value (*vs.* random chance)†	P value (*vs.* fused ensemble)‡	Max Youden J statistic	Predicted score threshold at max J statistic	Sensitivity (%)	Specificity (%)
Fused ensemble	0.910	0.853	0.967	*0.000*	NA	0.680	0.066	100.0	68.0
Combined feature ensemble	0.893	0.819	0.941	*0.000*	0.212	0.686	0.046	100.0	68.6
Clinical only ensemble	0.865	0.802	0.929	*0.000*	0.080	0.648	0.061	94.7	70.1
Radiomic ensemble (CT1 + Delta)	0.880	0.819	0.941	*0.000*	0.268	0.643	0.060	94.7	69.6
CT1 only ensemble	0.854	0.784	0.924	*0.000*	*0.026*	0.613	0.056	100.0	61.3
Delta only ensemble	0.867	0.803	0.930	*0.000*	0.150	0.613	0.054	100.0	61.3

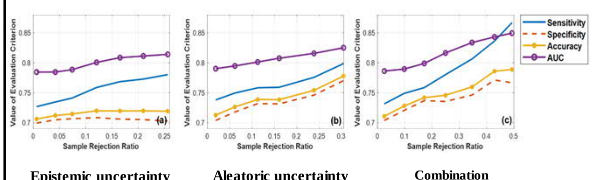
H Morgan et al, Quantitative Imaging in Medicine and Surgery, 2021



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



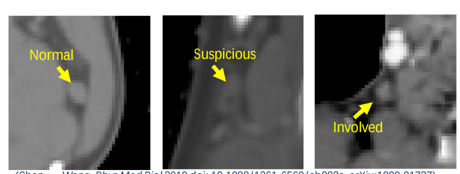
Kai Wang et al, under revision, Medical Physics

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

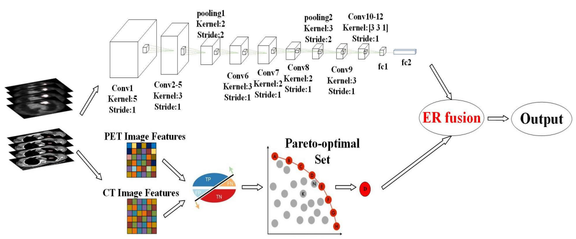
Cervical Lymph Node Malignancy Prediction

- There is often uncertainty about the malignant potential of small and less FDG avid lymph nodes (LNs) 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.



(Chen, ..., Wang, *Phys Med Biol* 2019 doi: 10.1088/1361-6560/ab083a, arXiv:1809.01737)

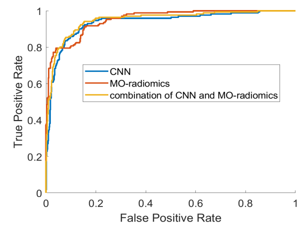
Combination of MO-Radiomics and CNN



(Chen, ..., Wang, *Phys Med Biol* 2019 doi: 10.1088/1361-6560/ab083a, arXiv:1809.01737)

LN malignancy prediction on surgical patients with pathological ground truth

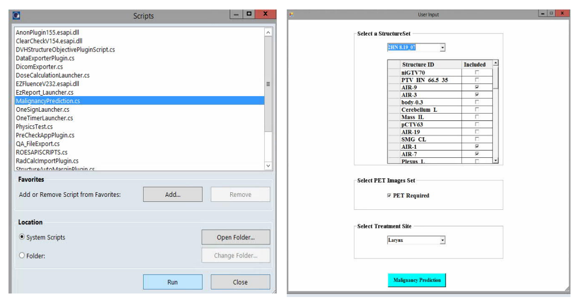
- Training nodes: 116 positive and 438 negative
- Testing nodes: 50 positive and 187 benign



AUC: 0.95 combined model

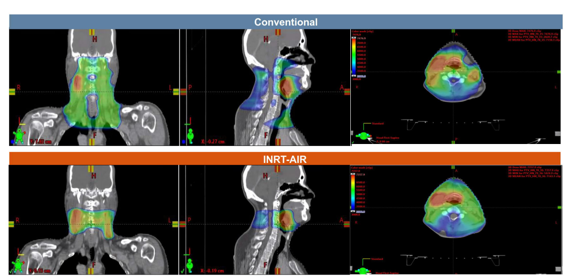
Model deployment for a phase II trial

- 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

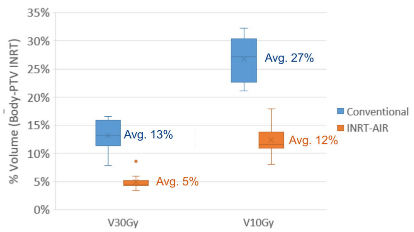


Row	CT Prediction Probability	PET CT Prediction Probability
AIR01	0	0.08
AIR02	0	0.05
AIR03	0.09	0.1
AIR04	0.23	0.51
AIR05	0.27	0.16
AIR06	0	0.01
AIR07	0	0.01
AIR08	0	0.06
AIR09	0.01	0.44
AIR10	0	0.02
AIR11	0	0.01
AIR12	0.06	0.21
AIR13	0.13	0.39
AIR14	0.32	0.82
AIR15	0	0.02
AIR16	0	0.01

Normal tissue dose sparing with INRT-AIR



Comparison of Dose bath of 30 Gy and 10 Gy



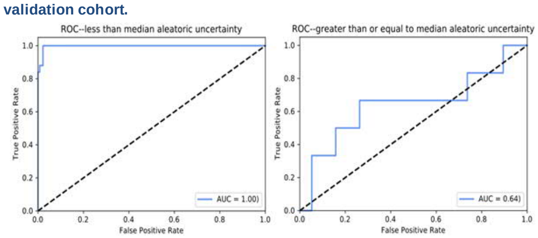
Preliminary results of INRT-AIR trial

- 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

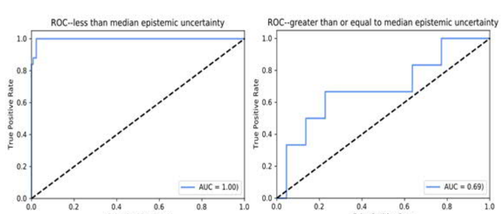
- **Model performance measured on the test data stratified by the median aleatoric uncertainty obtained from the incorrect predictions within the validation cohort.**



(M. Dohopolski, ..., J. Wang, Predicting Lymph Node Metastasis in Patients with Oropharyngeal Cancer by using Convolutional Neural Networks with associated Epistemic and Aleatoric Uncertainty, *PMB, vol. 65, 225002,* 2020)

Uncertainty quantification

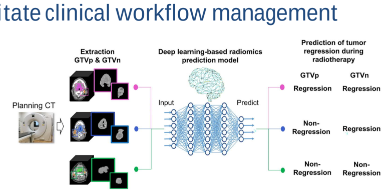
- **Model performance measured on the test data stratified by the median epistemic uncertainty obtained from the incorrect predictions within the validation cohort.**



(M. Dohopolski, ..., J. Wang, Predicting Lymph Node Metastasis in Patients with Oropharyngeal Cancer by using Convolutional Neural Networks with associated Epistemic and Aleatoric Uncertainty, *PMB, vol. 65, 225002,* 2020)

Anatomical change prediction

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



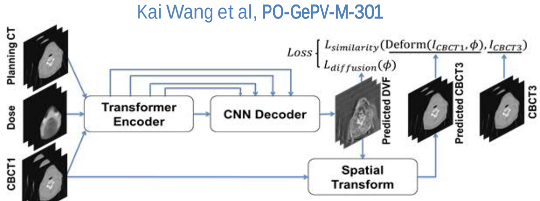
$$\Delta GTVp/treatment\ day = \frac{\text{Initial GTVp volume} - \text{Boost GTVp volume}}{\text{Initial GTVp volume}} / treatment\ day$$

GTVp regression, AUC=0.75
GTVn regression, AUC=0.73

Tanaka, Shohei, et al. "A deep learning-based radiomics approach to predict head and neck tumor regression for adaptive radiotherapy." *Scientific Reports* 12.1 (2022): 1-13.

Predicting Radiotherapy Induced Anatomic Change for Head and Neck Cancer Patients using Vision Transformer

Kai Wang et al, PO-GePV-M-301



$$Loss \begin{cases} L_{similarity}(\mathrm{Deform}(I_{CBCT1}, \phi), I_{CBCT3}) \\ L_{diffusion}(\phi) \end{cases}$$

$$Loss = L_{similarity}\left(\mathrm{Deform}(I_{CBCT1}, \phi), I_{CBCT3}\right) + \lambda L_{diffusion}(\phi)$$

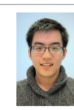
Median (95% confidence)	MSE	SSIM	DC	Volume Difference (cm^3)	ASD (mm)
Planning CT	0.0106 (0.0107, 0.0133)	0.8382 (0.8197, 0.8472)	0.9547 (0.9487, 0.9580)	103.92 (86.32, 159.70)	3.04 (2.94, 3.76)
Week-1 CBCT	0.0079 (0.0079, 0.010)	0.8690 (0.8690, 0.8945)	0.9712 (0.9663, 0.9722)	50.28 (30.24, 74.72)	1.40 (1.40, 1.66)
Predicted week-3 CBCT	0.0068 (0.0067, 0.0088)	0.8976 (0.8748, 0.8998)	0.9723 (0.9665, 0.9717)	0.14 (-14.76, 26.72)	1.28 (1.26, 1.58)

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

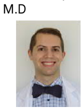
AIRT Lab (PI: Jing Wang/You Zhang)


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: RP110562, RP130109, RP160661
ACS: RSG-13-326-01-CCE