Machine Learning Approaches for Adaptive Radiotherapy in Head and Neck Squamous Cell Carcinoma (HNSCC)

Somiya Rauf, Abdul-Malik, Joshua Piña

July 14, 2025

Our Innovation

- Develop an ML framework to personalize/adapt RT dose fractionation.
- Input: Patient-specific clinical and imaging data.
- Goal: Predict optimal adaptive dosing schedule and dose.
- Outcome:
    - Maximize tumor control.
    - Minimize normal tissue toxicity.
- KBR-ART Incorporates:
    - Pre-treatment planning
    - Real-time model updates
    - Ongoing adaptation per response

Mission Statement

Tumor Control Probability (TCP):

- d: dose, n: fractions of doses, T_{pot}: the potential doubling time, N: initial population.

$$\text{TCP} = e^{-N \cdot e^{-n(\alpha d + \beta d^2) - t \ln 2 / T_{\text{pot}}}}$$

Normal Tissue Complication Probability (NTCP):

- Predicts risk of radiation-induced complications.

$$NTCP = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-u^2/2}\, du$$

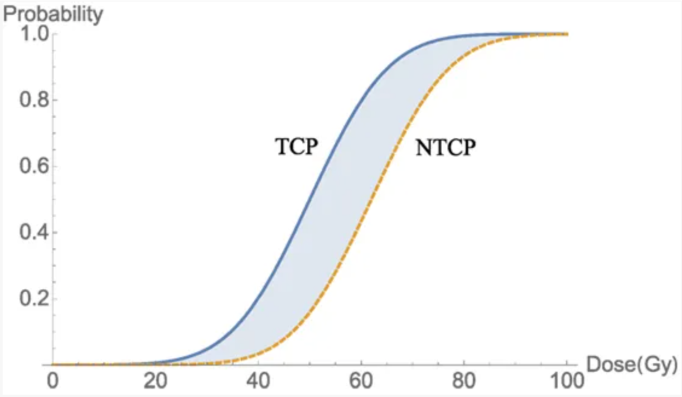
$$x = \frac{D - D_{50}}{m\, D_{50}}$$

D_{50} as the dose yielding 50% complication probability and m shaping the slope of the dose-response curve

Knowledge-Based Response-Adapted Radiotherapy (KBR-ART)

maximizing:

$$P = TCP \times (1 - NTCP)$$

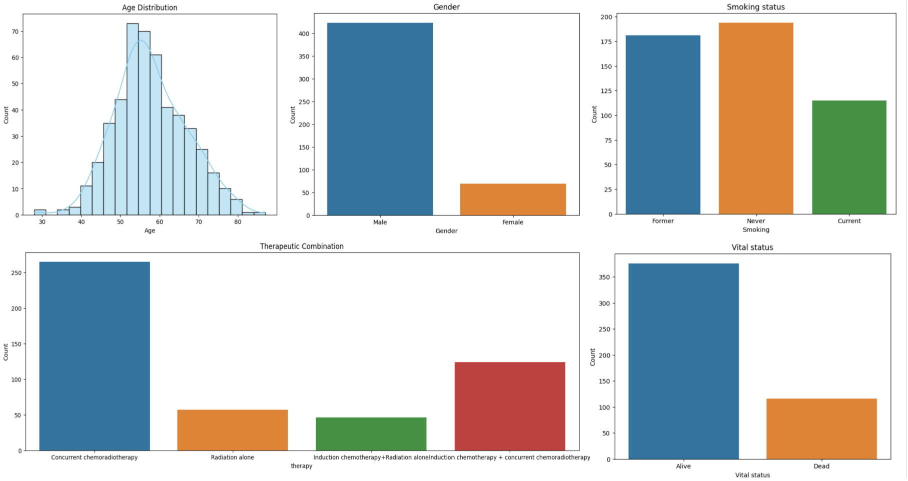


Datasets

- 493 HNSCC patients (TCIA-HNSCC, HN1 repositories). (492, 30)
- De-identified clinical data, CT imaging, RT plans.
- CT scans available for 336 patients.
- Data standardized by Zakari Aliti and Hamza Idmoudi.
- Publicly available on Kaggle:
  - `https://www.kaggle.com/datasets/hamzaidmoudi/hnscc-zip`

	count	mean	std	min	25%	50%	75%	max
Age at Diag	492.0	57.817073	9.205234	28.00	52.00	57.00	64.00	87.00
Smoking status (Packs-Years)	492.0	16.133130	23.414577	0.00	0.00	1.00	30.00	120.00
T-category	492.0	2.376016	0.967325	1.00	2.00	2.00	3.00	4.00
N-category	492.0	1.741870	0.663597	0.00	2.00	2.00	2.00	3.00
Radiation Treatment_duration	492.0	42.020325	4.639729	0.00	39.00	43.00	44.00	56.00
Total prescribed Radiation treatment dose	492.0	68.760000	2.329913	57.00	66.00	70.00	70.00	75.00
Radiation treatment_number of fractions	492.0	32.626016	3.028442	0.00	30.00	33.00	33.00	42.00
Radiation treatment_dose per fraction	491.0	2.111752	0.126464	1.71	2.12	2.12	2.20	3.27
Overall survival_duration of Merged updated ASRM V2	492.0	2370.018293	1050.606547	33.00	1765.00	2452.50	3149.25	4578.00
Local control_duration of Merged updated ASRM V2	491.0	2278.674134	1101.334133	72.00	1591.50	2384.00	3073.00	4578.00
Regional control_duration of Merged updated ASRM V2	491.0	2315.362525	1094.436384	72.00	1682.00	2417.00	3141.50	4578.00
Locoregional control_duration of Merged updated ASRM V2	491.0	2234.511202	1126.089231	72.00	1439.50	2345.00	3060.50	4578.00
Freedom from distant metastasis_duration of Merged updated ASRM V2	490.0	2332.069388	1080.829780	36.00	1738.50	2437.00	3141.75	4578.00
Days to last FU	492.0	2370.018293	1050.606547	33.00	1765.00	2452.50	3149.25	4578.00

Data Statistics



Features Overview

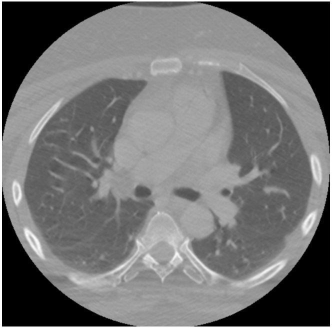


Figure: CT images of HNSCC (Base of tongue, Tonsil, NOS, and Soft palate)

- **Demographics:** Gender, Age at Diagnosis, Smoking status (+ packs-years), HPV Status
- **Tumor description:** Tumor laterality, Cancer subsite of origin, T-category, N-category, AJCC Stage
- **Treatment info:** Therapeutic Combination, Radiation Treatment duration, Total dose, Fractions, Dose per fraction
- **Outcomes:** Vital status, Overall survival, Local/Regional/Locoregional control + durations, Freedom from distant metastasis + duration, Relapse-free survival

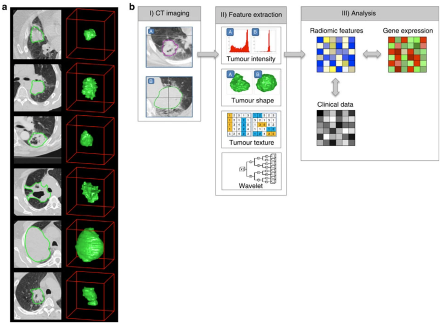
Model Steps

1. **Data Types for KBR-ART:**
   - Clinical data
   - Imaging data: CT scans. (CT images of tumours exhibit strong contrast reflecting differences in the intensity of a tumour on the image)
2. **Tumor Segmentation:**
   - Image processing + U-Net CNNs to auto-segment tumor vs normal tissue.
3. **Dose Prediction Model:**
   - Regression models using radiomic features + clinical variables.
   - Objective: Maximize $P = TCP \times (1 - NTCP)$.
4. **Adaptive Dose Scheduling:**
   - Rule-based triggers + linear feedback control.
   - Adjust doses dynamically after each fraction.

Progress So Far

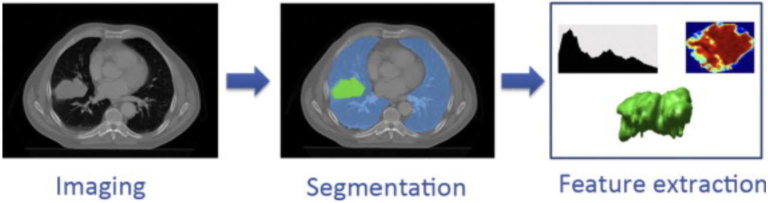
- Chose to use average of all CT image slices per patient for input (most representative).
  - Logs number of slices used for each average.
- Linked patients to their corresponding average CT images.
- Defined input and target variables:
  - Inputs: Demographic data (age, smoking status, etc.) + CT image averages.
  - Targets: Plan information (treatment dose, number of fractions, dose per fraction).
- Excluded patients without CT images from training data.
  - Consider using them as a potential test set.
- Encoded categorical features (e.g., smoking status) as binary.
- Discussed feature and target selection strategy.
- Removed 1 Patient due to missing values to improve dataset reliability.

Extracting radiomics data from images



The most widely used imaging modality in oncology is X-ray computed tomography (CT), which assesses tissue density. Indeed, CT images of lung cancer tumours exhibit strong contrast reflecting differences in the intensity of a tumour on the image, intratumour texture and tumour shape (Fig). However, in clinical practice, tumour response to therapy is only measured using one- or two-dimensional descriptors of tumour size (RECIST and WHO, respectively)

The workflow of Radiomics



- Standardized images.
- Define the tumor region in the image
- Extract measurable features from the tumor area (e.g., intensity, texture, shape, relation to nearby tissues).
- Perform feature selection.
- Analyze selected features for their links to treatment outcomes.

Required Steps

$$
\begin{array}{l}
N\ \text{samples} \\
\left\{ \left[\begin{array}{cccc|cccc|cccc}
x_{11}^{(0)} & x_{12}^{(0)} & \dots & x_{1n}^{(0)} & x_{11}^{(1)} & x_{12}^{(1)} & \dots & x_{1n}^{(1)} & x_{11}^{(2)} & x_{12}^{(2)} & \dots & x_{1n}^{(2)} \\
x_{21}^{(0)} & x_{22}^{(0)} & \dots & x_{2n}^{(0)} & x_{21}^{(1)} & x_{22}^{(1)} & \dots & x_{2n}^{(1)} & x_{21}^{(2)} & x_{22}^{(2)} & \dots & x_{2n}^{(2)} \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\
x_{N1}^{(0)} & x_{N2}^{(0)} & \dots & x_{Nn}^{(0)} & x_{N1}^{(1)} & x_{N2}^{(1)} & \dots & x_{Nn}^{(1)} & x_{N1}^{(2)} & x_{N2}^{(2)} & \dots & x_{Nn}^{(2)}
\end{array} \right] \right. ,
\end{array}
$$

Estimating Radiotherapy Outcome

- Analytical Model
- Data driven Models (Regression Model where y^i can be TCP or NTCP)

Adapting Plans for Radiotheraphy

- Before deciding on a treatment strategy, we want to evaluate possible results.
- This is like chess: you simulate future moves to choose the best one.
- In cancer treatment, the goal is to choose the plan that maximizes the patient's benefit.

Linear Feedback Control

$$\dot{\mathbf{x}}^{(t)} = A\mathbf{x}^{(t)} + B\mathbf{u}^{(t)}, \qquad \mathbf{y}^{(t)} = C\mathbf{x}^{(t)} \tag{19}$$

$$\dot{\tilde{\mathbf{x}}}^{(t)} = A\tilde{\mathbf{x}}^{(t)} + B\mathbf{u}^{(t)} + L\delta\mathbf{y}^{(t)}, \qquad \tilde{\mathbf{y}}^{(t)} = C\tilde{\mathbf{x}}^{(t)}, \tag{20}$$

Eq 19 describes how the actual system changes over time under treatment
A: intrinsic tumor dynamics B: how the control (dose) affects the state
C: how state maps to what you can observe (e.g., imaging)
Eq 20 is the estimated state, adjusting based on observed error.
L: a gain matrix that determines how strongly to respond to the difference (feedback).
$\delta y(t) = y(t) - y(t)'$ is the feedback signal (error in observation).
If L=0(no feedback), If L0 (feedback included).