Using Machine Learning for Treatment Plan Optimization in Head and Neck Cancer

By Nadeem Shaheen
Machine learning allows complex problems to be solved by algorithms.
Inverse planning is complex.

<table>
<thead>
<tr>
<th>Normal Structures &amp; Goals:</th>
<th>Max Dose</th>
<th>Mean Dose</th>
<th>Comments:</th>
</tr>
</thead>
<tbody>
<tr>
<td>Brain (Cerebellum)</td>
<td>&lt;54 Gy</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Brainstem</td>
<td>&lt;52 Gy</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Chiasm</td>
<td>&lt;45 Gy</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cochea_L</td>
<td>&lt;40 Gy</td>
<td>&lt;35 Gy</td>
<td></td>
</tr>
<tr>
<td>Cochea_R</td>
<td>&lt;40 Gy</td>
<td>&lt;35 Gy</td>
<td></td>
</tr>
<tr>
<td>OARpharynx</td>
<td></td>
<td>&lt;45 Gy</td>
<td></td>
</tr>
<tr>
<td>Esophagus</td>
<td></td>
<td>&lt;30 Gy</td>
<td></td>
</tr>
<tr>
<td>Globe_L</td>
<td>3 Gy</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Globe_R</td>
<td>3 Gy</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lens_L</td>
<td>2 Gy</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lens_R</td>
<td>2 Gy</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lips</td>
<td></td>
<td>&lt;20 Gy</td>
<td></td>
</tr>
<tr>
<td>Mandible</td>
<td>66 Gy</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nape</td>
<td></td>
<td>&lt;22 Gy</td>
<td></td>
</tr>
<tr>
<td>Optic Nerve_L</td>
<td>5 Gy</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Optic Nerve_R</td>
<td>5 Gy</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Oral Cavity</td>
<td>&lt;60 Gy</td>
<td>&lt;30 Gy</td>
<td></td>
</tr>
<tr>
<td>Parotid_R</td>
<td>&lt;12 Gy</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Spinal Cord</td>
<td>&lt;45 Gy</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Spinal Cord+5mm</td>
<td>&lt;50 Gy</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Submandibular_L</td>
<td>&lt;39 Gy</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Submandibular_R</td>
<td>&lt;20 Gy</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Unspecified Tissue</td>
<td>&lt;74 Gy</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ear</td>
<td>&lt;60</td>
<td>40</td>
<td></td>
</tr>
<tr>
<td>base of tongue</td>
<td></td>
<td>30</td>
<td></td>
</tr>
</tbody>
</table>
Inverse planning is complex.
We wanted our solution to meet a few criteria.

1. Easy to use, understand, and implement
We wanted our solution to meet a few criteria.

1. Easy to use, understand, and implement
2. Operate on easily accessible data that all patients have
We wanted our solution to meet a few criteria.

1. Easy to use, understand, and implement
2. Operate on easily accessible data that all patients have
3. Model an individual clinics data
In order to begin analyzing this problem, we need to categorize the plans.
In order to begin analyzing this problem, we need to categorize the plans.

40 Head and Neck Cancer Patients

DVH’s for OAR
In order to begin analyzing this problem, we need to categorize the plans.

40 Head and Neck Cancer Patients

DVH's for OAR

Visual Categorization
In order to begin analyzing this problem, we need to categorize the plans.

- 40 Head and Neck Cancer Patients
- DVH's for OAR
  - Visual Categorization
  - K Means Clustering
In order to begin analyzing this problem, we need to categorize the plans.

- 40 Head and Neck Cancer Patients
- DVH's for OAR
- K Means Clustering
- Visual Categorization
- Dose Groups
In order to begin analyzing this problem, we need to categorize the plans.

40 Head and Neck Cancer Patients

DVH’s for OAR

Visual Categorization

K Means Clustering

Dose Groups

Unsupervised Labeling
We extracted the DVH’s for Parotids and SMGs.
DVH’s can be clustered visually.
Another way to categorize data is with K-means clustering.
Another way to categorize data is with K-means clustering.

Generate a dataset

\[
X, Y = \text{make_blobs}(n_{\text{samples}}=1000, \text{centers}=3, n_{\text{features}}=10, \text{random_state}=0)
\]
Another way to categorize data is with K-means clustering.

Generate a dataset

Randomly assigned centers

\[ d(p, q) = \sqrt{(q_1 - p_1)^2 + (q_2 - p_2)^2} \]

\[ c_i = \frac{1}{|S_i|} \sum_{x_i \in S_i} x_i \]
Another way to categorize data is with K-means clustering.

Generate a dataset
```python
X, Y = make_blobs(n_samples=1000, centers=3, n_features=10, random_state=0)
```

Randomly assigned centers

\[ d(p, q) = \sqrt{(q_1 - p_1)^2 + (q_2 - p_2)^2} \]

\[ c_i = \frac{1}{|S_i|} \sum_{x \in S_i} x_i \]

Repeat distance calculations
Another way to categorize data is with K-means clustering.

Generate a dataset

Randomly assigned centers

\[ d(p, q) = \sqrt{(q_1 - p_1)^2 + (q_2 - p_2)^2} \]

\[ c_i = \frac{1}{|S_i|} \sum_{x \in S_i} x_i \]

Repeat distance calculations
Converge on a solution
We can apply k-means to our OAR data.
We determined the most representative number of clusters to be 3.

$$Error = \sum_{j=1}^{k} \sum_{i=1}^{n} |X_i^j - C_j|^2$$
We can cluster the parotids using $k = 3$. 
We can cluster the SMGs using $k = 3$. 

![Graph showing SMGs k = 3]
Our average DVH curves are improved with k-means.

Average Parotid DVH - Visual

Average Parotid DVH - k-means
Our average DVH curves are improved with k-means.
Our average DVH curves are improved with k-means.

Average Parotid DVH - k-means

\[-0.00367x^2 + 0.07057x + 0.554\]

Average SMG DVH - k-means

\[0.001914x^2 - 0.7419x + 1.67\]
We can apply these fits to influence the optimizer

Relative Volume
\[= -1.354 \times 10^{-8} \times Dose^5 + 2.572 \times 10^{-6} \times Dose^4 - .0001876 \times Dose^3 + .006547 \times Dose^2 - .1189 \times Dose + 1.301\]
We need to predict the dose groups.

40 Head and Neck Cancer Patients → DVH’s for OAR → Visual Categorization → K Means Clustering → Dose Groups → Unsupervised Labeling
We need to predict the dose groups.

- 40 Head and Neck Cancer Patients
- DVH's for OAR
- Visual Categorization
- K Means Clustering
- Dose Groups
- Unsupervised Labeling
- Pre Plan Data
We need to predict the dose groups.

- **40 Head and Neck Cancer Patients**
- **DVH's for OAR**
- **Visual Categorization**
- **K Means Clustering**
- **Dose Groups**

**Pre Plan Data**

**Supervised Method**

**Unsupervised Labeling**
We need to predict the dose groups.

- 40 Head and Neck Cancer Patients
- DVH's for OAR
- Visual Categorization
- K Means Clustering
- Dose Groups
- Unsupervised Labeling
- Pre Plan Data
- Supervised Method
- Prediction
We need to predict the dose groups.

40 Head and Neck Cancer Patients

DVH's for OAR

Visual Categorization

K Means Clustering

Dose Groups

Unsupervised Labeling

Pre Plan Data

Supervised Method

Prediction

Supervised Method
We need to predict the dose groups.

40 Head and Neck Cancer Patients

DVH’s for OAR

Visual Categorization

K Means Clustering

Dose Groups

Unsupervised Labeling

Pre Plan Data

Supervised Method

Prediction

Supervised Prediction
We need to predict the dose groups.

40 Head and Neck Cancer Patients

DVH’s for OAR

Visual Categorization

K Means Clustering

Dose Groups

Unsupervised Labeling

Pre Plan Data

Supervised Method

Supervised Prediction
Overlap volumes from PTV expansions can be used as a predictor.
### Parotid Data (cm³)

<table>
<thead>
<tr>
<th>Parotid</th>
<th>1 cm</th>
<th>2 cm</th>
<th>3 cm</th>
<th>4 cm</th>
<th>5 cm</th>
<th>6 cm</th>
<th>7 cm</th>
<th>8 cm</th>
<th>9 cm</th>
<th>10 cm</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0.26</td>
<td>8.6</td>
<td>23.19</td>
<td>36.86</td>
</tr>
<tr>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1.82</td>
<td>7.08</td>
<td>10.67</td>
<td>10.85</td>
<td>10.85</td>
</tr>
<tr>
<td>3</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0.43</td>
</tr>
<tr>
<td>...</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>68</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0.78</td>
<td>6.37</td>
<td>16.36</td>
<td>28.22</td>
<td>35.51</td>
</tr>
<tr>
<td>69</td>
<td>6.82</td>
<td>11.99</td>
<td>16.68</td>
<td>20.44</td>
<td>20.94</td>
<td>20.94</td>
<td>20.94</td>
<td>20.94</td>
<td>20.94</td>
<td>20.94</td>
</tr>
<tr>
<td>70</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0.07</td>
<td>1.19</td>
<td>7.67</td>
<td>19.51</td>
<td>20.49</td>
<td>20.49</td>
</tr>
</tbody>
</table>

We now have this multi-parameter dataset.
Multinomial Logistic Regression (MNR) can be used to classify our new data

Ordinary Least Square – Minimizes distance

\[ y = mx + b \]

Maximum “Likelihood” Estimation – Most likely to produce observed data

\[ f(x) = \frac{1}{1 + e^{-x}} \]
The MNR fit provides weighting coefficients for probability of classification.

<table>
<thead>
<tr>
<th>Group</th>
<th>1 cm</th>
<th>2 cm</th>
<th>3 cm</th>
<th>4 cm</th>
<th>5 cm</th>
<th>6 cm</th>
<th>7 cm</th>
<th>8 cm</th>
<th>9 cm</th>
<th>10 cm</th>
</tr>
</thead>
<tbody>
<tr>
<td>Low</td>
<td>0.46</td>
<td>-0.078</td>
<td>0.19</td>
<td>0.35</td>
<td>0.044</td>
<td>-0.026</td>
<td>-0.036</td>
<td>-0.045</td>
<td>-0.064</td>
<td>-0.36</td>
</tr>
<tr>
<td>Medium</td>
<td>-0.36</td>
<td>-0.69</td>
<td>0.32</td>
<td>0.16</td>
<td>-0.52</td>
<td>0.17</td>
<td>0.12</td>
<td>-0.23</td>
<td>0.052</td>
<td>0.27</td>
</tr>
<tr>
<td>High</td>
<td>-0.10</td>
<td>0.77</td>
<td>-0.52</td>
<td>-0.51</td>
<td>0.48</td>
<td>-0.15</td>
<td>-0.089</td>
<td>0.27</td>
<td>0.012</td>
<td>0.093</td>
</tr>
</tbody>
</table>
The MNR fit provides weighting coefficients for probability of classification.

<table>
<thead>
<tr>
<th>Group</th>
<th>1 cm</th>
<th>2 cm</th>
<th>3 cm</th>
<th>4 cm</th>
<th>5 cm</th>
<th>6 cm</th>
<th>7 cm</th>
<th>8 cm</th>
<th>9 cm</th>
<th>10 cm</th>
</tr>
</thead>
<tbody>
<tr>
<td>Low</td>
<td>0.46</td>
<td>-0.078</td>
<td>0.19</td>
<td>0.35</td>
<td>0.044</td>
<td>-0.026</td>
<td>-0.036</td>
<td>-0.045</td>
<td>-0.064</td>
<td>-0.36</td>
</tr>
<tr>
<td>Medium</td>
<td>-0.36</td>
<td>-0.69</td>
<td>0.32</td>
<td>0.16</td>
<td>-0.52</td>
<td>0.17</td>
<td>0.12</td>
<td>-0.23</td>
<td>0.052</td>
<td>0.27</td>
</tr>
<tr>
<td>High</td>
<td>-0.10</td>
<td>0.77</td>
<td>-0.52</td>
<td>-0.51</td>
<td>0.48</td>
<td>-0.15</td>
<td>-0.089</td>
<td>0.27</td>
<td>0.012</td>
<td>0.093</td>
</tr>
</tbody>
</table>

**Linear Predictor Function**

\[
f(k, i) = \beta_{0,k} + \beta_{1,k}X_{1,i} + \cdots + \beta_{M,k}X_{M,i}
\]

\[
f(k, i) = \overline{\beta_k} \cdot \overline{X_i}
\]
The MNR fit provides weighting coefficients for probability of classification.

### Parotid Weighting Coefficients

<table>
<thead>
<tr>
<th>Group</th>
<th>1 cm</th>
<th>2 cm</th>
<th>3 cm</th>
<th>4 cm</th>
<th>5 cm</th>
<th>6 cm</th>
<th>7 cm</th>
<th>8 cm</th>
<th>9 cm</th>
<th>10 cm</th>
</tr>
</thead>
<tbody>
<tr>
<td>Low</td>
<td>0.46</td>
<td>-0.078</td>
<td>0.19</td>
<td>0.35</td>
<td>0.044</td>
<td>-0.026</td>
<td>-0.036</td>
<td>-0.045</td>
<td>-0.064</td>
<td>-0.36</td>
</tr>
<tr>
<td>Medium</td>
<td>-0.36</td>
<td>-0.69</td>
<td>0.32</td>
<td>0.16</td>
<td>-0.52</td>
<td>0.17</td>
<td>0.12</td>
<td>-0.23</td>
<td>0.052</td>
<td>0.27</td>
</tr>
<tr>
<td>High</td>
<td>-0.10</td>
<td>0.77</td>
<td>-0.52</td>
<td>-0.51</td>
<td>-0.51</td>
<td>-0.15</td>
<td>-0.089</td>
<td>0.27</td>
<td>0.012</td>
<td>0.093</td>
</tr>
</tbody>
</table>

**Linear Predictor Function**

\[ f(k, i) = \beta_{0,k} + \beta_{1,k} \cdot X_{1,i} + \ldots + \beta_{M,k} \cdot X_{M,i} \]

\[ f(k, i) = \overline{\beta_k} \cdot \overline{X_i} \]

\[ P(Y_i = k - 1) = \frac{e^{\beta_{K-1} \cdot \overline{X_i}}}{1 + \sum_{k=1}^{K-1} e^{\beta_k \cdot \overline{X_i}}} \]
The predictive accuracy of our model can be visualized using a confusion matrix.

**Parotid Prediction Matrix**

- Predicted label:
  - Low: 3
  - Medium: 0
  - High: 1

- Actual label:
  - Low: 0
  - Medium: 5
  - High: 8

**Submandibular Prediction Matrix**

- Predicted label:
  - Low: 3
  - Medium: 1
  - High: 1

- Actual label:
  - Low: 1
  - Medium: 4
  - High: 4
An accuracy histogram gives a more precise accuracy.
The accuracy of our models can be improved with additional data.

1. Increase number of plans

2. Increase number of pre plan parameters
Basic machine learning algorithms can improve treatment planning.

1. Utilizes two basic machine learning methods
Basic machine learning algorithms can improve treatment planning.

1. Utilizes two basic machine learning methods
2. Operates on simple expansions to predict DVHs
Basic machine learning algorithms can improve treatment planning.

1. Utilizes two basic machine learning methods
2. Operates on simple expansions to predict DVHs
3. Maintains the patterns and trends of our clinic
I’d like to thank my advisors and the SPORE grant.

Adam Bayliss, PhD  
Patrick Hill, PhD  
Bryan Bednarz, PhD

This project was supported by the Specialized Program of Research Excellence (SPORE) program, through the NIH National Institute for Dental and Craniofacial Research (NIDCR) and National Cancer Institute (NCI), grant P50DE026787. The content is solely the responsibility of the authors and does not necessarily represent the official views of the NIH.”
Thank You