Robust Prediction of Radiotherapy-induced Injury to Ensure Patient Safety

Shiva Das, PhD

Department of Radiation Oncology
Duke University Medical Center

Safety

What do planning dose constraints have to do with safety?

Everything in terms of not harming the patient!

Safety is not just about preventing mistakes that can lead to a patient being harmed. It is also about judiciously adopting guidelines (e.g., dose constraints) that can be directly linked to patient injury.

Guidelines that we usually take for granted ………

But, it isn’t that something catastrophic can happen if we use incorrect/sub-optimal guidelines for critical organs constraints, right?

Maybe so. This may not be the sort of disaster scenarios we associate with the word “safety”. Nevertheless, adopting incorrect/sub-optimal guidelines can result in organ injury to the patient that compromises function. Also, it affects every single patient!

Adopting optimal guidelines can reduce injury and ensure quality of life, a broader interpretation of “safety”.

RT-induced Pneumonitis

- The main dose-limiting toxicity for thoracic RT
- 5 - 15% of patients developed pneumonitis
- Models are needed to predict and reduce risk

Literature

- Identifies volumes above doses from 5 Gy to 50 Gy as being predictive of radiation pneumonitis.
  - Approximately 25% of volume over 20 Gy is frequently used.
  - Volume > 5 Gy has recently become popular.
  - No consensus
- Non-dosimetric factors largely ignored.

What do we need?

A model that can be used ahead of treatment time, incorporating factors such as the dose distribution, chemotherapy, patient parameters (age, sex, etc.), which can output the probability of injury.

If the predicted probability of injury is high, change treatment parameters (dose, chemo).

What sort of model?

Complicated phenomena may be best modeled using machine-learning models (e.g., neural networks).

Isn’t this too complicated?

Do we think that modeling the weather is possible with just one or two variables? It is the same with biological injury – the simplest models have no consensus.

Rationale for Machine Learning

Can more easily extricate the underlying dose-effect relationship.

Disadvantages of machine learning:

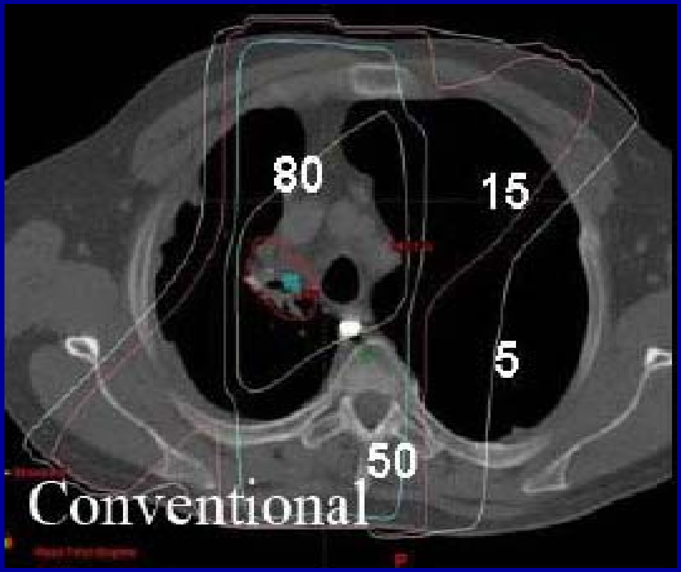
1. If not careful can overfit the model with too many features (fits the training data too well, and performs poorly on test data).

2. A particular machine learning method can have a certain bias (e.g., consistently over/underestimating injury risk in certain situations).

3. Can be hard to interpret cause-effect relationship in simple terms.

Patient Database

- 235 lung cancer patients: 34 pneumonitis (excluding 16 “hard-to-score”)
- 3D conformal radiotherapy



- **Primary fields: AP/P**
  - **40-45 Gy**
- **Boost fields: Off-cord oblique**
  - **20 Gy**

Radiation Pneumonitis

Grade	Definition
0	No increase in symptoms
1	Symptoms not requiring initiation or increase in steroids and/or oxygen
2	Symptoms requiring initiation or increase in steroids
3	Symptoms requiring oxygen
4	Symptoms requiring assisted ventilation or causing death.

Scheme

Use both dose (lung DVH and EUD) and clinical factors to develop non-parametric models for prediction of RT-induced pneumonitis:

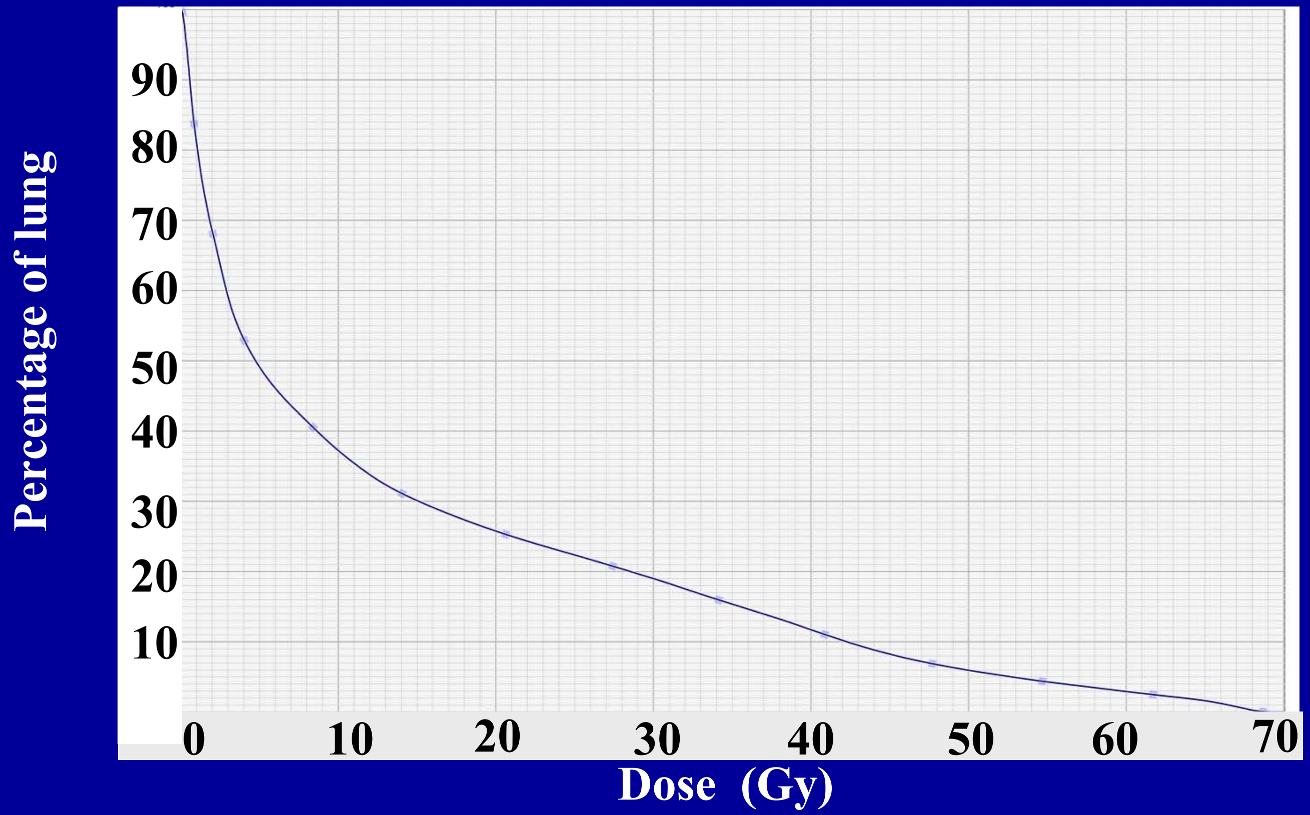
- Decision Trees
- Feed forward neural networks
- Self-organizing maps
- Support vector machines

Combine the four models (decision fusion) to predict injury.

- Reduce individual model bias.
- Extract features that are most important to all models.

Dose Factors

Cumulative dose-volume histogram (DVH)



Dose Factors

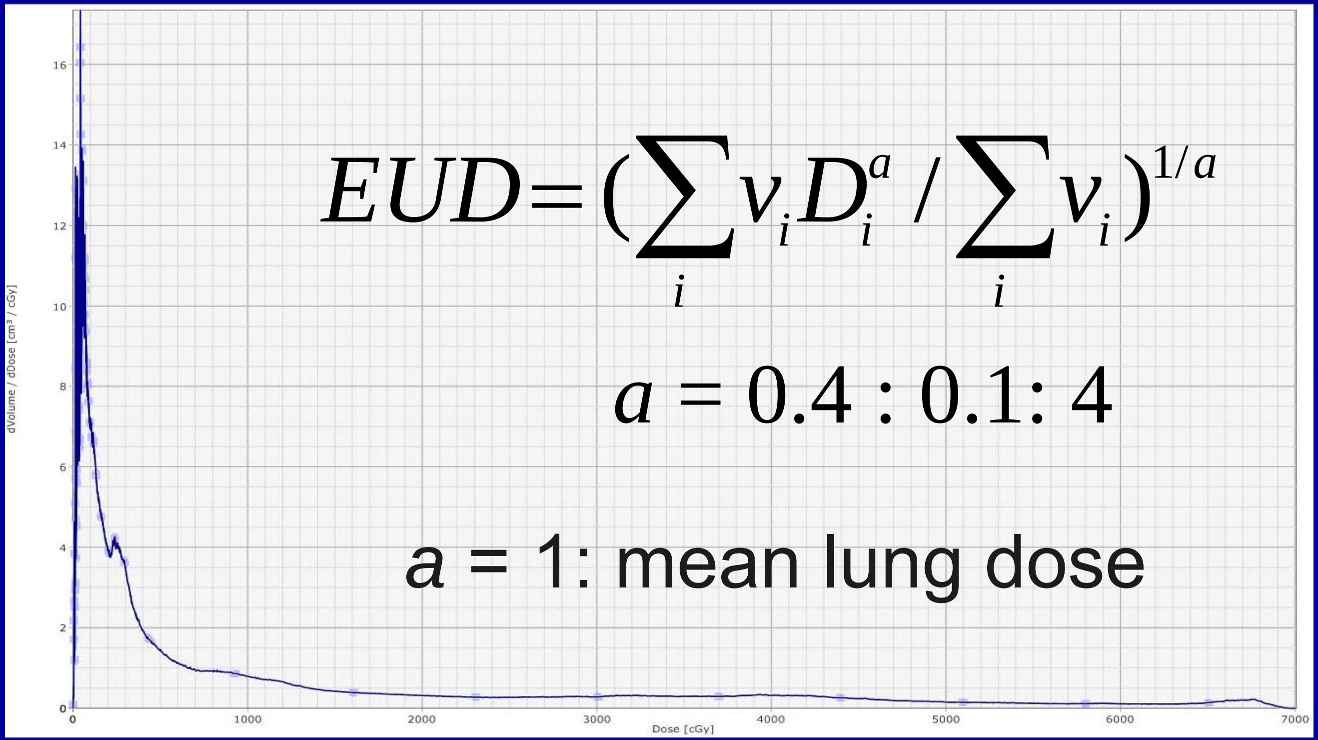
Equivalent Uniform Doses (EUD) are converted from differential DVH

$$EUD = \left(\sum_i v_i D_i^a \Big/ \sum_i v_i\right)^{1/a}$$

$$a = 0.4 : 0.1 : 4$$

a = 1: mean lung dose

V



D

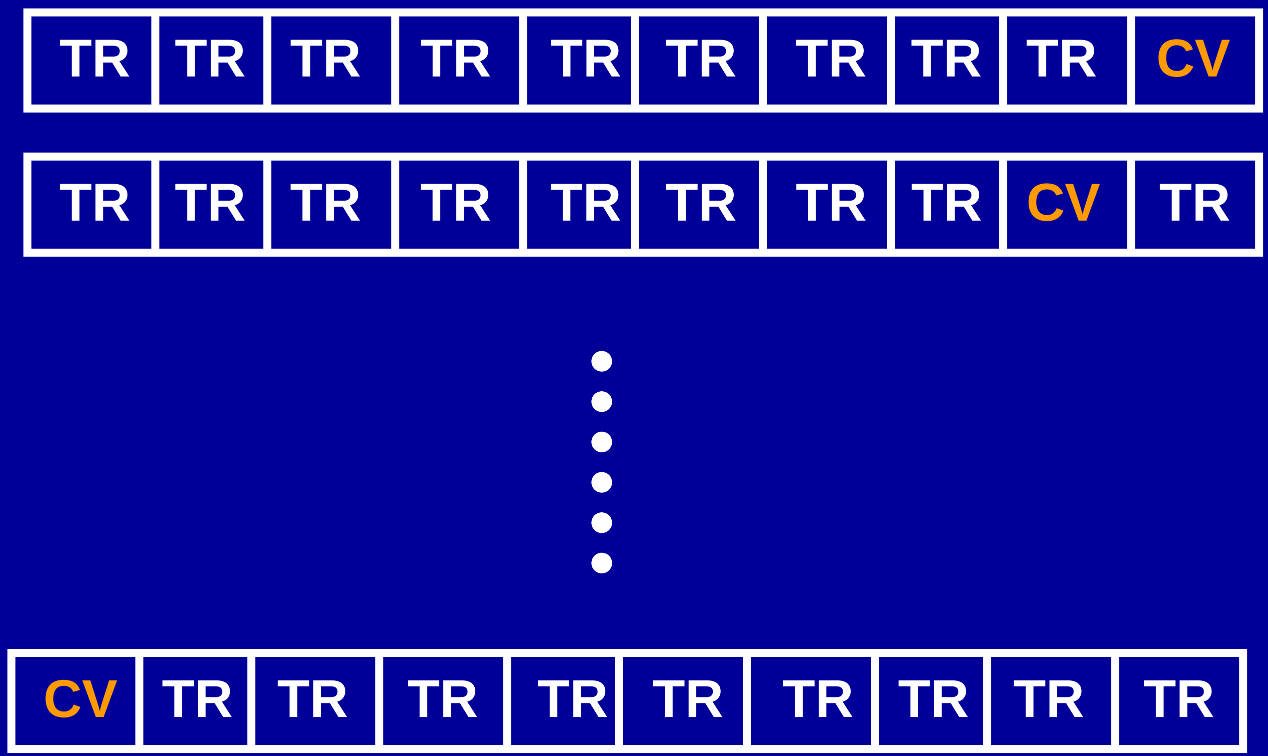
Non-dose Factors

- FEV1 (forced expiratory volume in 1 s)
- pre-RT DLCO (Carbon Monoxide diffusion capacity in lung)
- chemotherapy
- race, age, gender, tumor stage
- tumor location (right/left, up/middle/low, central/peripheral)
- histology (squamous/adenocarcinoma/non-small/small/large/other)
- once/twice daily RT
- surgery (yes or no)

- **Ten-fold cross-validation**
  - **Build and test the models**

- **Receiver Operating Characteristics (ROC) Curve**
  - **Assess model predictive ability**

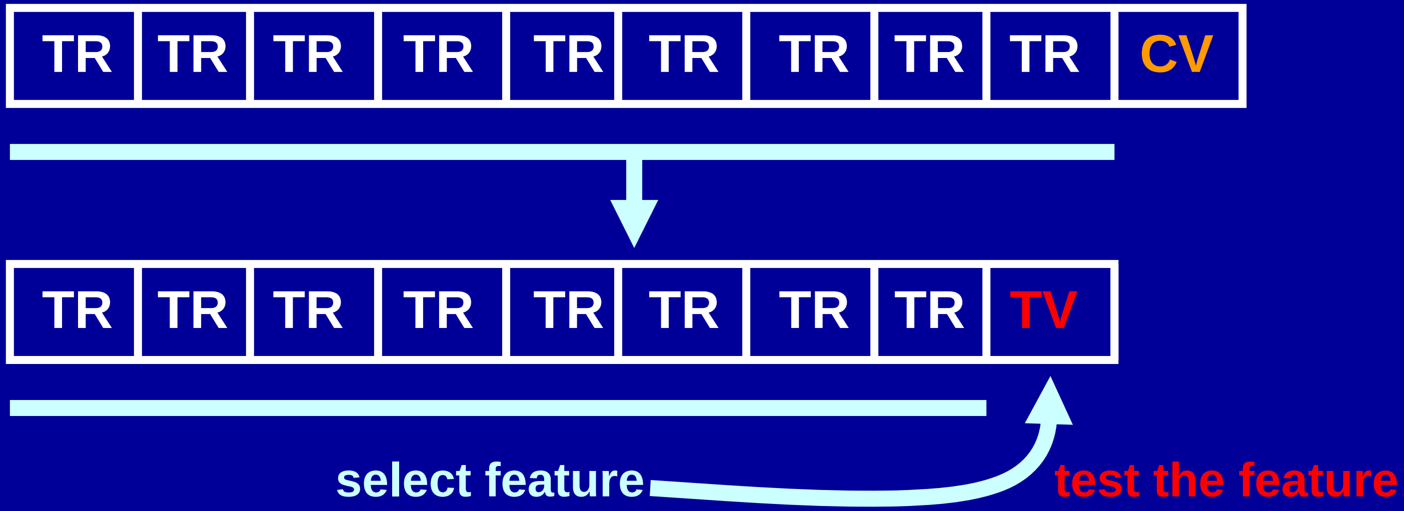
Ten-Fold Cross-Validation



TR: training data

CV: cross-validation

Feature selection



TR: training data, TV: training validation, CV: cross-validation

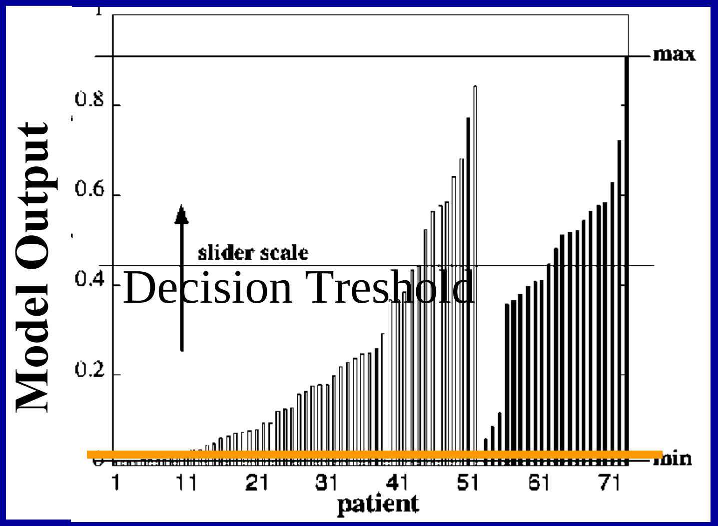
In general, the model was sequentially built by adding, substituting, deleting features.

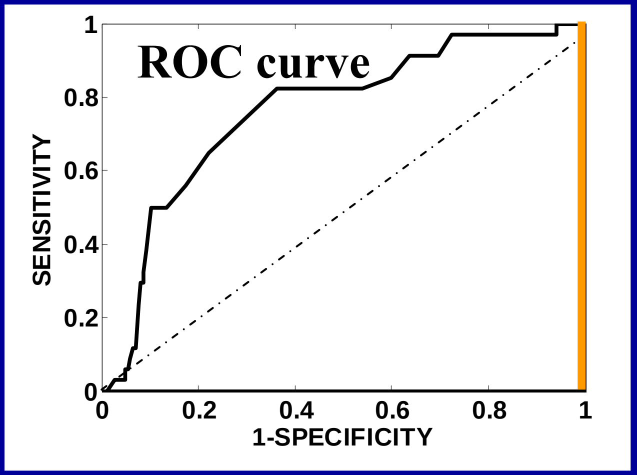
Receiver Operating Characteristics (ROC) Curve

Sensitivity = $\dfrac{\text{Model-identified True Positives}}{\text{True Positives}}$

Specificity = $\dfrac{\text{Model-identified True Negatives}}{\text{True Negatives}}$

- Area under ROC curve
- Perfect model: area = 1
- Model as good as chance: area = 0.5

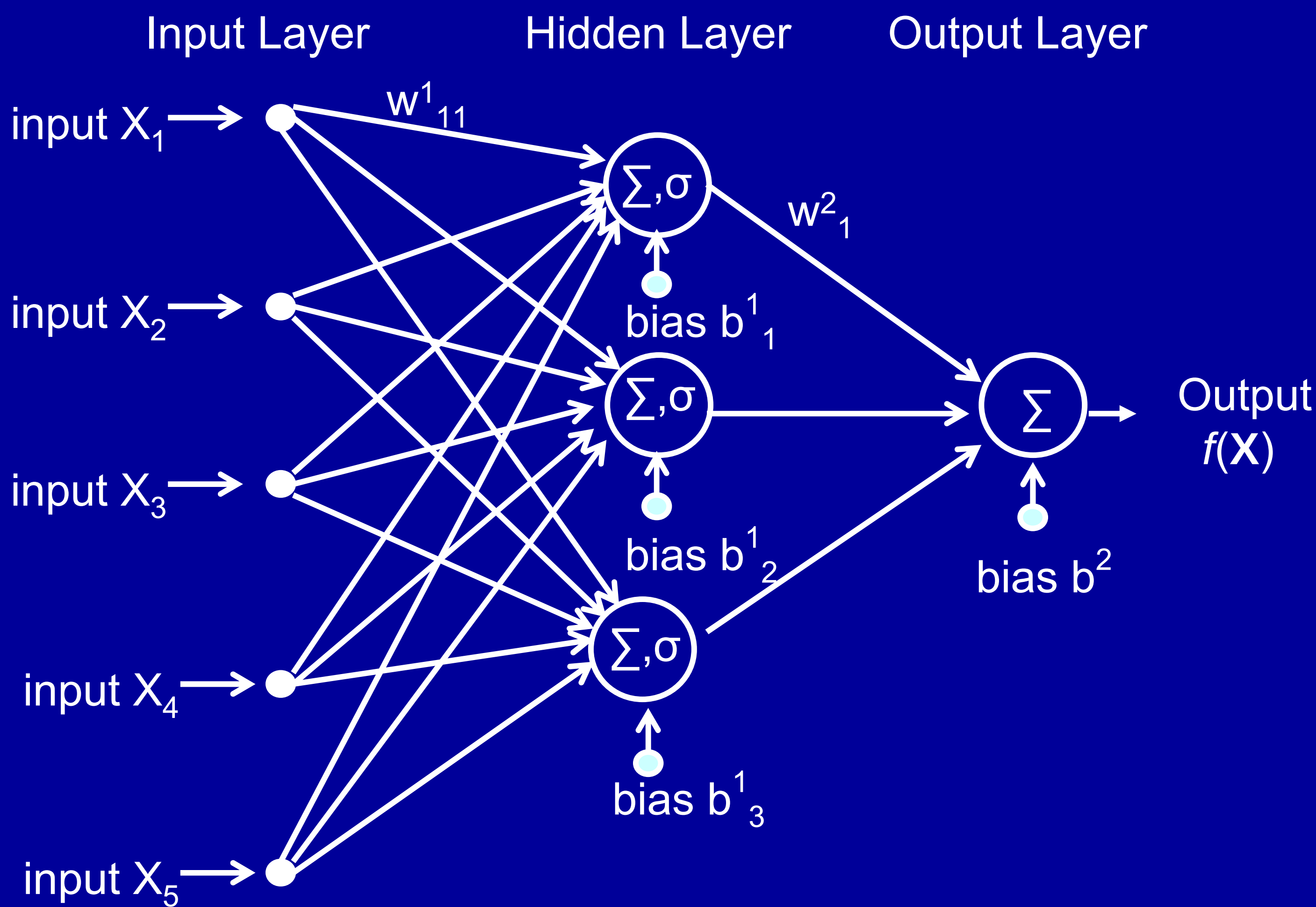




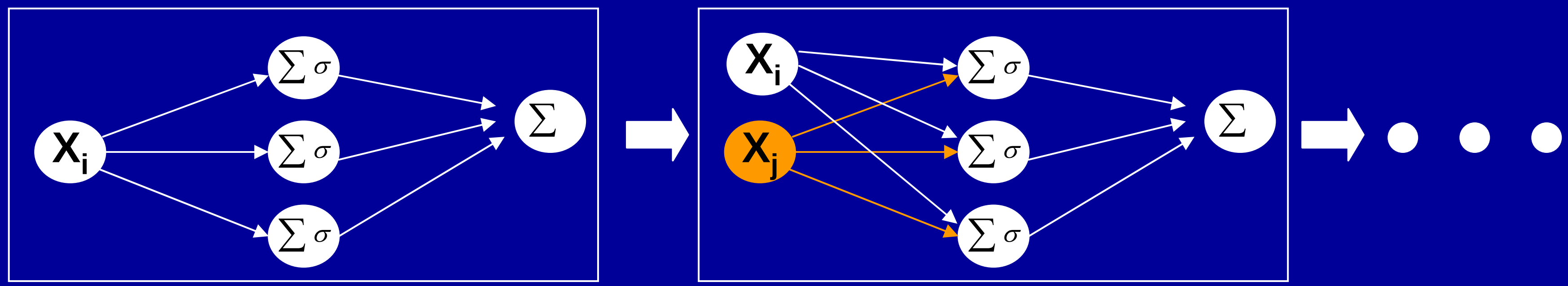
Sensitivity

1-specificity

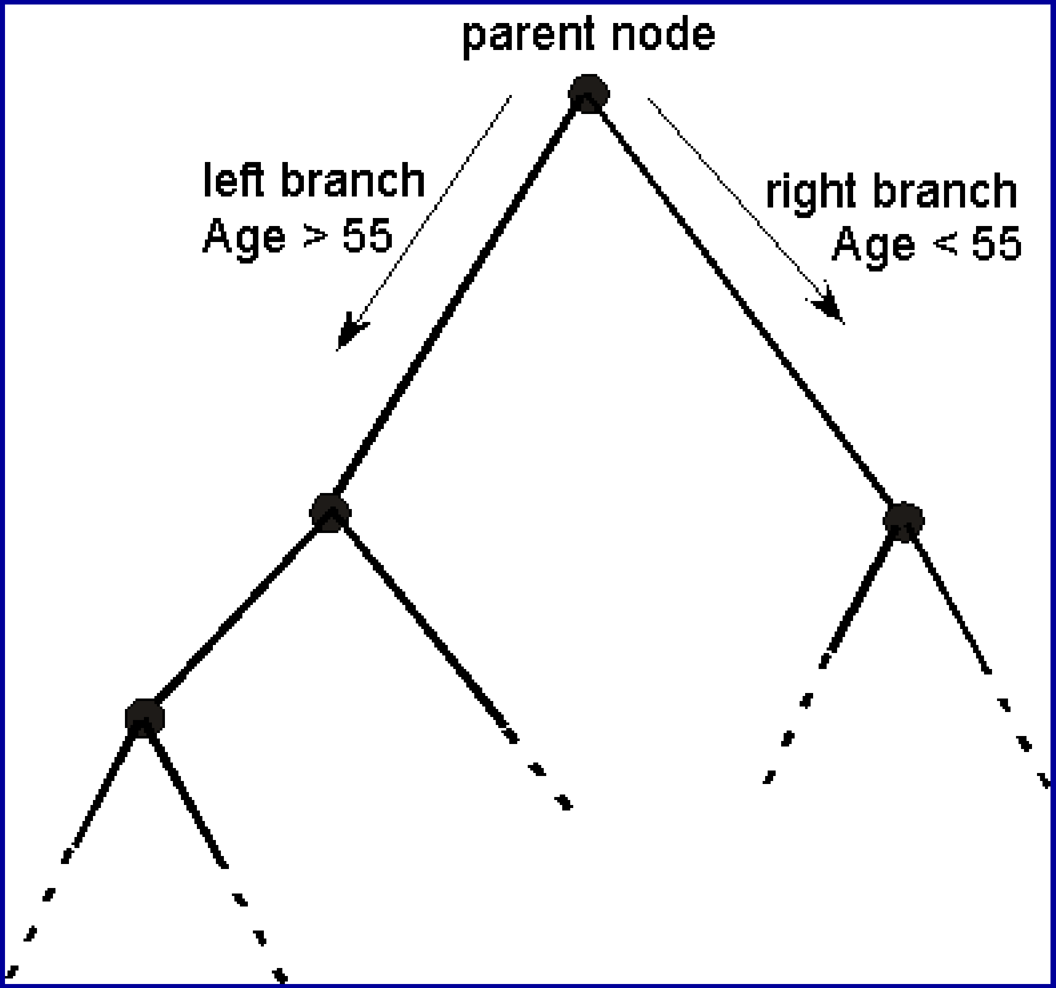
Feed-Forward Neural Network



Growing the Network: Construction Algorithm

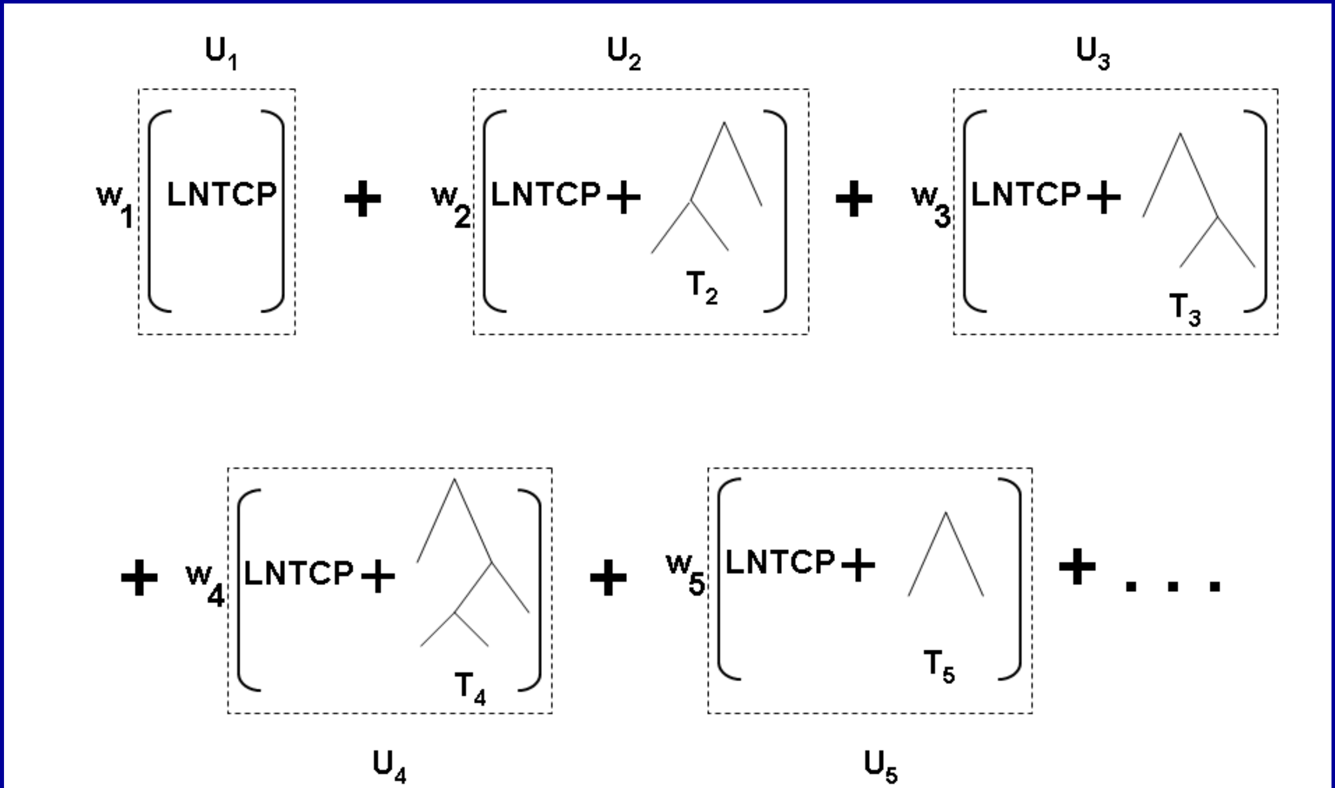


Decision Trees

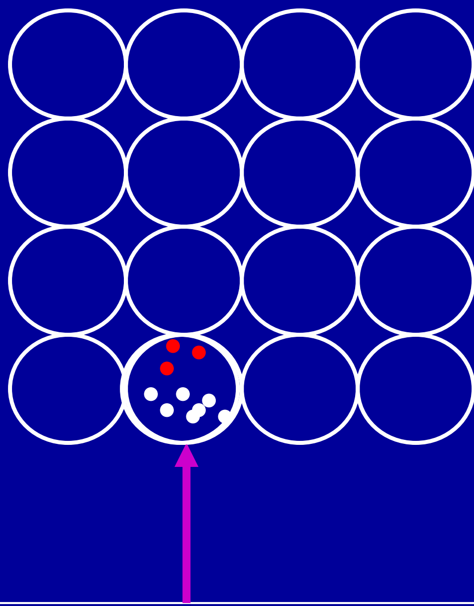


Model Building

Model is built by combining weighted predictive units. Each predictive unit is composed of the Lyman NTCP added to a decision tree. Weighting is achieved by a statistical methodology termed AdaBoost.



Self-Organizing Map (SOM)



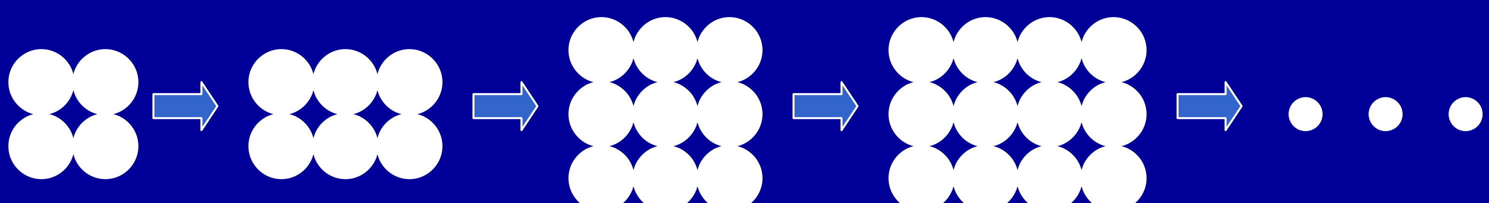
Patients with similar features are clustered in the same region.

$$risk = \frac{N_p}{N_n + N_p}$$

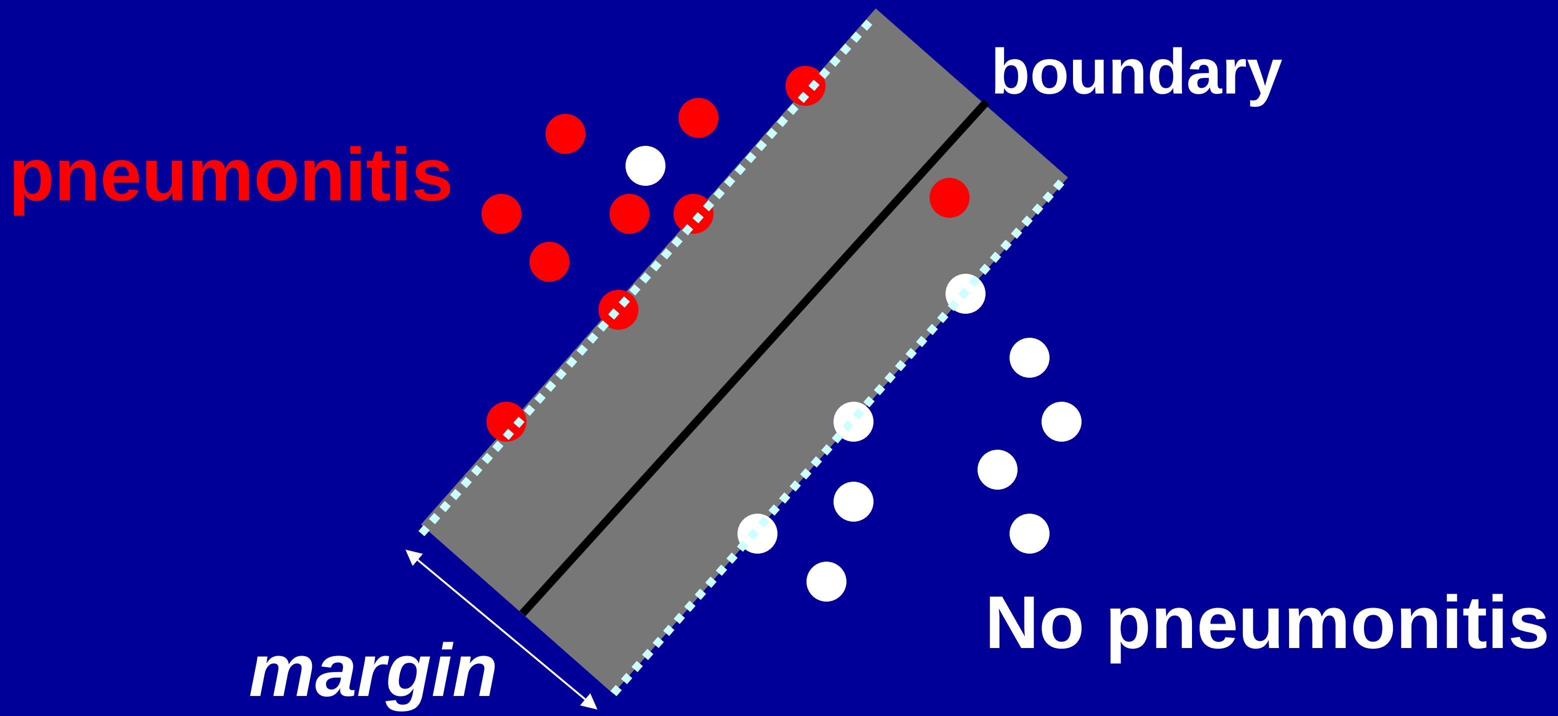
Mean lung dose ~ 15 Gy
Chemotherapy
Female

3: pneumonitis
7: no pnemonitis

$$\text{Risk} = \frac{3}{3+7} = 30\%$$



Support Vector Machine (SVM)



Boundary is optimized to maximize the width of margin.
SVM (unlike LDA) is not as affected by outliers.

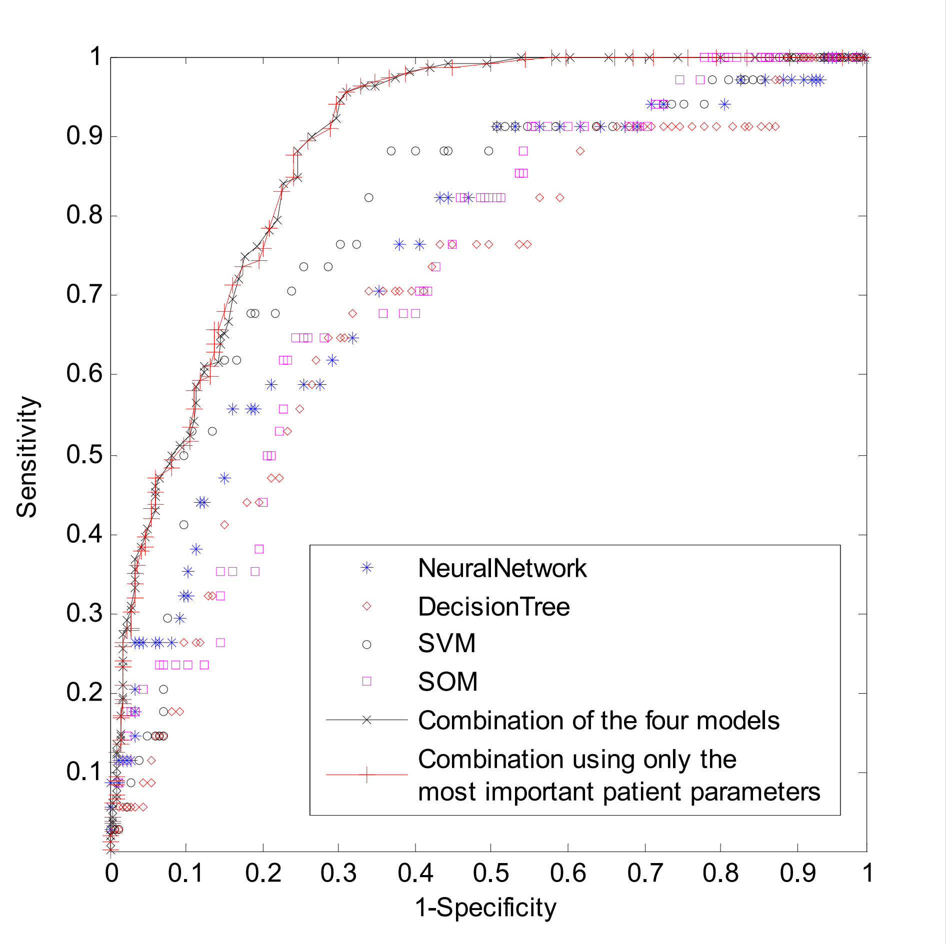
Decision Fusion of Four Models

Initially:

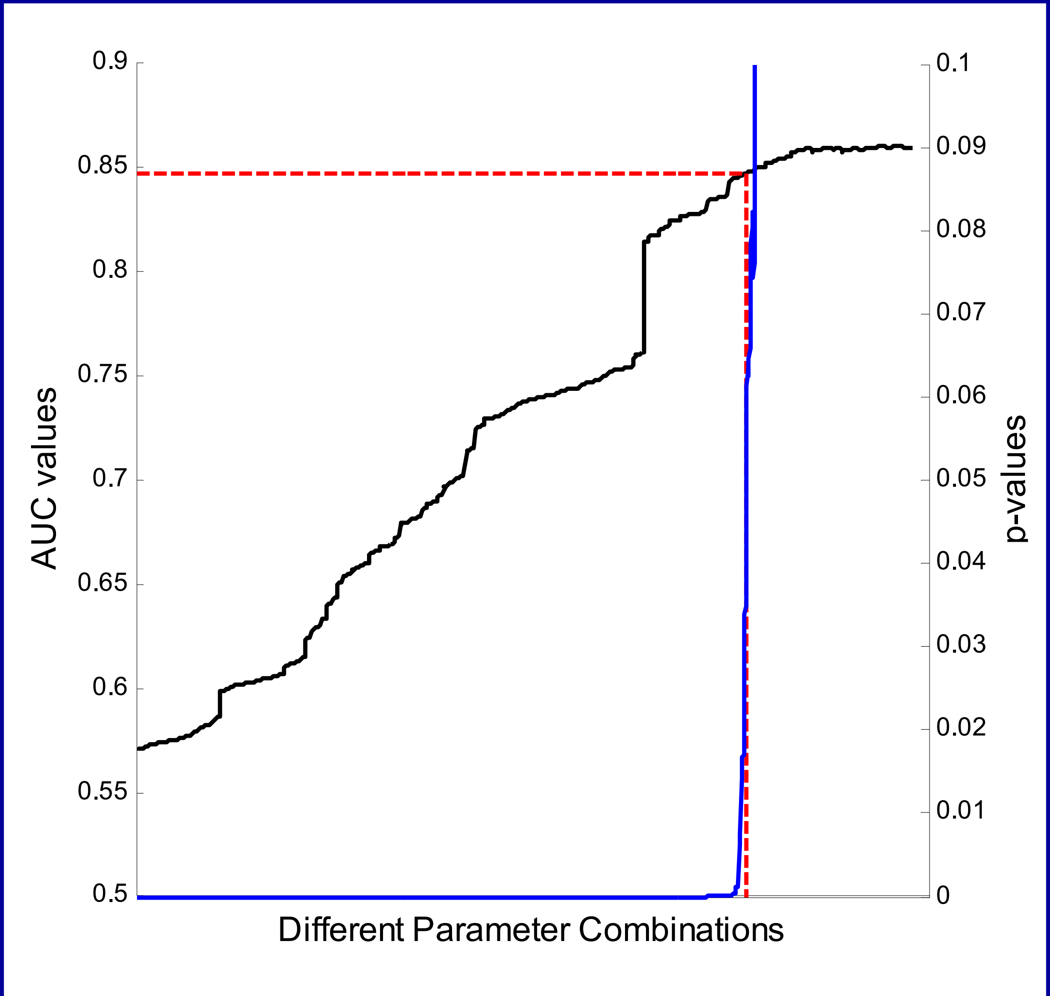
$$\textbf{Injury risk} = \frac{\text{NNET} + \text{DT} + \text{SOM} + \text{SVM}}{4}$$

Refined Later:

Injury risk = Bayesian



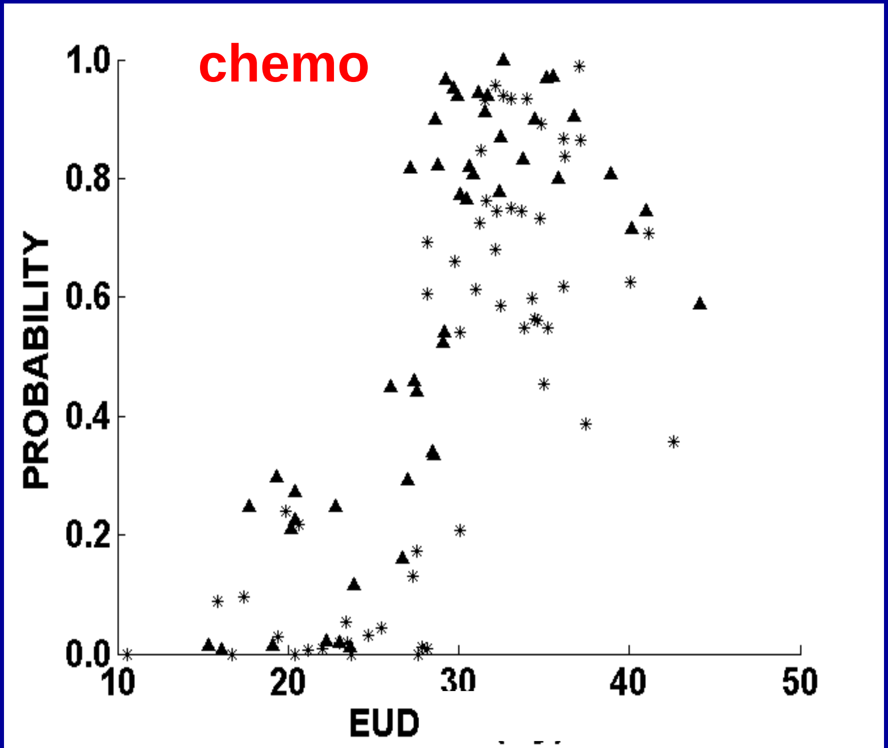
Fused Decision Fusion AUC = 0.85

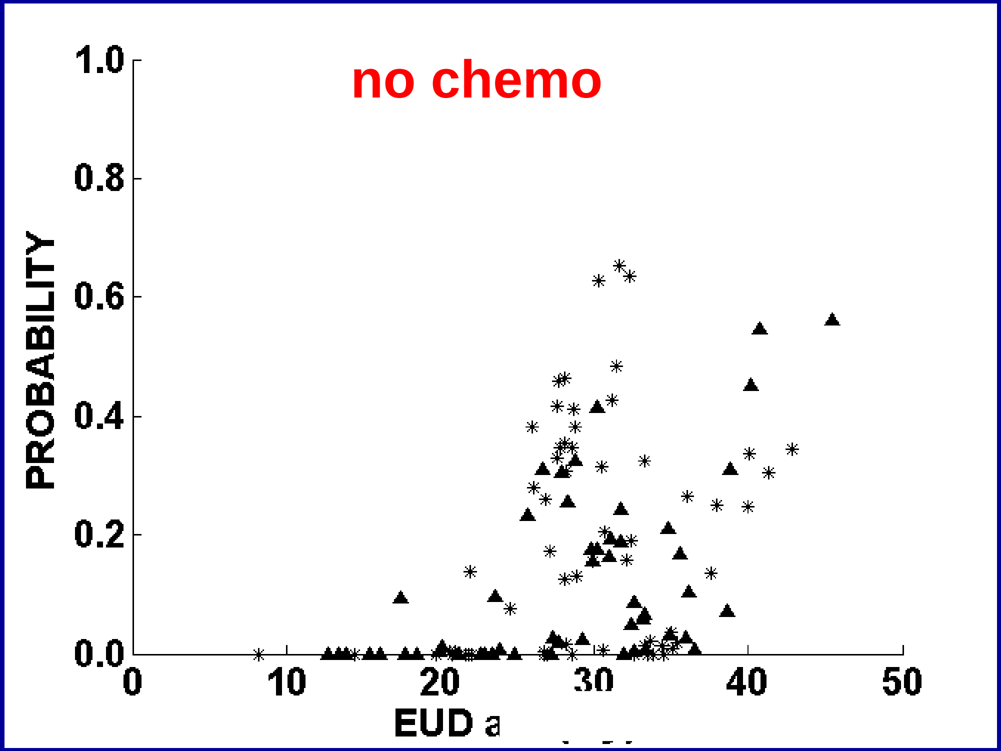


The different machine learning models constituting the fusion pick slightly different variables.
However, the most common and important variables are sufficient to almost fully characterize system.

Parameters	% Frequency	Parameters	% Frequency
$EUD_{1.3}$	50%	V_{30}	49%
Chemotherapy prior to Radiotherapy	52%	Female gender	44%
Non-small Cell	34%	Squamous Cell	38%
Adenocarcinoma	41%	Small Cell Histology	39%
Central Tumor Location	42%	Inferior Tumor location	37%
Number of fractions (BID)	27%	FEV_1	13%

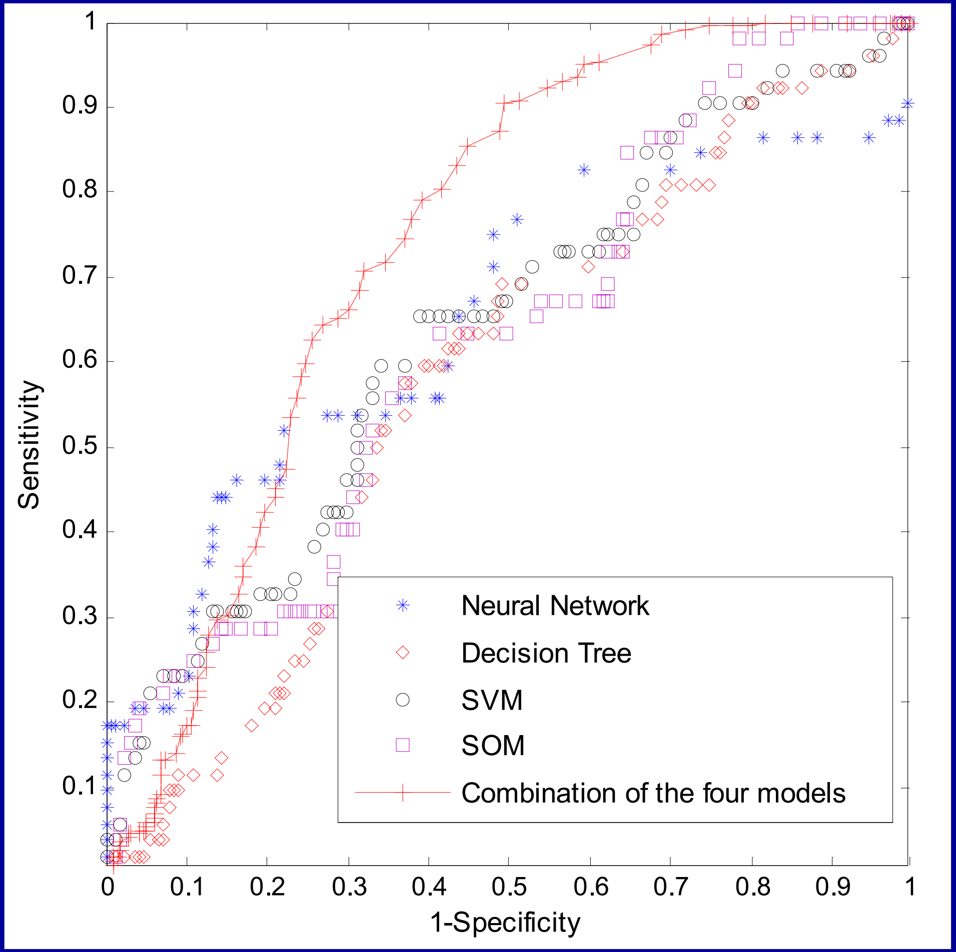
Fusion Identified Features: $EUD_{1.3}$, V_{30}, Tumor location, Histology, Female gender, and Chemotherapy schedule


chemo
1.0
0.8
0.6
0.4
0.2
0.0
PROBABILITY
10
20
30
40
50
EUD

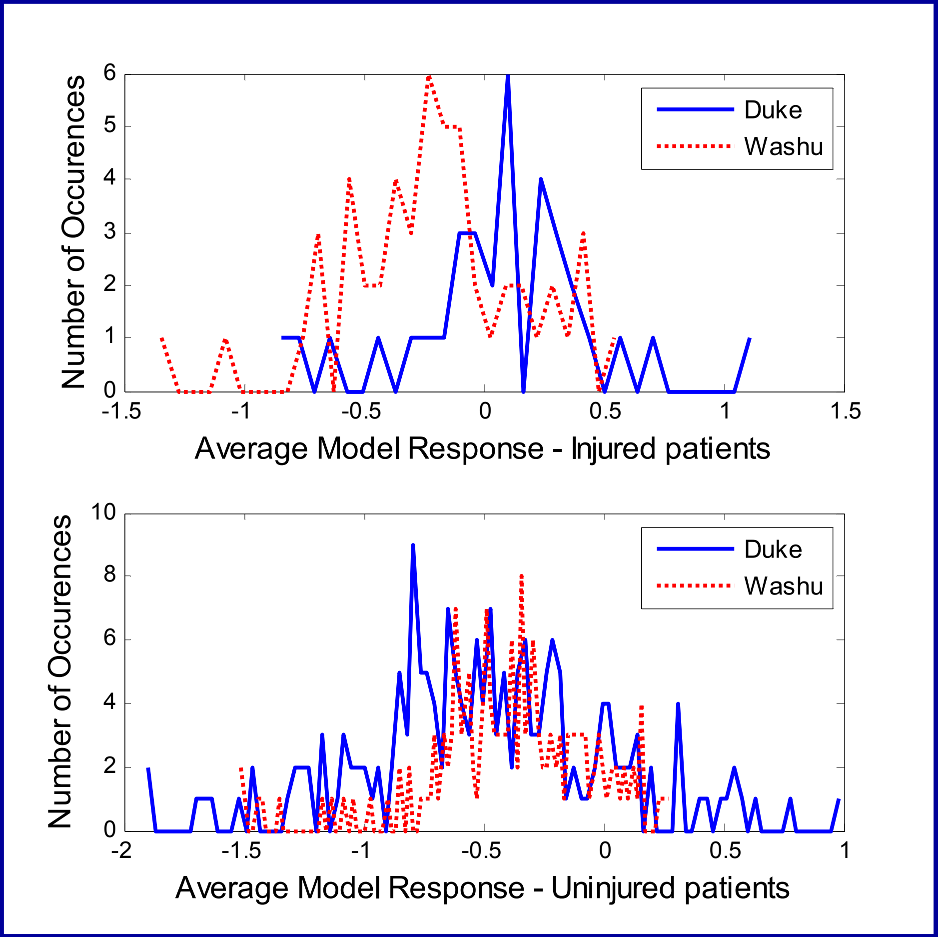

no chemo
1.0
0.8
0.6
0.4
0.2
0.0
PROBABILITY
0
10
20
30
40
50
EUD a

Testing on External Dataset

Washington University dataset:
219 patients (52 developed RP post-radiotherapy)



AUC = 0.73



There is a rough equivalence between the Duke and WashU responses from the fusion model.

Conclusions

- Fusing machine learning models can reduce individual model bias, i.e., function like an almost bias-free model.
- Different models can predict different features. Fusion allows us to extract common “consensus” features.
- Testing on an external dataset demonstrates robustness.

Ensuring reduced patient complications should be a primary safety goal!