The Institute of Cancer Research

in partnership with The Royal Marsden NHS Foundation Trust

Outcomes Models with Machine Learning

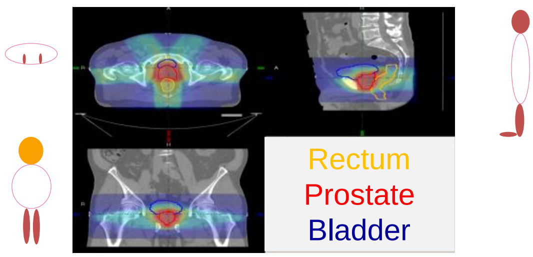
Sarah Gulliford PhD
Division of Radiotherapy & Imaging

sarahg@icr.ac.uk

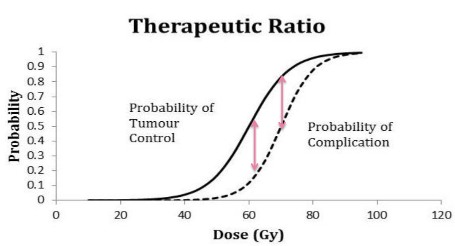
AAPM
31st July 2017

Making the discoveries that defeat cancer

Radiotherapy Outcomes



3



Reviews of Outcome Modelling in Radiotherapy

4

Kang, J., R. Schwartz, J. Flickinger, and S. Beriwal. 2015. "Machine Learning Approaches for Predicting Radiation Therapy Outcomes: A Clinician's Perspective." *Int J Radiat Oncol Biol Phys* 93 (5):1127-35.

El Naqa, I., J. D. Bradley, P. E. Lindsay, A. J. Hope, and J. O. Deasy. 2009. "Predicting radiotherapy outcomes using statistical learning techniques." *Phys Med Biol* 54 (18):S9-S30.

What is Machine Learning?

Original concept based on the way that a human brain learns

- Algorithms designed to learn from the data
- No a priori knowledge of the relationship between the data
- Training using example cases
- Ability to generalise to unseen cases

Unsupervised learning

Data grouped together using common features
No reference made to corresponding output
'Unlabelled data'

- Self organising maps (kohonen)
- Principal Component Analysis

Can be used for feature selection prior to a supervised learning approach

Supervised learning

Algorithms trained to relate input features to output (outcomes)
'Labelled' data
Iterative training using cost function to find best model

- Support Vector Machines
- Random Forest
- Neural Networks

Used for classification & regression*

Common considerations (1)

Data splitting:

Cross validation
Bootstrapping (sampling with replacement)
Independent test set

TRIPOD guidelines

Moons et al Ann Intern Med 162:W1-73 (2015)

Common considerations (2)

Assessment of results:

Receiver Operator Curve (ROC analysis)
Calibration Curves
Learning curves (bias/variance)

Common considerations (3)

Curse of Dimensionality:

High order data becomes sparse in a multidimensional space

http://www.visiondummy.com/2014/04/curse-dimensionality-affect-classification/

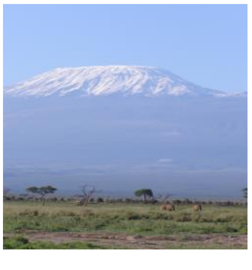
Common considerations (4)

Garbage in Garbage out:

Models are entirely dependent on the quality of the data

- Tumour/organ contouring consistency
- Intra/Inter fraction motion
- Adaptive planning
- Reporting of events using standardised scales
- Quality Assurance

The curse of dimensions



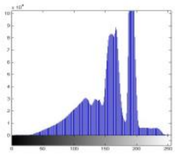
Data stored as a jpeg
3 dimensional array
2816x2112x3

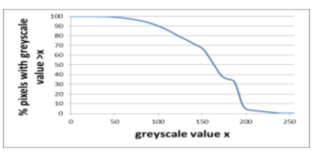
The curse of dimensions

The curse of dimensions

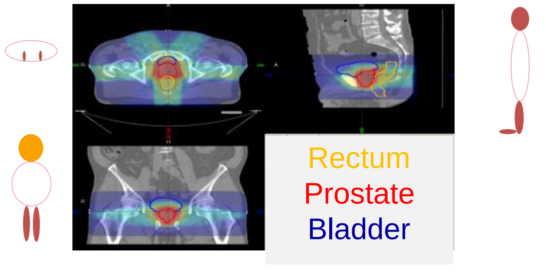
Data stored as 2D matrix 2816 by 2112

The curse of dimensions

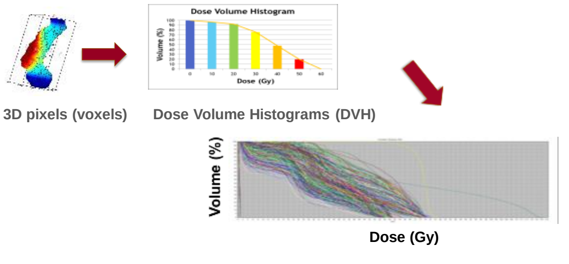




Radiotherapy Outcomes



Dosimetry Features



Challenges of modelling dose-volume effects

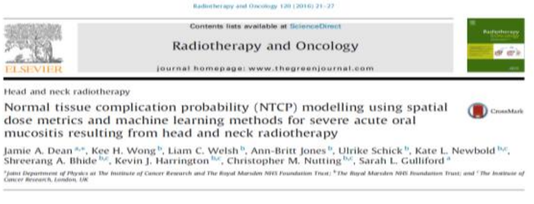
•Dose-volume relationship to toxicity is complex and not well understood

•Highly correlated data

•Toxicity related to a number of factors including dose-volume effects

Challenges of modelling dose-volume effects

- Dose-volume relationship to toxicity is complex and not well understood
- no a priori knowledge of model required
- Highly correlated data
- methods to deal with correlated data
- Toxicity related to a number of factors including dose-volume effects
- can include all types of data without knowing how the variables are related



Radiotherapy and Oncology 120 (2016) 21–27

Contents lists available at ScienceDirect

Radiotherapy and Oncology

journal homepage: www.thegreenjournal.com

Head and neck radiotherapy

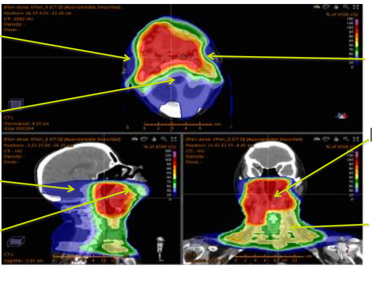
Normal tissue complication probability (NTCP) modelling using spatial dose metrics and machine learning methods for severe acute oral mucositis resulting from head and neck radiotherapy

Jamie A. Dean [a,*], Kee H. Wong [b], Liam C. Welsh [b], Ann-Britt Jones [b], Ulrike Schick [b], Kate L. Newbold [b,c], Shreerang A. Bhide [b,c], Kevin J. Harrington [b,c], Christopher M. Nutting [b,c], Sarah L. Gulliford [a]

[a] Joint Department of Physics at The Institute of Cancer Research and The Royal Marsden NHS Foundation Trust; [b] The Royal Marsden NHS Foundation Trust; and [c] The Institute of Cancer Research, London, UK

Radiotherapy planning

21



Patients

22

Trial	Number available	Primary disease site	Radiotherapy technique	Concurrent chemotherapy
PARSPORT	71	Oropharynx, hypopharynx	Bilateral; Conventional, IMRT	No
COSTAR	78	Parotid gland	Unilateral; Conventional, IMRT	No
Dose Escalation	30	Hypopharynx, larynx	Bilateral; IMRT	Yes
Midline	117	Oropharynx	Bilateral; IMRT	Yes
Nasopharynx	36	Nasopharynx	Bilateral; IMRT	Yes
Unknown Primary	19	Unknown primary	Bilateral; IMRT	Yes

Dean et al Rad Onc **120 (2016) 21–27**

Toxicity scoring

23

CTCAE toxicity	Grade 1	Grade 2	Grade 3	Grade 4
Clinical oral mucositis	Erythema of the mucosa	Patchy ulcerations	Confluent ulcerations	Tissue necrosis; significant spontaneous bleeding

Dose limiting toxicity
Treatment interruptions

Dean et al Rad Onc **120 (2016) 21–27**

Toxicity scoring

24

- Prospectively measured at baseline, weekly during and 1, 2, 3, 4 and 8 weeks post-radiotherapy
- 351 patients with data available
- Patients with baseline toxicity excluded
- Peak grade < 3 vs >= 3
- Patients with missing data excluded
- Final Dataset 183 patients

Dean et al Rad Onc **120 (2016) 21–27**

Clinical data

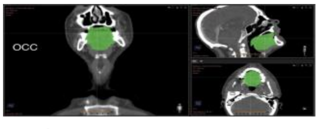
- Age
- Sex
- Primary disease site
- Definitive radiotherapy vs postoperative radiotherapy
- Concomitant treatments
  - Induction chemotherapy
  - Concurrent chemotherapy regime (cisplatin/carboplatin/both)
- No smoking, alcohol or genetic data

Dean et al Rad Onc **120 (2016) 21–27**

Oral mucositis modelling

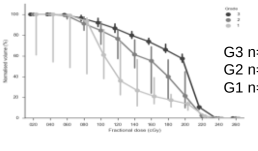
Grade 3 'Confluent ulceration'

Oral cavity



Dose-volume histogram

- fractional dose



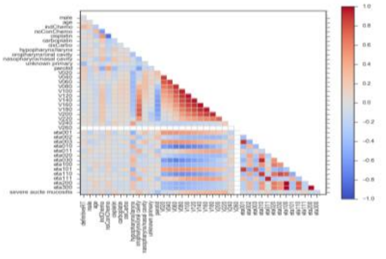
G3 n=134
G2 n=41
G1 n=8

Spatial features

- 3D moment invariants

Dean et al Rad Onc **120 (2016) 21–27**

Spearman's Correlation Matrix



Dean et al Rad Onc **120 (2016) 21–27**

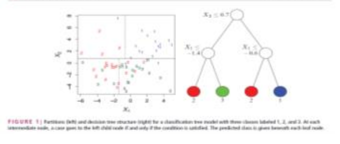
Penalised Logistic Regression

Logistic regression technique extended to mitigate for highly correlated data.

- Ridge Regression – some coefficients set to zero
- Least absolute shrinkage and selection operator LASSO regularisation. –coefficients reduced

Random Forests

Ensembles of decision trees created and initialised using a randomly selected subset of the available data cases.



Loh, W. Y. 2011. "Classification and regression trees." Wiley Interdisciplinary Reviews-Data Mining and Knowledge Discovery 1 (1):14-23. doi: 10.1002/widm.8.

Random Forests

The final result is aggregated from the contributions of each tree.

- outcome classification this will be the most votes (i.e.) class chosen by the most trees
- regression the outcome will be averaged across all the trees.

Support Vector Machines

Classify data by translating variables in to a higher dimensional space where they are linearly separable

Ideally a boundary can be found that completely separates the two possible classes and maximises the distance between them.

Mapping achieved using a Kernel function
- Radial Basis function
- Polynomial function

Support Vector Machines

Computationally intensive to solve however it is possible to characterise the prediction function using only a subset of training data (support vectors)

Predicting radiotherapy outcomes using statistical learning techniques S11

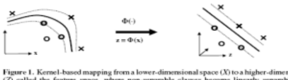
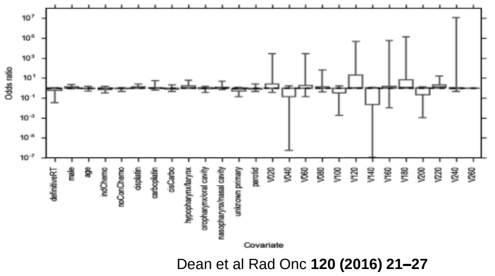


Figure 1. Kernel-based mapping from a lower-dimensional space (*X*) to a higher-dimensional space (*Z*) called the feature space, where non-separable classes become linearly separable. Already established linear theory could be used to estimate the separating hyperplane. Samples on the margin are denoted as support vectors and they define the prediction function, which could be implemented efficiently using the kernel trick.

El Naqa et al Phys. Med. Biol. **54 (2009) S9–S30**

PLR – no spatial features

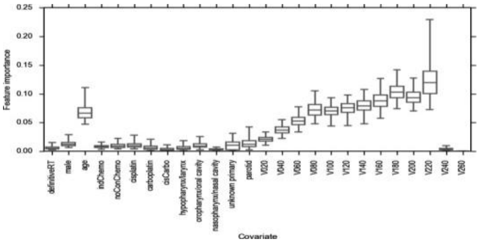
33



Dean et al Rad Onc **120 (2016) 21–27**

Random Forest – no spatial features

34



Dean et al Rad Onc **120 (2016) 21–27**

Results

Table 1
Performance of models on internal validation.

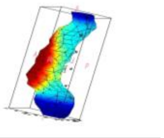
Model	Hyper-parameters	Mean AUC (s.d.)	Mean log loss (s.d.)	Mean Brier score (s.d.)	Mean calibration slope (s.d.)	Mean calibration intercept (s.d.)
$PLR_{standard}$	Regularisation = LASSO, C = 0.1	0.72 (0.09)	0.66 (0.03)	0.23 (0.02)	12.4 (10.9)	−5.0 (5.2)
$SVC_{standard}$	Kernel = radial basis function, C = 0.1, gamma = 0.01	0.72 (0.09)	–	–	–	–
$RFC_{standard}$	Max depth = 5, max features = square root	0.71 (0.09)	0.56 (0.08)	0.19 (0.03)	3.9 (2.2)	−1.5 (1.4)
$PLR_{spatial}$	Regularisation = LASSO, C = 0.1	0.72 (0.09)	0.66 (0.04)	0.23 (0.02)	11.9 (10.9)	−4.8 (5.2)
$SVC_{spatial}$	Kernel = radial basis function, C = 1.0, gamma = 0.001	0.71 (0.09)	–	–	–	–
$RFC_{spatial}$	Max depth = 5, max features = square root	0.70 (0.09)	0.56 (0.07)	0.18 (0.03)	4.2 (2.3)	−1.9 (1.6)

PLR – penalised logistic regression; SVC – support vector classification; RFC – random forest classification; s.d. – standard deviation; C – inverse regularisation strength.

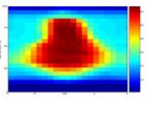
Spatial information did not improve predictive performance

Dean et al Rad Onc **120 (2016) 21–27**

Representing Dose distribution using dose surface Maps



3D dose distribution



Rectal Dose Surface Map

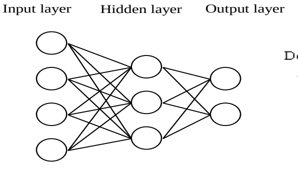
IOP Publishing — Physics in Medicine and Biology
Phys. Med. Biol. 54 (2009) 5139–5153 — doi:10.1088/0031-9155/54/17/005

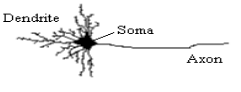
Using dose-surface maps to predict radiation-induced rectal bleeding: a neural network approach

Florian Buettner, Sarah L Gulliford, Steve Webb and Mike Partridge

Joint Department of Physics, Institute of Cancer Research and Royal Marsden NHS Foundation Trust, Sutton, Surrey SM2 5PT, UK

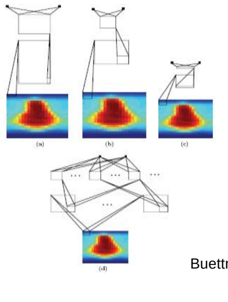
Artificial Neural Network





Weighted sum on each node $\Sigma_i w_{ij}$
Non linear activation function
Backpropagation of errors

Dose surface map ANN architecture



locally connected NN
2 hidden layers
a-c) individualised weights

d) shared weights

2 output nodes

Buettner et al Phys. Med. Biol. **54 (2009) 5139–5153**

Ensemble Learning

.

Ensemble learning incorporates groups of neural networks each with different starting conditions and selected subset of training data sets
250 NN initialised. Results aggregated

Expert ensemble
Results of each NN are assessed and if they improve the performance of the ensemble they are “voted in”.

Buettner et al Phys. Med. Biol. **54 (2009) 5139–5153**

Patients

Prostate cancer UK-MRC RT01 trial

Compared 64Gy vs 74Gy (circa 1998-2001)
388 patients with data
Used to predict rectal bleeding >= Grade 2
(RMH score) simple outpatient management/transfusion

Patients with baseline toxicity excluded
329 patients 53 patients with G2 Rectal Bleeding

Buettner et al Phys. Med. Biol. **54 (2009) 5139–5153**

Results

Table 3. Performances of all locally connected nets. AUC_{all} denotes the AUC calculated from all nets in the ensemble and AUC_{exp} the AUC derived from the experts only.

Architecture	No of hidden and output nodes	No of weights and biases [a]	AUC_{all}	AUC_{exp}
1	312	1442	0.57	0.57
2	111	1173	0.61	0.64
3	58	562	0.59	0.62
4	492	964	0.56	0.57

[a] The number of unique weights and biases was calculated by adding up the unique connections between the nodes: $1442 = 210\times3+100\times3+2\times100+312$, $1173 = 100\times9+9\times16+2\times9+111$, $562 = 49\times9+7\times7+2\times7+58$, $964 = 4\times9+16\times4+32\times6+2\times90+492$.

Compare results with fully connected NN using DSH data
AUC 0.59

Buettner et al Phys. Med. Biol. **54 (2009) 5139–5153**

Why such a low AUC?

- Incomplete characterisation of spatial information
- Model architecture
- Inter & Intra fraction rectal motion/filling
- Only dosimetry in the model

What's missing?

- ❑ Clinical factors (age, diabetes etc)
- ❑ Other therapies (hormones)
- ❑ Genetic variants(SNPS)

Why don't we use Machine Learning more?

Reputation mystical black box

Wide variety of techniques (which approach is appropriate?)

The road less trodden

Summary

•Evidence that Machine Learning approaches are complimentary to traditional statistical techniques and each other.

•Data hungry: more variables need more datasets

•Require rigorous methodology and independent validation

ICR The Institute of Cancer Research

CANCER RESEARCH UK

in partnership with The ROYAL MARSDEN NHS Foundation Trust

Acknowledgements

Prof David Dearnaley
Prof Christopher Nutting
Dr Florian Buettner
Dr Julia Murray
Dr Jamie Dean
Mr Matt Sydes
Prof Emma Hall

Patients and Colleagues who report, collect and share data

MRC Medical Research Council

Radiotherapy Trials QA

Making the discoveries that defeat cancer

ICR