Radiomics in Head and Neck Cancer

Severin Langberg

December 5, 2018

OUTLINE

1) **Personalized cancer medicine:**
   - Radiomics: Motivation and principles
2) **A head and neck cancer study:**
   - Comparing machine learning algorithms

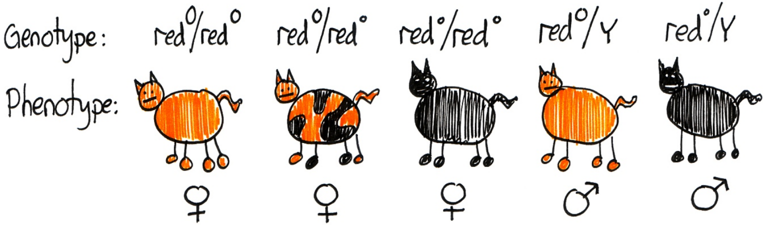
Personalized medicine:

- Predict patient response → tailor therapies

Impersonalized medicine:

Personalized medicine:

- Phenotypic differences in medical images



[1]

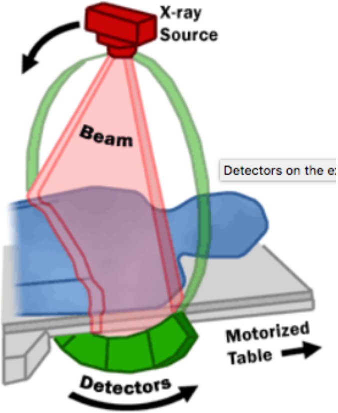
[1]http://cubocube.com/dashboard.php?a=1181b=1260c=103

Imaging for precision medicine:
Characterization and quantification of disease stage

- Non-invasive
- Different modalities in clinical practice

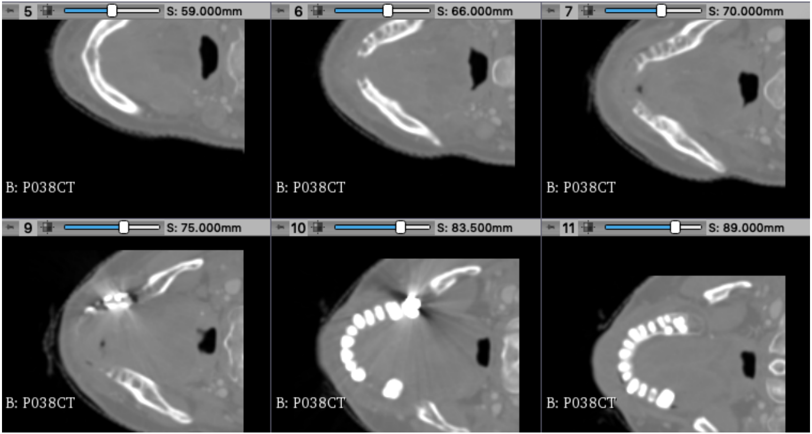
- Heterogenous aquisition protocols
- Qualitative not quantitative

Computed Tomography (CT)

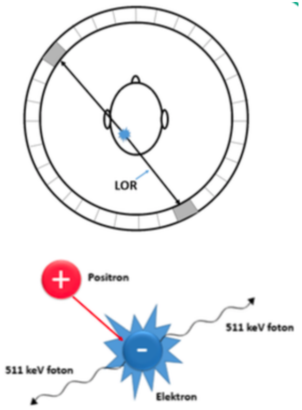
2


- Measure absorption of X-rays
- Displays bodily structures

[2] U.S. Food and Drug Administration. www.fda.gov. 11/27/2018

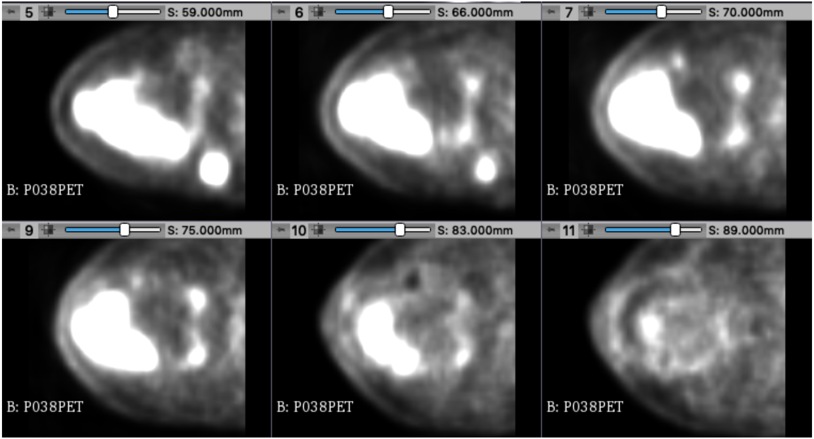

5
S: 59.000mm
B: P038CT
6
S: 66.000mm
B: P038CT
7
S: 70.000mm
B: P038CT
9
S: 75.000mm
B: P038CT
10
S: 83.500mm
B: P038CT
11
S: 89.000mm
B: P038CT

Positron Emission Tomography (PET)

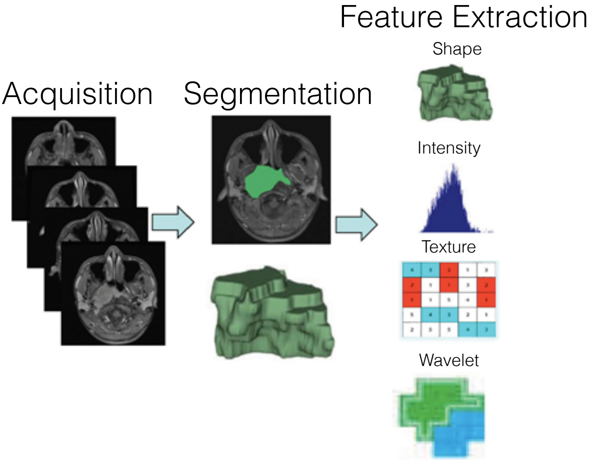
3 

- Inject radioactive tracers
- Displays metabolic activity

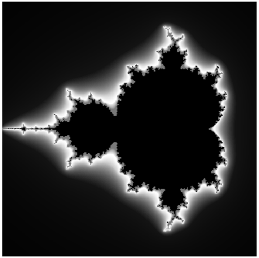
[3] U.S. Food and Drug Administration. www.fda.gov. 11/27/2018


5
S: 59.000mm
6
S: 66.000mm
7
S: 70.000mm
B: P038PET
B: P038PET
B: P038PET
9
S: 75.000mm
10
S: 83.000mm
11
S: 89.000mm
B: P038PET
B: P038PET
B: P038PET

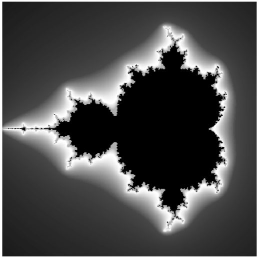
Radiomics


Feature Extraction
Shape
Acquisition
Segmentation
Intensity
Texture
Wavelet

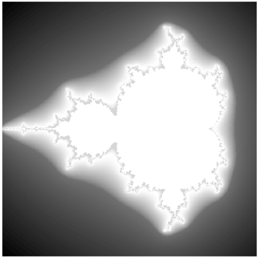
FILTERING



Original

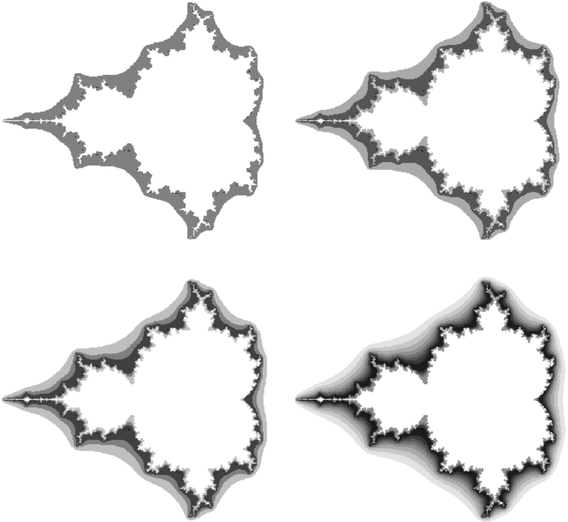


Squareroot



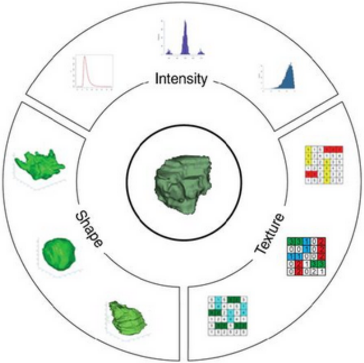
Logarithm

Discretization



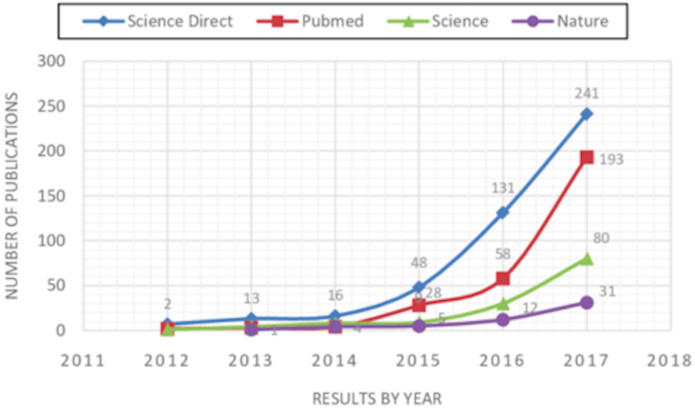
Radiomics

- Shape-based (16)
- First-order statistics (19)
- Texture (75)



[5] Zhang, Ouyang, & Gu et al (2017). Advanced nasopharyngeal carcinoma

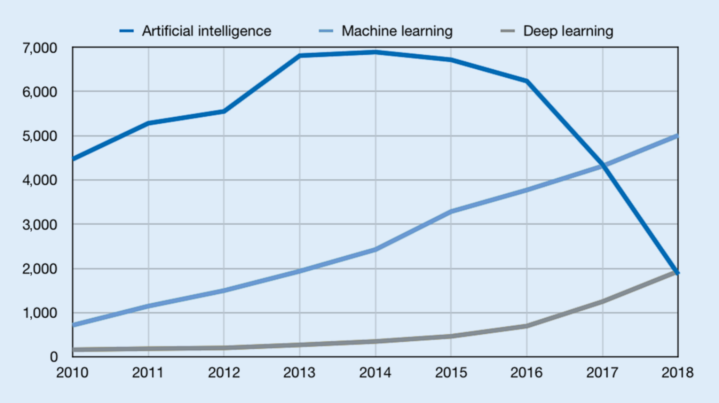
Radiomics



Radiomic publications by year. [6]

[6] Florez, Edward & Fatemi et al (2018). Emergence of Radiomics

Radiomics + AI

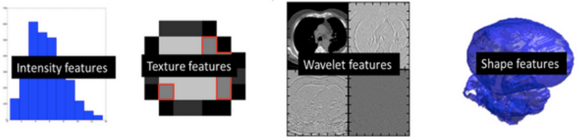


AI in PubMed publications by year. [7]

[7] Langs, Röhrich, Hofmanninger et al (2018). Radiologe

RADIOMICS + AI

Radiomic Features



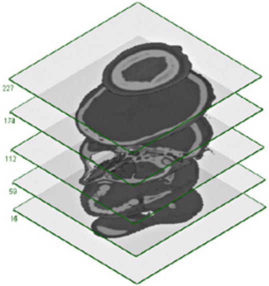
Machine Learning

Treatment outcome

CASE

Goal: Classify patients by radiotherapy response
Data set: 198 head and neck cancer patients

- Disease-free survival/locoregional relapse
- Clinical variables (sex, age, smoking habits, etc.)
- PET + CT stacks



CASE

Base study:

- Graylevel discretization (PET = 16 bins, CT = 128 bins)
- Square root transformation
- 4-fold nested cross-validation (CV)
  - PLSR, log reg (L2), LDA, QDA (AUC of 0.66)

Parmar & **Grossmann et al.**: [8]

- 464 lung cancer patients (CT)
- 14 feature selection and 12 classification methods
- AUC of 0.68 with random forest
- Adopted hyperparameter settings from another study

[8] Parmar, Grossmann et al. 2015. *Machine learning methods for quantitative radiomic biomarkers*

The .632+ bootstrap method

- K-fold CV: nearly unbiased, but can be
  - highly variable
  - pessimistic
- Based on .632 estimator by Bradley Efron

$$\varepsilon_{.632} = 0.368\varepsilon_T + 0.632\varepsilon_V$$

Issue: Cannot handle overfitting

The .632+ bootstrap method

- Improved: Efron strikes back (with Tibhirani)

$$\varepsilon_{.632+} = \omega \varepsilon_T + (1 - \omega) \varepsilon_V$$

- variable weighting

$$\omega = \frac{0.632}{1 - 0.368 R(\gamma)}$$

R: relative overfitting rate (degree of overfitting)
γ: no-information error rate ($\geq$ majority class)

The .632+ bootstrap method

- relative overfitting rate

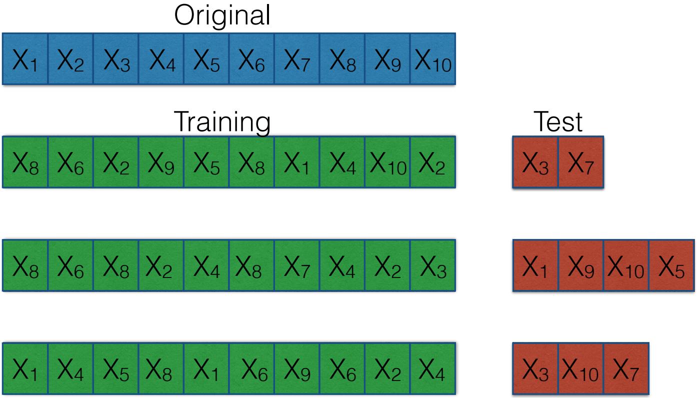
$$R = \frac{\varepsilon_T - \varepsilon_V}{\gamma - \varepsilon_V}$$

- No-information error rate (binary)

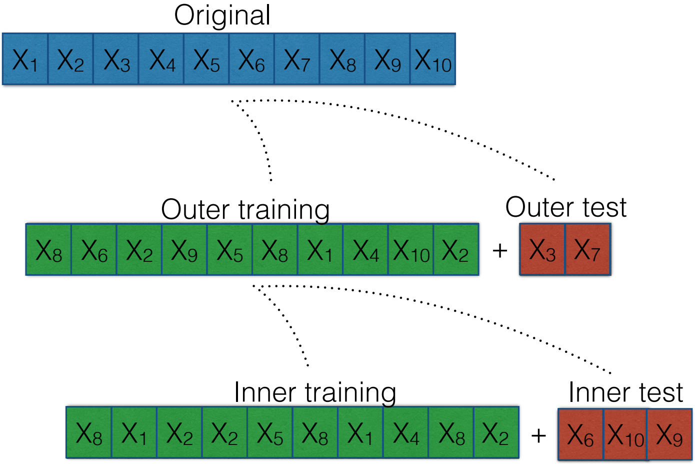
$$\gamma = p_1(1 - q_1) + (1 - p_1)q_1$$

p_1: # observations = 1
q_1: # predictions = 1

Bootstrap Out-of-Bag



Nested bootstrap Out-of-Bag



EXPERIMENTS

Preprocessing:

- Graylevel discretization (PET = 16 bins, CT = 128 bins)
- Square root transformation

Feature extraction:

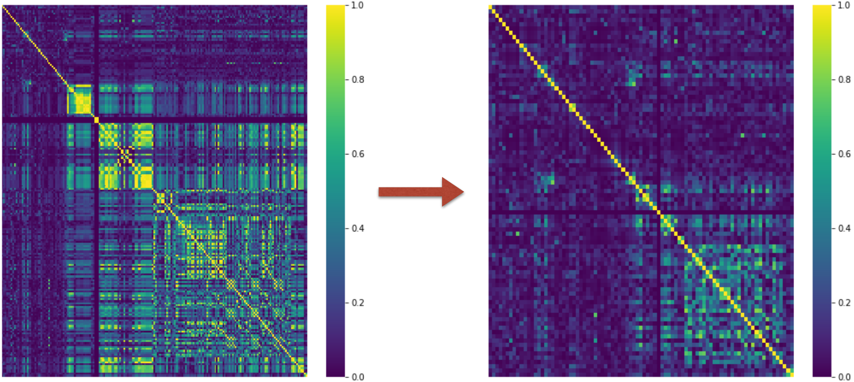
- 188 PyRadiomics image features
- 49 clinical variables (encoded)

EXPERIMENTS

Postprocessing: Remove unattainable + correlated features

- 37 PET textural features failed extraction (flat image regions)
- In total 87 features to analysis

Correlations (≥ 0.85)



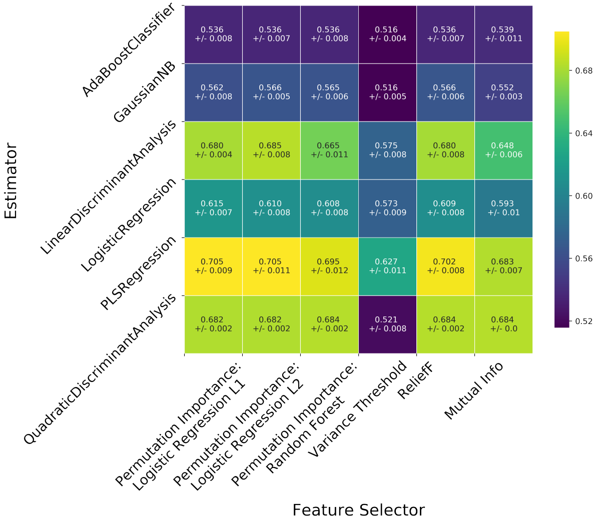
Model comparisons

- 10 experimental parallels (DFS/LRR)
- 50 out-of-bag nested iterations
- Area Under Curve (TPR / FPR) based .632+ score

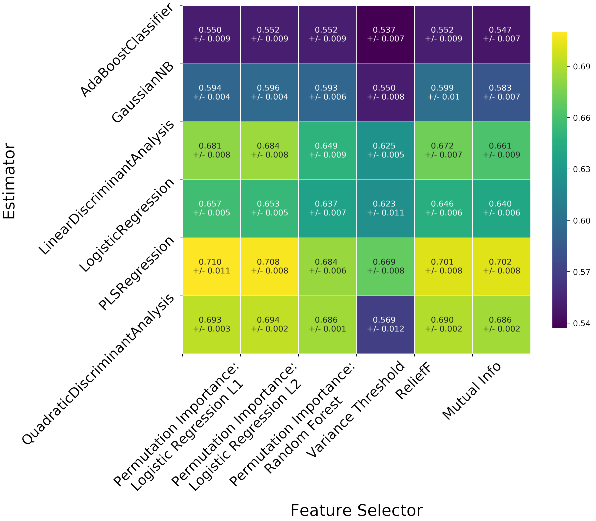
Note: majority of events

- Disease free survival: 0.67
- Locoregional relapse: 0.75

Locoregional relapse



Disease free survival



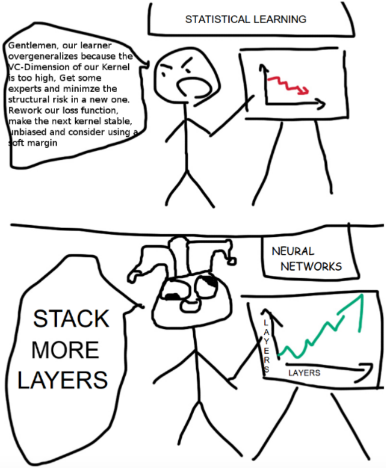
Observations

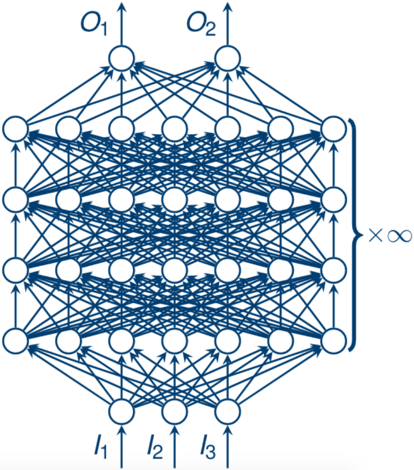
- PLSR (best performing):
  - Disease free survival: $0.67 < 0.710)$
  - Locoregional relapse: $0.75 > 0.705)$
- Improved performance compared to base study (0.66)

For the future

- Alternative image transformations
- Feature extraction/feature selection/Deep feature synthesis
- Increase model complexity (boosting/DL)

SUMMARY

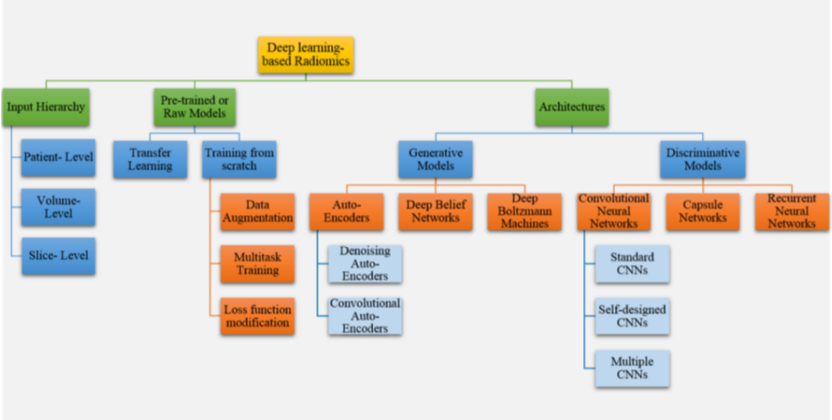




9

[9] https://www.reddit.com/r/ProgrammerHumor/comments/5si1f0/machine_learning_approaches/

Deep learning in radiomics



[10]

[10] Afshar, Parnian, et al. "From Hand-Crafted to Deep Learning-based Cancer Radiomics. (2018).

Feature reduction techniques

Category	Description	Methods
Supervised	Considers the relation of features with the class labels and features are selected mostly based on their contribution to distinguish classes.	
• Filtering (Univariate)	Test the relation between the features and the class label one by one.	Fisher score (FSCR); Wilcoxon rank sum test; Gini index (GINI); Mutual information feature selection (MIFS); Minimum redundancy maximum relevance (MRMR), and; Student t-test [44].
• Wrapper (Multivariate)	Considers both relevancy and redundancy.	Greedy forward selection, and Greedy backward elimination.
Unsupervised	Does not consider the class labels and its objective is to remove redundant features.	
• Linear	Features have linear correlations.	Principle Component Analysis (PCA), and; Multidimensional scaling (MDS)
• Nonlinear	Features are not assumed to be lied on a linear space.	Isometric mapping (Isomap), and; Locally linear embedding (LLE).

[11]

[11] Afshar, Parnian, et al. "From Hand-Crafted to Deep Learning-based Cancer Radiomics. (2018).

.632+ ESTIMATOR

The no information rate in a dichotomous classification problem

$$\gamma = p(1-q) + (1-p)q$$

- p: Proportion of responses $y_i = 1$.
- q: Proportion of predictions $\hat{y}_j = 1$.

.632+ ESTIMATOR

The relative overfitting rate, R

$$R = \frac{\epsilon_{(V)} - \epsilon_{(T)}}{\gamma - \epsilon_{(T)}}$$

- No overfitting: $\epsilon_{(V)} = \epsilon_{(T)} \Rightarrow R = 0$
- Overfitting equal to no information value $\gamma - \epsilon_{(T)} \Rightarrow R = 1.$

.632+ estimator

Alternative .632+

$$\varepsilon_{.632+} = \varepsilon_{.632} + (\overline{\varepsilon_r} - \varepsilon_V)\frac{0.367(1-0.367)\overline{R}}{1-0.367\overline{R}}$$

where $\overline{\varepsilon_T} = \min\{\varepsilon_T, \gamma\}$, and $\overline{R} = \varepsilon_T$ if $\gamma > \varepsilon_V$, else 0.

MODEL COMPARISONS

- Logistic regression
- Partial least squares regression
- Gaussian Naive Bayes classifier
- Linear discriminant classifier
- Quadratic discriminant classifier
- Adaboost classifier

MODEL COMPARISONS

- Variance threshold
- ReliefF (feature ranking approach)
- Mutual information (entropy measure)
- L1 or L2 regularized logistic regression permutation importance
- Random forest permutation importance