GÖTEBORGS UNIVERSITET

Machine learning for radiation biomarkers

“Development of a machine learning framework for radiation biomarker discovery and absorbed dose prediction”

BJÖRN ANDERSSON, BIONFORMATICS AND DATA CENTRE (BDC)

GÖTEBORGS UNIVERSITET

Bionformatics and Data Centre

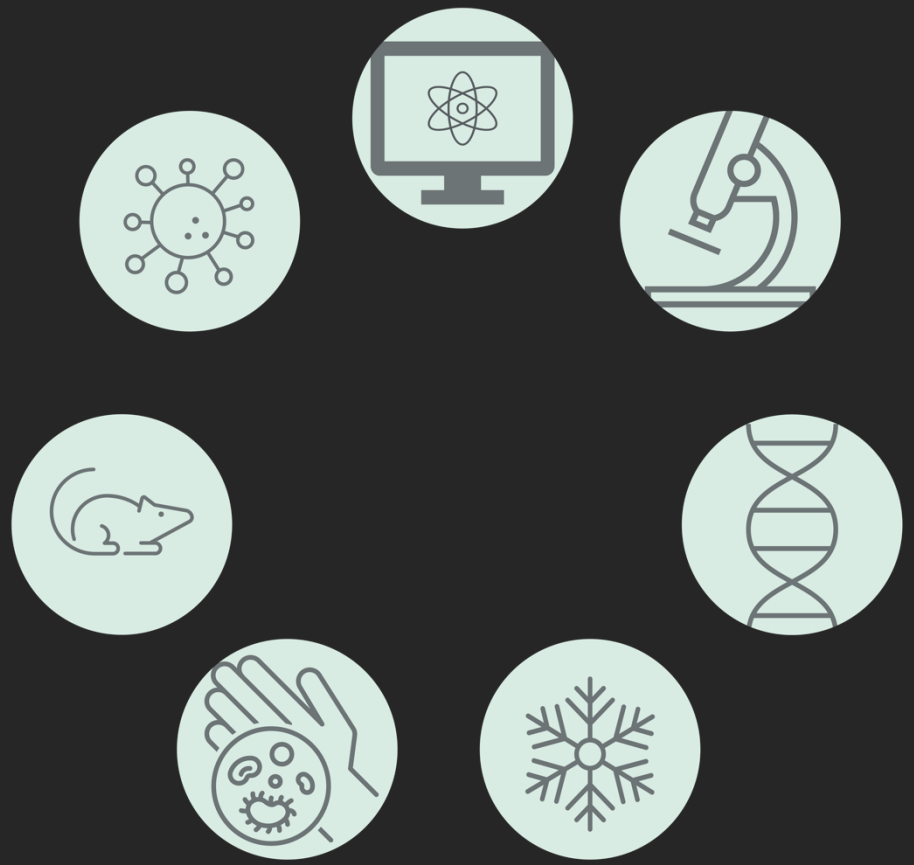
A bioinformatics hub supporting biomedical and clinical research

Biobank Core Facility
Sample Service Coordination Across Sweden

Bioinformatics and Data Centre (BDC)
Sequence, Analyze, Visualize, and Integrate 'Big Data'

Centre for Cellular Imaging (CCI)
Advanced Multimodal Imaging

Experimental Biomedicine (EBM)
State-of-the-Art Support for Animal Research & Education



Mammalian Protein Expression Core Facility (MPE)
Complex Recombinant Proteins & Monoclonal Antibodies

OligoNova Hub
Design & Synthesis of Therapeutic Oligonucleotides

Proteomics Core Facility (PCF)
Cutting-Edge Mass Spectrometry- Based Protein Analysis

Outline

Background & aim
Method
Results
What learned & what to do today

Background

A Swedish Foundation for Strategic Research (SSF) grant
Develop pipelines/apps + education

Radiation researcher
Radiation research conservative
1 statistician, 1 bioinformatician

Background - radiation biomarkers

- Radiation biomarkers - cancer radiotherapy
- Dose of ionizing radiation versus biological effect
- Over/under treatment
- Markers tissue damage

Aim

- To develop a resource-efficient ML framework for radiation biomarker discovery
- To identify biomarker panels predicting radiation dose response with tissue specificity

Method - messy -> “simple”

- Multiple transcriptomic dataset from different sites (radiation types, doses, time points, tissues, gender and strains) – “impossible” to combine
- Ineffective Genetic Algorithm/k-Nearest Neighbor (GA/KNN) – good markers, didn’t scale well

- Simpler - KISS rule

Method

- Mice tissue
- Transcriptomics - normal tissues radiation response
- 36 samples, kidney cortex and medulla
- Controls + 5 doses from low to very high, 1 time-point
- Analyzes, visualization and pre-processing done in R
- Recursive Feature Elimination (RFE), caret package
- 13 ML models were evaluated using the caret package

Data cleaning/preparation

- 25,697 features -> 9783
- Rows with not significant detection p-value
- All NAs and zero variance predictors were removed
- Quantile normalization
- Scaling and centering

Feature selection

- Trade off – biological insight vs prediction accuracy
- Modelling time
- To reduce dimensionality (non-informative + redundant features)
- Minimize noise and prevent collinearity
- ML model with less features but with similar model performance.
- Sizes – (1:10, 15, 20, 25) - optimal subset size/performance
- 10-fold cross-validation, repeated 5 times
- Random forest variable importance selection
- Two sets of features had similar metrics the smaller one was selected

Modelling

- 70/30 train/test split – avoid imbalance!!!
- Randomness & reproducibility – seed/multicore
- Trade off – biological insight vs prediction accuracy
- 5-fold CV, repeated 20 times
- Logistic regression – reference
- 13 ML algorithms – performance test/train (time)

TABLE 2 Performance results of different machine learning algorithms.

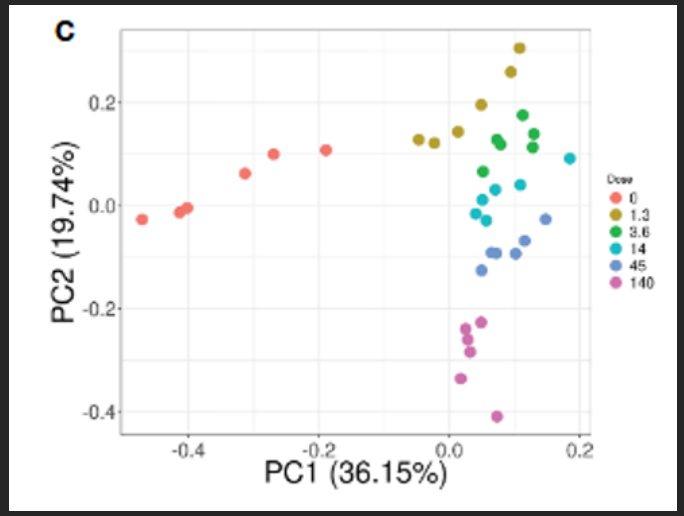
	Logistic Regression	k-Nearest Neighbors (KNN)	Penalized Discriminant Analysis (PDA)	Lasso/ Elastic Net	C5.0 Decision Tree	Soft Independent Modelling by Class Analogy (SIMCA)	Partial Least Squares (PLS)	Naive Bayes	Random Forest	Neural Networks (NNET)	Gradient Boosting Machine (GBM)	Classification and Regression Trees (CART)	Support Vector Machine (SVM)
Measure	m_glm	m_kknn	m_pda	m_glmnet	m_C5	m_CSimca	m_pls	m_nb	m_parRF	m_nnet	m_gbm	m_rpart	m_svm
Train data													
Accuracy	0.56	0.86	0.87	0.87	0.74	0.42	0.87	0.87	0.80	0.78	0.78	0.75	0.79
Kappa	0.11	0.47	0.54	0.53	0.24	0.10	0.54	0.48	0.00	0.40	0.40	0.20	0.00
F1	0.43	0.67	0.73	0.68	0.52	0.40	0.70	0.70	N/A	0.59	0.59	0.49	N/A
Sensitivity	0.62	0.43	0.56	0.49	0.42	0.85	0.53	0.42	0.00	0.62	0.62	0.36	0.00
Specificity	0.54	0.97	0.96	0.98	0.83	0.31	0.97	0.99	1.00	0.82	0.82	0.86	1.00
Pos_Pred_Value	0.27	0.85	0.82	0.90	0.42	0.28	0.86	0.98	N/A	0.53	0.53	0.40	N/A
Neg_Pred_Value	0.86	0.87	0.89	0.88	0.84	0.93	0.89	0.87	0.80	0.89	0.89	0.84	0.79
Precision	0.27	0.85	0.82	0.90	0.42	0.28	0.86	0.98	N/A	0.53	0.53	0.40	N/A
Recall	0.62	0.43	0.56	0.49	0.42	0.85	0.53	0.42	0.00	0.62	0.62	0.36	0.00
Detection_Rate	0.13	0.09	0.12	0.10	0.09	0.18	0.11	0.09	0.00	0.13	0.13	0.08	0.00
Balanced_Accuracy	0.58	0.70	0.76	0.73	0.63	0.58	0.75	0.71	0.50	0.72	0.72	0.61	0.50
Test data													
Accuracy	0.48	0.86	0.81	0.86	0.67	0.52	0.86	0.86	0.86	0.81	0.81	0.81	0.81
Kappa	-0.02	0.49	0.00	0.35	0.01	0.21	0.35	0.35	0.35	0.00	0.00	0.24	0.00
Sensitivity	0.50	0.50	0.00	0.25	0.25	1.00	0.25	0.25	0.25	0.00	0.00	0.25	0.00
Specificity	0.47	0.94	1.00	1.00	0.76	0.41	1.00	1.00	1.00	1.00	1.00	0.94	1.00

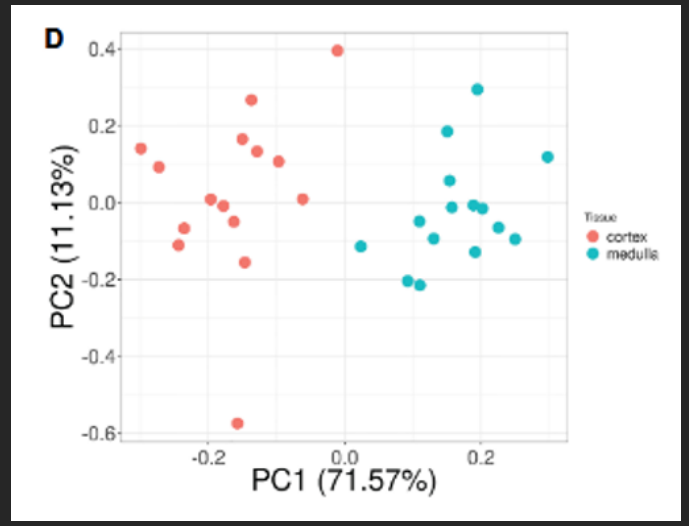
Thirteen different ML approaches were tested using train and test data subsets created from transcriptomic microarray (GSE44762) data from kidney cortex and medulla irradiated with 0.13 – 13 Gy absorbed dose from i.v. injected ^{177}Lu-octreotate over 24 hours including unirradiated controls.
N/A, not available.

Results

- K-nearest neighbors (KNN)
- Dose response markers
- Tissue specific marker
- Radiation markers

- Known & novel radiation markers





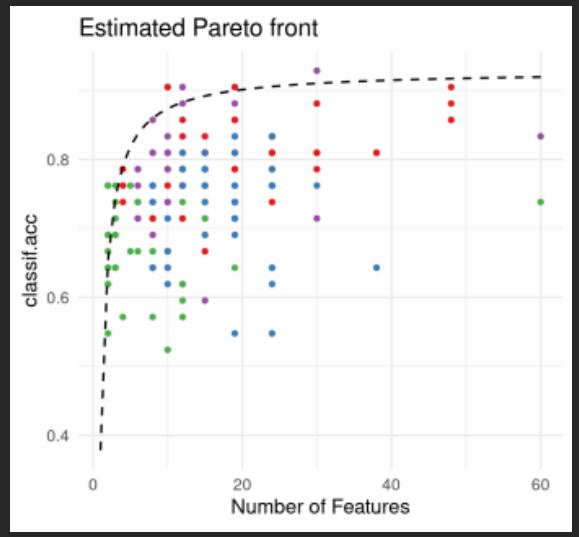
Lesson learned

- Inherit code - worth fixing?
- “One size fits all”?
- More data = more filtering?
- Where is my favorite biomarker? - interchangeable
- Back track results – log file/naming
- “good biomarker”?- clinically relevant vs prediction
- Cohorts are different – different sites
- Cohort bias - unintentionally
- Small dataset – train/test?
- Stable & reproducible models
- Multiple model confirmation markers

What to do today?

- Easier today! – less Nas & normalized
- Assign a PhD student

Workflow today



- Feature selection (- correlated?)
- Caret/tidy models/wrappers
- Fever models (include RF) – check performance scores
- Pareto Plot/Knee Point identification – “optimal trade-off between performance and number of selected features”
- Stability (e.g. Jaccard Index) – which model was the most stable in identifying same predictive features
- Feature Ranking - feature selection frequency to satisfaction approval voting (SAV), models with fewer selected features receive higher weights
- Compare to old workflow