Sparse Bayesian Predictive Modelling of Tumor Response from Radiomic Data

Shirin Golchi

McGill University, Montréal, QC.

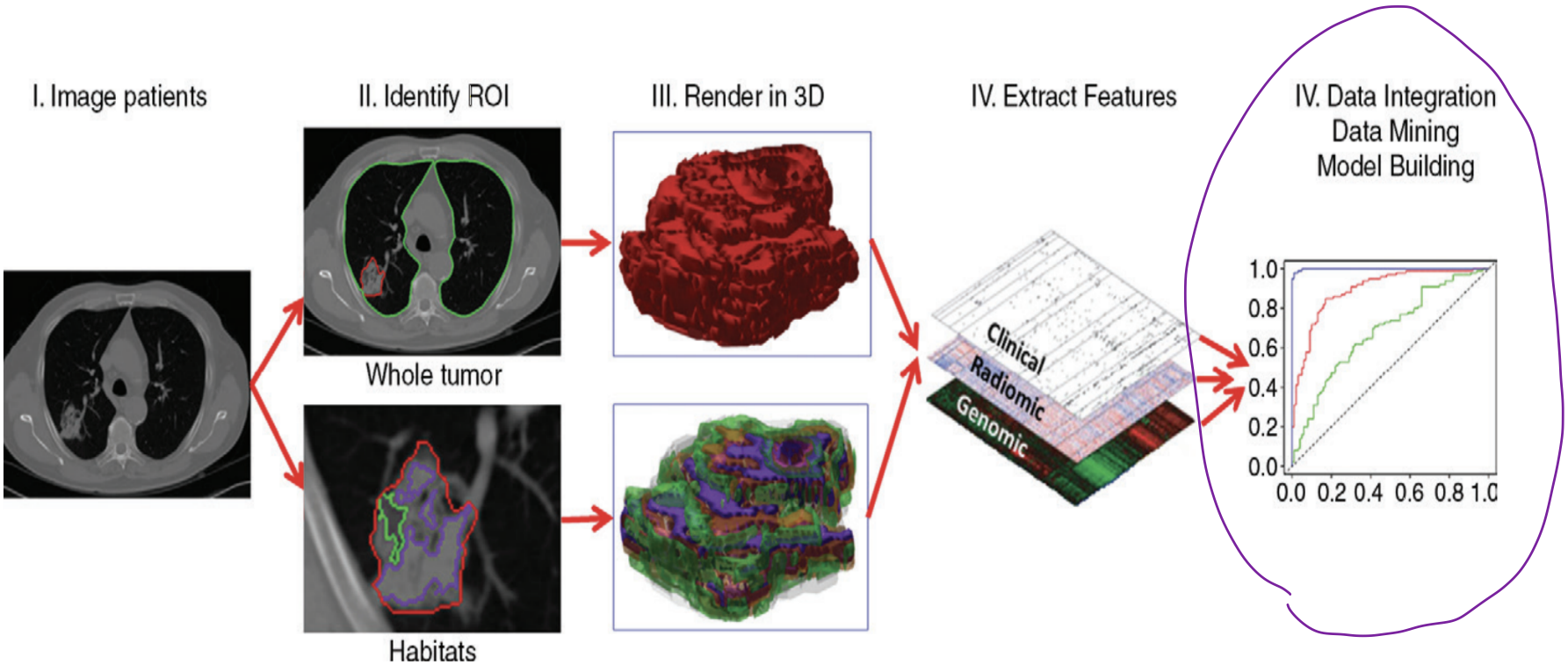
Collaborators: Sahir Bhatnagar, Janet Fu, Reza Forghani

Context and objectives

- Context: radiomic data for characterization of head and neck squamous cell carcinoma (HNSCC)
- A main objective in oncology: creation of a standardized set of criteria to predict tumor response to treatment;
- Radiomic features for characterization of HNSCC in addition to the traditional use of imaging;
- Image analysis algorithms extract mathematically defined features of the tumor's appearance giving rise to high-dimensional matrix covariates

Radiomic features

"Radiomics: Images Are More than Pictures, They Are Data" [Gillies et al., 2015]



Data

- Retrospective study of 605 patients diagnosed with primary pathology proven HNSCC with
- Tumor sites: 164 arising from oral cavity (OC), 200 oropharynx (OP), and 241 from larynx or hypopharynx (LHP)
- Outcome: lymph node metastasis (LN) (+/-)
- An important risk factor: Human Papilloma Virus (HPV) status (+/-)
- Covariates: smoking, drinking, T stage
- Radiomic features: 36

Data structure and challenges

- Dimensionality of the features matrix; only a few may be useful predictors – need efficient variable selection techniques
- Heterogeneity among tumor sites – site-stratified variable selection and inference
- Disproportionately large volume of missing data – variable selection and inference while dealing with missing data

What's currently done

- Random Forest classifiers/Lasso: no uncertainty estimates
- Decision of which features are the "most" important requires an arbitrary threshold
- Site-stratified analyses: not capable of borrowing information
- Imputation of missing values in a pre-processing step: under-representation of uncertainty

The model

- Bayesian hierarchical model for $n = 1, \ldots, N$ patient data,

$$\overbrace{y_{1n}}^{\text{LN}} \sim \text{Bernoulli}(\pi_{1n}), \quad \overbrace{y_{2n}}^{\text{HPV}} \sim \text{Bernoulli}(\pi_{2n})$$

$$\text{logit}(\pi_{1n}) = \phi\pi_{2n} + \mathbf{z}_n\boldsymbol{\eta}_1 + \mathbf{x}_n\boldsymbol{\beta}_{1s_n},$$

$$\text{logit}(\pi_{2n}) = \mathbf{z}_n\boldsymbol{\eta}_2 + \mathbf{x}_n\boldsymbol{\beta}_{2s_n},$$

- $s_n = 1, \ldots, S$ are the tumor sites,
- $\mathbf{z}_n$ are a set of covariates (drinking, smoking and T-stage group)
- $\mathbf{x}_n$ is the $F \times 1$ vector of radiomic features

A word on sparse Bayesian regression

- A useful case study by Michael Betancourt [Betancourt, 2018]
- The direct translation of Lasso penalty to a prior (Laplace) doesn't quite equivalent;
- The Horseshoe prior [Carvalho et al., 2009] solves this by introducing global and a local scale parameters,

$$\beta_j \sim \mathcal{N}(0, \tau^2 \lambda_j^2),$$

$$\lambda_j \sim \mathcal{C}^+(0, 1),$$

$$\tau \sim \mathcal{C}^+(0, \tau_0).$$

- Problem: non-zero slopes are unregularized, can result in nonidentification or weak identification.

The prior

- Feature selection via the regularized Horseshoe prior [Piironen and Vehtari, 2017]

$$\beta_{j,s_n} \sim \mathcal{N}(0, \tau^2_{s_n}\tilde{\lambda}^2_{j,s_n}), \qquad \tilde{\lambda}^2_{j,s_n} = \frac{c^2\lambda^2_{j,s_n}}{c^2 + c^2\lambda^2_{j,s_n}},$$

$$\lambda_{j,s_n} \sim \mathcal{C}^+(0,1),$$

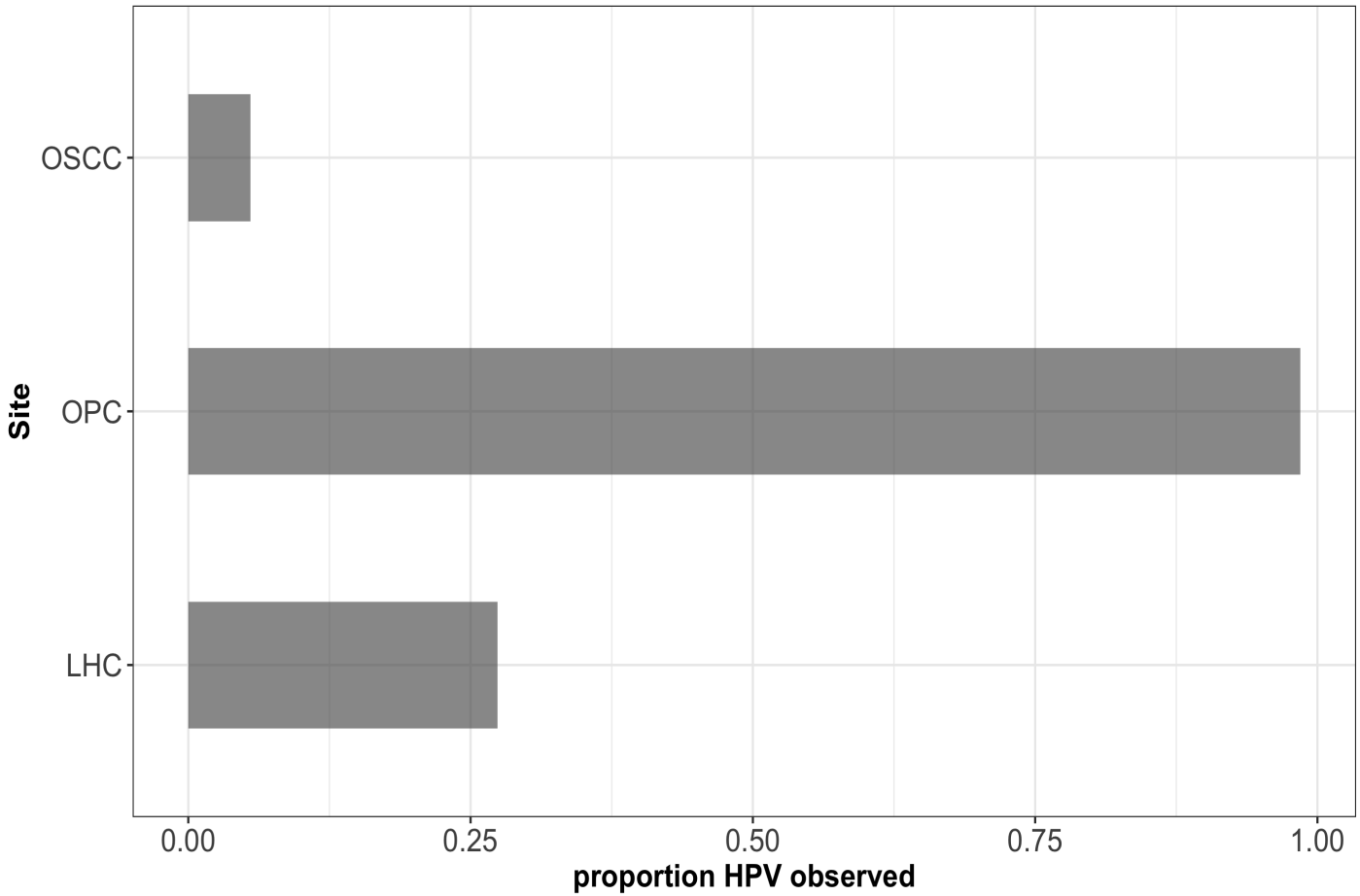
$$c^2 \sim \mathcal{IG}(\frac{\nu}{2}, \frac{\nu}{2}s^2),$$

$$\tau_{s_n} \sim \mathcal{C}^+(0, \tau_0).$$

- Note: the subscripts s_n represent the site-specific variances for β_{j,s_n}.

Model capabilities

– Site-specific feature selection while information is borrowed across sites;



Model capabilities

- The risk associated with missing HPV values are augmented and estimated as parameters;
- Automatically incorporating prediction uncertainty associated with missing HPV values;
- Can improve coverage by 20% according to simulation results.

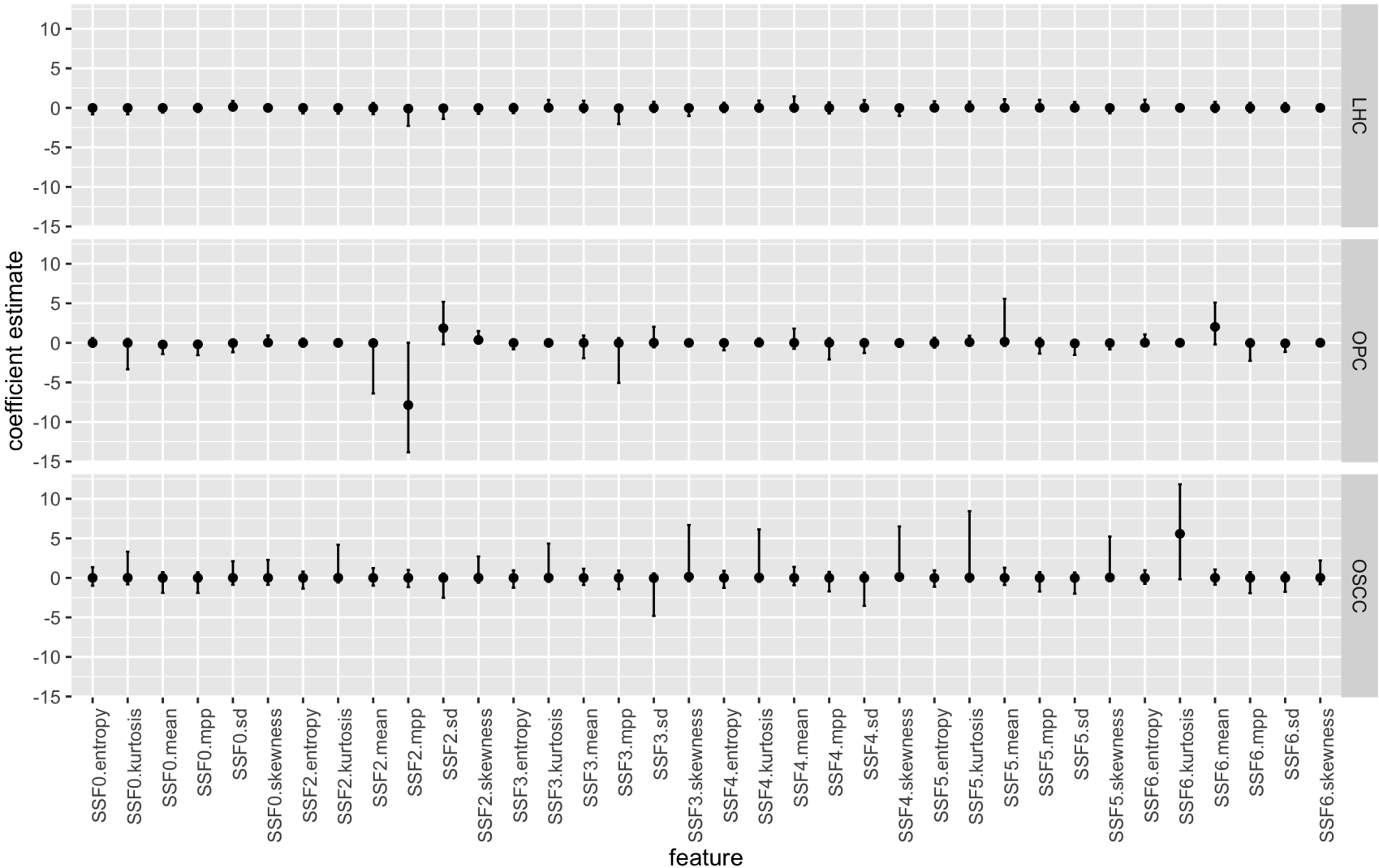
Model capabilities

Bayesian feature selection

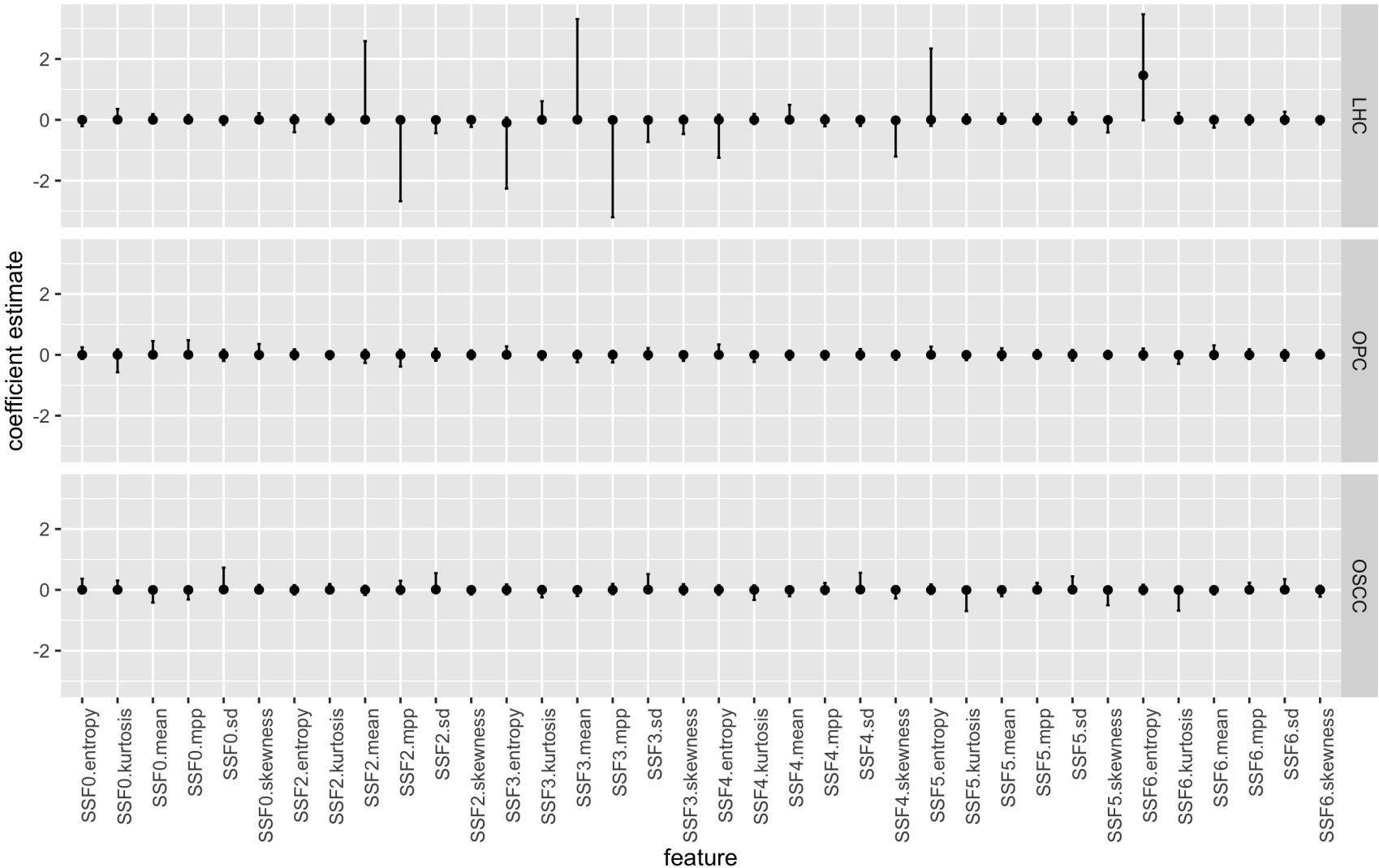
- The Bayesian shrinkage priors provide uncertainty estimates for selected feature coefficients unlike their frequentist counterparts currently used;
- The regularized Horseshoe prior achieves reliable sparse inference.

Multiple steps unified within one hierarchical model implemented in Stan.

Features predictive of LN

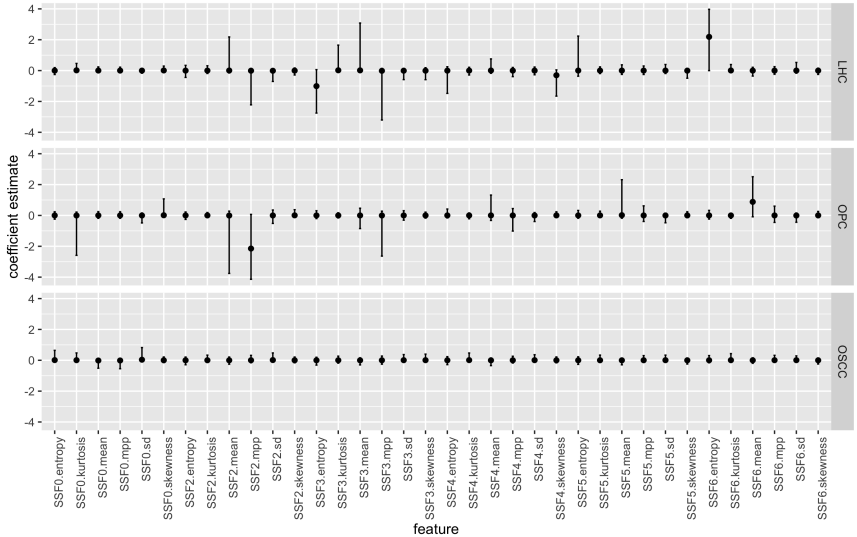


Features predictive of HPV

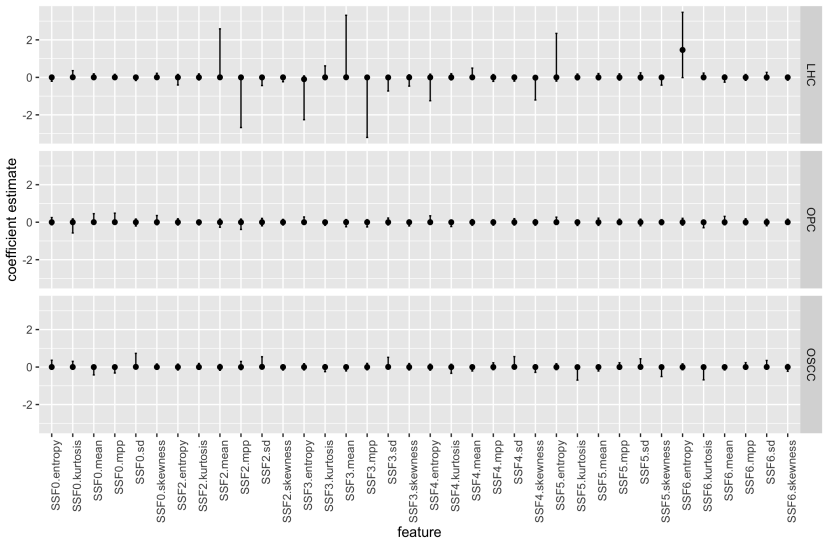


The difference inclusion of HPV can make

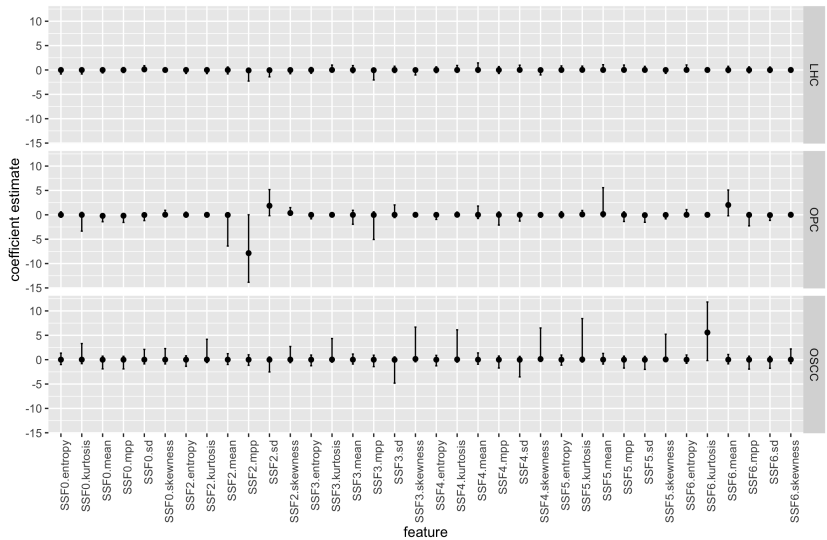
Model without HPV



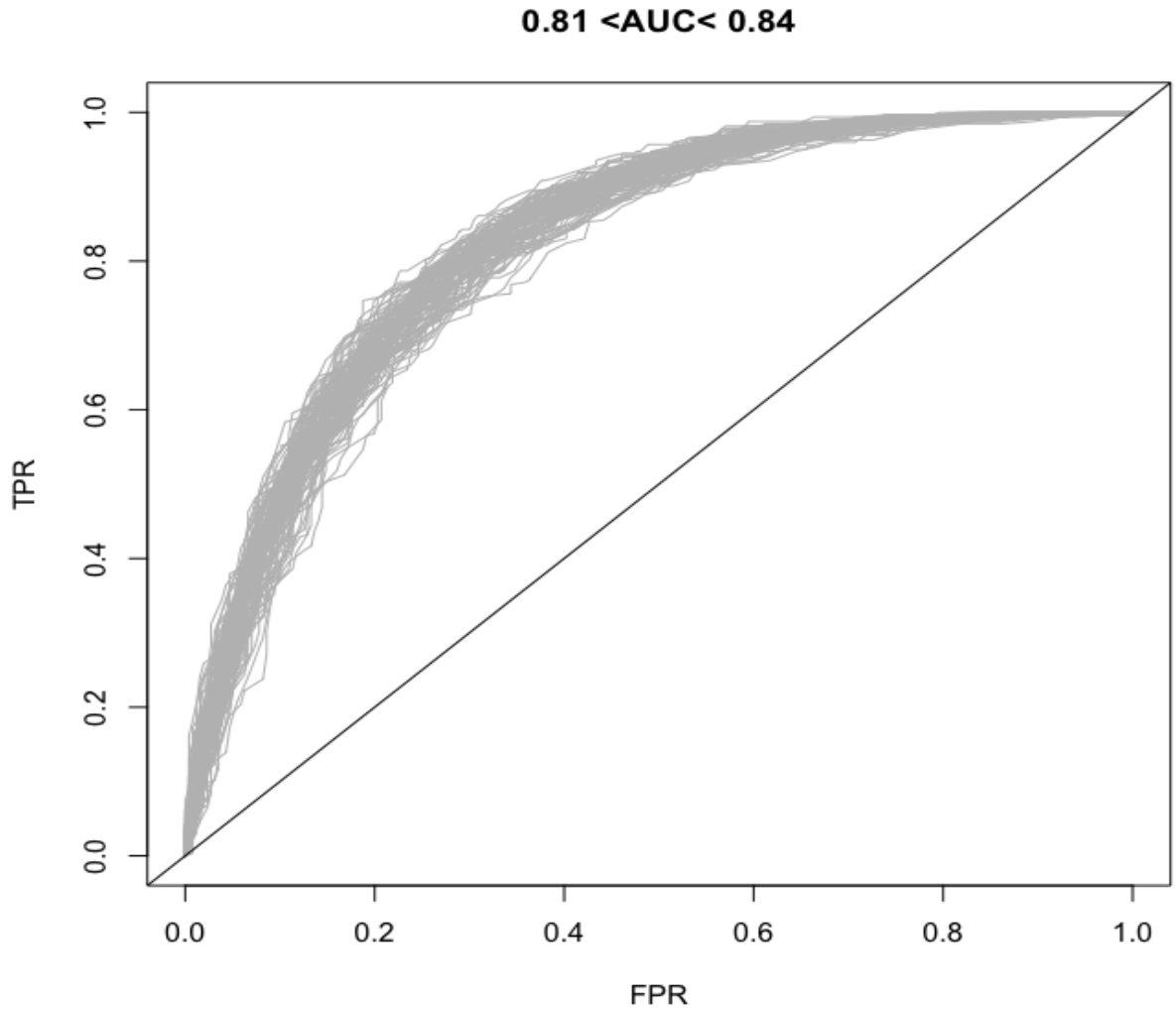
HPV



LN with HPV in the model

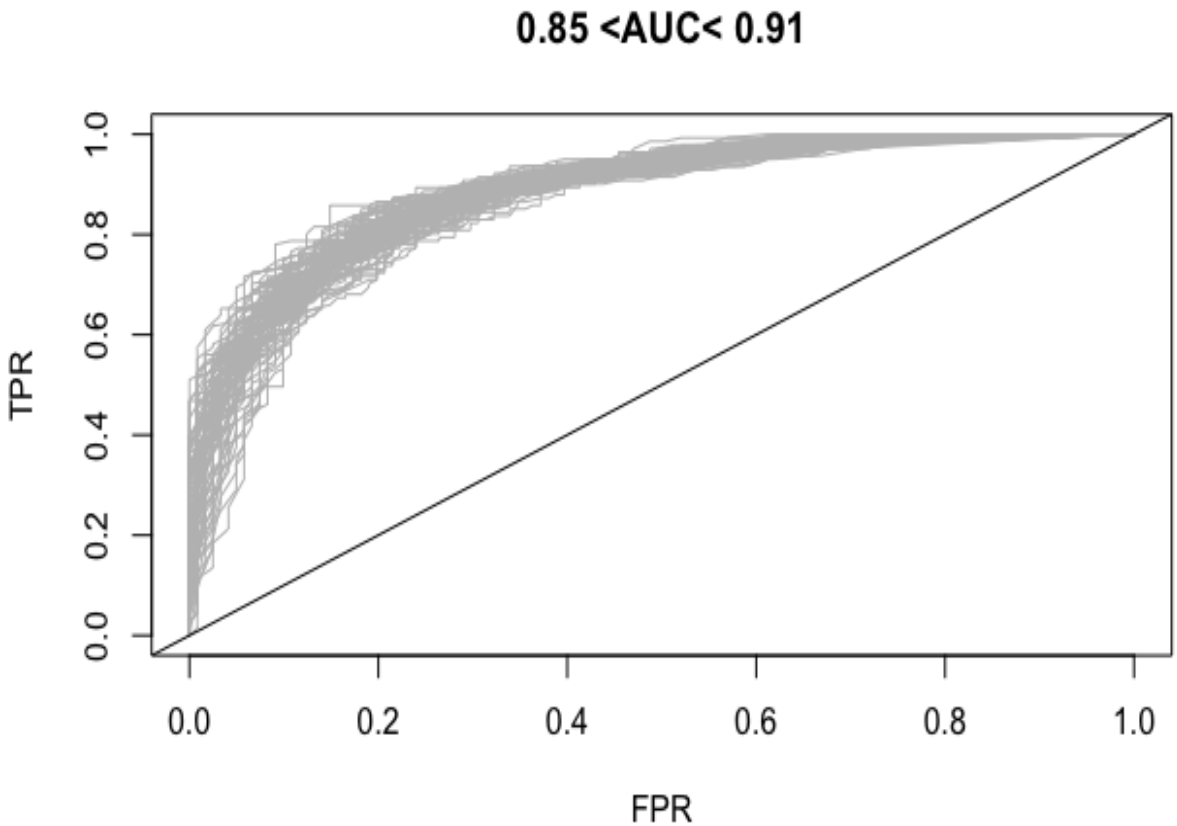


LN prediction Accuracy



The CV AUC obtained from random forest was 0.76.

HPV prediction Accuracy



Work in progress and concluding remarks

Simulation study addressing

- To what extent the integrated inference can improve the results over a multi-step analysis?
- What are the problem settings (dimensionality, correlation structure, etc) that the proposed model works best within?

Main message

- Uncertainty quantification is important even when the main objective is prediction;
- The cost is the computation, especially as the dimensionality increases.

References

Betancourt, M. (2018).

Bayes sparse regression.

https://betanalpha.github.io/assets/case-studies/bayes-sparse-regression.html.

Carvalho, C. M., Polson, N. G., and Scott., J. G. (2009).

Handling sparsity via the horseshoe.

In Proceedings of the Twelth International Conference on Artificial Intelligence and Statistics.

Gillies, R. J., Kinahan, P. E., and Hricak, H. (2015).

Radiomics: Images are more than pictures, they are data.

Radiology, 278(2).

Piironen, J. and Vehtari, A. (2017).

Sparsity information and regularization in the horseshoe and other shrinkage priors.

Electronic Journal of Statistics, 11(2):5018–5051.