


Sparse Bayesian predictive modelling of tumour response using radiomic features

Shirin Golchi¹  | Jingyan Fu¹ | Xiaoyang Liu^{2,3,4} | Eugene Yu^{2,4} |
Reza Forghani⁵ | Sahir Bhatnagar^{1,5}

¹Department of Epidemiology, Biostatistics and Occupational Health, McGill University, Montreal, Quebec, Canada

²Princess Margaret Hospital, University of Toronto, Toronto, Ontario, Canada

³Department of Radiology, Brigham and Women's Hospital, Harvard University, Boston, Massachusetts, USA

⁴Department of Medical Imaging, University of Toronto, Toronto, Ontario, Canada

⁵Department of Diagnostic Radiology, McGill University, Montreal, Quebec, Canada

Correspondence

Shirin Golchi, Purvis Hall 1020 Pine Ave. West
Montreal, QC, Canada H3A 1A2.
Email: shirin.golchi@mcgill.ca

Funding information

Natural Sciences and Engineering Research Council of Canada

We propose a sparse Bayesian hierarchical model for the analysis of data including radiomic features for characterization of head and neck squamous cell carcinoma. The proposed model facilitates radiomic feature selection, handling of missing values in key predictors as well as prediction in a unified framework. The fully Bayesian approach enables adequate incorporation of uncertainty arising from various aspects of the inference and prediction procedure. The prediction performance of the model is assessed via cross validation and compared with two frequentist methods.

KEYWORDS

hierarchical model, horseshoe prior, missing data, radiomics

1 | INTRODUCTION

Medical imaging analysis plays a central role in the personalized treatment decisions and outcome prognosis of cancer patients. For example, in head and neck squamous cell carcinoma (HNSCC), a heterogeneous malignancy constituting more than 95% of head and neck cancers (Liu et al. 2021), clinical management is based on manually measured features such as the tumour size, local disease extension and presence of distant metastasis (Deschler & Day, 2008). Unfortunately, there is a considerable time cost to collecting these semantic features, since they must be quantified by a trained radiologist. There is also potential added value to including agnostic features (e.g., image descriptors not part of the radiologist's lexicon, Savadjiev et al. 2019) to augment the mainly qualitative interpretation currently done towards precision medicine.

With the need for a more scalable approach, there has been renewed interest in extracting large amounts of imaging biomarkers, that is, mathematically defined quantitative descriptors of the tumour which may capture information not currently used by experts (Lambin et al. 2012). Indeed, image-based biomarkers are non-invasive and can quantify regional intratumour heterogeneity at a very detailed level (Aerts, 2016). This automatic extraction of features is referred to as radiomics and has been shown to improve classification accuracy for some cancer types over clinical characteristics alone (Gillies et al. 2016). The main idea is to use a supervised machine learning model to create a radiomic signature which is based on a potentially non-linear combination of the most discriminative features (Savadjiev et al. 2019).

Examples of work in this area include Ji et al. (2020) who implemented a lasso Cox regression using radiomics features to predict tumour recurrence after resection of early stage hepatocellular carcinoma. Vallieres et al. (2017) used a random forest classifier to predict individuals who would respond to treatment for head-and-neck cancer, while Liu et al. (2021) used a similar approach to predict nodal metastasis in patients with HNSCC.

Due to the large number of radiomics features extracted from the computed tomography (CT) images, an initial univariate screening procedure is often employed prior to fitting the classifier (Parmar et al. 2015). A limitation of this approach is that this initial selection is not incorporated into the uncertainty estimates of the prediction accuracy. Another limitation of the current approaches is that when using random forests, feature importance is used as a proxy for significance of a feature. However, the uncertainty associated with the feature importance measures is not accounted for in the prediction process. Every feature has a feature importance, and determining the “most important” features requires specification of an arbitrary threshold. Penalized regression approaches (e.g., lasso) do not directly provide confidence intervals for the predictors, and often a non-zero coefficient estimate is incorrectly equated with statistical significance.

Another challenging aspect of the problem is handling of missing data. Important predictors such as human papillomavirus (HPV) status are commonly missing for a significantly large number of patients. While existing imputation techniques may be employed, accounting for the uncertainty arising from imputation is not straightforward in frequentist-based machine learning models.

Considering the multiple sources of uncertainty in the analysis of such data, a fully Bayesian approach that can accommodate variable selection, missing data handling and information borrowing while accounting for the uncertainty associated with each step is preferred. However, little has been done under the Bayesian framework within the radiomics literature. In a recent review paper of 97 articles with a minimum sample size of 50 patients, the only Bayesian method used was a naive Bayesian network based classifier (Avanzo et al. 2017).

We propose a Bayesian hierarchical model that can address radiomic feature selection and prediction in a unified framework while dealing with complexities such as missing values in predictors. The hierarchical nature of the model enables information borrowing across tumour sites while allowing site-specific variable selection and parameter estimation. Integrating variable selection and missing data handling, together with inference, results in predictions with adequate representation of uncertainty associated with each of these procedures. We present the results of the analysis as Bayesian feature selection outcomes across sites and the accuracy for predictions of lymph node metastasis.

The remaining of the paper is organized as follows. In Section 2 we describe the data structure and variables and the statistical analysis that is currently used. In Section 3, we proposed the Bayesian hierarchical model and the regularized shrinkage prior. Section 4 follows with the results of applying the proposed model to the data set and a discussion is provided in Section 5.

2 | DATA

In the following we describe the specific data set that is used to illustrate the capabilities of the proposed model. The main characteristics of this data set include the presence of a large number of radiomics features, the stratification of patient data with respect to tumour sites and the disproportionate amount of missing values of an important predictor; these characteristics are common among data arising in similar context in oncology.

The data comprise 603 contrast enhanced pre-treatment neck CT scans evaluated from patients diagnosed with HNSCC, with tumours arising in three sites: 241 from the larynx or hypopharynx (LHC), 162 oral cavity (OSCC), and 200 oropharynx (OPC), further stratified based on the HPV status to avoid its confounding effects. The tumour region in each CT scan was first manually segmented by an expert radiologist. First order texture features with additional filtrations were then extracted from each segmentation using the TexRAD software (TexRAD; University of Sussex, Falmer, England) and used in conjunction with patient age, smoking status, drinking status, and tumour T-stage to construct models for predicting lymph node metastasis. Examples of tumour segmentation with region of interest (ROI) drawn using TexRad are illustrated in Figure 1 for each primary anatomical location of larynx or hypopharynx, oral cavity and oropharynx.

Many challenges arise from the structure of radiomics data. The number of radiomics features is often very large but only a few are expected to be significant predictors of the tumour response. Therefore, an efficient and reliable variable selection technique is required to select a reasonable number of radiomics features to be used in prediction of key outcomes such as lymph node metastasis. Variable selection and prediction should reflect the heterogeneity among tumour sites which has been observed at the histological level and is associated with adverse clinical outcomes (Sala et al. 2017). In addition, there is a considerable amount of missing data among important predictors such as the HPV status.

The proportion of missing HPV values is much higher among patients with specific tumour sites which further complicates a stratified analyses. Specifically, HPV status is missing in 55% of the sample, with 175 missing from the LHC, 3 from the OPC and 155 from the OSCC, that is, about 95% of the total observations with site OSCC (see Figure 2). Note that HPV is considered an important predictor for lymph node metastasis (Liu et al., 2021). Therefore, given the disproportionate missingness structure, an independent site-stratified inference and prediction can result in severe loss of information and low statistical power.

3 | MODEL

As discussed above, the radiomics data structure creates unique challenges. These include the size of the feature matrix, the need for site-specific analysis and the disproportionately large amount of missing data in one of the sites. Therefore, an appropriate analysis should include an effective

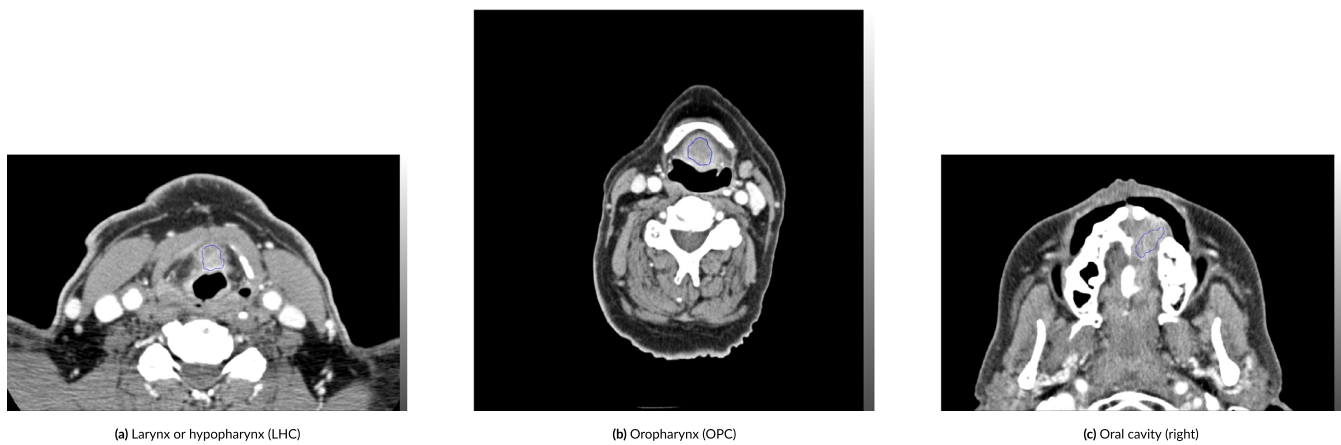


FIGURE 1 Examples of tumour segmentation using manually placed region of interest (ROI) in three lesions of squamous cell carcinoma: larynx or hypopharynx (left), oropharynx (middle) and oral cavity (right). The radiomic features are computed from the segmented region in the image (outlined in a solid blue line), which defines the spatial extent of the tumour

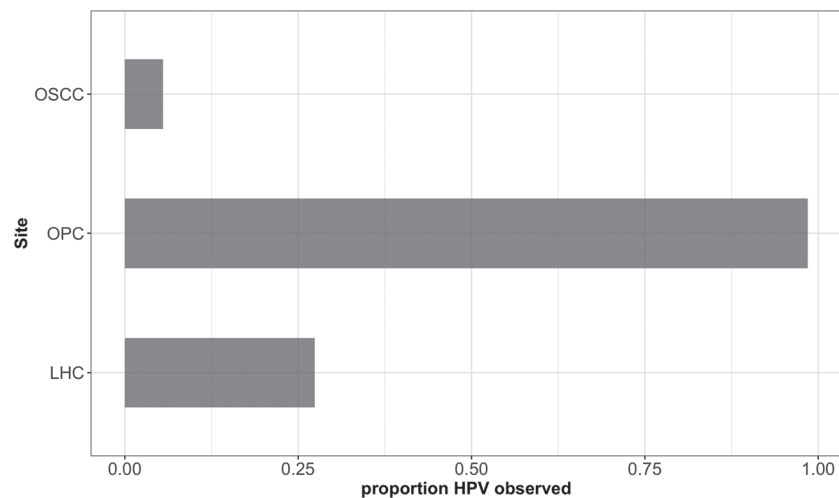


FIGURE 2 Proportion of observed human papillomavirus (HPV) status by tumour site

and efficient variable selection technique to select the features predictive of the outcome; it should be capable of performing variable selection separately within sites while borrowing information across sites; and it should deal with missing data such that the uncertainty associated with the handling of missing values is well-represented in the predictions.

The model presented in this section is a Bayesian hierarchical model with components that deal with the above mentioned challenges. Missing data are handled by augmenting missing values as unknown parameters that are estimated together with the rest of the model parameters and the estimation uncertainty is automatically incorporated into the predictions.

3.1 | Site specific inference with information borrowing

The following Bayesian hierarchical model is proposed. Let y_{1n} and y_{2n} denote the binary outcomes lymph node metastasis and HPV for patient $n = 1, \dots, N$, respectively.

$$y_{1n} \sim \text{Bernoulli}(p_{1n}), \quad y_{2n} \sim \text{Bernoulli}(p_{2n})$$

where,

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

where p_{2n} is the risk of HPV for patient n that is in turn modelled as

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

where $s_n = 1, \dots, S$ are the tumour sites, \mathbf{z}_n are a set of C covariates and \mathbf{x}_n is the vector of F radiomic features for patient n . The reasoning behind the above model is that the same set of covariates and features are expected to be predictors of lymph node metastasis and HPV. In addition, HPV is considered an important predictor for lymph node metastasis. The proposed model aims at capturing these relationships in the simplest way possible while the dependence of $\text{logit}(p_{2n})$ on the same set of variables has the added benefit of capturing non-linear effects of variables on lymph node metastasis.

Allowing for feature selection to be performed separately across the three sites introduces $F \times S$ coefficients. The notation $\boldsymbol{\beta}_{s_n}$ is used to represent the site-specific set of the radiomics feature coefficients matrix.

Note that, the above model implies that a feature may be a direct or indirect predictor for lymph node metastasis. However, the goal of the model is to predict lymph node metastasis via direct or indirect predictors and interpretation of selected features under either model is not of interest.

Inference is made through the joint posterior of model parameters, $(\phi, \boldsymbol{\eta}_1, \boldsymbol{\eta}_2, \boldsymbol{\beta}_1, \boldsymbol{\beta}_2)$, where $\boldsymbol{\beta}_1$ and $\boldsymbol{\beta}_2$ include all feature coefficients across the three sites.

$$\begin{aligned} \pi(\phi, \boldsymbol{\eta}_1, \boldsymbol{\eta}_2, \boldsymbol{\beta}_1, \boldsymbol{\beta}_2 | \mathbf{y}_1, \mathbf{y}_2, \mathbf{X}, \mathbf{Z}) &\propto \pi(\mathbf{y}_1 | \phi, \boldsymbol{\eta}_1, \boldsymbol{\beta}_1, \boldsymbol{\eta}_2, \boldsymbol{\beta}_2, \mathbf{X}, \mathbf{Z}) \\ &\quad \pi(\mathbf{y}_2 | \boldsymbol{\eta}_2, \boldsymbol{\beta}_2, \mathbf{X}, \mathbf{Z}) \\ &\quad \pi(\phi, \boldsymbol{\eta}_1, \boldsymbol{\eta}_2, \boldsymbol{\beta}_1, \boldsymbol{\beta}_2), \end{aligned}$$

where the first two terms on the right hand side constitute the likelihood; \mathbf{y}_1 is the vector of lymph node metastasis outcomes that depends on $(\phi, \boldsymbol{\eta}_1, \boldsymbol{\beta}_1)$ through the associated risks \mathbf{p}_1 , and on $(\boldsymbol{\eta}_2, \boldsymbol{\beta}_2)$ through the HPV risks \mathbf{p}_2 . Likewise, \mathbf{y}_2 is the vector of HPV outcomes that depends only on $(\boldsymbol{\eta}_2, \boldsymbol{\beta}_2)$ through \mathbf{p}_2 . The third term on the right hand side is the prior which can be written as the product of the prior distributions over individual model parameters;

$$\pi(\phi, \boldsymbol{\eta}_1, \boldsymbol{\eta}_2, \boldsymbol{\beta}_1, \boldsymbol{\beta}_2) \propto \pi(\phi) \pi(\boldsymbol{\eta}_1) \pi(\boldsymbol{\eta}_2) \pi(\boldsymbol{\beta}_1) \pi(\boldsymbol{\beta}_2).$$

The prior distributions for $\phi, \boldsymbol{\eta}_1$ and $\boldsymbol{\eta}_2$ are defined as weakly informative normal distributions. The prior distributions of $\boldsymbol{\beta}_1$ and $\boldsymbol{\beta}_2$ are specified as continuous shrinkage priors that are described in the following section.

3.2 | Feature selection via regularized horseshoe prior

As explained earlier, radiomics feature selection is an important aspect of the problem. While the current data set includes only 36 radiomics features, the size of the feature matrix can be much larger in other similar data sets. It is therefore crucial to be able to identify the few features that are useful predictors of the outcome. Another interesting aspect of the problem is that the predictive features can be different across tumour sites.

Continuous shrinkage priors (Polson & Scott, 2011) are commonly used to perform sparse regression under the Bayesian framework. One of the most popular shrinkage priors is the horseshoe prior due to its desirable theoretical properties and its good performance in most problems (Carvalho et al. 2009; 2010). Suppose that $\theta_j, j = 1, \dots, J$ represent a large set of model parameters, only a few of which are expected to be non-zero. The horseshoe prior includes a two-component variance for the model parameters:

$$\begin{aligned} \theta_j &\sim \mathcal{N}(0, \tau^2 \lambda_j^2) \\ \lambda_j &\sim \mathcal{C}^+(0, 1), \end{aligned}$$

where τ is a global parameter that controls the shrinkage towards zero and λ_j are local scales that allow some of the coefficients to deviate from zero.

As Piironen and Vehtari (2017) discuss, however, the horseshoe prior has two main shortcomings - first, lack of a recipe for specifying τ or defining a prior for it to determine the shrinkage level; second, lack of regularization of the non-zero parameter estimates that can cause convergence issues in case of weak identifiability. The latter issue is in fact a concern in the present application where data are sparse in some sites, specially as the number of features increases.

Therefore, we use the regularized horseshoe priors proposed by Piironen and Vehtari (2017), but allow for site-specific hyper-parameters to perform site-specific radiomics feature selection while borrowing information across sites. The coefficients, $\beta_{f_{sn}}, f = 1, \dots, F, s_n = 1, 2, 3$, of the radiomics features are assigned the following prior distribution,

$$\beta_{f_{sn}} \sim \mathcal{N}(\mathbf{0}, \tau^2 \tilde{\lambda}_{f_{sn}}^2)$$

$$\tilde{\lambda}_{f_{sn}}^2 = \frac{c^2 \lambda_{f_{sn}}^2}{c^2 + \tau^2 \lambda_{f_{sn}}^2}$$

where

$$\begin{aligned} \lambda_{f_{sn}} &\sim \mathcal{C}^+(\mathbf{0}, 1), \\ c^2 &\sim \text{IG}\left(\frac{\nu}{2}, \frac{\nu s^2}{2}\right), \\ \tau &\sim \mathcal{C}^+(\mathbf{0}, \tau_0). \end{aligned} \quad (1)$$

Similar to the horseshoe priors, the shrinkage towards zero is imposed by the global parameter τ which is shared across sites as there is no reason to define varying shrinkage parameters under different sites. The parameters $\lambda_{f_{sn}}$, however, are local scales that allow different sets of coefficients to be far from zero under different sites. The regularization is defined by an additional level of hierarchy over the local scales which prevents the non-zero estimates from taking extremely large values; for a given value of c^2 , if $\tau^2 \lambda_{f_{sn}}^2 > c^2$, the corresponding coefficient is non-zero and the prior approaches a Gaussian distribution with variance c^2 . The prior will therefore behave like the horseshoe for large values of c^2 while small c^2 imposes more regularization. In absence of information that could be used to specify c^2 , Piironen and Vehtari (2017) recommend placing a prior on c^2 as in (1) which when integrated over c^2 results in a Student- $t_\nu(0, s^2)$ distribution over the $\beta_{f_{sn}}$.

3.3 | The handling of missing data

As mentioned earlier, a considerable portion of the patients have missing HPV outcomes. The percentage of missing HPV values is disproportionately high in two of the sites (see Figure 2). We approach inference in the presence of missing data in a fully Bayesian framework, that is, by augmenting the HPV data with the unknown values of HPV and estimating the missing values together with the rest of the model parameters.

Since posterior sampling is performed in Stan where it is currently not possible to sample discrete parameters, the missing HPV values are marginalized out of the posterior distribution. The posterior uncertainty of the risks associated with missing values is then automatically incorporated into the estimation of lymph node metastasis regression parameters and predictions.

4 | ANALYSIS RESULTS

We apply the proposed model to the data set described in Section 2. The results are presented as the posterior estimates for the model parameters together with 95% credible intervals; as well as summary of prediction performance.

The first set of model parameters that we focus on are the coefficients of radiomics features, β 's, in the HPV and lymph node metastasis model. Figures 3 and 4 show the point estimates and 95% credible intervals of these parameters by tumour site for the HPV and lymph node metastasis, respectively. The greater estimation uncertainty for the coefficient estimates of the HPV model associated with the OSCC site reflects the high portion of missing values under this site. Overall, only a handful of radiomics features are selected by the shrinkage prior to predict lymph node metastasis and HPV. The selected features vary by tumour site.

Figure 5 shows point estimates (posterior median) and 95% credible intervals for the covariate coefficients in the lymph node metastasis model representing the log odds ratio. For drinking and smoking variables, "heavy drinking" and "current smoking" are selected as reference categories, respectively. According to these estimates only non-drinkers have a significantly lower odds of lymph node metastasis in comparison to heavy drinkers. HPV status is clearly a significant predictor; positive HPV status appears to be associated with three times higher odds of lymph node metastasis. Finally, the patients in the first two tumour T-stage groups have a significantly lower risk of lymph node metastasis.

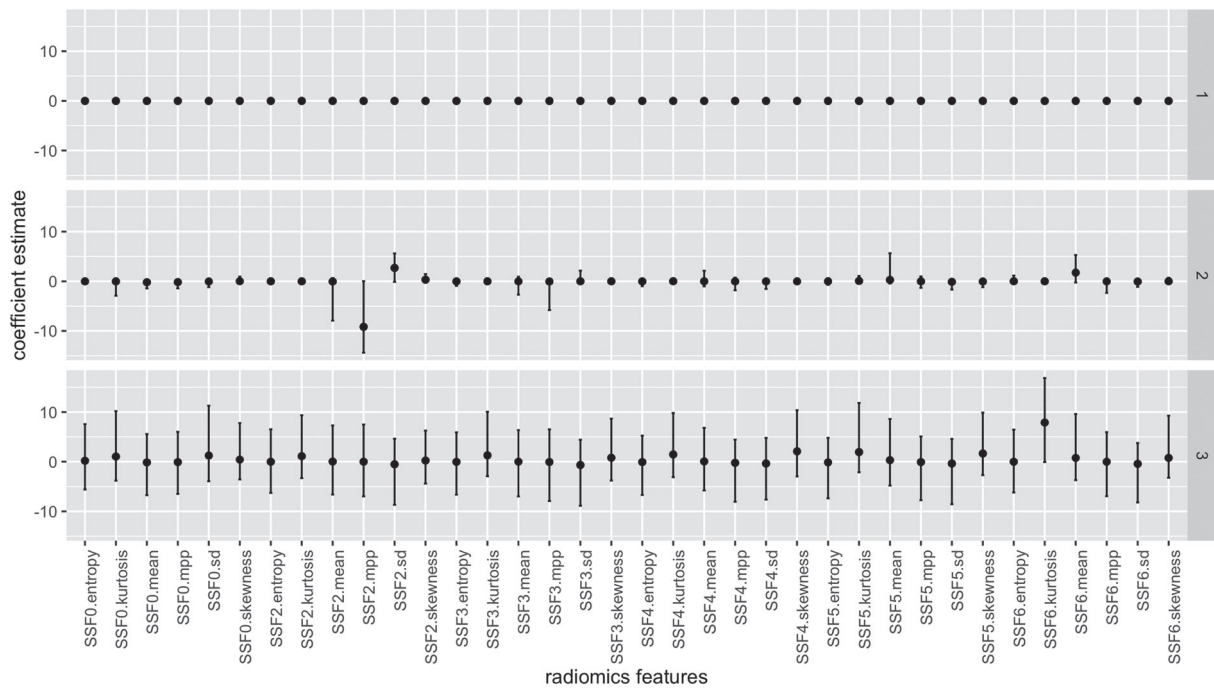


FIGURE 3 Site-specific point estimates and 95% credible intervals for the radiomics feature coefficients in the predictive model for human papillomavirus (HPV)

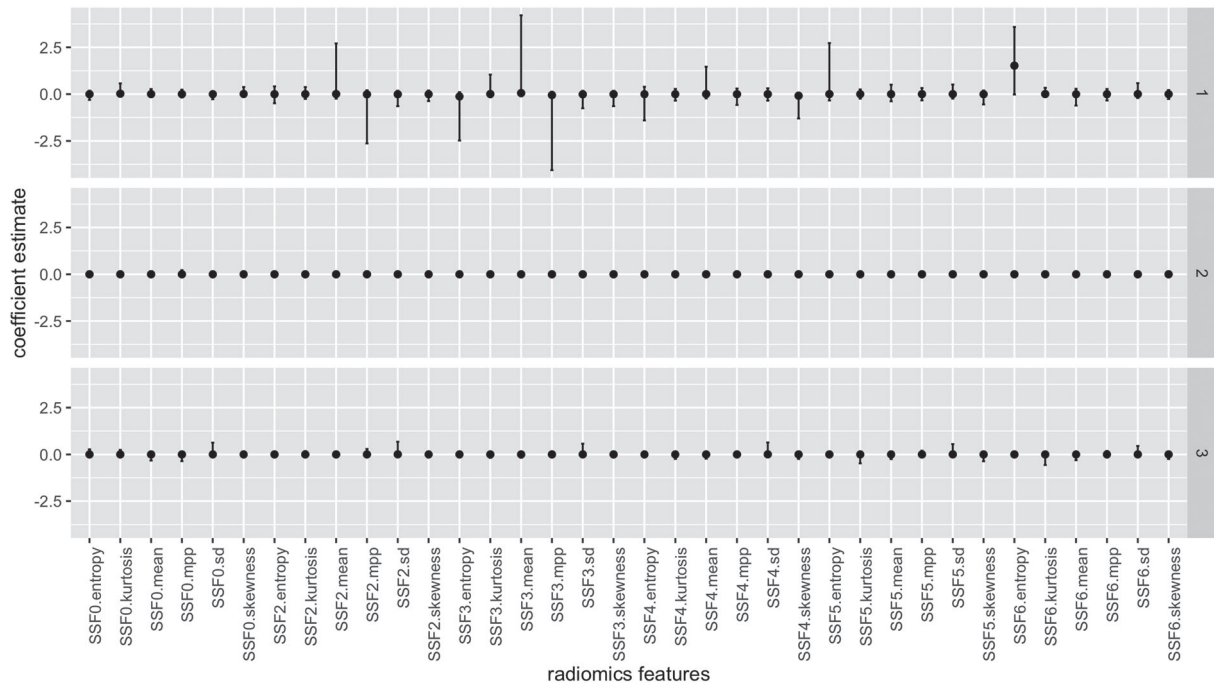


FIGURE 4 Site-specific point estimates and 95% credible intervals for the radiomics feature coefficients in the predictive model for lymph node metastasis

The most important aspect of the model, however, is its prediction performance which is illustrated in Figures 6 in form of Receiver Operating Characteristic (ROC) curves that correspond to 1000 draws from the posterior predictive distribution, representing the posterior uncertainty in the prediction performance. The ROC curves are obtained by a 50-fold cross validation procedure. Specifically, the data are randomly partitioned

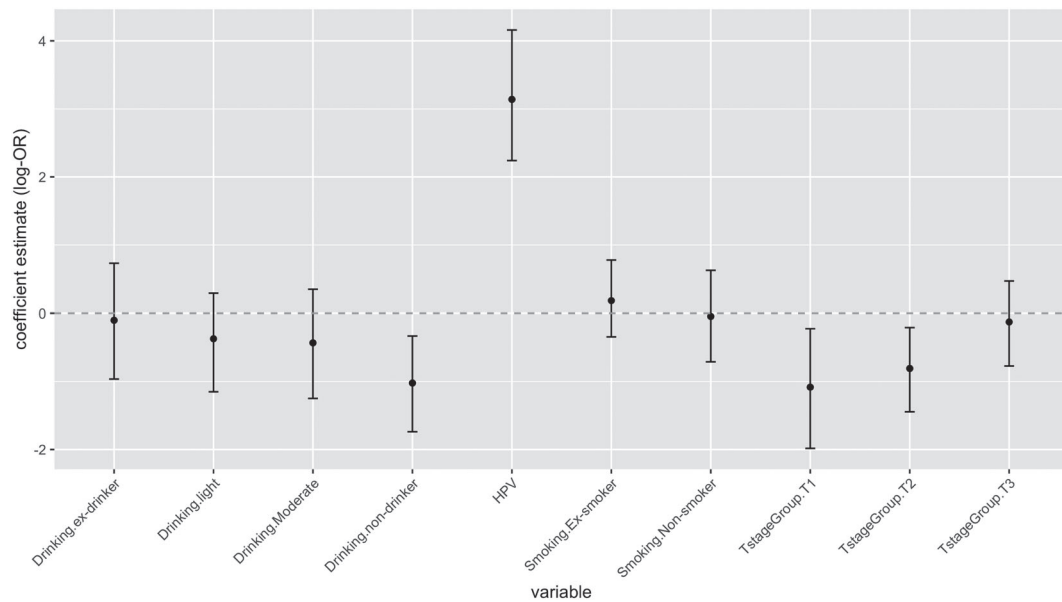


FIGURE 5 Point estimates (posterior median) and 95% credible intervals for the covariate coefficients in the lymph node metastasis model representing the log odds ratio. For drinking and smoking variables, “heavy drinking” and “current smoking” are selected as reference categories, respectively

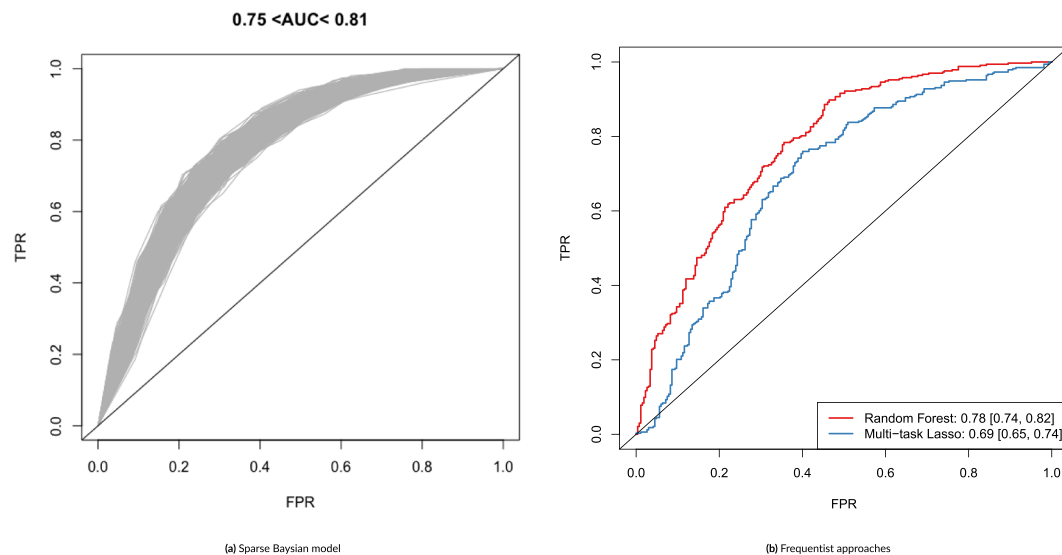


FIGURE 6 Comparison of the cross-validated Receiver Operating Curves for predicting lymph node metastasis. FPR: false positive rate; TPR: true positive rate. (a) Posterior draws using our proposed sparse Bayesian model. The AUC varies from 0.75 to 0.81. (b) The cross-validated AUC [95% CI] is 0.78 [0.74, 0.82] for Random Forest, and 0.69 [0.65, 0.74] for the Multi-task Lasso. Confidence intervals were calculated using the DeLong approach (DeLong et al. 1988)

into 50 disjoint subsets such that the proportion of observed to missing HPV status remains approximately the same as that in the full data set in all subsets. This, in turn, guarantees similar missingness structure in the training sets. At every cross validation iteration one of the 50 subsets is held out as a test set while the other 49 are used to train the model and predict the lymph node metastasis for patients in the test set. The 50 predicted subsets are then combined and compared to the observed data resulting in the ROC curves that show the percentages of false positive and true positive predictions for a sequence of prediction thresholds. The Area Under the Curve (AUC) corresponding to this sample of ROC curves varies between 75% and 81% and is reported in the title of the graph as a summary measure of prediction performance.

4.1 | Comparison with two frequentist methods

We compared our approach with two frequentist methods: (1) the Multi-task Lasso (Caruana, 1997) and (2) Random forests (Breiman, 2001). The Multi-task Lasso extends the lasso (Tibshirani, 1996) to borrow information across a set of different learning tasks. This approach allows the effect of each feature to vary by task while controlling the overall sparsity of the model. In our application, we used the three tumour sites as different tasks along with the exact same radiomics features and covariates described in Section 2. The model was fit using the RMTL package version 0.9 (Cao et al. 2019). We also applied random forests to our data which is a machine learning approach that can account for nonlinearity and different levels of interactions between features. Since there is no explicit modelling of site specific effects, we included the tumour site in the model (in addition to radiomics features and covariates). The model was fit using the ranger package version 0.12.1 (Wright & Ziegler, 2017). Note that both of these approaches do not allow for missing values; therefore, we performed single imputation using the mice package version 3.11.0 (Van Buuren & Groothuis-Oudshoorn, 2011). We report the prediction performance using the same cross-validation approach in Figure 6b. Confidence intervals were calculated using the DeLong approach (DeLong et al. 1988). The cross-validated AUC [95% CI] is 0.69 [0.65, 0.74] for the Multi-task Lasso, and 0.78 [0.74, 0.82] for Random Forests. While the performance for Random Forests is similar to our proposed approach, we note that these confidence intervals do not account for the imputation step and thus are likely to be overoptimistic. This also suggests that the proposed functional form of the relationship between the response and features is reasonable.

5 | DISCUSSION

We have proposed a Bayesian hierarchical predictive model to analyse complex and high dimensional radiomics features together with a set of covariates to predict lymph node metastasis in HNSCC. The novelty of the proposed model is in dealing with multiple challenging issues within a unified framework. Namely, variable selection, missing data imputation and prediction are performed simultaneously within the proposed model.

The main advantage of integrating variable (feature) selection and missing data handling together with estimation and prediction, over existing ad-hoc multi-step procedures, is adequate representation of uncertainty in the final results. The two main sources of uncertainty are the feature selection step and imputation of missing data. Both of these layers of uncertainty are ignored in the machine learning procedures currently in use due to lack of a formal mechanism for incorporating the uncertainty.

The performance of the model is assessed through a 50-fold cross-validation which partitions the data set into 50 subsets that have similar proportions of missing to observed HPV status. The reason for the specific partitioning method is that the proportion of missing to observed data can drastically affect the estimation and prediction uncertainty. This, of course, means that for a different data set with different missingness structure the model may have a better or worse prediction performance.

We compared the prediction performance of our sparse Bayesian model with the Multi-task Lasso and Random Forests. We performed single imputation on the data set because both frequentist methods required complete data. The Multi-task Lasso performed significantly worse, while Random Forest had a similar AUC to our approach. These results suggest that a simple linear relationship between the response and features is comparable to a more complex one. A limitation of both methods is that there is no clear way to incorporate uncertainty of the imputation procedure into parameter estimation, likely leading to overoptimistic measures of uncertainty. Furthermore, Random Forests do not provide site specific effect estimates; they give only feature importance measures which require arbitrary thresholds for determining significance.

As stated earlier, the proposed model is expected to handle high-dimensional feature matrices. This is not demonstrated via the present application, however, since the feature matrix includes only 36 radiomics features. To verify the capability of the model, we artificially generated additional 164 features as select two-way interactions of the 36 features, resulting in 200 features in total, and assessed the convergence as well as prediction performance of the model that incorporated these additional features. The posterior AUC of this model was between 0.81 and 0.87 pointing to a satisfactory prediction performance. Note, however, that the higher AUC in comparison to that presented in the paper is due to the fact that these are not cross-validated due to higher computational burden of a model with a larger number of features.

ACKNOWLEDGEMENTS

Golchi and Bhatnagar gratefully acknowledge funding via a Discovery Grant from the Natural Sciences and Engineering Research Council of Canada (NSERC).

DATA AVAILABILITY STATEMENT

The data and scripts to reproduce the results in this article are publicly available at <https://github.com/sahirbhatnagar/radbayes>.

ORCID

Shirin Golchi  <https://orcid.org/0000-0003-3382-9563>

REFERENCES

- Aerts, H. J. W. L. (2016). The potential of radiomic-based phenotyping in precision medicine: A review. *JAMA Oncology*, 2(12), 1636–1642.
- Avanzo, M., Stancanello, J., & El Naqa, I. (2017). Beyond imaging: The promise of radiomics. *Physica Medica*, 38, 122–139.
- Breiman, L. (2001). Random forests. *Machine Learning*, 45(1), 5–32.
- Cao, H., Zhou, J., & Schwarz, E. (2019). Rmtl: An R library for multi-task learning. *Bioinformatics*, 35(10), 1797–1798.
- Caruana, R. (1997). Multitask learning. *Machine Learning*, 28(1), 41–75.
- Carvalho, C. M., Polson, N. G., & Scott, J. G. (2009). Handling sparsity via the horseshoe. In *In Proceedings of the 12th International Conference on Artificial Intelligence and Statistics (D. van Dyk and M. Welling, eds.)*. *Proceedings of Machine Learning Research*, 5, pp. 73–80.
- Carvalho, C. M., Polson, N. G., & Scott, J. G. (2010). The horseshoe estimator for sparse signals. *Biometrika*, 97, 465–480.
- DeLong, E. R., DeLong, D. M., & Clarke-Pearson, D. L. (1988). Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. *Biometrics*, 44(3), 837–845.
- Deschler, D. G., & Day, T. (2008). Tnm staging of head and neck cancer and neck dissection classification. *American Academy of Otolaryngology-Head and Neck Surgery Foundation*, 10–23.
- Gillies, R. J., Kinahan, P. E., & Hricak, H. (2016). Radiomics: Images are more than pictures, they are data. *Radiology*, 278(2), 563–577.
- Ji, G.-W., Zhu, F.-P., Xu, Q., Wang, K., Wu, M.-Y., Tang, W.-W., Li, X.-C., & Wang, X.-H. (2020). Radiomic features at contrast-enhanced CT predict recurrence in early stage hepatocellular carcinoma: a multi-institutional study. *Radiology*, 294(3), 568–579.
- Lambin, P., Rios-Velazquez, E., Leijenaar, R., Carvalho, S., Van Stiphout, R. u.d.G. P. M., Granton, P., Zegers, C.M.L., Gillies, R., Boellard, R., Dekker, A., & Aerts, H. J. (2012). Radiomics: Extracting more information from medical images using advanced feature analysis. *European Journal of Cancer*, 48(4), 441–446.
- Liu, X., Maleki, F., Muthukrishnan, N., Ovens, K., Huang, S. H., Pérez-Lara, A., Romero-Sanchez, G., Bhatnagar, S. R., Chatterjee, A., Pusztaszeri, M. P., Spatz, A., Batist, G., Payabvash, S., Haider, S., Mahajan, A., Reinhold, C., Forghani, B., O'Sullivan, B., Yu, E., & Forghani, R. (2021). Site-specific variation in radiomic features of head and neck squamous cell carcinoma and its impact on machine learning models. *Cancers*, 13, 3723. <https://www.mdpi.com/2072-6694/13/15/3723#cite>
- Parmar, C., Grossmann, P., Bussink, J., Lambin, P., & Aerts, H. J. W. L. (2015). Machine learning methods for quantitative radiomic biomarkers. *Scientific Reports*, 5(1), 1–11.
- Piironen, J., & Vehtari, A. (2017). Sparsity information and regularization in the horseshoe and other shrinkage priors. *Electronic Journal of Statistics*, 11, 5018–5051.
- Polson, N. G., & Scott, J. G. (2011). Shrink globally, act locally: Sparse bayesian regularization and prediction. In *Bayesian statistics 9 (J. M. Bernardo, M. J. Bayarri, J. O. Berger, A. P. Dawid, D. Heckerman, A. F. M. Smith and M. West, eds.)* 501–538.
- Sala, E., Mema, E., Himoto, Y., Veeraraghavan, H., Brenton, J. D., Snyder, A., Weigelt, B., & Vargas, H. A. (2017). Unravelling tumour heterogeneity using next-generation imaging: radiomics, radiogenomics, and habitat imaging. *Clinical Radiology*, 72(1), 3–10.
- Savadjiev, P., Chong, J., Dohan, A., Agnus, V., Forghani, R., Reinhold, C., & Gallix, B. (2019). Image-based biomarkers for solid tumor quantification. *European Radiology*, 29(10), 5431–5440.
- Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B (Methodological)*, 58(1), 267–288.
- Vallieres, M., Kay-Rivest, E., Perrin, L. J., Liem, X., Furstoss, C., Aerts, H. J. W. L., Khaouam, N., Nguyen-Tan, P. F., Wang, C.-S., Sultanem, K., & Seuntjens, J. (2017). Radiomics strategies for risk assessment of tumour failure in head-and-neck cancer. *Scientific Reports*, 7(1), 1–14.
- Van Buuren, S., & Groothuis-Oudshoorn, K. (2011). mice: Multivariate imputation by chained equations in r. *Journal of Statistical Software*, 45, 1–67.
- Wright, M. N., & Ziegler, A. (2017). Ranger: A fast implementation of random forests for high dimensional data in C++ and R. *Journal of Statistical Software*, 77(1), 1–17.

How to cite this article: Golchi, S., Fu, J., Liu, X., Yu, E., Forghani, R., & Bhatnagar, S. (2022). Sparse Bayesian predictive modelling of tumour response using radiomic features. *Stat*, 11(1), e450. <https://doi.org/10.1002/sta4.450>