Web-based Supplemental Material for "Predictive Classification of Correlated Targets with Application to Detection of Metastatic Cancer using Functional CT Imaging"

Yuan Wang, Brian P. Hobbs, Jianhua Hu, Chaan S. Ng, and Kim-Anh Do

A Proofs

A.1 Proof of (9) :Quadratic form of the Likelihood Function

Let $\mathcal{Y}^0$ denote all the measurements from normal ROIs and $\mathcal{Y}^1$ that from tumor. Because measurements from different classes are independent, we have

$$p(\mathcal{Y}|\boldsymbol{\theta}) = \prod_{z=0}^{1} p(\mathcal{Y}^z|\boldsymbol{\theta}) \tag{A.1}$$

where

$$p(\mathcal{Y}^z|\boldsymbol{\theta}) = \prod_{i=1}^{N} \Big\{ (2\pi)^{-\frac{1}{2}pn_i^z} |\boldsymbol{\Sigma}_z|^{-\frac{1}{2}n_i^z} |\boldsymbol{\Psi}_i^z|^{-\frac{1}{2}m} \times \; etr\{-\frac{1}{2}(\boldsymbol{\Psi}_i^z)^{-1}(\boldsymbol{Y}_i^z - \boldsymbol{M}_i^z)^T \boldsymbol{\Sigma}_z^{-1}(\boldsymbol{Y}_i^z - \boldsymbol{M}_i^z)\} \Big\}$$

For written simplicity, we focus on one group and simplify $\boldsymbol{Y}_i^z$, $\boldsymbol{\Psi}_i^z$, $\boldsymbol{M}_i^z$, n_i^z, $\boldsymbol{\mu}_z$, and $\boldsymbol{\Sigma}_z$ to $\boldsymbol{Y}_i$, $\boldsymbol{\Psi}_i$, $\boldsymbol{M}_i$, n_i, $\boldsymbol{\mu}$, and $\boldsymbol{\Sigma}$, correspondingly. Moreover, let $\boldsymbol{R}_i = \boldsymbol{\Psi}_i^{-1}$ with columns means $r_{i1}, \ldots, r_{in_i}$. Given $\boldsymbol{M}_i = \mathbf{1}_{n_i}^T \otimes \boldsymbol{\mu}$, it is straightforward to show that

$$\begin{aligned} \mathrm{tr}(\boldsymbol{R}_i \boldsymbol{Y}_i^T \boldsymbol{\Sigma}^{-1} \boldsymbol{M}_i) &= \mathrm{tr}(\boldsymbol{Y}_i^T \boldsymbol{\Sigma}^{-1} (\mathbf{1}_{n_i}^T \otimes \boldsymbol{\mu}) \boldsymbol{R}_i) \\ &= \mathrm{tr}(\boldsymbol{Y}_i^T \boldsymbol{\Sigma}^{-1} ((\mathbf{1}_{n_i}^T \boldsymbol{R}_i) \otimes \boldsymbol{\mu})) = \sum_{j=1}^{n_i} r_{ij} \boldsymbol{y}_{ij}^T \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu} \end{aligned}$$

Similarly, one have $\mathrm{tr}(\boldsymbol{R}_i \boldsymbol{M}_i^T \boldsymbol{\Sigma}^{-1} \boldsymbol{M}_i) = \sum_{j=1}^{n_i} r_{ij} \boldsymbol{\mu}^T \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu}$ Therefore, one can derive the following expression for the previous exponential trace

$$\begin{aligned} &\prod_{i=1}^{N} etr\{-\frac{1}{2}\boldsymbol{\Psi}_i^{-1}(\boldsymbol{Y}_i - \boldsymbol{M}_i)^T \boldsymbol{\Sigma}^{-1}(\boldsymbol{Y}_i - \boldsymbol{M}_i)\} \\ &= \exp\{-\frac{1}{2}\sum_{i=1}^{N}[\sum_{j=1}^{n_i} r_{ij} \boldsymbol{\mu}^T \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu} - 2\sum_{j=1}^{n_i} r_{ij} \boldsymbol{y}_{ij}^T \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu} + \mathrm{tr}(\boldsymbol{\Psi}_i^{-1} \boldsymbol{Y}_i^T \boldsymbol{\Sigma}^{-1} \boldsymbol{Y}_i)]\} \\ &= \exp\{-\frac{\sum_{i=1}^{N}\sum_{j=1}^{n_i} r_{ij}}{2}(\boldsymbol{\mu} - \tilde{\boldsymbol{y}})^T \boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu} - \tilde{\boldsymbol{y}}) - \frac{1}{2}\mathrm{tr}(\boldsymbol{\Sigma}^{-1}\tilde{\boldsymbol{S}})\}, \text{ where} \end{aligned} \tag{A.2}$$

$$\tilde{\boldsymbol{y}}^z = \frac{1}{\sum_{i=1}^{N}\sum_{j=1}^{n_i} r_{ij}} \sum_{i=1}^{N}\sum_{j=1}^{n_i^z} r_{ij}^z \boldsymbol{y}_{ij}^z, \text{ and}$$

$$\tilde{\boldsymbol{S}}^z = \sum_{i=1}^{N} \boldsymbol{Y}_i^z (\boldsymbol{\Psi}_i^z)^{-1} (\boldsymbol{Y}_i^z)^T - \sum_{i=1}^{N}\sum_{j=1}^{n_i^z} r_{ij}^z \tilde{\boldsymbol{y}}^z (\tilde{\boldsymbol{y}}^z)^T.$$

Combining measurements from both classes, the likelihood function can be expressed as

$$p(\mathcal{Y}|\theta) = \prod_{z=0}^{1} \left\{ \prod_{i=1}^{N} \left[(2\pi)^{-\frac{1}{2}mn_i^z} |\boldsymbol{\Sigma}_z|^{-\frac{1}{2}n_i^z} |\boldsymbol{\Psi}_i^z|^{-\frac{1}{2}m} \right] \times \right.$$
$$\left. \exp\{ -\frac{\sum_{i=1}^{N}\sum_{j=1}^{n_i} r_{ij}^z}{2} (\boldsymbol{\mu}_z - \tilde{\boldsymbol{y}}^z)^T \boldsymbol{\Sigma}_z^{-1} (\boldsymbol{\mu}_z - \tilde{\boldsymbol{y}}^z) - \frac{1}{2}\operatorname{tr}(\boldsymbol{\Sigma}_z^{-1}\tilde{\boldsymbol{S}}^z) \} \right\}. \tag{A.3}$$

A.2 Proof of (14): Conditional Predictive Distribution

Let $\boldsymbol{\theta}_1 = \{\boldsymbol{u}_z, \boldsymbol{\Sigma}_z\}_{z=0}^1$ the collection of mean and covariance parameters, and $\boldsymbol{\theta} = \{\boldsymbol{\theta}_1, \rho\}$. The conditional predictive density follows as

$$p(\boldsymbol{Y}_{N+1}|\mathcal{Y}, \boldsymbol{z}_{N+1}, \rho) = \int p(\boldsymbol{Y}_{N+1}|\boldsymbol{\theta}_1, \mathcal{Y}, \boldsymbol{z}_{N+1}, \rho)\pi(\boldsymbol{\theta}_1|\mathcal{Y}, \boldsymbol{z}_{N+1}, \rho)d\boldsymbol{\theta}_1 \tag{A.4}$$

$$= \int p(\boldsymbol{Y}_{N+1}|\boldsymbol{\theta}_1, \boldsymbol{z}_{N+1}, \rho)\pi(\boldsymbol{\theta}_1|\mathcal{Y}, \rho)d\boldsymbol{\theta}_1 \tag{A.5}$$

$$= \int p(\boldsymbol{Y}_{N+1}|\boldsymbol{\theta}_1, \boldsymbol{z}_{N+1}, \rho)\frac{p(\mathcal{Y}|\boldsymbol{\theta}_1, \rho)\pi(\boldsymbol{\theta}_1)}{p(\mathcal{Y}, \rho)}d\boldsymbol{\theta}_1 \tag{A.6}$$

$$= \int p(\boldsymbol{Y}_{N+1}|\boldsymbol{\theta}, \boldsymbol{z}_{N+1})p(\mathcal{Y}|\boldsymbol{\theta})\pi(\boldsymbol{\theta}_1)d\boldsymbol{\theta}_1/p(\mathcal{Y}, \rho) \tag{A.7}$$

Given the class indicator $\boldsymbol{z}_{N+1}$ for the new patient to be predicted, we let $\boldsymbol{Y}_{N+1}^0 = [\boldsymbol{y}_{N+1,1}^0, \ldots, \boldsymbol{y}_{N+1,n_{N+1}^0}^0]$ denote the observables for normal ROIs and $\boldsymbol{Y}_{N+1}^1 = [\boldsymbol{y}_{N+1,1}^1, \ldots, \boldsymbol{y}_{N+1,n_{N+1}^1}^1]$ for tumor, respectively. Applying the quadratic form for the likelihood function derived in Appendix A.1, it is straightforward to show that

$$p(\boldsymbol{Y}_{N+1}|\boldsymbol{\theta}, \boldsymbol{z}_{N+1})p(\mathcal{Y}|\boldsymbol{\theta}) = \prod_{z=0}^{1} \left\{ \prod_{i=1}^{N+1} \left[(2\pi)^{-\frac{1}{2}pn_i^z} |\boldsymbol{\Sigma}_z|^{-\frac{1}{2}n_i^z} |\boldsymbol{\Psi}_i^z|^{-\frac{1}{2}m} \right] \times \right.$$
$$\left. \exp\{ -\frac{\sum_{i=1}^{N+1}\sum_{j=1}^{n_i} r_{ij}^z}{2} (\boldsymbol{\mu}_z - \tilde{\boldsymbol{y}}_{N+1}^z)^T \boldsymbol{\Sigma}_z^{-1} (\boldsymbol{\mu}_z - \tilde{\boldsymbol{y}}_{N+1}^z) - \frac{1}{2}\operatorname{tr}(\boldsymbol{\Sigma}_z^{-1}\tilde{\boldsymbol{S}}_{N+1}^z) \} \right\}, \tag{A.8}$$

where r_{ij}^z is the jth column sum of $(\boldsymbol{\Psi}_i^z)^{-1}$,

$$\tilde{\boldsymbol{y}}_{N+1}^z = \frac{1}{\sum_{i=1}^{N+1}\sum_{j=1}^{n_i} r_{ij}^z} \sum_{i=1}^{N+1} \sum_{j, z_{ij}=z} r_{ij}^z \boldsymbol{y}_{ij}, \text{ and} \tag{A.9}$$

$$\tilde{\boldsymbol{S}}_{N+1}^z = \sum_{i=1}^{N+1} \boldsymbol{Y}_i^z (\boldsymbol{\Psi}_i^z)^{-1} (\boldsymbol{Y}_i^z)^T - \sum_{i=1}^{N+1}\sum_{j=1}^{n_i} r_{ij}^z \tilde{\boldsymbol{y}}_{N+1}^z (\tilde{\boldsymbol{y}}_{N+1}^z)^T \tag{A.10}$$

Notice that values of n^z_{N+1}, $\boldsymbol{\Psi}^z_{N+1}$, and c^z_{N+1} depend on the group index $\boldsymbol{z}_{N+1}$ since different $\boldsymbol{z}_{N+1}$ may correspond to different numbers of tumor and normal ROIs. Because the prior density $\pi(\boldsymbol{\theta}_1)$ in (10) is a conjugate prior for (A.8), the conditional predictive density may be expressed in closed form as

$$p(\boldsymbol{Y}_{N+1}|\mathcal{Y}, \boldsymbol{z}_{N+1}, \rho) = \prod_{z=0}^{1} \left\{ \prod_{i=1}^{N+1} \left[(2\pi)^{-\frac{1}{2}m(n_i^z-1)} |\boldsymbol{\Psi}_i^z|^{-\frac{1}{2}m} \right] (\sum_{i=1}^{N+1}\sum_{j=1}^{n_i} r_{ij}^z)^{-\frac{1}{2}m} \right.$$
$$\left. \frac{h(\delta, \boldsymbol{\Omega}_z)}{h(\sum_{i=1}^{N+1} n_i^z + \delta - 1, \boldsymbol{\Omega}^z + \tilde{\boldsymbol{S}}^z_{N+1})} \right\} \tag{A.11}$$
$$\propto \prod_{z=0}^{1} \left\{ \frac{|\boldsymbol{\Psi}^z_{N+1}|^{-\frac{1}{2}m} (\sum_{i=1}^{N+1}\sum_{j=1}^{n_i} r_{ij}^z)^{-\frac{1}{2}m}}{h(\sum_{i=1}^{N+1} n_i^z + \delta - 1, \boldsymbol{\Omega}^z + \tilde{\boldsymbol{S}}^z_{N+1})} \right\} \tag{A.12}$$

A.3 Proof of (18): Maximization of Marginal Likelihood

First, we derive the conditional marginal likelihood function given correlation parameter ρ

$$p(\mathcal{Y}|\rho) = \int p(\boldsymbol{Y}|\boldsymbol{\mu}, \boldsymbol{\Sigma}, \rho) \prod_{z=0}^{1} \pi(\boldsymbol{\mu}_z, \boldsymbol{\Sigma}_z|\delta, \boldsymbol{\Omega}_z) d\boldsymbol{\mu}_z d\boldsymbol{\Sigma}_z,$$

where the likelihood $p(\boldsymbol{Y}|\boldsymbol{\mu}, \boldsymbol{\Sigma}, \rho)$ is given in (9) and the prior $\pi(\boldsymbol{\mu}_z, \boldsymbol{\Sigma}_z|\delta, \boldsymbol{\Omega}_z)$ is in (10). By pluggin in these two functions, we obtain

$$p(\boldsymbol{Y}|\boldsymbol{\mu}, \boldsymbol{\Sigma}, \rho) \prod_{z=0}^{1} \pi(\boldsymbol{\mu}_z, \boldsymbol{\Sigma}_z|\delta, \boldsymbol{\Omega}_z)$$
$$\propto \prod_{z=0}^{1} |\boldsymbol{\Sigma}_z|^{-\frac{1}{2}(n+\delta+m)} etr\{-\frac{1}{2}(\boldsymbol{\Omega}_z + \tilde{\boldsymbol{S}}^z)\boldsymbol{\Sigma}_z^{-1}\} h(\delta, \boldsymbol{\Omega}_z) d\boldsymbol{\Sigma}_z$$
$$\propto \prod_{z=0}^{1} \frac{h(\delta, \boldsymbol{\Omega}_z)}{h(n-1+\delta, \boldsymbol{\Omega}_z + \tilde{\boldsymbol{S}}^z)}$$

Next, we obtain the maximizer of $\boldsymbol{\Omega}_z$ for a given class z. Omitting the index z, define $f(\boldsymbol{\Omega}) = \frac{|\boldsymbol{\Omega}|^{\frac{1}{2}\delta}}{|\boldsymbol{\Omega}+\tilde{\boldsymbol{S}}|^{\frac{1}{2}(n^z-1+\delta)}}$. Consider the eigenvalue decompositions $\tilde{\boldsymbol{S}} = \boldsymbol{U}_S \boldsymbol{D}_S \boldsymbol{U}_S^T$ and $\boldsymbol{\Omega} = \boldsymbol{U}_\Omega \boldsymbol{D}_\Omega \boldsymbol{U}_\Omega^T$, where $\boldsymbol{U}_S, \boldsymbol{U}_\Omega$ are orthogonal matrices, and $\boldsymbol{D}_S$, $\boldsymbol{D}_\Omega$ are diagonal matrices with diagonal elements $\lambda_{S,1} \geq \ldots \lambda_{S,m} > 0$, $\lambda_{\Omega,1} \geq \ldots \lambda_{\Omega,m} > 0$ respectively. The results in Fiedler (1971) yield the following bound

$$f(\Omega) \leq \prod_{k=1}^{m} \frac{\lambda_{\Omega,k}^{\frac{1}{2}\delta}}{(\lambda_{S,k} + \lambda_{\Omega,k})^{\frac{1}{2}(\delta+n-1)}}. \tag{A.13}$$

The equality holds when $\boldsymbol{U}_S = \boldsymbol{U}_\Omega$. Moreover, using the Karush-Kuhn-Tucker conditions, it is easy to show that the maximum of the right hand side of (A.13) is attained when

$$\lambda_{\Omega,k} = \frac{\delta}{n-1} \lambda_{S,k}, \tag{A.14}$$

for $k = 1, \ldots, m$. Thus the empirical Bayesian estimate for the hyperparameter $\mathbf{\Omega}$ conditional on ρ is

$$\hat{\mathbf{\Omega}} = \frac{\delta}{n-1}\tilde{\boldsymbol{S}}. \tag{A.15}$$

B Posterior Inference

Markov Chain Monte Carlo was used to conduct full posterior inference on the CT perfusion data. Given the posterior density function of ρ, the random walk Metropolis algorithm was used to generate 25000 posterior samples of ρ using the *MCMCmetrop1R* function in the R package "MCMCpack". Gibbs sampling was then used to generate posterior samples for the mean and covariance parameters using the conditional posterior distributions given in (13). Fig. A plots the resultant kernel smooth posterior density for the three correlation structures we've considered. For all the structures, the posterior distribution of ρ is unimodal. Moreover, the model reflects high interdependence among observables derived from disparate intra-patient ROIs. For compound symmetry, the posterior distribution of ρ has mean of 0.734 and standard deviation of 0.048, which suggests the inter-region correlation is 0.734 regardless distance between two regions. For exponential and spherical correlation, the parameter ρ does not reveal the inter-region correlation strength directly. In fact, for exponential correlation, the posterior distribution of ρ has mean of 2.04 and standard deviation of 0.441, which suggests that two regions with distance 0.39cm (median distance for CTp data) have correlation 0.826. For spherical correlation, the posterior distribution of ρ has mean of 2.99 and standard deviation of 0.558, which suggests that two regions with distance 0.39cm (median distance for CTp data) have correlation 0.834. The posterior distribution of ρ are provided in Fig. A. The resultant posterior mean vectors of perfusion characteristics in normal and tumor regions, for compound symmetry (CS), exponential (Exp), and spherical (Sph) are provided below

$$\begin{aligned}
CS: &\quad \hat{\mu}_0 = (4.906, 2.680, 4.381, 2.129, -1.856), \hat{\mu}_1 = (5.378, 2.730, 3.877, 1.740, -0.805),\\
Exp: &\quad \hat{\mu}_0 = (4.910, 2.682, 4.380, 2.128, -1.850), \hat{\mu}_1 = (5.382, 2.733, 3.878, 1.739, -0.806),\\
Sph: &\quad \hat{\mu}_0 = (4.910, 2.682, 4.380, 2.128, -1.849), \hat{\mu}_1 = (5.381, 2.733, 3.878, 1.739, -0.806).
\end{aligned}$$

C Sensitivity to the Prior Distribution

We run the classification under the compound symmetric structure with three different prior distributions for ρ, using four predictors without BV. The results are given in Table A. Generally, the results are similar in different choices of the prior distribution.

D Selection of Predictors

Using different choice of predictors, the results for classification varied, as shown in Table B. It can be seen that the classification accuracy is improved when adding more predictors up to

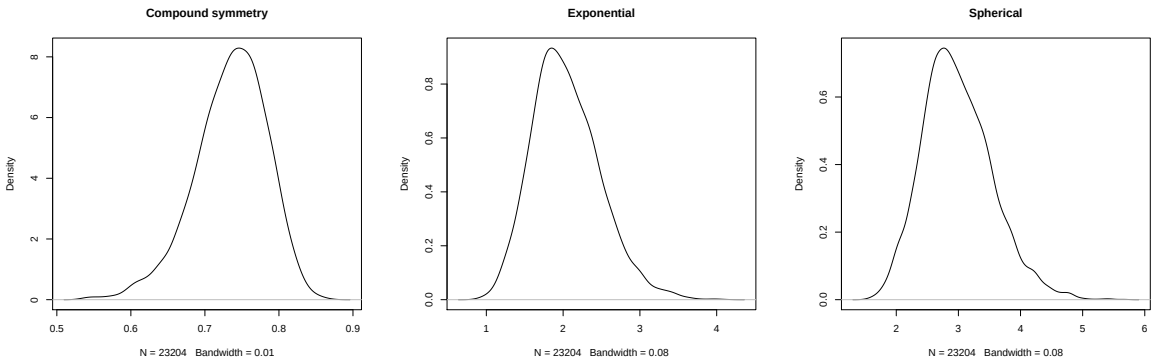


Fig. A: The posterior density for ρ fit to the CT perfusion data when inter-region correlation is compound symmetric (left), exponential (middle), and spherical (right).

	beta(1,1)		beta(10, 10)		beta(0.2, 0.2)	
	TPR	FPR	TPR	FPR	TPR	FPR
mrFB($\alpha = 0.5$)	0.96	0.04	0.92	0.04	0.96	0.04
mrFB($\alpha = 0.8$)	1.00	0.07	0.96	0.19	1.00	0.04
lda	0.92	0.15	0.92	0.15	0.92	0.15
qda	0.80	0.19	0.80	0.19	0.80	0.19

Table A: Classification results with different prior of the correlation parameter ρ.

4. The results for using all 5 CTp characteristics failed to yield improvement when compared to results obtained for the selected 4 characteristic. This may be due to the strong correlation between blood flow and blood volume.

	BF		BF, PS		BF, PS, MTT		w.o. BV		all 5	
	TPR	FPR	TPR	FPR	TPR	FPR	TPR	FPR	TPR	FPR
mrEB($\alpha = 0.5$)	0.60	0.26	0.92	0.04	0.96	0.07	0.96	0.04	0.96	0.07
mrEB($\alpha = 0.8$)	0.96	0.70	0.96	0.11	0.96	0.15	1.00	0.04	0.96	0.15
lda	0.64	0.22	0.84	0.11	0.80	0.11	0.92	0.15	0.88	0.15
qda	0.52	0.15	0.80	0.15	0.80	0.15	0.80	0.19	0.84	0.22

Table B: classification with the compound symmetric structure using selected predictors.

E Sensitivity to Compound Symmetric Covariance

In this section, we conduct another simulation study to investigate sensitivity to misspecification of inter-region correlation, $\mathbf{\Psi}$. Specifically, we evaluated the extent to which classification performance is diminished for simultaneous classification using compound symmetry, when the true underlying correlation structure exhibits spatial dependence with exponential decay. In

	BF		BF, PS		BF, PS, MTT		w.o. BV		all 5	
	TPR	FPR	TPR	FPR	TPR	FPR	TPR	FPR	TPR	FPR
mrEB($\alpha = 0.5$)	0.64	0.33	0.96	0.07	0.96	0.04	0.96	0.04	0.96	0.04
mrEB($\alpha = 0.8$)	0.96	0.59	0.96	0.11	0.96	0.11	0.96	0.04	0.96	0.15
lda	0.64	0.22	0.84	0.11	0.80	0.11	0.92	0.15	0.88	0.15
qda	0.52	0.15	0.80	0.15	0.80	0.15	0.80	0.19	0.84	0.22

Table C: classification with the spatial structure (power decay) using selected predictors.

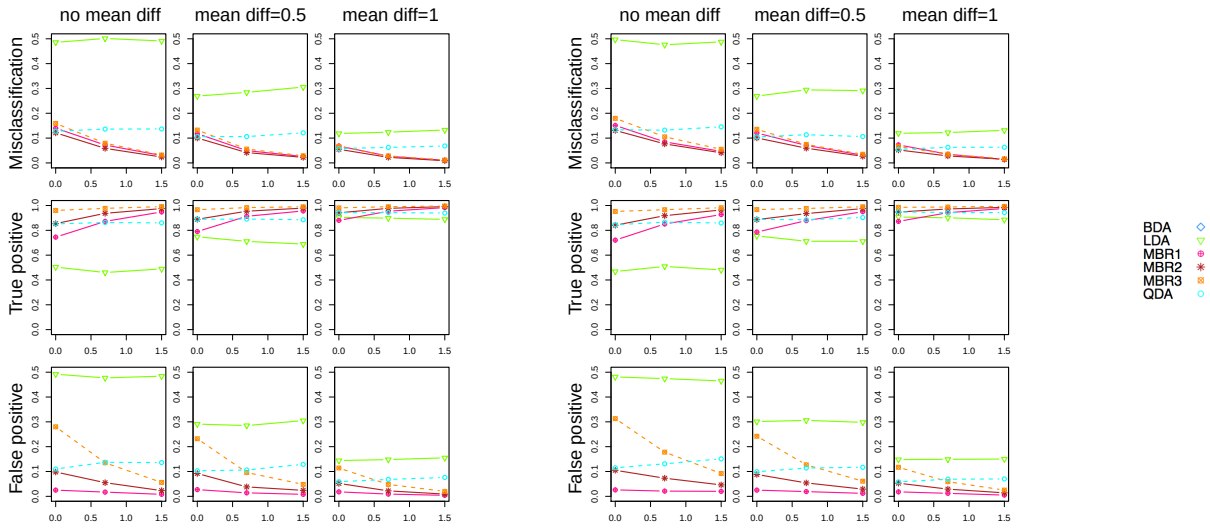
the simulation, the mean and covariance parameters are identical to the simulation in Section 5 scenario 2, where we mimic the real CT perfusion data by fixing the covariance matrices for each group to be their posterior mean estimates. The mean for normal group is also fixed at its posterior mean $\hat{\boldsymbol{\mu}}_0$ under the compound symmetric structure. The mean CTp characteristics for tumor are specified to reflect a Δ% shift from normal: $\hat{\boldsymbol{\mu}}_0(\Delta + 1)$. Three values of Δ are considered: $0, 0.5$, and 1. The location of regions are randomly drawn from the unit square and the average distance between regions is around 0.5. Moreover, three different values of the decay rate ϕ, $(0.1, 0.7, 1.5)$ are considered reflecting weak, moderate, and strong average spatial dependence. The results are shown in Fig. B. When $\phi = 0.1$, the correlation is weak between all regions, performance for simultaneous classification using compound symmetry is robust to the misspecification. When $\phi = 0.7$, correlation is highly spatially dependent. For example, correlation for two regions that are 0.1 cm (1 cm) apart attain correlation 0.87(0.18). For this scenario, assuming compound symmetry diminished performance when compared to exponential as observed by the 1.6% in increased misclassification rate and 1.8% in decreased true positive rate for out MBR method with $\alpha = 0.5$ and $\Delta = 0$. When $\phi = 1.5$, there exists strong correlation between both nearby and distant regions, the classifier using the compound symmetric structure yields comparable results with the true exponential model. Generally, results for the simultaneous approaches outperform conventional classification methods in the presence of increasing inter-region correlation regardless of the assumed inter-region correlation, $\boldsymbol{\Psi}$.

F Software

Software for implementing simultaneous classification as well as the various conventional classification methods discussed in the paper is included with supplementary materials. The code was written using statistical software R (R Development Core Team, http://www.r-project.org) version 3.0.1.

References

Fiedler, M. (1971). Bounds for the determinant of the sum of Hermitian matrices. *Proceedings of the American Mathematical Society* **30,** 27–31.



(a) exponential (b) compound symmetric

Fig. B: Misclassification rate (top row), true positive rate (middle row) and false positive rate (bottom row) when (a) the exponential structure is fitted; (b) the compound symmetric structure is fitted. Results for the simultaneous classification by minimizing Bayesian risk (MBR1, $\alpha = 0.2$, shown by "$\oplus$"; MBR2, $\alpha = 0.5$, shown by "$*$"; MBR3, $\alpha = 0.8$, shown by "$\otimes$"), and conventional classification using quadratic discriminant analysis (QDA, shown by "$\circ$"), and linear discriminant analysis (LDA, shown by "$\triangledown$") are provided. As the correlation between ROIs increases, simultaneous MBR yields increasing true positive rate and decreasing false positive rate, while the conventional methods are static.