

ROBUST FEATURE SELECTION FOR BLOCK COVARIANCE BAYESIAN MODELS

Ali Foroughi pour and Lori A. Dalton

Department of Electrical and Computer Engineering
The Ohio State University, Columbus, OH, 43210, USA

ABSTRACT

Recent work proposes new algorithms for feature selection based on a Bayesian hierarchical model that places priors on both the identity of all features, and the identity-conditioned feature-label distribution. Given training data, Bayesian inference can be used to predict the feature identities. While algorithms developed in prior work rely on certain independence assumptions, in this work we present a new algorithm, with low computational complexity, designed for a family of Bayesian models that each assume different block covariance structures. We show the new algorithm, and the previous algorithm assuming independent features, have robust performance across the family of models under synthetic data, and provide results from real colon cancer microarray data.

Index Terms— Bayesian Feature Selection, Robustness Analysis, Biomarker Discovery

1. INTRODUCTION

Many applications require feature selection on small-sample high-dimensional data, such as biomarker discovery, where it is desired to find biomarkers involved in the biological mechanism of the disease under study, which can be further utilized for diagnosis, drug development, etc. Filter and wrapper methods are two popular feature selection methodologies for biomarker discovery [1, 2, 3]. Filter methods assess each feature individually, while wrapper methods minimize a cost function over all feature sets. Most wrapper methods use heuristic objective functions, such as a classification error estimate, and require suboptimal search heuristics like sequential forward search (SFS). Although high-throughput technology provides a deluge of data per sample point, research is usually constrained to small samples with no performance guarantees, thus impeding reliable biomarker discovery [4, 1].

Recent work proposes a Bayesian framework, where the probability of each feature being “good”, that we wish to select, is computed. Assuming independent features, a closed-form optimal Bayesian filter (OBF) has been found [5]. A fast suboptimal wrapper method, here called 2MNC-DGIB, assumes good features are dependent with each other, while the remaining “bad” features are independent, and exhibits outstanding performance under synthetic microarray data [6].

Here, we extend the model to consider a family of Bayesian models with more complex covariance structures, and propose an approximate posterior. Although the approximation does not correspond to any block structures, we demonstrate 2MNC using the approximation, called 2MNC-Robust, and OBF are robust under a family of Bayesian models each assuming different covariance structure.

2. MODEL

Consider a binary feature selection problem with labels $y \in \{0, 1\}$, where F is the set of feature indices. Assume features are partitioned into blocks, where features in the same block are dependent, and features in different blocks are independent of each other. Blocks are good or bad: a good block has different class conditional distributions, and a bad block has the same distribution in both classes. Denote a *feature partition*, i.e., a partitioning of F to good and bad blocks, by $P = (P_G, P_B)$, where $P_G = \{G_1, \dots, G_u\}$ and $P_B = \{B_1, \dots, B_v\}$ are the set of good and bad blocks, respectively. Define the set of good features as $G = \cup G_i$. We denote the true feature partition and true good features by \bar{P} and \bar{G} , respectively. Define $\pi(P) = P(\bar{P} = P)$ as the prior on the true feature partition. If $\pi(P)$ is only non-zero when $u = v = 1$ we have a *Dependent Good Dependent Bad* (DGDB) block structure, if $\pi(P)$ assumes all blocks are of size 1 we have an *Independent Good Independent Bad* (IGIB) block structure, and if $\pi(P)$ is non-zero only if $u = 1$ and all bad blocks are of size 1 we have an *Dependent Good Independent Bad* (DGIB) block structure.

Fix P and let $\theta = [\theta_0^{G_1}, \dots, \theta_0^{G_u}, \theta_1^{G_1}, \dots, \theta_1^{G_u}, \theta^{B_1}, \dots, \theta^{B_v}]$, where $\theta_y^{G_i}$ parametrizes the distribution of class y features in G_i , $f_{\theta_y^{G_i}}(\cdot|y)$, and θ^{B_j} parametrizes the distribution of features in B_j , $f_{\theta^{B_j}}(\cdot)$. Assume $\theta_y^{G_i}$ and θ^{B_j} are independent given P , i.e., $f(\theta|P) = \prod_{i=1}^u f(\theta_y^{G_i}) \prod_{j=1}^v f(\theta^{B_j})$. Finally, let P and θ be fixed, and let \mathbb{R}^F be the sample space, and x be a sample point in class y . We assume

$$f(x|y, P, \theta) = \prod_{i=1}^u f_{\theta_y^{G_i}}(x^{G_i}|y) \prod_{j=1}^v f_{\theta^{B_j}}(x^{B_j}),$$

where for each block A , x^A contains feature values in A .

Let S be a training set of size n with n_y points per class. For each block A , let S^A denote feature values in

A for all sample points, and S_y^A denote feature values in A for class y sample points. We have the likelihood functions $f(S_y^{G_i}|\theta_y^{G_i}) = \prod_{x^{G_i} \in S_y^{G_i}} f_{\theta_y^{G_i}}(x^{G_i}|y)$, and likewise for $f(S^{B_j}|\theta^{B_j})$. Using techniques from [6, 7], it can be shown that the posterior on the true feature partition is,

$$\begin{aligned} \pi^*(P) &= \mathbb{P}(\bar{P} = P|S) \\ &\propto \pi(P) \prod_{y=0}^1 \prod_{i=1}^u \int f(\theta_y^{G_i}) f(S_y^{G_i}|\theta_y^{G_i}) d\theta_y^{G_i} \\ &\quad \times \prod_{j=1}^v \int f(\theta^{B_j}) f(S^{B_j}|\theta^{B_j}) d\theta^{B_j}. \end{aligned}$$

Observe that the posterior probability that feature $g \in F$ is a good feature is $\mathbb{P}(g \in \bar{G}|S) = \sum_{P: g \in \cup G_i} \pi^*(P)$.

3. GAUSSIAN MODEL

Here we consider the case where blocks are jointly Gaussian. Let A be a good block. We have $\theta_y^A = [\mu_y^A, \Sigma_y^A]$, where μ_y^A is the mean and Σ_y^A is the covariance. Assume $f(\theta_y^A)$ is normal-inverse-Wishart, i.e., $f(\theta_y^A) = f(\Sigma_y^A) f(\mu_y^A|\Sigma_y^A)$, where

$$\begin{aligned} f(\Sigma_y^A) &= K_y^A |\Sigma_y^A|^{-\frac{\kappa_y^A + |A| + 1}{2}} \exp(-0.5 \text{Tr}(\Sigma_y^A (\Sigma_y^A)^{-1})), \\ f(\mu_y^A|\Sigma_y^A) &= L_y^A |\Sigma_y^A|^{-0.5} \\ &\quad \times \exp(-0.5 \nu_y^A (\mu_y^A - m_y^A)^T (\Sigma_y^A)^{-1} (\mu_y^A - m_y^A)). \end{aligned}$$

S_y^A, κ_y^A, m_y^A , and ν_y^A are hyperparameters. For a proper prior, S_y^A is an $|A| \times |A|$ symmetric positive-definite matrix, m_y^A is a length $|A|$ vector, $\nu_y^A > 0$, $L_y^A = (2\pi/\nu_y^A)^{-0.5|A|}$, $K_y^A = |S_y^A|^{0.5} \kappa_y^A 2^{-0.5\kappa_y^A |A|} / \Gamma_{|A|}(0.5\kappa_y^A)$, and $\kappa_y^A > |A| - 1$, where Γ_d is the multivariate gamma function. Improper priors may also be used, for instance Jeffreys prior assigns $K_y^A = L_y^A = 1$, S_y^A an all-zero matrix, and $\kappa_y^A = \nu_y^A = 0$. The posterior is normal-inverse-Wishart with updated hyperparameters $\kappa_y^{A*} = \kappa_y^A + n_y$, $\nu_y^{A*} = \nu_y^A + n_y$, $m_y^{A*} = \frac{\nu_y^A m_y^A + n_y \hat{\mu}_y^A}{\nu_y^{A*}}$, and

$$S_y^{A*} = S_y^A + (n_y - 1) \hat{\Sigma}_y^A + \frac{\nu_y^A n_y}{\nu_y^A + n_y} (\hat{\mu}_y^A - m_y^A) (\hat{\mu}_y^A - m_y^A)^T,$$

where $\hat{\mu}_y^A$ and $\hat{\Sigma}_y^A$ are the sample mean and covariance of feature values in A for class y sample points, respectively [8]. Suppose A is a bad block, and $f(\theta^A)$ is normal-inverse-Wishart with hyperparameters S^A, κ^A, m^A , and ν^A , and normalization constants K^A and L^A . The posterior on $f(\theta^A)$ has updated hyperparameters $\kappa^{A*} = \kappa^A + n$, $\nu^{A*} = \nu^A + n$, $m^{A*} = \frac{\nu^A m^A + n \hat{\mu}^A}{\nu^{A*}}$, and

$$S^{A*} = S^A + (n - 1) \hat{\Sigma}^A + \frac{\nu^A n}{\nu^A + n} (\hat{\mu}^A - m^A) (\hat{\mu}^A - m^A)^T,$$

where $\hat{\mu}^A$ and $\hat{\Sigma}^A$ are the sample mean and covariance of feature values in A for all sample points, respectively [8].

Assuming: (1) $\pi(P)$ is such that the block structure, i.e., the number and size of good and bad blocks, is fixed, (2) for each good block $A \in P_G$, K_y^A, L_y^A, κ_y^A , and ν_y^A do not depend

on the index of features assigned to A , and (3) for each bad block $A \in P_B$ the hyperparameters K^A, L^A, κ^A , and ν^A do not depend on the index of features in A , it can be shown that

$$\pi^*(P) \propto \pi(P) \left(\prod_{i=1}^u |S_0^{G_i^*}|^{\kappa_0^{G_i^*}} |S_1^{G_i^*}|^{\kappa_1^{G_i^*}} \prod_{j=1}^v |S^{B_j^*}|^{\kappa^{B_j^*}} \right)^{-0.5}.$$

4. SET SELECTION

The *Maximum Number Correct* (MNC) criterion labels all features as good or bad with a maximal expected number of correctly labeled features, resulting in the predicted set of good features $\{g \in F : \mathbb{P}(g \in \bar{G}|S) > 0.5\}$ [6]. The *Constrained MNC* (CMNC) criterion adds a constraint of selecting a fixed number of features, and is equivalent to computing $\mathbb{P}(g \in \bar{G}|S)$ and reporting the top ranked features.

Under IGIB block structures, $\mathbb{P}(g \in \bar{G}|S)$ can be easily found in closed form, and we call the optimal set selection algorithm under the MNC or CMNC criteria OBF [5]. However, under general block structures computing $\mathbb{P}(g \in \bar{G}|S)$ is infeasible when $|F|$ is large. 2MNC is a fast suboptimal algorithm that computes $\mathbb{P}(g \in \bar{G}|S)$ under a Bayesian model assuming 2 good features, and reports the top ranked features [6]. 2MNC can be used with any block structure assuming 2 good features, and with Jeffreys prior has been shown to have outstanding performance under DGIB block structures.

We propose a robust algorithm, called 2MNC-Robust, implementing 2MNC with the following approximate posterior:

$$\tilde{\pi}^*(G) \propto \pi(G) \left(|S_0^{G^*}|^{\kappa_0^{G^*}} |S_1^{G^*}|^{\kappa_1^{G^*}} / |S^{G^*}|^{\kappa^{G^*}} \right)^{-0.5},$$

which generalizes the exact posterior under the DGIB model. The approximate marginal is $\mathbb{P}(g \in \bar{G}|S) = \sum_{G: g \in G} \tilde{\pi}^*(G)$.

5. SYNTHETIC DATA SIMULATIONS

Here we present synthetic simulations to demonstrate the robustness of OBF and 2MNC-Robust across a complete family of block structures, and a family of synthetic microarray models. We then present results on colon cancer microarray data.

5.1. Small Feature Simulation

We assume $|F| = 8$, $|\bar{G}| = 4$, blocks are jointly Gaussian, and consider all 25 possible block structures (note 4 features can be grouped 5 ways, with group sizes (1, 1, 1, 1), (2, 1, 1), (2, 2), (3, 1) or (4)). Bad features and class 0 good features have 0 mean. We have 3 mean types for class 1 good features: redundant, synergetic, and marginal. Redundant features have mean [0.3, 0.3, 0.3, 0.3]. Synergetic features have mean [-0.3, -0.1, 0.1, 0.3], where the structure (2, 1, 1) assigns blocks to means [-0.3, -0.1], 0.1, and 0.3, structure (2, 2) assigns means [-0.3, -0.1] and [0.1, 0.3], and structure (3, 1) assigns means [-0.3, -0.1, 0.1] and 0.3. Each marginal

good block contains 1 feature with mean 0.3, while other features in that block have 0 mean. All features have class conditioned variances of 0.5. Features of a bad block have the same correlation coefficient c^b , and similarly features in a good block in class y have correlation coefficient c_y^g , which each take values 0.1, 0.5, or 0.9.

For each block structure, mean type, combination of correlation coefficients, and sample size, we randomly draw a feature partition, draw a sample with equal points per class, and run 2MNC-Robust, as well as 25 CMNC algorithms corresponding to each of the possible block structures, including CMNC-DGDB, CMNC-DGIB, and CMNC-OBF. All of these use Jeffrey’s prior and select 4 features. This is iterated 1000 times. Finally, for a fixed mean type, combination of correlation coefficients and sample size, we define the minimax algorithm to be the CMNC algorithm with best worst-case performance across all 25 block structures, and define the model constrained robust algorithm to be the CMNC algorithm with the best average performance across all 25 block structures, where performance is evaluated as the average number of correctly labeled features over iterations.

Figures 1 and 2 plot examples of worst-case performance and average performance over all 25 block structures as sample size increases from 10 to 100, assuming various mean types and correlations. CMNC-OBF has the best worst-case performance under redundant and synergetic means, as in Fig. 1(a), and is hence minimax. Under marginal features, all algorithms tend to have equally poor worst-case performance when good features of both classes have low correlation or a mix of low to moderate correlations, as in Fig. 1(b), CMNC-DGDB tends to outperform under moderate correlations, as in Fig. 1(c), and 2MNC-Robust excels when good features of both classes have high correlation or a mix of moderate to high correlations, as in Fig. 1(d). CMNC-DGDB is typically minimax for marginal features under moderate and high correlations, as in Figs. 1(c) and 1(d). Considering average performance and redundant means, CMNC-OBF outperforms when $c_0^g = c_1^g$, as in Fig. 2(a), otherwise it is typically beat by CMNC-DGDB and 2MNC-Robust, which perform similarly, as in Fig. 2(b). Under synergetic and marginal means, CMNC-OBF has slightly better average performance when $c_0^g = c_1^g = 0.1$, as in Fig. 2(g), otherwise CMNC-DGDB is either the best or close to the best algorithm, as in Fig. 2(c)-(f). 2MNC-Robust often performs close to either CMNC-DGDB or CMNC-OBF, and there are cases where 2MNC-Robust is the best performer of the three, most notably under very small sample size, as in Figs. 2(d) and (f), or when $c^b = 0.1$ and either $c_0^g = 0.9$ or $c_1^g = 0.9$, as in Fig. 2(h). In Fig. 2, CMNC-OBF is the model constrained algorithm in parts (a) and (g), and CMNC-DGDB is typically the model-constrained algorithm otherwise. These figures are representative of the trends seen among the 3 mean types and 27 combinations of correlation coefficients we studied. CMNC-OBF is the fastest, followed by 2MNC-Robust, CMNC-DGIB, and CMNC-DGDB

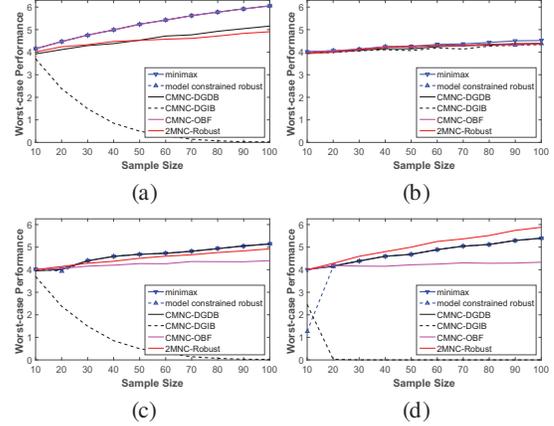


Fig. 1: Worst-case performance, $c = c_0^g = c_1^g = c^b$: (a) redundant, $c = 0.5$; (b) marginal, $c = 0.1$; (c) marginal, $c = 0.5$; (d) marginal, $c = 0.9$.

with 1.1 and 25, and 27 times its runtime, respectively.

Overall, CMNC-OBF is extremely fast and has superior robustness in terms of worst-case performance, as long as features are not too weak as in the marginal mean models. CMNC-DGDB is perhaps the most robust in terms of average performance, but it cannot be used in high-dimensional small-sample problems because of issues with improper (invalid) posteriors and computation time. 2MNC-Robust is thus an attractive alternative, with similar robust performance across many models and outstanding computation time. Although CMNC-DGIB has very competitive performance when $c^b = 0.1$, it can exhibit extremely poor performance in some settings, see for instance Figs. 1(a), (c) and (d).

5.2. Synthetic Microarray Model

A synthetic model to mimic microarrays is proposed in [3]. Features are divided into four types: global markers, heterogeneous markers, low variance non-markers and high variance non-markers. Global markers are Gaussian and homogeneous within each class. Heterogeneous markers are each associated with one of c subclasses in class 1, and behave as class 1 global markers for sample points in the corresponding subclass, and as a class 0 global markers for all other points. Markers comprise blocks of size k , and can be of three types: redundant, synergetic, and marginal with class 1 means $[1, \dots, 1]$, $[1, 1/2, \dots, 1/k]$, and $[1, 0, \dots, 0]$, respectively, for each block. All markers have 0 mean in class 0. Markers in each block of class y have covariance $\sigma_y \Sigma_y$, where diagonal elements of Σ_y are 1, and non-diagonal elements are ρ_y . We have extended the original model to allow $\rho_0 \neq \rho_1$. High-variance non-markers are independent with distribution $pN(0, \sigma_0) + (1 - p)N(0, \sigma_1)$, where $p \sim \text{Uniform}[0, 1]$ for each feature. In [3], low-variance non-markers are independent with distribution $N(0, \sigma_0)$; here they comprise size k blocks, with distribution $N(0, \sigma_0 \Sigma_0)$.

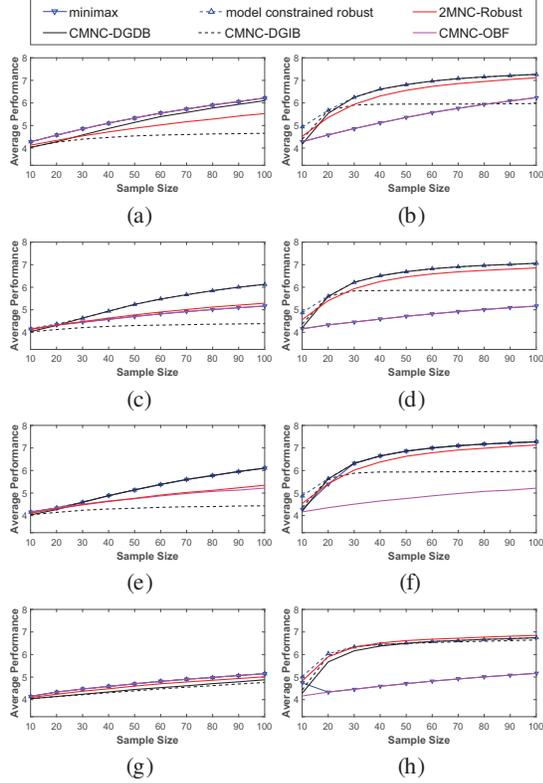


Fig. 2: Average performance, c_0^g, c_1^g, c^b are 0.5 unless otherwise stated: (a) redundant; (b) redundant, $c_1^g = 0.9$; (c) synergetic; (d) synergetic, $c_1^g = 0.9$; (e) marginal; (f) marginal, $c_1^g = 0.9$; (g) synergetic, $c_0^g = c_1^g = c^b = 0.1$; (h) synergetic, $c_0^g = 0.9, c_1^g = c^b = 0.1$.

Fix $|F| = 5000$, with 20 global and 80 heterogeneous markers, 2000 high-variance non-markers, $c = 2, k = 5, 10$, or 20, $\rho_0, \rho_1 = 0.1, 0.5$, or 0.9, $\sigma_0 = 0.25$, and $\sigma_1 = 0.64$. We use CMNC-OBF, 2MNC-Robust, t-test, Mutual Information (MI), Bhattacharyya distance (BD), and a two-stage method selecting 300 features using BD, then using SFS with bolstered error estimation [9] and regularized linear discriminant analysis (BD300-SFS-RLDA) to detect markers. Bayesian methods use Jeffreys prior, and all methods select 100 features. Figure 3 plots the average performance under each mean type across all 9 combinations of correlation values and all 3 values of k over 1000 iterations as sample size increases from 10 to 100, and the worst-case performance across all settings. CMNC-OBF has the best average performance under redundant markers and the best worst-case performance, with similar performance to BD. 2MNC-Robust has the best average performance under synergetic and marginal markers, and remains competitive in other cases.

6. REAL DATA ANALYSIS

Data obtained in [10] is curated on Gene Expression Omnibus (GEO) [11] with accession number GSE17538, containing

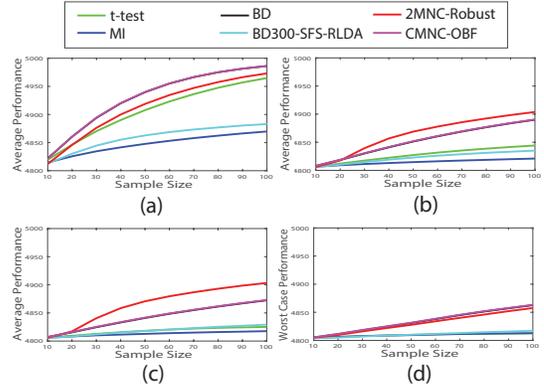


Fig. 3: Average and worst-case performance on synthetic microarray data: (a) redundant, average; (b) synergetic, average; (c) marginal, average; (d) all means types, worst-case.

Table 1: Over-represented pathways of colon cancer dataset

CMNC-OBF gene set		2MNC-Robust gene set	
Pathway Name	p-value	Pathway Name	p-value
Cadherin sig. p.w.	$1.14e-10$	Ionotropic glutamate rec. p.w.	$3.24e-3$
WNT sig. p.w.	$3.55e-7$	Axon guidance med. by Slit/Robo	$8.35e-3$
Plasminogen activating cascade	$2.54e-6$	EGF receptor sig. p.w.	$2.82e-2$
Integrin sig. p.w.	$1.02e-4$	Cadherin sig. p.w.	$4.13e-2$
Angiogenesis	$2.53e-4$	Heterotrimeric G-prot. sig. p.w....	$4.18e-2$
Gonadotropin-rel. hor. rec. p.w.	$3.99e-4$	ATP synthesis	$4.56e-2$
Blood coagulation	$4.79e-4$	Gamma-aminobutyric acid synthesis	$5.06e-2$
CCKR sig. map	$8.94e-4$	Histidine biosynthesis	$6.91e-2$
Alzheimer disease-presenilin p.w.	$1.21e-3$	Blood coagulation	$7.73e-2$
Beta2 adrenergic rec. sig. p.w.	$4.27e-3$	Betal adrenergic rec. sig. p.w.	$8.25e-2$

238 patients in 4 stages of colon cancer. We assign 28 stage 1 patients to class 0, and the remaining 210 patients to class 1. 2MNC-Robust and CMNC-OBF using Jeffreys prior pick the top 2000 genes, and enrichment analysis is performed using PANTHER [12, 13]. PANTHER pathways recognize 176 and 251 of the genes selected by 2MNC-Robust, and CMNC-OBF, respectively. Table 1 lists the top 10 pathways from CMNC-OBF and 2MNC-Robust gene sets. Literature review indicates most of the top pathways, such as the cadherin signaling and Ionotropic glutamate receptor pathways, are suggested to be involved in colon cancer [14, 15, 16, 17].

7. CONCLUSION

Bayesian feature selection is a promising framework for small-sample high-dimensional data. We proposed a new fast suboptimal feature selection algorithm, 2MNC-Robust, and demonstrated robust performance of CMNC-OBF and 2MNC-Robust compared with other popular algorithms. CMNC-OBF stands out in identifying individually strong (redundant mean) or low correlation features, whereas 2MNC-Robust can identify individually weak (marginal mean) features with high correlation or marked differences in correlation between classes, which are believed to be common in microarray studies. Future work includes devising robust suboptimal methods adaptively tuning to the feature block structure.

8. REFERENCES

- [1] Chao Sima and Edward R. Dougherty, “The peaking phenomenon in the presence of feature-selection,” *Pattern Recognition Letters*, vol. 29, no. 11, pp. 1667–1674, 2008.
- [2] Paulo J. S. Silva, Ronaldo F. Hashimoto, Seungchan Kim, Junior Barrera, Leônidas O. Brandão, Edward Suh, and Edward R. Dougherty, “Feature selection algorithms to find strong genes,” *Pattern Recognition Letters*, vol. 26, no. 10, pp. 1444–1453, 2005.
- [3] Jianping Hua, Waibhav D. Tembe, and Edward R. Dougherty, “Performance of feature-selection methods in the classification of high-dimension data,” *Pattern Recognition*, vol. 42, no. 3, pp. 409–424, 2009.
- [4] Chao Sima and Edward R. Dougherty, “What should be expected from feature selection in small-sample settings,” *Bioinformatics*, vol. 22, no. 19, pp. 2430–2436, 2006.
- [5] Ali Foroughi pour and Lori A. Dalton, “Optimal Bayesian feature filtering,” in *Proceedings of the 6th ACM Conference on Bioinformatics, Computational Biology and Health Informatics*, 2015, pp. 651–652.
- [6] Ali Foroughi pour and Lori A. Dalton, “Optimal Bayesian feature selection on high dimensional gene expression data,” in *Proceedings of the 2014 IEEE Global Conference on Signal and Information Processing (GlobalSIP)*, 2014, pp. 1402–1405.
- [7] Lori A. Dalton, “Optimal Bayesian feature selection,” in *Proceedings of the 2013 IEEE Global Conference on Signal and Information Processing (GlobalSIP)*, 2013, pp. 65–68.
- [8] Kevin P. Murphy, “Conjugate Bayesian analysis of the Gaussian distribution,” Tech. Rep., University of British Columbia, Canada, 2007.
- [9] Ulisses Braga-Neto and Edward R. Dougherty, “Bolstered error estimation,” *Pattern Recognition*, vol. 37, no. 6, pp. 1267–1281, 2004.
- [10] J. Joshua Smith, Natasha G. Deane, Fei Wu, Nipun B. Merchant, Bing Zhang, Aixiang Jiang, Pengcheng Lu, J. C. Johnson, Carl Schmidt, Christina E. Bailey, Steven Eschrich, Christian Kis, Shawn Levy, M. K. Washington, Martin J. Heslin, Robert J. Coffey, Timothy J. Yeatman, Yu Shyr, and R. Daniel Beauchamp, “Experimentally derived metastasis gene expression profile predicts recurrence and death in patients with colon cancer,” *Gastroenterology*, vol. 138, no. 3, pp. 958–968, 2010.
- [11] Ron Edgar, Michael Domrachev, and Alex E. Lash, “Gene Expression Omnibus: NCBI gene expression and hybridization array data repository,” *Nucleic Acids Research*, vol. 30, no. 1, pp. 207–210, 2002.
- [12] Huaiyu Mi and Paul Thomas, “PANTHER pathway: An ontology-based pathway database coupled with data analysis tools,” *Protein Networks and Pathway Analysis*, pp. 123–140, 2009.
- [13] Huaiyu Mi, Sagar Poudel, Anushya Muruganujan, John T. Casagrande, and Paul D. Thomas, “PANTHER version 10: Expanded protein families and functions, and analysis tools,” *Nucleic Acids Research*, vol. 44, no. 1, pp. 336–342, 2016.
- [14] A. Jeanes, C. J. Gottardi, and A. S. Yap, “Cadherins and cancer: How does cadherin dysfunction promote tumor progression?,” *Oncogene*, vol. 27, no. 55, pp. 6920–6929, 2008.
- [15] Maralice Conacci-Sorrell, Inbal Simcha, Tamar Ben-Yedidia, Janna Blechman, Pierre Savagner, and Avri Ben-Ze’ev, “Autoregulation of E-cadherin expression by cadherin–cadherin interactions the roles of β -catenin signaling, Slug, and MAPK,” *The Journal of Cell Biology*, vol. 163, no. 4, pp. 847–857, 2003.
- [16] Andrzej Stepulak, Radoslaw Rola, Krzysztof Polberg, and Chrysanthy Ikonomidou, “Glutamate and its receptors in cancer,” *Journal of Neural Transmission*, vol. 121, no. 8, pp. 933–944, 2014.
- [17] Hella Luksch, Ortrud Uckermann, Andrzej Stepulak, Sandy Hendruschk, Jenny Marzahn, Susanne Bastian, Christian Staufner, Achim Temme, and Chrysanthy Ikonomidou, “Silencing of selected glutamate receptor subunits modulates cancer growth,” *Anticancer Research*, vol. 31, no. 10, pp. 3181–3192, 2011.