IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received February 12, 2019, accepted March 20, 2019, date of current version April 15, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2909071

Hi-LASSO: High-Dimensional LASSO

YOUNGSOON KIM[1], JIE HAO[2], TEJASWINI MALLAVARAPU[2], JOONGYANG PARK[3], AND MINGON KANG[2]

[1]Department of Computer Science, Kennesaw State University, Marietta, GA 30067, USA
[2]Analytics and Data Science Institute, Kennesaw State University, Kennesaw, GA 30144, USA
[3]Department of Information and Statistics, Gyeongsang National University, Jinju 660-701, South Korea

Corresponding author: Mingon Kang (mkang9@kennesaw.edu)

ABSTRACT High-throughput genomic technologies are leading to a paradigm shift in research of computational biology. Computational analysis with high-dimensional data and its interpretation are essential for the understanding of complex biological systems. Most biological data (e.g., gene expression and DNA sequence data) are high-dimensional, but consist of much fewer samples than predictors. Such high-dimension, low sample size (HDLSS) data often cause computational challenges in biological data analysis. A number of least absolute shrinkage and selection operator (LASSO) methods have been widely used for identifying biomarkers or prognostic factors in the field of bioinformatics. The LASSO solution has been improved through the development of the LASSO derivatives, including elastic-net, adaptive LASSO, relaxed LASSO, VISA, random LASSO, and recursive LASSO. However, there are several known limitations of the existing LASSO solutions: multicollinearity (particularly with different signs), subset size limitation, and the lack of the statistical test of significance. We propose a high-dimensional LASSO (Hi-LASSO) that theoretically improves a LASSO model providing better performance of both prediction and feature selection on extremely high-dimensional data. The Hi-LASSO alleviates bias introduced from bootstrapping, refines importance scores, improves the performance taking advantage of *global* oracle property, provides a statistical strategy to determine the number of bootstrapping, and allows tests of significance for feature selection with appropriate distribution. The performance of Hi-LASSO was assessed by comparing the existing state-of-the-art LASSO methods in extensive simulation experiments with multiple data settings. The Hi-LASSO was also applied for survival analysis with GBM gene expression data.

INDEX TERMS Hi-LASSO, LASSO, random LASSO, high-dimensional data, variable selection.

I. INTRODUCTION

High-throughput genomic technologies are leading to a paradigm shift in research of computational biology. Complex and diverse collections of high-dimensional genomic data sets have been generated in various large omics projects, e.g., The Cancer Genome Atlas (TCGA) and the Cancer Genome Project in Wellcome Trust Sanger Institute (WTSI). Specifically, TCGA provides various types of genomic and sequencing data of more than 33 cancers, including gene expression, copy number variation, DNA variation, DNA methylation, and microRNA. Most of such biological data are high-dimensional. However, there are often less samples available than predictors in the biological data. For instance, gene expression data include more than ten thousands of gene profiles from hundreds of patients. Although much research has been conducted with such high-dimensional biological data, effective and computationally feasible analysis of such data has still remained challenging.

Least Absolute Shrinkage and Selection Operator (LASSO) [1] and its derivatives, such as elastic-net [2] and adaptive LASSO [3], have been widely considered for the high-dimensional data analysis. Given a data set that consists of n observations $\{(\boldsymbol{x_i}, y_i)|1 \leq i \leq n\}$, where $\boldsymbol{x_i} = (x_{i1}, \ldots, x_{ip})$ is a p-dimensional vector of predictors and y_i is a response variable, a linear regression model is written as:

$$y_i = \boldsymbol{\beta x_i} + \epsilon_i, \quad i = 1, \ldots, n, \tag{1}$$

where $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)$ is a p-dimensional vector of regression coefficients and ϵ_i is a random error term which is assumed to be independently and identically normally distributed with mean of zero and variance of σ^2. It assumes that the response is mean-corrected and the predictors are

The associate editor coordinating the review of this manuscript and approving it for publication was Dongxiao Yu.

standardized, so the intercept term is not included in the model. The notations in Table 1 are used throughout this paper.

TABLE 1. Notations.

Notation	Description
n	a number of observations
p	a number of predictors (features)
$\boldsymbol{y}$	a vector of response variable (target)
$y_i, i = \{1, 2, \cdots, n\}$	an observed value of $\boldsymbol{y}$
$x_j, j = \{1, 2, \cdots, p\}$	a predictor (feature)
$x_{ij}, i = 1, 2, \cdots, n$	a value of predictor x_j in the i-th observation
$\boldsymbol{x}_i = (x_{i1}, x_{i2}, \ldots, x_{ip})$	a feature vector of i-th observation
$(y_i, \boldsymbol{x}_i), i = 1, 2, \ldots, n$	a data set
$\boldsymbol{\beta} = \{\beta_1, ..., \beta_p\}$	a vector of coefficients
$\beta_i, i = \{1, 2, \cdots, p\}$	a coefficient value of $\boldsymbol{\beta}$

LASSO is a feature selection approach based on a linear regression model with L_1-norm regularization:

$$\min_{\boldsymbol{\beta}} \sum_{i=1}^{n} \left(y_i - \sum_{j=1}^{p} x_{ij}\beta_j \right)^2 + \lambda \sum_{j=1}^{p} \left|\beta_j\right|, \tag{2}$$

where λ is a non-negative hyper-parameter. Although LASSO has been successfully used in high-dimensional data, LASSO has the two limitations in practice [2]. Firstly, LASSO selects features at most sample size n when sample size is less than the feature size, which consequently some critical features may be excluded in the model. Secondly, LASSO tends to identify only one or a few features from variables highly correlated with each other in the model. However, highly correlated variables are commonly observed in biological systems. Most biological components involve complex interactions with others. For instance, genes in the same pathway may be highly correlated.

To tackle the limitations of LASSO, several LASSO derivatives, such as elastic-net [2], adaptive LASSO [3], relaxed LASSO [4] and VISA [5], have been proposed. Elastic-net is a penalized regression with the mixture of the L_1-norm and L_2-norm penalties [2]:

$$\min_{\boldsymbol{\beta}} \sum_{i=1}^{n} \left(y_i - \sum_{j=1}^{p} x_{ij}\beta_j \right)^2 + \lambda_1 \sum_{j=1}^{p} \left|\beta_j\right| + \lambda_2 \sum_{j=1}^{p} {\beta_j}^2, \tag{3}$$

where λ_1 and λ_2 are non-negative hyper-parameters. Owing to the nature of the L_2-norm regularization, which is a ridge regression penalty, the number of selected variables is no longer limited by the sample size. However, the ridge penalty forces the coefficient estimation of highly correlated variables with different signs to be close to each other.

Adaptive LASSO with weighted L_1-norm penalization is proposed [3]:

$$\min_{\boldsymbol{\beta}} \sum_{i=1}^{n} \left(y_i - \sum_{j=1}^{p} x_{ij}\beta_j \right)^2 + \lambda \sum_{j=1}^{p} w_j \left|\beta_j\right|, \tag{4}$$

where w_j is the weight for the predictor j. Adaptive LASSO enjoys the oracle properties by utilizing the adaptively weighted L_1 penalty. The weight computed by $1/|\hat{\beta}|^r$, where r is a positive number and $\hat{\beta}$ can be estimated by ordinary least squares (OLS) estimator, ridge regression estimator, or univariate estimator. Adaptive LASSO shrinkage leads to a near-minimax optimal estimator. However, adaptive LASSO also suffers from the multicollinearity problem, when the weights are estimated based on OLS estimators.

Relaxed LASSO performs model selection and coefficient shrinkage with the two hyper-parameters of λ and ϕ [4]:

$$\min_{\boldsymbol{\beta}} \sum_{i=1}^{n} \left(y_i - \sum_{j=1}^{p} x_{ij}\left\{\beta_j \times M_j\right\} \right)^2 + \phi\lambda \sum_{j=1}^{p} \left|\beta_j\right|, \tag{5}$$

where M_j indicates that the predictor j was estimated as non-zero or zero by LASSO. The hyper-parameter λ controls the number of predictors of non-zero coefficients in the model, whereas the hyper-parameter ϕ determines a level of shrinkage on the selected predictors. Relaxed LASSO produces a sparse model avoiding overshrinkage on non-zero coefficients and outperforms LASSO when the number of predictors is large relative to the sample size.

Variable Inclusion and Shrinkage Algorithms (VISA) improves overshrinkage problems of LASSO using a path algorithm [5]. VISA adapts two tuning parameters for variable selection and coefficient shrinkage as well as relaxed LASSO. A parameter divides variables into two groups. The first group has higher preference than the second for model inclusion, but variables in the second group may still have a chance to be selected if significant.

Recently, LASSO solutions based on bootstrapping, such as random LASSO and recursive random LASSO, have been proposed for extremely high-dimensional data analysis where the number of predictors are much greater than the number of samples [6], [7]. Random LASSO implements the solution based on a bootstrapping regression modeling [6]. Random LASSO consists of two procedures: (a) approximating weights of variables by drawing bootstrap samples and randomly selecting subset of variables and then (b) estimating the coefficients by weighted bootstrapping techniques. Random LASSO deals with multicollinearity of different signs and is able to select more variables than the sample size. On the other hand, random LASSO has an extremely high-computational cost due to the intensive bootstrapping procedures and lack of solid statistics to determine the optimal threshold for feature selection.

Recursive random LASSO is implemented based on recursive bootstrapping, where it simultaneously generates the importance scores and performs regression modeling [7]. It also proposed a parametric statistical test to statistically select predictor variables in bootstrap regression modeling. However, the first random bootstrapping tends to introduce extreme bias for feature selection, which makes it often fail to identify significant features. Moreover, the probability that predictors are non-zero is underestimated in the significance test, due to setting the unselected predictors to zero.

In this paper, we develop a novel LASSO method, named Hi-LASSO, to improve the LASSO solutions for extremely high-dimensional data. The main contributions of Hi-LASSO are as following:

- Rectifying systematic bias introduced by bootstrapping,
- Refining the computation for importance scores,
- Providing a statistical strategy to determine the number of bootstrapping,
- Taking advantage of *global* oracle property, and
- Allowing tests of significance for feature selection with appropriate distribution.

The manuscript is organized as follows. In Section II, we propose our method Hi-LASSO comparing with random LASSO. In Section III, we present the experimental results with simulation data and compare the performance with state-of-the-art LASSO methods. We demonstrate our biological findings analyzed with gene expression data in Glioblastoma Multiforme (GBM) by Hi-LASSO in Section IV.

II. METHODS

In this section, we first describe random LASSO that our proposed method is based on in detail and then elucidate our proposed method, High-dimensional LASSO (Hi-LASSO), comparing with random LASSO.

A. RANDOM LASSO

The two-step bootstrap procedures of random LASSO are described in Algorithm 1. The importance scores of predictors are computed by bootstrapping in the Procedure I. Then, Procedure II estimates coefficients of a linear model using weighted bootstrapping with the importance scores, where predictors having higher importance scores have higher chances to be selected than lower ones. The final estimation of the coefficients is computed by taking the averages of multiple estimates from bootstrapping.

Although random LASSO advances the LASSO solution for extremely high-dimensional data, there are still several issues in question. Firstly, random LASSO sets the coefficients of unselected predictors to zero, while bootstrapping at **1.c** and **2.c** in Algorithm 1. The unselected predictors can be possibly estimated as non-zero coefficients if they are selected in the bootstrapping. Therefore, it would introduce systematic bias regardless its importance of the predictors. Moreover, the lower bootstrapping number of q_1 or q_2 would generate the more systematic bias. Note that q_1 and q_2 directly affect computational costs in random LASSO.

Secondly, random LASSO does not take advantage of *global* oracle property. Although random LASSO uses bootstrapping with weights being proportional to importance scores of predictors in **2.b**, the final coefficients are estimated without the weights (i.e. oracle property). Random LASSO may be able to adopt adaptive LASSO to fully get oracle property. However, adaptive LASSO takes *local* weights of each bootstrapping sample in random LASSO, where the *local* oracle property may vary depending on what other predictors are considered together in the model.

Algorithm 1 Random LASSO

Procedure I: Computing importance scores for predictors.

1.a: Draw B numbers of bootstrap samples of size $n^{(1)}$ by sampling with replacement, i.e., $n^{(1)} \leq n$.
1.b: Randomly select q_1 predictors on each bootstrap sample.
1.c: Apply LASSO to estimate the optimal coefficients $\{\hat{b}_{ij}^{(1)} | i = 1, \ldots, B, j = 1, ..., p\}$, where the coefficients of $(p - q_1)$ numbers of unselected predictors are considered as zeros.
1.d: Compute the importance of the predictors by $I_j = |\sum_{i=1}^{B} \hat{b}_{ij}^{(1)}|/B$.
Procedure II: Selecting variables.
2.a: Draw another set of B bootstrap samples with size $n^{(2)} \leq n$ by sampling with replacement.
2.b: Randomly select q_2 predictors with the probability which is proportional to its importance scores I_j and
2.c: Apply LASSO (or Adaptive LASSO) to estimate the optimal coefficients $\{\hat{b}_{ij}^{(2)} | i = 1, \ldots, B, j = 1, ..., p\}$, where the coefficients of $(p - q_2)$ numbers of unselected predictors are considered as zeros.
2.d: Finally, compute the estimate of the coefficients by $\hat{\beta}_j = \sum_{i=1}^{B} \hat{b}_{ij}^{(2)}/B$.
2.e: Select non-zero coefficient predictors such that $|\hat{\beta}_j| > t$ for some threshold t.

Finally, random LASSO does not provide a statistical approach to select statistically significant predictors. Random LASSO considers a heuristic threshold without statistical test, although the results of the feature selection substantially depend on the threshold.

B. HIGH-DIMENSIONAL LASSO

We develop Hi-LASSO that tackles the aforementioned limitations of random LASSO and improves the LASSO model. We elucidate the contributions of Hi-LASSO comparing to random LASSO in this section.

1) IMPORTANCE SCORES AND COEFFICIENT ESTIMATES

Hi-LASSO rectifies the systematic bias that random LASSO presents, while refining the computation of the importance scores. Random LASSO forces the coefficient estimates of unselected predictors to zeros at **1.c** in Algorithm 1, which introduces systematic bias to compute the importance scores at **1.d**. To prevent the systematic bias, Hi-LASSO considers the coefficient estimates of the unselected predictors as missing values on each bootstrap sample in the Procedure I.

Furthermore, Hi-LASSO amplifies importance scores of significant variables by averaging absolute coefficients, so that the variables can fully enjoy the *global* oracle property. The coefficient estimation of a predictor varies depending on what other predictors are considered together in the model, because a regression coefficient describes the effect of a

predictor with other predictors in the model. Thus, the coefficient of a predictor may have a different value or opposite sign with its estimate in different linear models with other predictors. Specifically, multicollinearity with different signs may often cause coefficient estimates of different signs over bootstrap samples. For instance, suppose that a linear model contains three variables $\boldsymbol{x}_1$, $\boldsymbol{x}_2$, and $\boldsymbol{x}_3$, and $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$ are highly correlated but with opposite signs. When $q_1 = 2$, suppose the first bootstrap sample includes the two variables of $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$, and the second bootstrap sample includes $\boldsymbol{x}_2$ and $\boldsymbol{x}_3$. Then, regression models for the bootstrap samples can be written as:

$$\boldsymbol{y} = \beta_0 + \beta_1 \boldsymbol{x}_1 + \beta_2 \boldsymbol{x}_2 + \boldsymbol{\epsilon}, \tag{6}$$

$$\boldsymbol{y} = \beta_0 + \beta_2 \boldsymbol{x}_2 + \beta_3 \boldsymbol{x}_3 + \boldsymbol{\epsilon}. \tag{7}$$

Although both bootstrap samples share the variable of $\boldsymbol{x}_2$, the estimate of the coefficient of β_2 would be different in the two models. Strictly speaking, the above models should be written as:

$$\boldsymbol{y} = \beta_0 + \beta_1 \boldsymbol{x}_1 + \beta_2 \boldsymbol{x}_2 + \boldsymbol{\epsilon}, \tag{8}$$

$$\boldsymbol{y} = \beta_0 + \beta_2' \boldsymbol{x}_2 + \beta_3' \boldsymbol{x}_3 + \boldsymbol{\epsilon}, \tag{9}$$

where $\beta_2 \neq \beta_2'$ or $\text{sign}(\beta_2) \neq \text{sign}(\beta_2')$. Moreover, elastic-net estimates the highly correlated variables with the same sign, and LASSO picks only one dominant variable while making the another one zero. Therefore, taking the absolute value of the sum of the coefficient estimates of bootstrap samples may reduce the importance score.

Hence, Hi-LASSO computes the importance score by taking the sum of absolute coefficient estimates of bootstrap samples. Suppose that $\boldsymbol{\kappa}_j^{(\ell)}$ denote a set of indices of bootstrap samples that include the j-th predictor in the ℓ-th procedure ($\ell = 1$ or 2), and $|\boldsymbol{\kappa}_j^{(\ell)}|$ be the size of $\boldsymbol{\kappa}_j^{(\ell)}$. Then, the importance score of a variable in Hi-LASSO is computed by the average of absolute coefficients avoiding systematic bias:

$$I_j = \sum_{i \in \boldsymbol{\kappa}_j^{(1)}} \left| \hat{b}_{ij}^{(1)} \right| \Big/ \left| \boldsymbol{\kappa}_j^{(1)} \right|. \tag{10}$$

Then, the final estimate of coefficients is finally computed by:

$$\hat{\beta}_j = \sum_{i \in \boldsymbol{\kappa}_j^{(2)}} \hat{b}_{ij}^{(2)} \Big/ \left| \boldsymbol{\kappa}_j^{(2)} \right|. \tag{11}$$

Note that the final estimate of the coefficient does not take the absolute coefficient values of the bootstrapping.

2) *GLOBAL* ORACLE PROPERTY

Hi-LASSO adopts adaptive LASSO in Procedure II to take advantage of *global* oracle property obtained from the importance score in Procedure I. Adaptive LASSO [3] optimizes with the weights of the coefficients as:

$$\min_{\beta_1,\ldots,\beta_p} \sum_{i=1}^{n} \left(y_i - \sum_{j=1}^{p} x_{ij} \beta_j \right)^2 + \lambda \sum_{j=1}^{p} w_j \left| \beta_j \right|, \tag{12}$$

where $w_j = 1/\left|\hat{\beta}_j^{(OLS)}\right|^{\gamma}$ and $\hat{\beta}_j^{(OLS)}$ is the ordinary least squares estimate of the j-th predictor. Adaptive LASSO tends to exclude insignificant predictors of low weights from the model.

Random LASSO selects q_2 predictors with weights of I_j at **2.b** in Algorithm 1, so the predictors of the higher importance scores would have a higher chance to be selected in the bootstrap samples. Hi-LASSO also randomly selects q_2 predictors with selection probabilities proportional to the importance scores, but Hi-LASSO takes the importance scores into consideration for the final coefficient estimates. The importance scores, which measure the global importances of the variables, provide *global* oracle property to adaptive LASSO in Procedure II. In Hi-LASSO, the weights of the adaptive LASSO are introduced as $w_j = 1/I_j$.

3) THE NUMBER OF BOOTSTRAP SAMPLE (B)

The determination of the number of bootstrap samples (B) is crucial to ensure the performance, specifically for high-dimensional data. However, there has been no statistical guideline suggested yet. For instance, some predictors may never be considered due to the nature of random sampling, no matter how important they are in the model. Therefore, B should be large enough so that all predictors are considered at least sufficient times (L).

The probability that a predictor is included in the bootstrap sample of q predictors is:

$$\frac{\binom{p-1}{q-1}}{\binom{p}{q}} = \frac{q}{p}. \tag{13}$$

Then, the expected value how many times a variable is selected with B is:

$$E\left[\left| \boldsymbol{\kappa}_j^{(\ell)} \right| \right] = B \times \frac{q}{p}, \quad \text{for all } j. \tag{14}$$

In order to make a predictor selected at least L times on average ($E\left[\left| \boldsymbol{\kappa}_j^{(\ell)} \right| \right] \geq L$), $B \geq \frac{Lp}{q}$.

Moreover, $\left| \boldsymbol{\kappa}_j^{(\ell)} \right| / B$ can be considered as an estimate of q/p from Eq. (14). Then, it is equivalent to decide sampling size of population proportion. Hence, we can obtain B with a confidence level $(1-\alpha)$ given an estimate error d so that $\left| \frac{\left| \boldsymbol{\kappa}_j^{(\ell)} \right|}{B} - \frac{q}{p} \right| \leq d$ as:

$$B \geq \frac{z_{\alpha/2}^2 \frac{q}{p} \left(1 - \frac{q}{p}\right)}{d^2}, \tag{15}$$

where $z_{\alpha/2}$ is the $(1-\alpha/2)\,100\%$ percentile of the standard normal distribution.

4) TESTS OF STATISTICAL SIGNIFICANCE FOR FEATURE SELECTION

Tests of significance are essential for identifying statistically significant factors. However, conventional statistics typically

does not provide a solution for high-dimensional data, specifically when the number of samples is much less than the number of predictors.

Recursive random LASSO has suggested a parametric statistical test for gene selection based on the results of bootstrap regressions [7]. Let $\boldsymbol{D} = \{d_{ij} | 1 \leq i \leq B, 1 \leq j \leq p\}$ be a $B \times p$ binary matrix obtained from bootstrap samples, each of which is one if coefficient estimate is non-zero and zero otherwise:

$$d_{ij} = \begin{cases} 1, & \hat{b}_{ij} \neq 0 \\ 0, & \text{otherwise.} \end{cases} \tag{16}$$

Recursive random LASSO assumes that d_{ij} is a random variable that follows Bernoulli distribution, $B(1, \pi)$, where π is a probability that $\hat{b}_{ij}$ is nonzero. The probability π can be estimated as follow:

$$\hat{\pi} = \frac{1}{p \times B} \sum_{j=1}^{p} \sum_{i=1}^{B} d_{ij}, \tag{17}$$

which indicates an average of the selection ratio of all predictor variables in B bootstrap samples. Then, the sum of d_{ij} for the j-th variable, i.e., $\sum_{i=1}^{B} d_{ij}$, may follow binomial distribution $B(B, \hat{\pi})$. Finally, p-value can be obtained from the binomial distribution.

However, binomial distribution assumes that all independent trials have an identical probability π. In recursive random LASSO, π is a selection probability of the predictor, which does not follow the assumption. The initial selection in recursive random LASSO determines the selection probability in the remaining bootstrap procedures, so the selection probabilities of initially selected predictors are extremely higher than others. On the other hand, predictors, which are not initially selected, may be seldom chosen in the remaining bootstrapping. Hi-LASSO relieves the problems when performing the test of statistical significance, although Hi-LASSO also does not have same selection probabilities on all predictors due to weighted bootstrapping in Procedure II. The detail procedures of Hi-LASSO is described in Algorithm 2.

Algorithm 2 Hi-LASSO

Procedure I: Computing importance scores for predictors.

1.a: Draw B bootstrap samples with size n by sampling with replacement, where $B \geq \left(z^2_{\alpha/2} \frac{q_1}{p} \left(1 - \frac{q_1}{p}\right)\right) / d^2$.
1.b: Randomly select $q1$ predictors ($q1 \leq n$) on each bootstrap sample.
1.c: Estimate $\{\hat{b}^1_{ij} | i = 1, \ldots, B, j = 1, ..., p\}$, where the coefficients of the unselected predictors ($p - q_1$) are considered as missing.
1.d: Compute the importance scores by $I_j = \sum_{i \in \kappa^1_j} \left|\hat{b}^1_{ij}\right| / \left|\kappa^1_j\right|$.
Procedure II: Selecting variables.
2.a: Draw another set of B bootstrap samples with size n by sampling with replacement.
2.b: Randomly select q_2 predictors with a probability proportional to its importance scores I_j on each bootstrap sample.
2.c: Apply Adaptive LASSO with $w_j = I_j^{-1}$ to estimate $\{\hat{b}^2_{ij} | i = 1, \ldots, B, j = 1, ..., p\}$, where the coefficients of the unselected predictors ($p - q_2$) are considered as missing.
2.d: Finally, compute the estimate of the coefficients by $\hat{\beta}_j = \sum_{i \in \kappa^2_j} \hat{b}^2_{ij} / \left|\kappa^2_j\right|$.
2.e: Select significant predictors with the significance level α (e.g., 0.05 or 0.01).

III. SIMULATION STUDIES

We conducted extensive simulation experiments to assess Hi-LASSO and to compare the performance with several state-of-the-art LASSO methods. In the simulation studies, we focused on the two characteristics for the setting of data; (a) the number of features is much larger than the number of samples, and (b) features are highly correlated but with identical or opposite signs. We carried out experiments with various simulation data with respect to the measurements: Relative Model Error (RME), Root Mean Square Error (RMSE), and F1 score. We repeated each simulation experiment ten times for reproducibility.

We considered the simulation data with four hypothesis models. The simulated datasets were basically generated from the following linear regression model,

$$\boldsymbol{y} = \beta_1 \boldsymbol{x_1} + \beta_2 \boldsymbol{x_2} + \cdots + \beta_p \boldsymbol{x_p} + \boldsymbol{\epsilon}, \tag{18}$$

where $\epsilon \sim \mathcal{N}(0, \sigma^2)$ and $\boldsymbol{x}_i \sim \mathcal{N}(0, 1)$. Coefficient parameters of ground truth are given in the simulation data, which makes it possible to evaluate the performance. For the multicollinearity, we set a covariance matrix (Σ) on each simulation data, so the data sets contain a number of variables that are highly correlated with both identical and opposite sign.

Dataset I consists of 100 variables, where the first ten coefficients are non-zeros and the remaining coefficients are all zeros. The regression coefficients of ground truth were defined as:

$$\boldsymbol{\beta} = (3, 3, -3, 2, 2, -2, 1.5, 1.5, 1.5, -1.5, 0, \cdots, 0). \tag{19}$$

The pairwise correlations between the first three variables were set to be 0.9, and the same correlation structure was repeatedly set for the next three and four variables. The remaining 90 variables were designed independent from each other. Then, the independent variables $(\boldsymbol{x_1}, \ldots, \boldsymbol{x_p})$ were generated from the multivariate normal distribution with zero mean and the covariance matrix of:

$$\begin{bmatrix} \boldsymbol{\Sigma}^3_{0.9} & 0 & 0 & 0 \\ 0 & \boldsymbol{\Sigma}^3_{0.9} & 0 & 0 \\ 0 & 0 & \boldsymbol{\Sigma}^4_{0.9} & 0 \\ 0 & 0 & 0 & \mathbf{I}^{90} \end{bmatrix}, \tag{20}$$

TABLE 2. Description of the simulation data.

	p	β	Training (n_{tr})	Validation (n_{val})	Test (n_{te})	q_1, q_2	B	σ
Dataset I	100	(3, 3, -3, 2, 2, -2, 1.5, 1.5, 1.5, -1.5, 0, $\cdots$, 0)	50	10	10	50	40	3
Dataset II	1,000	$\beta_1, \cdots, \beta_{50} \sim \mathcal{N}(0, 4)$, others zero	100	20	20	100	198	3
Dataset III	10,000	$\beta_1, \cdots, \beta_{50} \sim \mathcal{N}(0, 4)$, others zero	200	40	40	200	1,000	3
Dataset IV	10,000	$\beta_1, \cdots, \beta_{50} \sim \mathcal{N}(0, 4)$, others zero	400	80	80	400	500	3

where $\mathbf{\Sigma}_v^k$ is a $k \times k$ matrix with unit diagonal elements and off-diagonal elements of value v, and $\mathbf{I}^k$ is an identity matrix of size k. The dependent variable ($\boldsymbol{y}$) was generated by the linear combination of the independent variables ($\boldsymbol{x_1}, \ldots, \boldsymbol{x_p}$) and the coefficients of ground truth ($\boldsymbol{\beta}$) and noise ($\boldsymbol{\epsilon}$). The Signal-to-Noise Ratios (SNR) of Dataset I was 2.9, where SNR is defined as $Var(\boldsymbol{X}'\boldsymbol{\beta})/Var(\boldsymbol{\epsilon})$.

Dataset II includes 1,000 variables, where the first 50 coefficients of non-zero were drawn from $\mathcal{N}(0, 4)$, and the remaining 950 coefficients were set to zeros. Then, the independent variables were generated from a multivariate normal distribution with zero mean and the covariance matrix of:

$$\begin{bmatrix} \mathbf{\Sigma}_{0.9}^{15} & 0 & 0 & 0 \\ 0 & \mathbf{\Sigma}_{0.9}^{15} & \mathbf{J}_{0.3} & 0 \\ 0 & \mathbf{J}_{0.3}^{T} & \Sigma_{0.9}^{20} & 0 \\ 0 & 0 & 0 & \mathbf{I}^{950} \end{bmatrix}, \tag{21}$$

where $\mathbf{J}_v$ is a matrix with all unit elements of a value v. The corresponding SNR was 4.2.

Dataset III is comprised of 10,000 variables, where the first 50 non-zero coefficients were drawn from $\mathcal{N}(0, 4)$ and the remaining 9,950 coefficients were sets to zeros. The setting of Dataset III is identical to Dataset II, but considers much more variables. The corresponding SNR was 14.4. Dataset IV considered double samples in the same model of Dataset III. The corresponding SNR was 9.5.

We compared the performance of Hi-LASSO to state-of-the-art LASSO methods by repeating ten times. We generated n samples and split the data into training, validation, and test data, where the sizes of the validation (n_{val}) and test (n_{te}) data were 20% of the training data (n_{tr}), respectively. All datasets were normalized and the dependent variables were centered. The validation data were used to find the optimal hyper-parameters. The four simulation data sets are briefly summarized in Table 2.

We considered the benchmark LASSO methods including LASSO, adaptive LASSO (Adaptive), elastic-net (Elastic), relaxed LASSO (Relaxed), random LASSO (Random$_{EE}$ and Random$_{EA}$), and recursive random LASSO (Recursive). The two letters of the subscript on random LASSO (Random) indicate regression methods used to estimate coefficients with bootstrap samples in Procedure I and II respectively, where we consider elastic-net (E) and adaptive LASSO (A). For instance, Random$_{EA}$ denotes random LASSO that uses elastic-net (E) in Procedure I and adaptive LASSO (A) in Procedure II. Hi-LASSO considers elastic-net (E) in Procedure I and adaptive LASSO (A) in Procedure II. Note that Hi-LASSO uses only adaptive LASSO in the second procedure to take advantage of the *global* oracle property. For adaptive LASSO in Random$_{EA}$, we computed the weights by ridge regression with bootstrap samples, whereas Hi-LASSO introduced the inverse of the importance scores as weights on adaptive LASSO.

The optimal hyper-parameter of L_1/L_2-norm regularization (λ) was obtained to minimize the prediction error with the validation data. The number of bootstrapping (B) was set so that each variable is selected at least twenty times on average by Eq. (15) where $\alpha = 0.05, z_{\alpha/2} = 1.96$ for random LASSO, recursive random LASSO, and Hi-LASSO. The statistical test for significance of feature was not considered in the simulation studies.

Importance scores on random LASSO, recursive random LASSO, and Hi-LASSO can be zero for some variables. The coefficients of zero force the variables not to be selected in Procedure II in random LASSO and the next iteration in recursive LASSO. Also, it causes the issue of *division by zero* when applying adaptive LASSO in the methods. We replaced the importance scores of zero by 10^{-10} to prevent the problem.

To evaluate the performance, we measured RME, RMSE and F1 score. RME shows the error between coefficients of ground truth and estimates of predictors, defined as $(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})^\top \mathbf{\Sigma}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})/\sigma^2$, where $\hat{\boldsymbol{\beta}}$ is an estimated coefficient vector, $\boldsymbol{\beta}$ is a vector of coefficients of ground truth, Σ is the covariance matrix of the predictors, and σ is the standard deviation of the error terms in the linear regression model. We measured RME$_{All}$ and RME$_{Nonzeros}$; RME$_{All}$ was computed with all predictors, whereas RME$_{Nonzeros}$ was with only variables of non-zero coefficient in ground truth. LASSO methods without bootstrapping produce the number of non-zero coefficients less than the sample size, so most coefficient estimates are zeros in high-dimensional data. RME$_{Nonzeros}$ can evaluate their performance with only the variables of non-zero coefficient in ground truth rather than the overall performance including all variables in RME$_{All}$. Note that Dataset I contains the first ten variables of non-zero coefficient and Dataset II-IV consists of the first fifty variables of non-zero. RMSE presents prediction errors between the given observation ($\boldsymbol{y}$) and the prediction from the linear model. Finally, F1 score evaluates the performance of variable selection. Confusion matrices were computed with estimates of coefficients and their ground truth. The confusion matrix is defined as:

- True Positive (TP): correctly identified non-zero coefficients as non-zero,
- False Positive (FP): incorrectly identified zero coefficients as non-zero,

TABLE 3. Experimental results with simulation data.

		LASSO	Adaptive	Elastic	Relxed	Random$_{EE}$	Random$_{EA}$	Recursive	Hi-LASSO
Dataset I	RME$_{All}$	0.9727 (0.3691)	1.1755 (0.2874)	1.0476 (0.4162)	1.0198 (0.4908)	1.4688 (0.4071)	1.0147 (0.3770)	1.2752 (0.5026)	**0.8513 (0.2837)**
	RME$_{Nonzeros}$	0.8619 (0.3796)	0.6261 (0.2478)	0.8186 (0.3508)	0.8257 (0.4283)	1.4238 (0.4024)	0.8908 (0.3564)	**0.6764 (0.2244)**	0.6949 (0.2764)
	RMSE	3.8003 (0.7547)	3.7842 (0.9627)	3.7936 (0.7171)	3.9159 (0.8523)	4.1216 (0.6056)	3.5869 (0.7351)	3.7973 (0.9150)	**3.3108 (0.7586)**
	F1 Score	0.4399 (0.1332)	0.4302 (0.0992)	0.4238 (0.1566)	0.3969 (0.1492)	0.4237 (0.9255)	0.4635 (0.0896)	0.4476 (0.1468)	**0.4781 (0.1009)**
Dataset II	RME$_{All}$	2.2148 (0.1978)	2.6570 (0.3672)	2.4190 (0.3509)	1.9685 (0.4677)	2.4894 (0.1749)	2.1987 (0.1912)	2.7472 (0.4742)	**1.8608 (0.2091)**
	RME$_{Nonzeros}$	2.1858 (0.2050)	2.3318 (0.4433)	2.3822 (0.3619)	1.9270 (0.4691)	2.4206 (0.1660)	2.1617 (0.1988)	2.4161 (0.4740)	**1.6953 (0.1651)**
	RMSE	5.4900 (0.6989)	5.7736 (0.6604)	5.5973 (0.7530)	5.2609 (0.8956)	5.7314 (0.7019)	5.5392 (0.6688)	5.7810 (0.6604)	**5.3235 (0.6611)**
	F1 Score	0.1169 (0.0310)	0.0897 (0.0360)	0.1578 (0.0696)	0.1014 (0.0417)	**0.2869 (0.1816)**	0.1183 (0.0289)	0.0858 (0.0463)	0.2291 (0.1190)
Dataset III	RME$_{All}$	1.7904 (0.2321)	3.2938 (0.7374)	1.8154 (0.2363)	**1.3929 (0.3063)**	4.7696 (0.4310)	1.9002 (0.2410)	3.4341 (0.5015)	1.7877 (0.2573)
	RME$_{Nonzeros}$	1.7432 (0.2090)	2.9248 (0.4644)	1.7366 (0.2228)	**1.3519 (0.2956)**	4.1757 (0.6076)	3.5786 (0.7165)	3.2745 (0.4557)	1.6626 (0.3630)
	RMSE	5.0761 (0.4248)	6.2527 (0.7388)	5.0877 (0.5101)	**4.6850 (0.3502)**	7.3016 (0.9653)	5.1917 (0.4767)	6.4037 (0.7044)	5.1963 (0.5177)
	F1 Score	0.3206 (0.0467)	0.2447 (0.0495)	0.3258 (0.0664)	0.3223 (0.0415)	0.4114 (0.1664)	0.3825 (0.1854)	0.1081 (0.0112)	**0.5147 (0.0664)**
Dataset IV	RME$_{All}$	1.7360 (0.0980)	1.8232 (0.2795)	1.7373 (0.1031)	1.5143 (0.1150)	4.3766 (0.3483)	1.5573 (0.1226)	3.3670 (2.1756)	**1.3592 (0.2291)**
	RME$_{Nonzeros}$	1.7275 (0.1007)	1.7067 (0.2002)	1.7232 (0.0998)	1.5097 (0.1125)	4.3766 (0.3483)	1.5162 (0.1196)	3.2787 (1.9625)	**1.2997 (0.1430)**
	RMSE	4.9911 (0.4385)	5.0371 (0.5949)	4.9791 (0.4502)	4.7476 (0.3583)	7.0131 (0.8054)	4.8094 (0.4572)	6.2664 (1.8079)	**4.5800 (0.4066)**
	F1 Score	0.2930 (0.0503)	0.3823 (0.0775)	0.3145 (0.0528)	0.2775 (0.0411)	0.5860 (0.1290)	0.4220 (0.1746)	0.1000 (0.0299)	**0.6085 (0.0483)**
Total	RME$_{All}$	1.6799 (0.5109)	2.2374 (0.9263)	1.7548 (0.5711)	1.4739 (0.4985)	3.2631 (1.4024)	1.6989 (0.5223)	2.7059 (1.4261)	**1.4648 (0.4716)**
	RME$_{Nonzeros}$	1.6296 (0.5400)	1.8974 (0.9281)	1.6652 (0.6245)	1.4036 (0.5250)	3.2346 (1.4276)	1.6389 (0.5477)	2.4114 (1.4683)	**1.3381 (0.4746)**
	RMSE	4.8394 (0.8578)	5.2119 (1.1880)	4.8644 (0.9163)	4.6523 (0.8045)	6.0291 (1.4597)	4.7909 (0.9540)	5.5621 (1.5107)	**4.6026 (0.9927)**
	F1 Score	0.2926 (0.1381)	0.2867 (0.1501)	0.3055 (0.1332)	0.2745 (0.1358)	0.4356 (0.1890)	0.3270 (0.1750)	0.1854 (0.1711)	**0.4576 (0.1653)**

Note: The average of the experiments are shown, where bold-face indicates the best performance and parenthesis indicates the standard error.

TABLE 4. Statistical assessment by Wilcoxon signed-rank test.

	RME$_{All}$	RME$_{Nonzeros}$	RMSE	F1 Score
Hi-LASSO vs Lasso	1.50e-04*	1.54e-06*	1.05e-03*	8.09e-07*
Hi-LASSO vs Adaptive	4.55E-11*	1.16e-07*	9.75e-10*	2.43e-07*
Hi-LASSO vs Elastic	8.85e-06*	2.92e-06*	7.14e-04*	1.70e-06*
Hi-LASSO vs Relaxed	6.27e-01	7.03e-02	4.85e-01	4.73e-07*
Hi-LASSO vs Random$_{EE}$	3.64e-12*	1.82e-12*	3.07e-10*	3.37e-01
Hi-LASSO vs Random$_{EA}$	1.46e-07*	5.58e-10*	2.64e-04*	1.39e-05*
Hi-LASSO vs Recursive	6.00e-11*	2.36e-08*	4.55e-11*	5.23e-07*

Note: the asterisk (*) indicates statistical significance, i.e., p-value < 0.05.

- False Negative (FN): incorrectly identified non-zero coefficients as zero, and
- True Negative (TN): correctly identified zero coefficients as zero.

Then, F1 score was calculated by 2(PPV $\times$ TPR) / (PPV + TPR), where TPR $=$ TP/(TP + FN) and PPV $=$ TP/(TP + FP).

The experimental results are shown in Table 3. The average of the experiments are shown, where bold-face indicates the best performance and parenthesis indicates the standard error. Overall, Hi-LASSO outperformed others on most of the datasets. Hi-LASSO produced the lowest RME$_{All}$ of 1.4648$\pm$0.4716, RME$_{Nonzeros}$ of 1.3381$\pm$0.4746, and RMSE of 4.6026$\pm$0.9927 and the highest F1 score of 0.4576$\pm$0.1653 on average in all of the experiments.

The outstanding performance of Hi-LASSO was also statistically assessed by Wilcoxon signed-rank test which is a non-parametric paired two sided test. Table 4 shows the p-values of Wilcoxon signed-rank test between Hi-LASSO and other benchmark methods, where the

TABLE 5. Estimates of coefficient and coefficient sign in Dataset I.

		Random$_{EE}$			Random$_{EA}$			Hi-LASSO		
No.	True	Estimate	+	-	Estimate	+	-	Estimate	+	-
β_1	3	0.6084 (0.3344)	10	0	1.1833 (0.9964)	8	0	**1.5324** (0.9811)	9	0
β_2	3	0.6199 (0.2681)	10	0	1.1043 (0.6241)	9	0	**1.3097** (1.0469)	8	0
β_3	-3	0.0703 (0.0726)	6	0	−0.0543 (0.2491)	1	2	**−0.1802** (0.5698)	0	1
β_4	2	0.1827 (0.1659)	7	0	0.4331 (0.5502)	5	0	**0.6110** (0.7246)	5	0
β_5	2	0.2036 (0.2548)	7	0	0.3206 (0.5280)	4	0	**0.4852** (0.7010)	6	0
β_6	-2	0.0341 (0.0396)	5	0	0.0344 (0.1089)	1	0	0	0	0
β_7	1.5	0.3457 (0.2712)	8	0	0.8146 (0.8508)	7	0	**1.1057** (1.1685)	6	0
β_8	1.5	0.3464 (0.2503)	8	0	0.5852 (0.5000)	8	0	**0.8117** (0.7463)	8	0
β_9	1.5	0.2675 (0.1946)	8	0	**0.3786** (0.4549)	6	0	0.3094 (0.5059)	4	0
β_{10}	-1.5	0.0990 (0.1329)	7	0	0.0159 (0.0504)	1	0	**−0.0302** (0.0956)	0	1

TABLE 6. Estimates of coefficient and coefficient sign in Dataset II.

		Random$_{EE}$			Random$_{EA}$			Hi-LASSO		
No.	True	Estimate	+	-	Estimate	+	-	Estimate	+	-
β_1	0.5285	**0.0078** (0.0131)	3	0	0	0	0	0	0	0
β_2	-2.3780	0.0029 (0.0090)	1	0	0	0	0	0	0	0
β_3	-0.8379	0.0042 (0.0133)	1	0	0	0	0	0	0	0
β_4	-1.2074	0.0000 (0.0201)	2	1	0	0	0	0	0	0
β_5	-0.7173	0.0039 (0.0094)	2	0	0	0	0	0	0	0
β_6	-0.1747	0.0092 (0.0132)	4	0	0	0	0	0	0	0
β_7	-2.4928	0.0008 (0.0079)	1	1	0	0	0	0	0	0
β_8	-1.0483	0.0109 (0.0182)	3	0	0	0	0	0	0	0
β_9	-2.7500	**−0.0056** (0.0198)	1	3	0	0	0	0	0	0
β_{10}	2.1934	**0.0103** (0.0171)	3	0	0	0	0	0	0	0
β_{11}	2.9586	0.0186 (0.0248)	4	0	0	0	0	**0.0610** (0.1928)	1	0
β_{12}	-1.2497	0.0050 (0.0122)	2	0	0	0	0	0	0	0
β_{13}	0.8943	**0.0119** (0.0168)	4	0	0	0	0	0	0	0
β_{14}	4.4044	**0.0171** (0.0225)	4	0	0	0	0	0	0	0
β_{15}	2.2559	**0.0141** (0.0191)	4	0	0	0	0	0	0	0
β_{16}	-0.1860	**−0.0827** (0.0167)	0	10	0	0	0	−0.0498 (0.2930)	1	1
β_{17}	-1.2135	**−0.0892** (0.0173)	0	10	−0.0428 (0.1354)	0	1	0	0	0
β_{18}	0.6491	−0.0841 (0.0163)	0	10	0	0	0	0	0	0
β_{19}	-0.1325	**−0.0779** (0.0300)	0	9	0	0	0	−0.0218 (0.0691)	0	1
β_{20}	-3.4773	−0.1043 (0.0160)	0	10	−0.7266 (0.7908)	0	5	**−1.1771** (1.1366)	0	6
β_{21}	0.8682	−0.0758 (0.0166)	0	10	0	0	0	**0.0839** (0.2654)	1	0
β_{22}	-3.3421	−0.1026 (0.0127)	0	10	−0.4090 (0.6476)	0	4	**−0.8676** (0.9740)	0	6
β_{23}	1.9083	−0.0733 (0.0140)	0	10	0	0	0	**0.1167** (0.2594)	2	0
β_{24}	0.2358	**−0.0827** (0.0183)	0	10	0	0	0	0	0	0
β_{25}	-1.3433	−0.0872 (0.0103)	0	10	0	0	0	**−0.1291** (0.3023)	0	2
β_{26}	1.8962	−0.0774 (0.0206)	0	10	−0.1042 (0.3296)	0	1	**0.0317** (0.5132)	2	1
β_{27}	-0.7924	**−0.0870** (0.0175)	0	10	0	0	0	0.0118 (0.0374)	1	0
β_{28}	1.9262	−0.0656 (0.0266)	0	9	0	0	0	**0.1664** (0.3708)	2	0
β_{29}	1.6013	−0.0637 (0.0263)	0	9	0	0	0	**0.0816** (0.2181)	2	0
β_{30}	-1.0147	−0.0914 (0.0225)	0	10	−0.1639 (0.4378)	0	2	**−0.2684** (0.4332)	0	3
β_{31}	2.1436	−0.0773 (0.0141)	0	10	0	0	0	**0.2322** (0.4910)	2	0
β_{32}	-1.1531	−0.0996 (0.0167)	0	10	−0.2007 (0.4844)	0	3	**−0.5073** (0.7380)	0	4
β_{33}	-1.1430	**−0.0933** (0.0170)	0	10	0	0	0	−0.0468 (0.1480)	0	1
β_{34}	1.1434	−0.0745 (0.0125)	0	10	0	0	0	**0.1735** (0.4410)	2	0
β_{35}	-2.9320	−0.1053 (0.0170)	0	10	−0.5269 (0.7895)	0	4	**−0.9914** (1.0633)	0	6
β_{36}	-1.7806	−0.0936 (0.0186)	0	10	0	0	0	**−0.1526** (0.4407)	0	2
β_{37}	1.1054	−0.0750 (0.0173)	0	10	0	0	0	**0.1664** (0.4370)	2	0
β_{38}	1.4061	−0.0856 (0.0190)	0	10	0	0	0	**−0.0581** (0.1838)	0	1
β_{39}	1.0178	−0.0816 (0.0150)	0	10	0	0	0	**0.1096** (0.3464)	1	0
β_{40}	-1.1676	−0.0932 (0.0173)	0	10	0	0	0	**−0.1321** (0.3373)	1	2
β_{41}	0.5193	−0.0848 (0.0351)	0	9	−0.0703 (0.2223)	0	1	0	0	0
β_{42}	1.5576	−0.0736 (0.0151)	0	10	0	0	0	**0.0674** (0.2133)	1	0
β_{43}	-2.9148	−0.1058 (0.0194)	0	10	−0.6335 (0.8554)	0	4	**−1.0599** (1.0475)	0	6
β_{44}	1.3486	−0.0747 (0.0158)	0	10	0	0	0	**0.2102** (0.3860)	3	0
β_{45}	1.7686	−0.0836 (0.0183)	0	10	0	0	0	−0.0186 (0.0587)	0	1
β_{46}	-0.9674	**−0.0954** (0.0097)	0	10	−0.0191 (0.0604)	0	1	0.0523 (0.4378)	1	1
β_{47}	-2.4344	−0.1014 (0.0155)	0	10	−0.2642 (0.6326)	0	2	**−0.2882** (0.6058)	0	3
β_{48}	-1.2931	**−0.0926** (0.0161)	0	10	0	0	0	−0.0361 (0.1141)	0	1
β_{49}	-1.8384	−0.0989 (0.0197)	0	10	−0.3062 (0.4242)	0	5	**−0.8396** (0.9188)	0	6
β_{50}	2.0790	−0.0800 (0.0225)	0	10	0	0	0	**0.0795** (0.2513)	1	0

statistical significances with a significance level of 0.05 (p-value $<$ 0.05) were proved for most of the tasks. Hi-LASSO showed statistically significant outperformance against other benchmark methods on most of the experiments. Although Hi-LASSO was not statistically significant against relaxed LASSO on RME$_{All}$ (p-value $=$ 0.627),

TABLE 7. Estimates of coefficient and coefficient sign in Dataset III.

		Random$_{EE}$			Random$_{EA}$			Hi-LASSO		
No.	True	Estimate	+	-	Estimate	+	-	Estimate	+	-
β_1	0.0479	0.1464 (0.0650)	10	0	0.0401 (0.0620)	8	1	**0.0482** (0.3251)	2	1
β_2	1.3255	0.2066 (0.0844)	10	0	0.3866 (0.3660)	9	1	**0.7988** (0.9565)	6	0
β_3	-1.5152	0.0673 (0.0419)	10	0	−0.0835 (0.0980)	0	9	**−0.7198** (0.6106)	0	7
β_4	0.7677	0.2074 (0.0895)	10	0	0.2555 (0.2128)	8	1	**0.5058** (0.8046)	5	2
β_5	-1.3114	0.1174 (0.0717)	10	0	−0.0050 (0.2599)	5	4	**−0.2844** (1.0504)	2	3
β_6	5.6381	0.5884 (0.1621)	10	0	3.0094 (0.5874)	10	0	**4.4338** (0.8946)	10	0
β_7	-0.1895	0.1707 (0.0568)	10	0	0.1631 (0.1511)	8	1	**0.2718** (0.4432)	3	0
β_8	-0.0132	0.1188 (0.0331)	10	0	0.0249 (0.0612)	5	3	**−0.0488** (0.3423)	1	1
β_9	2.8316	0.2815 (0.1533)	10	0	0.8613 (0.8223)	9	0	**1.6828** (1.3741)	8	0
β_{10}	0.0734	0.1227 (0.0803)	10	0	**0.0814** (0.1715)	4	2	0.1842 (0.4750)	2	1
β_{11}	0.6283	0.1369 (0.0681)	10	0	0.0980 (0.1419)	8	2	**0.2318** (0.5972)	2	0
β_{12}	-0.9740	0.0836 (0.0419)	10	0	−0.0841 (0.1369)	1	6	**−0.6336** (0.7403)	0	5
β_{13}	0.8331	0.1556 (0.0709)	10	0	0.1935 (0.1950)	9	0	**0.5236** (0.5965)	5	1
β_{14}	-2.3539	0.0560 (0.0253)	10	0	−0.1097 (0.0680)	0	9	**−1.1431** (0.7800)	0	8
β_{15}	-0.3433	0.1163 (0.0471)	10	0	0.0067 (0.0716)	5	5	**−0.2063** (0.5864)	1	2
β_{16}	-1.9762	−0.4654 (0.1339)	0	10	−1.1469 (0.5745)	0	10	**−2.0313** (0.8694)	0	9
β_{17}	-3.4923	−0.6163 (0.1771)	0	10	−2.1877 (0.8580)	0	10	**−3.2001** (0.9795)	0	10
β_{18}	-1.8792	−0.4344 (0.1434)	0	10	−0.9141 (0.7962)	0	10	**−1.0959** (1.1241)	0	6
β_{19}	-1.9041	−0.4602 (0.1011)	0	10	−0.9327 (0.5383)	0	10	**−1.4690** (1.2261)	0	7
β_{20}	-0.5524	−0.3157 (0.1023)	0	10	−0.3047 (0.4242)	2	7	**−0.3766** (0.9700)	2	6
β_{21}	-1.3737	−0.4175 (0.1098)	0	10	−0.6928 (0.3737)	0	10	**−1.0582** (0.6654)	0	8
β_{22}	0.8986	−0.2323 (0.0652)	0	10	0.1707 (0.3243)	7	3	**1.0639** (0.9973)	7	1
β_{23}	0.6483	−0.2241 (0.0456)	0	10	0.0932 (0.2173)	7	2	**0.4779** (0.8403)	4	0
β_{24}	2.5387	−0.2224 (0.0840)	0	10	0.3219 (0.4677)	9	1	**1.7168** (1.0654)	9	0
β_{25}	-0.9872	−0.4249 (0.1423)	0	10	**−0.8597** (0.6487)	0	9	−1.3749 (1.1434)	1	9
β_{26}	-1.5920	−0.4215 (0.1277)	0	10	−0.8469 (0.5885)	0	10	**−1.3615** (1.0356)	0	8
β_{27}	-1.3512	−0.4288 (0.1297)	0	10	−0.8794 (0.6489)	0	10	**−1.1701** (0.9533)	0	8
β_{28}	-1.2319	−0.3432 (0.0771)	0	10	−0.3974 (0.3531)	0	10	**−0.6008** (0.7653)	0	6
β_{29}	2.3743	−0.2083 (0.0797)	0	10	0.4920 (0.4916)	9	1	**2.1741** (1.3416)	9	0
β_{30}	-0.4158	**−0.2973** (0.0579)	0	10	−0.2028 (0.1871)	0	9	−0.1508 (0.5274)	1	3
β_{31}	2.2840	0.0009 (0.0026)	2	0	**0.0143** (0.0286)	3	0	0	0	0
β_{32}	-1.2235	0.0001 (0.0002)	1	0	0.0027 (0.0056)	2	0	0.1745 (0.3701)	2	0
β_{33}	1.7933	0.0017 (0.0049)	2	0	0.0517 (0.0889)	5	0	**0.5318** (0.9466)	3	0
β_{34}	-0.3724	0.0017 (0.0052)	2	0	0.0184 (0.0389)	2	0	0.3468 (0.5852)	3	0
β_{35}	-2.3045	0.0000 (0.0001)	1	0	0.0207 (0.0492)	4	0	0.0413 (0.1305)	1	0
β_{36}	-0.9752	0.0002 (0.0005)	1	0	0.0162 (0.0391)	3	0	0.0000 (0.0000)	0	0
β_{37}	2.5499	0.0161 (0.0510)	1	0	0.0053 (0.0129)	2	0	**0.3792** (0.6112)	3	0
β_{38}	-2.2791	0.0009 (0.0028)	1	0	0.0065 (0.0111)	3	0	0	0	0
β_{39}	2.3359	0.0073 (0.0202)	3	0	0.0413 (0.0699)	5	0	**0.1684** (0.5324)	1	0
β_{40}	-0.8053	0.0000 (0.0001)	1	0	0.0007 (0.0021)	1	0	0	0	0
β_{41}	-0.4587	0.0007 (0.0019)	2	0	0.0050 (0.0085)	3	0	0.2290 (0.5036)	2	0
β_{42}	-0.8091	0.0002 (0.0007)	1	0	0.0051 (0.0160)	1	0	0	0	0
β_{43}	0.1418	0.0003 (0.0010)	1	0	**0.0160** (0.0217)	4	0	0	0	0
β_{44}	-2.6257	0.0008 (0.0025)	1	0	0.0197 (0.0385)	3	0	0	0	0
β_{45}	0.5124	0.0042 (0.0122)	3	0	0.0510 (0.0805)	5	0	**0.1025** (0.3242)	1	0
β_{46}	4.1545	0.0030 (0.0084)	3	0	0.0552 (0.1417)	4	0	**0.1268** (0.4011)	1	0
β_{47}	2.9346	0.0010 (0.0024)	3	0	0.0153 (0.0308)	3	0	**0.1306** (0.4131)	1	0
β_{48}	-1.5871	0.0001 (0.0002)	1	0	0.0234 (0.0423)	3	0	0.0869 (0.1937)	2	0
β_{49}	-1.0047	0.0001 (0.0002)	1	0	0.0086 (0.0126)	4	0	0	0	0
β_{50}	0.4792	0.0002 (0.0005)	1	0	**0.0133** (0.0234)	3	0	0	0	0

RME$_{Nonzeros}$ (p-value = 0.0703), and RMSE (p-value = 0.485), Hi-LASSO was statistically significant on F1 score (p-value = 4.73e-7). In particular, the significance of Hi-LASSO on RME$_{Nonzeros}$ was assessed with a significance level of 0.1 (p-value < 0.1). This is because relaxed LASSO produces less non-zero coefficients than Hi-LASSO in high-dimensional data.

Furthermore, we examined coefficient estimates and signs in Random$_{EE}$, Random$_{EA}$, and Hi-LASSO in detail. Table 5 – 8 summarizes the estimates of the non-zero coefficients with Datasets I – IV, respectively. The coefficients of the first ten variables are non-zeros in ground truth in Dataset I, and the other datasets contain non-zero coefficients in the first fifty variables. '+' and '−' in the tables show how many numbers of positive or negative coefficients were estimated in the experiments repeated ten times, and the best coefficient estimates are indicated as a bold-face in the tables. Dataset I includes three groups of highly correlated variables ($\beta_1 - \beta_3$, $\beta_4 - \beta_6$, and $\beta_7 - \beta_{10}$) with different signs. The ground truth of β_3, β_6, and β_{10} are negative, whereas others

TABLE 8. Estimates of coefficient and coefficient sign in Dataset IV.

		Random$_{EE}$			Random$_{EA}$			Hi-LASSO		
No.	True	Estimate	+	-	Estimate	+	-	Estimate	+	-
β_1	0.8963	0.2383 (0.0341)	10	0	0.5086 (0.4946)	9	1	**0.6446** (0.6446)	5	0
β_2	-0.3388	0.1991 (0.0972)	10	0	−0.0679 (0.3799)	4	4	**−0.3485** (0.3485)	1	3
β_3	3.4622	0.4330 (0.0999)	10	0	2.3283 (0.5722)	10	0	**2.9893** (2.9893)	10	0
β_4	-1.9280	0.1546 (0.0249)	10	0	−0.8342 (0.5232)	0	10	**−1.5755** (1.5755)	0	10
β_5	0.4280	0.2248 (0.0166)	10	0	0.1903 (0.3398)	6	2	**0.3302** (0.3302)	4	0
β_6	2.5957	0.3432 (0.0734)	10	0	1.6220 (0.6555)	10	0	**2.3297** (2.3297)	10	0
β_7	0.1296	0.2098 (0.0394)	10	0	**0.0555** (0.3381)	5	5	0.0505 (0.0505)	3	2
β_8	-0.8355	0.1887 (0.0239)	10	0	−0.2211 (0.3124)	1	6	**−0.6352** (0.6352)	1	6
β_9	4.0314	0.5209 (0.1501)	10	0	2.9866 (0.8580)	10	0	**3.5330** (3.5330)	10	0
β_{10}	-1.5423	0.1576 (0.0249)	10	0	−0.8880 (0.5458)	0	10	**−1.5605** (1.5605)	0	9
β_{11}	-2.4993	0.1405 (0.0240)	10	0	−1.0399 (0.5737)	0	10	**−1.7320** (1.7320)	0	10
β_{12}	2.9854	0.3917 (0.0856)	10	0	2.0294 (0.5392)	10	0	**2.8651** (2.8651)	10	0
β_{13}	2.2903	0.3140 (0.0780)	10	0	1.2118 (0.6183)	9	0	**1.8630** (1.8630)	9	0
β_{14}	-0.8247	0.1778 (0.0291)	10	0	−0.3539 (0.3530)	1	8	**−0.6840** (0.6840)	0	6
β_{15}	-1.0764	0.1798 (0.0397)	10	0	−0.4207 (0.2926)	1	9	**−0.9477** (0.9477)	0	7
β_{16}	-0.1810	0.0268 (0.0156)	10	0	**−0.0068** (0.0162)	1	2	−0.3883 (0.3883)	1	4
β_{17}	-0.0336	0.0308 (0.0146)	10	0	0.0135 (0.0404)	3	1	0.1086 (0.1086)	2	0
β_{18}	-1.8431	0.0147 (0.0119)	10	0	−0.2271 (0.2619)	0	6	**−1.7572** (1.7572)	0	9
β_{19}	2.5745	0.1027 (0.0651)	10	0	0.6931 (0.4822)	10	0	**1.7580** (1.7580)	10	0
β_{20}	0.9166	0.0662 (0.0301)	10	0	0.1969 (0.2761)	6	0	**0.8154** (0.8154)	7	1
β_{21}	1.2671	0.0437 (0.0229)	10	0	0.0942 (0.1254)	5	1	**0.7021** (0.7021)	7	0
β_{22}	2.1809	0.0755 (0.0411)	10	0	0.5039 (0.5998)	9	0	**1.5966** (1.5966)	10	0
β_{23}	4.3567	0.1848 (0.0761)	10	0	2.1936 (0.7802)	10	0	**3.5489** (1.0812)	10	0
β_{24}	-0.3840	0.0312 (0.0186)	10	0	−0.0328 (0.0962)	1	2	**−0.1146** (.06143)	2	2
β_{25}	1.3360	0.0527 (0.0226)	10	0	0.1822 (0.2182)	7	0	**0.8397** (0.6235)	7	0
β_{26}	-1.2351	0.0171 (0.0091)	10	0	−0.0854 (0.1312)	0	5	**−0.9334** (0.8186)	0	7
β_{27}	-1.0592	0.0184 (0.0095)	10	0	−0.0913 (0.1890)	0	4	**−0.7609** (0.6710)	0	7
β_{28}	1.6842	0.0938 (0.0523)	10	0	0.5152 (0.3429)	9	0	**1.3417** (0.4649)	10	0
β_{29}	-3.9505	0.0062 (0.0049)	10	0	−0.7996 (0.5342)	0	9	**−2.4225** (1.8175)	0	7
β_{30}	-1.7889	0.0186 (0.0184)	10	0	−0.2212 (0.2090)	0	8	**−1.1939** (1.0036)	0	7
β_{31}	1.2657	0	0	0	−0.0028 (0.0054)	0	3	0	0	0
β_{32}	-2.1514	**−0.0002** (0.0006)	0	1	−0.0127 (0.0259)	0	4	0	0	0
β_{33}	1.3114	0	0	0	0	0	0	0	0	0
β_{34}	-0.1235	0	0	0	**−0.0292** (0.0855)	0	3	0	0	0
β_{35}	-0.3155	0	0	0	0	0	0	**−0.1232** (0.3896)	0	1
β_{36}	-1.3677	0	0	0	**−0.0006** (0.0020)	0	1	0	0	0
β_{37}	1.7765	0	0	0	0	0	0	0	0	0
β_{38}	-0.7332	0	0	0	**−0.0006** (0.0019)	0	1	0	0	0
β_{39}	1.9905	0	0	0	0	0	0	0	0	0
β_{40}	4.5254	0	0	0	0	0	0	0	0	0
β_{41}	-2.2227	−0.0000 (0.0001)	0	1	**−0.0098** (0.0271)	0	2	0	0	0
β_{42}	-2.4738	−0.0001 (0.0003)	0	2	**−0.0077** (0.0243)	0	1	0	0	0
β_{43}	1.2153	0	0	0	−0.0009 (0.0029)	0	1	0	0	0
β_{44}	-1.6102	0	0	0	**−0.0082** (0.0202)	0	2	0	0	0
β_{45}	-0.4553	0	0	0	**−0.0053** (0.0112)	0	2	0	0	0
β_{46}	-0.5033	0.0000 (0.0001)	0	2	**−0.0126** (0.0211	0	3	0	0	0
β_{47}	0.7253	0	0	0	−0.0050 (0.0158)	0	1	0	0	0
β_{48}	-0.2993	0	0	0	0	0	0	0	0	0
β_{49}	0.5579	0	0	0	−0.0006 (0.0018)	0	1	0	0	0
β_{50}	-2.7347	0	0	0	**−0.0062** (0.0198)	0	1	0	0	0

are all positive. Similarly, Datasets II – IV consist of three groups of highly correlated variables ($\beta_1 - \beta_{15}$, $\beta_{16} - \beta_{30}$, and $\beta_{31} - \beta_{50}$) with different signs.

Random$_{EE}$ estimated highly correlated variables with the same sign simultaneously, although some variables of them are with opposite signs. For instance, all estimates are all positive in Table 5, whereas the ground truths of β_3, β_6, and β_{10} are negatives. The coefficient estimates with the same sign are also shown with other datasets, which is a well-known limitation of elastic-net. On the other hand, Random$_{EA}$ tended to make non-dominant variables with opposite sign in multicollinearity to shrink toward zero. For instance, β_1 and β_2 may dominate the effects with a positive sign while being also highly correlated to β_3, but the ground truth of β_3 is negative in Table 5. Such variables (e.g., β_3, β_6, and β_{10}) are easily shrunk toward zeros in adaptive LASSO in Random$_{EA}$, because the coefficient estimates of the variables are relatively smaller than others in bootstrap samples and used as weights in adaptive LASSO. Most coefficient estimates of Hi-LASSO appeared to be the closest to the ground truths among the benchmark methods, which corroborates the least RME score with Hi-LASSO. Moreover, Hi-LASSO

accurately estimated the coefficients of different signs in highly correlated variables, e.g., β_3 and β_{10} in Table 5. The *global* oracle property of Hi-LASSO may relieve the problem with adaptive LASSO.

IV. GLIOBLASTOMA GENE EXPRESSION DATA ANALYSIS

Furthermore, we conducted experiments with gene expression data in Glioblastoma Multiforme (GBM) for assessing Hi-LASSO with real-world biological data. GBM is the most common primary malignant brain tumor in adults and one of the most lethal of all cancers [8]. We used the gene expression data of GBM patients at The Cancer Genome Atlas (http://cancergenome.nih.gov). The gene expression data in GBM consist of 447 patients and 12,042 genes excluding censored patients. The average survival time was 16.8 months.

Hi-LASSO was applied to the GBM data, where q_1 and q_2 were set to 447 (sample size of GBM) and B was set to 1,400 so that each gene is selected at least fifty on average in the bootstrap samples. We considered the logarithm of time to death as a response variable and standardized all of the variables. For the feature selection, we used a significance level of $\alpha = 0.05$ for testing of statistical significance on Hi-LASSO.

Hi-LASSO identified 139 genes of significance out of 12,042 genes, and we examined the genes in the biological literature. According to the biological literature, a number of genes are shown as significantly associated to survivals in GBM. Table 9 shows twenty top-ranked genes in descending order of absolute value of estimated coefficients. A Single Nucleotide Polymorphism (SNP) in the gene ZNF208 (rs8105767) was reported associated with the overall survival rates of low-grade glioma patients in Kaplan-Meier analysis [9]. PARD6B was identified as a significant gene by the methylated DNA immunoprecipitation microarray chip (MeDIP-Chip) analysis in human gliomas [10]. DEC1 appeared as a prognostic factor of glioma and response especially to temozolomide chemotherapy in high-grade glioma patients [11]. Overexpressing $VEGF_A$ and PDE4A was suggested to warrant treatment of GBM [12]. The effect of phosphodiesterase (PDE) inhibitor on glioblastoma cells was studied, and the transcripts of PDE4A and PDE4B were detected in A172 and U87MG human glioblastoma cells [13]. PDE4A was expressed in medulloblastoma, glioblastoma, oligodendroglioma, ependymoma, and meningioma [14]. Moreover, when PDE4A1 was overexpressed in Daoy medulloblastoma and U87 glioblastoma cells, in vivo doubling times were significantly shorter for PDE4A1-overexpressing xenografts compared with controls. In-vitro data showed that the purified rFGF23 can induce the phosphorylation of mitogen-activated protein kinases in the glioma U251 cell [15]. The results of in-vivo animal experiments also showed that rFGF23 could decrease the concentration in the plasma of normal rats fed with a fixed formula diet [15]. MAGEC1 expression was shown as up-regulated in the TMZ-resistant (TMZ-R) U87 glioblastoma cell line [16]. Competitive BET bromodomain inhibitors (BBIs) targeting BET proteins (BRD2, BRD3, BRD4, and BRDT) have showed promising preclinical activities against brain cancers [18].

TABLE 9. Top-20 ranked genes by Hi-LASSO in GBM.

Gene	Coefficient	Reference
ZNF208	1.0258	[9]
PARD6B	0.5334	[10]
FSHB	-0.4348	-
LIPF	-0.4181	-
NYX	0.3909	-
HCRTR2	0.3718	-
MAP3K7IP1	-0.3711	-
OSBP2	-0.3676	-
KCNQ1	0.3564	-
DEC1	0.3514	[11]
C10orf59	-0.3487	-
PDE4A	0.3366	[12], [13], [14]
FGF23	0.3189	[15]
MAGEC1	0.2696	[16], [17]
HIST1H4L	0.2693	-
BRDT	0.2661	[18]
IFNA5	-0.2424	-
PCDH11X	0.2417	
DNAH2	-0.2366	-
VAV2	0.2304	-

V. CONCLUSION

We proposed a novel High-dimensional LASSO (Hi-LASSO) for variable selection in a linear regression model with extremely high-dimensional data. Hi-LASSO refines the random LASSO solution not only to improve the performance but also to provide a theoretically elaborate LASSO method. We also suggested a statistical strategy to determine the optimal number of bootstrapping and to perform tests of significance for selecting statistically significant features.

Generating an enough number of bootstrap samples is critical for Hi-LASSO to produce reliable results. In random LASSO, predictors with higher importance scores are more often selected in the second procedure, which causes the variables to have a higher priority to estimate the coefficients. However, the biased selection distribution causes a problem of applying for test of significance. On the other hand, Hi-LASSO may require more bootstrapping numbers B, specifically in the second procedure. This would add additional computational costs. The computations with bootstrap samples are independent, so implementation of Hi-LASSO using parallel computing systems would provide efficient solutions for extremely high-dimensional data.

Hi-LASSO outperformed the state-of-the-art LASSO methods, including adaptive, elastic, relaxed, random, and recursive LASSO, with respect to relative model error, root mean square error, and F1 score in the extensive experiments with simulation data and Glioblastoma gene expression data. Through the experiments, Hi-LASSO showed that Hi-LASSO not only estimates the true model accurately but also performs feature selection effectively.

ACKNOWLEDGMENT

(Joongyang Park and Mingon Kang are co-senior authors.)

REFERENCES

[1] R. Tibshirani, "Regression selection and shrinkage via the lasso," *J. Roy. Stat. Soc. B*, vol. 58, no. 1, pp. 267–288, 1996.
[2] H. Zou and T. Hastie, "Regularization and variable selection via the elastic net," *J. Roy. Stat. Soc.*, vol. 67, no. 2, pp. 301–320, 2005.
[3] H. Zou, "The adaptive lasso and its oracle properties," *J. Amer. Stat. Assoc.*, vol. 101, no. 476, pp. 1418–1429, Dec. 2006.
[4] N. Meinshausen, "Relaxed lasso," *Comput. Stat. Data Anal.*, vol. 52, no. 1, pp. 374–393, 2007.
[5] P. Radchenko and G. M. James, "Variable inclusion and shrinkage algorithms," *J. Amer. Stat. Assoc.*, vol. 103, no. 483, pp. 1304–1315, 2008.
[6] S. Wang, B. Nan, S. Rosset, and J. Zhu, "Random lasso," *Ann. Appl. Statist.*, vol. 5, no. 1, pp. 468–485, 2011.
[7] H. Park, S. Imoto, and S. Miyano, "Recursive random lasso (RRLasso) for identifying anti-cancer drug targets," *PLoS One*, vol. 10, no. 11, 2015, Art. no. e0141869.
[8] F. Hanif, K. Muzaffar, K. Perveen, S. M. Malhi, and S. U. Simjee, "Glioblastoma multiforme: A review of its epidemiology and pathogenesis through clinical presentation and treatment," *Asian Pacific J. Cancer Prevention*, vol. 18, no. 1, pp. 3–9, 2017.
[9] Y. Cui *et al.*, "The effects of gene polymorphisms on glioma prognosis," *J. Gene Med.*, vol. 19, no. 11, pp. 345–352, 2017.
[10] J. Chen, Z.-Y. Xu, and F. Wang, "Association between DNA methylation and multidrug resistance in human glioma SHG-44 cells," *Mol. Med. Rep.*, vol. 11, no. 1, pp. 43–52, 2015.
[11] X.-M. Li *et al.*, "Dec1 expression predicts prognosis and the response to temozolomide chemotherapy in patients with glioma," *Mol. Med. Rep.*, vol. 14, no. 6, pp. 5626–5636, 2016.
[12] S. Ramezani *et al.*, "Rolipram potentiates bevacizumab-induced cell death in human glioblastoma stem-like cells," *Life Sci.*, vol. 173, pp. 11–19, Mar. 2017.
[13] E.-Y. Moon, G.-H. Lee, M.-S. Lee, H.-M. Kim, and J.-W. Lee, "Phosphodiesterase inhibitors control A172 human glioblastoma cell death through cAMP-mediated activation of protein kinase A and Epac1/Rap1 pathways," *Life Sci.*, vol. 90, nos. 9–10, pp. 373–380, 2012.
[14] P. Goldhoff *et al.*, "Targeted inhibition of cyclic AMP phosphodiesterase-4 promotes brain tumor regression," *Clin. Cancer Res.*, vol. 14, no. 23, pp. 7717–7725, 2008.
[15] X. Liu *et al.*, "SUMO fusion system facilitates soluble expression and high production of bioactive human fibroblast growth factor 23 (FGF23)," *Appl. Microbiol. Biotechnol.*, vol. 96, no. 1, pp. 103–111, 2012.
[16] Y. Akiyama *et al.*, "YKL-40 downregulation is a key factor to overcome temozolomide resistance in a glioblastoma cell line," *Oncol. Rep.*, vol. 32, no. 1, pp. 159–166, 2014.
[17] T. Ashizawa *et al.*, "Effect of the STAT3 inhibitor STX-0119 on the proliferation of a temozolomide-resistant glioblastoma cell line," *Int. J. Oncol.*, vol. 45, no. 1, pp. 411–418, 2014.
[18] L. Xu *et al.*, "Targetable BET proteins-and E2F1-dependent transcriptional program maintains the malignancy of glioblastoma," *Proc. Nat. Acad. Sci.*, vol. 115, no. 22, pp. E5086–E5095, 2018.

YOUNGSOON KIM received the B.S., M.S., and Ph.D. degrees in statistics from Kyeongsang National University, South Korea, in 1994, 1998, and 2005, respectively.

From 2009 to 2017, she was a Senior Researcher with the Gyeongnam Development Institute, South Korea. She is continuing the research as a Postdoctoral Researcher with Kennesaw State University. Her research interests include bioinformatics, machine learning, data mining, and big data analytics.

JIE HAO received the B.S. degree in statistics from the North China University of Technology, China, in 2012, and the M.S. degree in mathematics from East Tennessee State University, Johnson City, TN, USA, in 2014. She is currently pursuing the Ph.D. degree in analytics and data science with Kennesaw State University. Her research interests include machine learning, deep learning, and bioinformatics.

TEJASWINI MALLAVARAPU received the B.S. degree in pharmacy from Acharya Nagarjuna University, India, in 2010, and the M.S. degree in computer science from Kennesaw State University (KSU), GA, USA, in 2018. She is currently pursuing the Ph.D. degree in analytical and data science with KSU. Her research interests include machine learning, deep learning, bioinformatics, and big data analytics.

JOONGYANG PARK received the B.E. degree in applied statistics from Yonsei Univerity, South Korea, and the M.S. and Ph.D. degrees in industrial engineering from the Korea Advanced Institute of Science and Technology, South Korea, in 1984 and 1994, respectively.

He is currently a Professor with the Department of Information and Statistics, Gyeongsang National University, South Korea. His research interests include econometrics and machine learning.

MINGON KANG received the B.S. degree in computer engineering from Hanyang University, South Korea, and the M.S. and Ph.D. degrees in computer science from The University of Texas at Arlington, Arlington, TX, USA, in 2010 and 2015, respectively.

From 2015 to 2016, he was an Ad-Interim Assistant with Texas A&M University-Commerce. He is currently an Assistant Professor with the Department of Computer Science, Kennesaw State University. He has authored around 40 journal and conference articles. His research interests include bioinformatics, machine learning, data mining, and big data analytics.