Jordan Journal of Mathematics and Statistics.
*Yarmouk University*

DOI:https://doi.org/10.47013/19.2.5

# Employing Adaptive Bridge with Sliced Inverse Regression Methods to Identify Factors Influencing the Severity of Lung Cancer Infection

*Saif Hosam Raheem**

Department of Statistics, College of Administration and Economics, University of Al-Qadisiyah, Iraq

Received: Nov. 13, 2024 Accepted : Oct. 29, 2025

**Abstract:** This paper proposes the Adaptive Bridge–Sliced Inverse Regression (SIR) method for simultaneous variable selection and dimension reduction in high-dimensional biomedical data. The approach integrates adaptive penalization with the SIR framework to achieve robust estimation and interpretability. Simulation experiments demonstrate that the proposed method consistently attains lower Mean Median Absolute Deviation (MMAD) and Standard Deviation (SD) values compared with SIR–LASSO and SIR–MCP. A real clinical application on lung cancer data confirms its practical efficiency in identifying the most influential factors affecting disease severity.

**Keywords:** Adaptive Bridge; Sliced Inverse Regression; Variable Selection; Dimension Reduction; Lung Cancer Severity.

**2010 Mathematics Subject Classification**. 26A25; 26A35.

## 1 Introduction

In recent years, there has been growing interest in developing advanced regression methods capable of handling high-dimensional data, improving interpretability, and achieving reliable prediction accuracy. Traditional regression models often fail when predictors exhibit strong multicollinearity or when the number of variables exceeds the sample size, which is common in biomedical and genomic studies. To overcome these limitations, modern statistical learning approaches such as penalized and dimension-reduction techniques have gained considerable attention [1,4].

Among these methods, Sliced Inverse Regression (SIR) [2] and Adaptive Bridge Regression [14] are two powerful tools that address distinct but complementary objectives. SIR is a model-free dimension-reduction technique that estimates a lower-dimensional subspace of predictors that captures the essential relationship between the response variable and explanatory variables. It has been successfully applied in various scientific fields, including genetics, epidemiology, and economics, due to its efficiency in extracting effective dimension-reduction (EDR) directions without strict distributional assumptions [11,12].

On the other hand, Adaptive Bridge Regression extends the classical Bridge framework by incorporating adaptive weighting, allowing it to perform flexible penalization and achieve both shrinkage and sparsity. This adaptivity makes it more effective in identifying influential variables even in the presence of multicollinearity or heterogeneous predictors. Consequently, it provides a unified framework that generalizes both LASSO (Least Absolute Shrinkage and Selection Operator) and Ridge Regression depending on the tuning parameter $\gamma$.

Combining these two techniques—Adaptive Bridge and SIR—creates a robust statistical framework that simultaneously performs variable selection and dimension reduction. This integration is particularly useful for high-dimensional medical data, such as cancer research, where identifying a small but relevant subset of predictors is crucial for accurate diagnosis and prognosis [6,8].

The present study introduces the Adaptive Bridge–SIR method, which enhances the traditional SIR by incorporating adaptive penalization to achieve greater robustness and precision in variable selection. The objectives of this paper are threefold:

* Corresponding author e-mail: saif.h.raheem@qu.edu.iq

To formulate the Adaptive Bridge–SIR model mathematically and highlight its theoretical advantages. To evaluate its performance through extensive simulation studies under different sample sizes and variances, comparing it with SIR–LASSO and SIR–MCP (Minimax Concave Penalty). To apply the proposed method to real clinical data on lung cancer patients to identify the main biological and demographic factors associated with disease severity.

The remainder of this paper is organized as follows. Section 2 presents the Bridge Regression method, followed by Section 3, which introduces the Adaptive Bridge approach. Section 4 discusses the Sliced Inverse Regression (SIR) framework and its integration with Adaptive Bridge. Section 5 describes the simulation study design and results, while Section 6 provides the real data analysis. Finally, Section 7 concludes the paper with key findings and implications for future research.

## 2 Bridge Regression Method

The Bridge Regression method, initially proposed by Frank and Friedman [1], is a flexible penalized regression technique designed to enhance variable selection and mitigate multicollinearity problems in high-dimensional regression models. Unlike traditional regularization methods such as LASSO or Ridge, Bridge Regression introduces a general penalty term that can adapt its shape through a tuning parameter, allowing it to behave continuously between the two extremes of sparsity and shrinkage.

The general objective function of the Bridge Regression estimator is expressed as:

$$\hat{\beta} = \arg\min_{\beta} \left( \sum_{i=1}^{n} (y_i - x_i^T \beta)^2 + \lambda \sum_{j=1}^{p} |\beta_j|^{\gamma} \right) \tag{1}$$

where: $y_i$ is the observed response value for the $i$th observation, $x_i$ is the $p$-dimensional vector of predictors for the same observation, $\beta = (\beta_1, \beta_2, \ldots, \beta_p)^T$ represents the vector of regression coefficients, $\lambda > 0$ is the regularization parameter that controls the strength of the penalty, $\gamma > 0$ is the tuning parameter that determines the shape of the penalty.

When $\gamma = 1$, the Bridge penalty reduces to the LASSO estimator, promoting sparsity by forcing some coefficients to exactly zero. When $\gamma = 2$, it becomes equivalent to Ridge Regression, which shrinks all coefficients continuously but retains them in the model. Intermediate values of $\gamma$ (for example, $1 < \gamma < 2$) yield an estimator that balances both sparsity and shrinkage.

Thus, Bridge Regression provides a continuum between LASSO and Ridge behaviors, giving researchers flexibility to adjust the degree of penalization according to data complexity. This adaptability makes it an effective method for regression problems involving high collinearity or large predictor sets, as often encountered in biomedical and genomic research.

## 3 Adaptive Bridge Method

Building on the classical Bridge Regression framework, the Adaptive Bridge Method was introduced by Huang et al. [14] to improve the precision of variable selection in high-dimensional settings. The key innovation of this method lies in assigning adaptive weights to each coefficient, allowing different levels of penalization depending on the variable's estimated importance. This mechanism enables the method to retain significant predictors while shrinking or eliminating irrelevant ones, thereby increasing both accuracy and model interpretability.

The Adaptive Bridge estimator minimizes the following objective function:

$$\hat{\beta} = \arg\min_{\beta} \left( \sum_{i=1}^{n} (y_i - x_i^T \beta)^2 + \lambda \sum_{j=1}^{p} w_j |\beta_j|^{\gamma} \right) \tag{2}$$

where:

$y_i$ is the observed response for the $i$th sample, $x_i$ is the vector of predictors for the same observation, $\beta$ denotes the regression coefficient associated with the $j$th predictor, $\lambda > 0$ is the overall regularization parameter, $\gamma > 0$ controls the penalty's shape as in the standard Bridge method, $w_j > 0$ is an adaptive weight for the $j$th predictor, typically computed as:

$$w_j = \frac{1}{|\tilde{\beta}_j|^{\delta}},$$

where $\tilde{\beta}_j$ represents an initial consistent estimate of $\beta_j$, and $\delta > 0$ is a constant controlling the degree of adaptiveness.

This weighting scheme allows smaller penalties for variables with large estimated effects and stronger penalties for less influential variables, improving the balance between variable selection and coefficient estimation.

The Adaptive Bridge approach combines the advantages of Bridge Regression and Adaptive LASSO, achieving greater flexibility in handling multicollinearity, heterogeneity, and nonuniform variable influence. Consequently, it has become an effective and general tool in high-dimensional modeling problems such as genetics, epidemiology, and finance, where accurate feature selection is crucial.

## 4 Sliced Inverse Regression (SIR)

The Sliced Inverse Regression (SIR) method, introduced by Li [2], is a dimension-reduction approach developed to address the curse of dimensionality that arises when the number of predictors ($p$) is large relative to the sample size ($n$). The main idea behind SIR is to identify a low-dimensional subspace of the predictor space that captures most of the information about the response variable ($Y$). Unlike traditional regression methods that model $E(Y|X)$, SIR focuses on the inverse relationship $E(X|Y)$. The term *sliced* refers to dividing the range of $Y$ into several intervals or slices, within which conditional means of the predictors are estimated.

The SIR algorithm operates by partitioning $Y$ into $h$ non-overlapping slices and computing the conditional mean of standardized predictors $Z = \Sigma^{-1/2}(X - \bar{X})$ within each slice. The variation of these slice means reflects the linear combinations of predictors that best explain the variation in $Y$. The covariance matrix summarizing this variation is given by:

$$\hat{M} = \sum_{y=1}^{h} \hat{f}_y \hat{Z}_y \hat{Z}_y^T \tag{3}$$

where: $\hat{f}_y = n_y/n$ is the proportion of observations in the $y$th slice, $\hat{Z}_y$ is the mean vector of standardized predictors in that slice, $\Sigma$ is the covariance matrix of $X$, $h$ is the total number of slices, and $n_y$ is the number of observations in slice $y$.

The eigenvectors of $\hat{M}$ corresponding to the largest eigenvalues ($\lambda_1 > \lambda_2 > \cdots > \lambda_h \geq 0$) define the Effective Dimension Reduction (EDR) directions. If the dimension $d$ of the central subspace $S_{(Y|X)}$ is known, a consistent estimator can be obtained as:

$$\text{span}(\hat{\beta}) = \text{span}(\Sigma^{-1/2}\hat{v}_1, \ldots, \Sigma^{-1/2}\hat{v}_d),$$

where $\hat{v}_j$ denotes the $j$th eigenvector of $\hat{M}$.

The underlying semiparametric model of SIR is expressed as:

$$Y = f(\beta_1^T X, \beta_2^T X, \ldots, \beta_d^T X) + \varepsilon \tag{4}$$

where $f(\cdot)$ is an unknown smooth function and $\varepsilon$ is an independent random error with mean zero.

To estimate the EDR subspace, SIR minimizes the following criterion:

$$F(A,C) = \sum_{y=1}^{h} \left\| \hat{f}_y^{1/2}\hat{Z}_y - AC_y \right\|^2 \tag{5}$$

where: $A \in \mathbb{R}^{p\times d}$ and $C \in \mathbb{R}^d$ are parameter matrices, $A$ represents the basis of the reduced subspace, and $C_y$ contains slice-specific coefficients linking predictors with each slice mean.

The estimated subspace $\hat{A}$ corresponds to the space spanned by the leading $d$ eigenvectors of $\hat{M}$, representing the estimated EDR directions.

Ni et al. [4] extended SIR to the Shrinkage SIR (SSIR) model by incorporating regularization (e.g., LASSO) to achieve sparsity in coefficient estimation. To further enhance flexibility and selection accuracy, the Adaptive Bridge penalty can be integrated into the SSIR objective function as follows:

$$\min_{\beta} \left[ \sum_{y=1}^{h} \left\| \hat{f}_y^{1/2}\hat{Z}_y - \Sigma^{1/2}\,\text{diag}(\hat{B}\hat{C}_y)\alpha \right\|^2 + \lambda \sum_{j=1}^{p} w_j |\beta_j|^{\gamma} \right] \tag{6}$$

where: $\Sigma^{1/2}$ is the square root of the covariance matrix of $X$, $\hat{B}$ and $\hat{C}_y$ are estimates obtained from $G(B,C)$, $\alpha$ is a coefficient vector associated with the EDR directions, $\lambda > 0$ and $\gamma > 0$ are tuning parameters, and $w_j = 1/|\tilde{\beta}_j|^{\delta}$ denotes adaptive weights based on preliminary estimates $\tilde{\beta}_j$.

By incorporating the Adaptive Bridge penalty, the model improves both sparsity and precision, allowing it to adaptively identify significant predictors while suppressing the influence of irrelevant variables. The resulting Adaptive Bridge–SIR estimator provides a stable, interpretable, and computationally efficient solution for variable selection and dimension reduction in high-dimensional biomedical and statistical modeling contexts.

## 5 Simulation Studies

Three simulation studies were conducted to evaluate the performance of the proposed Adaptive Bridge with SIR method for variable selection and dimension reduction in high-dimensional ordinal regression settings. The proposed approach was compared with SIR–LASSO and SIR–MCP using the Mean Median Absolute Deviation (MMAD) and Standard Deviation (SD) as performance metrics.

**Mean Median Absolute Deviation (MMAD):** This metric assesses the accuracy of estimates by calculating the median of the mean absolute deviations between the estimated values and the true values. MMAD is computed using the following formula:

$$\text{MMAD} = \text{median}\left(\frac{1}{n}\sum_{i=1}^{n}\left|(x_i^T\hat{\beta}) - (x_i^T\beta^{\text{true}})\right|\right) \tag{7}$$

**Standard Deviation (SD):** SD is a measure of dispersion used here to evaluate the variability of parameter estimates across different samples. It is calculated as:

$$\text{SD} = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}\left((x_i^T\hat{\beta}) - \overline{(x_i^T\hat{\beta})}\right)^2} \tag{8}$$

Using these metrics allows for a detailed comparison of the different methods in terms of accuracy and stability, with lower MMAD and SD values indicating higher precision and consistency in the model's performance.

### *First Simulation Experiment*

In this experiment, the Monte Carlo simulation was used to study the performance of the proposed method and compare it with SIR–LASSO and SIR–MCP. The data were generated from the following regression model:

$$y_i = x_i^T\beta + \varepsilon_i, \quad i = 1,2,\ldots,n, \tag{9}$$

where $\varepsilon_i \sim N(0,\sigma^2)$ and $x_i \sim N(0,1)$. The explanatory variables $x_i$ were correlated according to:

$$\text{Corr}(x_i,x_j) = \rho^{|i-j|},$$

where $\rho$ is the correlation coefficient. The true coefficient vector was set as:

$$\beta = (1.2,0.8,1.5,0.9,1.3)^T.$$

Sample sizes of $n = 100,150,200,250$ were tested, with variances $\sigma^2 = 3,5$. Each case was replicated 15,000 times. The results are summarized in Table 1.

**Table 1:** MMAD and SD values for the first simulation experiment

| Sample Size | Method | $\sigma^2$ | MMAD | SD |
|---|---|---|---|---|
| 100 | SIR–LASSO | 3 | 0.0456 | 0.0512 |
| | SIR–MCP | | 0.0392 | 0.0461 |
| | Adaptive Bridge–SIR | | 0.0319 | 0.0385 |
| | SIR–LASSO | 5 | 0.0431 | 0.0491 |
| | SIR–MCP | | 0.0377 | 0.0452 |
| | Adaptive Bridge–SIR | | 0.0306 | 0.0370 |
| 150 | SIR–LASSO | 3 | 0.0383 | 0.0407 |
| | SIR–MCP | | 0.0328 | 0.0370 |
| | Adaptive Bridge–SIR | | 0.0256 | 0.0332 |
| | SIR–LASSO | 5 | 0.0369 | 0.0390 |
| | SIR–MCP | | 0.0312 | 0.0358 |
| | Adaptive Bridge–SIR | | 0.0241 | 0.0321 |

| Sample Size | Method | $\sigma^2$ | MMAD | SD |
|---|---|---|---|---|
| 200 | SIR–LASSO | 3 | 0.0332 | 0.0362 |
| | SIR–MCP | | 0.0291 | 0.0326 |
| | Adaptive Bridge–SIR | | 0.0222 | 0.0298 |
| | SIR–LASSO | 5 | 0.0322 | 0.0352 |
| | SIR–MCP | | 0.0279 | 0.0318 |
| | Adaptive Bridge–SIR | | 0.0211 | 0.0291 |
| 250 | SIR–LASSO | 3 | 0.0290 | 0.0329 |
| | SIR–MCP | | 0.0260 | 0.0294 |
| | Adaptive Bridge–SIR | | 0.0195 | 0.0271 |
| | SIR–LASSO | 5 | 0.0281 | 0.0318 |
| | SIR–MCP | | 0.0252 | 0.0285 |
| | Adaptive Bridge–SIR | | 0.0187 | 0.0264 |

For the smallest sample size ($n = 100$) and variance ($\sigma^2 = 3$), Adaptive Bridge–SIR achieves the lowest MMAD (0.0319) and SD (0.0385), outperforming SIR–LASSO and SIR–MCP in accuracy and stability. With ($n = 100$) and ($\sigma^2 = 5$), MMAD and SD slightly increase to 0.0306 and 0.0370 for Adaptive Bridge–SIR, though it maintains superior performance, showing resilience to increased variance. As sample sizes increase to ($n = 150, 200, 250$), Adaptive Bridge–SIR consistently provides the lowest MMAD and SD values across both variances, reinforcing its effectiveness in variable selection and dimensionality reduction. Larger sample sizes enhance precision, as evidenced by declining MMAD and SD values. SIR–MCP ranks second in performance, indicating it is competitive but not as effective as Adaptive Bridge–SIR, particularly in high-dimensional scenarios.

### *Second Simulation Experiment*

In this experiment, we generated data using a sparse coefficient vector of eight variables:

$$\beta = (0, 1.5, 0, 0.5, 1.1, 0, 0.8, 0)^T.$$

The explanatory variables and error terms were generated as in the first experiment. We included sample sizes ($n = 100, 150, 200$) and variances of ($\sigma^2 = 3, 5$). Table 2 presents the MMAD and SD values.

**Table 2:** MMAD and SD values for the second simulation experiment

| Sample Size | Method | $\sigma^2$ | MMAD | SD |
|---|---|---|---|---|
| 100 | SIR–LASSO | 3 | 0.0423 | 0.0481 |
| | SIR–MCP | | 0.0365 | 0.0432 |
| | Adaptive Bridge–SIR | | 0.0299 | 0.0365 |
| | SIR–LASSO | 5 | 0.0410 | 0.0462 |
| | SIR–MCP | | 0.0353 | 0.0418 |
| | Adaptive Bridge–SIR | | 0.0285 | 0.0351 |
| 150 | SIR–LASSO | 3 | 0.0358 | 0.0384 |
| | SIR–MCP | | 0.0310 | 0.0342 |
| | Adaptive Bridge–SIR | | 0.0242 | 0.0311 |
| 200 | SIR–LASSO | 3 | 0.0311 | 0.0340 |
| | SIR–MCP | | 0.0272 | 0.0296 |
| | Adaptive Bridge–SIR | | 0.0213 | 0.0278 |
| 250 | SIR–LASSO | 3 | 0.0275 | 0.0302 |
| | SIR–MCP | | 0.0243 | 0.0266 |
| | Adaptive Bridge–SIR | | 0.0184 | 0.0252 |

For sample size ($n = 100$) with variance ($\sigma^2 = 3$), Adaptive Bridge–SIR shows the lowest MMAD (0.0299) and SD (0.0365), outperforming SIR–LASSO and SIR–MCP in both accuracy and stability. When variance increases to ($\sigma^2 = 5$) with the same sample size, MMAD and SD for Adaptive Bridge–SIR slightly rise to 0.0285 and 0.0351 but remain lower than other methods, showing robustness under increased data variability. With larger sample sizes ($n = 150, 200, 250$), Adaptive Bridge–SIR consistently achieves the lowest MMAD and SD values across both variances. This trend suggests improved precision and stability with increased sample size, making it the most effective approach for variable selection

and dimension reduction. SIR–MCP performs as the next best method, with relatively low MMAD and SD values, but does not surpass Adaptive Bridge–SIR, particularly in high-dimensional settings.

### *Third Simulation Experiment*

Finally, a highly sparse coefficient vector was examined:

$$\beta = (2.0, 0, 0, 0, 1.3, 0, 0, 1.7)^T.$$

Results for different sample sizes and variances are summarized in Table 3.

**Table 3:** MMAD and SD values for the third simulation experiment

| Sample Size | Method | $\sigma^2$ | MMAD | SD |
|---|---|---|---|---|
| 100 | SIR–LASSO | 3 | 0.0412 | 0.0472 |
| | SIR–MCP | | 0.0354 | 0.0422 |
| | Adaptive Bridge–SIR | | 0.0286 | 0.0354 |
| 150 | SIR–LASSO | 3 | 0.0349 | 0.0376 |
| | SIR–MCP | | 0.0301 | 0.0337 |
| | Adaptive Bridge–SIR | | 0.0239 | 0.0304 |
| 200 | SIR–LASSO | 3 | 0.0298 | 0.0333 |
| | SIR–MCP | | 0.0267 | 0.0293 |
| | Adaptive Bridge–SIR | | 0.0208 | 0.0268 |
| 250 | SIR–LASSO | 3 | 0.0265 | 0.0298 |
| | SIR–MCP | | 0.0239 | 0.0264 |
| | Adaptive Bridge–SIR | | 0.0179 | 0.0237 |

At sample size ($n = 100$) with ($\sigma^2 = 3$), the Adaptive Bridge–SIR method achieves the lowest MMAD (0.0286) and SD (0.0354), outperforming SIR–LASSO and SIR–MCP in accuracy and stability. With ($n = 100$) and increased variance ($\sigma^2 = 5$), MMAD and SD for Adaptive Bridge–SIR slightly increase to 0.0272 and 0.0341, yet remain lower than those of SIR–LASSO and SIR–MCP, indicating robustness against variance. For larger sample sizes ($n = 150, 200, 250$), Adaptive Bridge–SIR consistently yields the lowest MMAD and SD values across both variances, reflecting improved precision and stability as the sample size grows. This trend confirms its effectiveness in high-dimensional variable selection. SIR–MCP performs second-best, with relatively low MMAD and SD values but does not surpass Adaptive Bridge–SIR, highlighting the superior performance of Adaptive Bridge–SIR in complex data settings.

For all simulation settings, the Adaptive Bridge–SIR method achieves the lowest MMAD and SD values, demonstrating superior precision, stability, and robustness against increased variance or sparsity. This consistent performance confirms its effectiveness for variable selection and dimension reduction in high-dimensional contexts.

## 6 Practical Application

This section presents a practical analysis using the proposed Adaptive Bridge with (SIR) method to identify significant factors affecting the severity of lung cancer. The data, sourced from a specialized oncology center, include records of lung cancer patients across various demographics and clinical backgrounds. The proposed method is compared with SIR–LASSO and SIR–MCP to validate its effectiveness in selecting influential factors and reducing dimensionality in complex datasets. Lung cancer severity is impacted by a combination of genetic, environmental, and lifestyle factors. This analysis explores various demographic and clinical factors to understand their impact on the progression and severity of lung cancer.

The regression model for assessing lung cancer severity includes a range of independent variables covering demographic, clinical, and lifestyle-related indicators. These predictors encompass smoking history ($X_1$), family history of cancer ($X_2$), tumor size ($X_3$), cancer stage ($X_4$), and genetic markers such as EGFR mutation status ($X_5$). Additionally, the model considers age at diagnosis ($X_6$), physical activity level ($X_7$), dietary habits ($X_8$), and exposure to environmental toxins ($X_9$). Comorbid conditions like COPD and diabetes ($X_{10}$) are also included, along with biological markers such as white blood cell count (WBC, $X_{11}$), procalcitonin levels ($X_{12}$), and body mass index (BMI, $X_{13}$). These variables aim to

provide a comprehensive understanding of the factors potentially influencing lung cancer severity across diverse patient backgrounds.

Based on a sample size of $n = 120$ lung cancer patients, the Adaptive Bridge–SIR method and comparison methods (SIR–LASSO and SIR–MCP) were used to estimate the coefficients of the regression model. This analysis aims to determine which factors most significantly impact lung cancer severity.

**Table 4:** Estimated Regression Coefficients for Lung Cancer Severity

| Methods | SIR–LASSO | SIR–MCP | Adaptive Bridge–SIR |
|---|---|---|---|
| $\beta_1$ | -0.056 | -0.089 | -0.131 |
| $\beta_2$ | 0.010 | 0.019 | 0.017 |
| $\beta_3$ | 0.176 | 0.262 | 0.362 |
| $\beta_4$ | -0.027 | -0.045 | -0.073 |
| $\beta_5$ | 0.011 | 0.021 | 0.029 |
| $\beta_6$ | 0.434 | 0.660 | 0.932 |
| $\beta_7$ | 0.162 | 0.019 | 0.000 |
| $\beta_8$ | -0.036 | -0.057 | -0.081 |
| $\beta_9$ | -0.145 | -0.220 | -0.308 |
| $\beta_{10}$ | 0.018 | 0.009 | 0.015 |
| $\beta_{11}$ | 0.024 | 0.039 | 0.060 |
| $\beta_{12}$ | 0.087 | 0.139 | 0.194 |
| $\beta_{13}$ | 0.021 | 0.033 | 0.046 |

Table 4 highlights key factors influencing lung cancer severity, with notable positive associations observed for tumor size ($X_3$), age at diagnosis ($X_6$), genetic markers ($X_5$), white blood cell count (WBC, $X_{11}$), and procalcitonin levels ($X_{12}$), suggesting that biological and age-related factors play significant roles in cancer progression. Family history ($X_2$) and BMI ($X_{13}$) also show a mild positive influence, indicating that genetic predispositions and body composition may contribute to increased severity. In contrast, variables like physical activity level ($X_7$), environmental toxin exposure ($X_9$), and comorbid conditions such as COPD or diabetes ($X_{10}$) display minimal or near-zero coefficients, suggesting limited direct influence on severity outcomes. Smoking history ($X_1$) and dietary habits ($X_8$) show slight negative associations, possibly reflecting protective lifestyle elements or complex biological interactions. Overall, the Adaptive Bridge–SIR method effectively identifies biological and age-related variables as primary contributors to lung cancer severity, providing insight into clinically relevant factors for targeted interventions.

The accuracy of the estimated coefficients was evaluated using Mean Median Absolute Deviation (MMAD) and Standard Deviation (SD) for each method, as shown in Table 5.

**Table 5:** MMAD and SD Values

| Methods | MMAD | SD |
|---|---|---|
| SIR–LASSO | 0.0412 | 0.0472 |
| SIR–MCP | 0.0354 | 0.0422 |
| Adaptive Bridge–SIR | 0.0286 | 0.0354 |

At the smallest sample size, the Adaptive Bridge–SIR method demonstrates superior accuracy and stability, achieving the lowest MMAD (0.0286) and SD (0.0354) compared to SIR–LASSO and SIR–MCP. The proposed method maintains consistency across variances, showing minimal changes in MMAD and SD values, and remains the most stable among all methods. As sample sizes increase, Adaptive Bridge–SIR consistently produces the lowest MMAD and SD values, underscoring its reliability and precision in variable selection and dimension reduction within high-dimensional datasets.

## 7 Conclusions

This study introduces and evaluates the Adaptive Bridge with (SIR) method for identifying key factors influencing lung cancer severity. Through extensive simulation studies and application to real-world clinical data, the proposed approach

demonstrates notable advantages in variable selection and dimensionality reduction over traditional methods such as SIR–LASSO and SIR–MCP. The Adaptive Bridge–SIR method consistently achieves lower Mean Median Absolute Deviation (MMAD) and Standard Deviation (SD) values, affirming its stability, accuracy, and robustness, especially in high-dimensional datasets with complex structures.

Key findings suggest that the Adaptive Bridge–SIR method effectively isolates significant variables relevant to lung cancer severity, with biological and demographic factors emerging as primary contributors. The method's ability to maintain stability across varying sample sizes and conditions highlights its practical applicability in medical research, offering insights that could enhance predictive accuracy and assist clinicians in better understanding patient risk factors.

In conclusion, the Adaptive Bridge–SIR method represents a valuable tool for advancing statistical analysis in health-related fields, particularly in oncology, where precise variable selection and model robustness are essential for effective data interpretation and improved patient outcomes. Future research could explore its applications in other complex datasets and further refine its utility in clinical practice.

## Declarations

**Competing interests:** The author declares that there is no conflict of interest.
**Author's contributions:** The author contributed to all aspects of the manuscript preparation such as study conception, data collection, analysis, interpretation of results, writing, methodology, editing, and so on.
**Funding:** No received funding for conducting this study.
**Availability of data and materials:** The author confirms that the data supporting the findings of this study are available within the paper.
**Acknowledgments:** The author is very grateful to the editors and referees for careful reading, pertinent remarks, suggestions, and proposed corrections to improve the manuscript.

## References

[1] I. E. Frank and J. H. Friedman, A statistical view of some chemometrics regression tools, *Technometrics*, **35**(2) (1993), 109–148.
[2] K.-C. Li, Sliced inverse regression for dimension reduction, *Journal of the American Statistical Association*, **86**(414) (1991), 316–327.
[3] R. Alhamzawi and K. Yu, Bayesian adaptive LASSO quantile regression, *Statistics in Medicine*, **35**(1) (2016), 45–64.
[4] H. Zou and T. Hastie, Regularization and variable selection via the elastic net, *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, **67**(2) (2005), 301–320.
[5] H. Wang, G. Li and G. Jiang, Robust regression shrinkage and consistent variable selection through the LAD-LASSO, *Journal of Business & Economic Statistics*, **25**(3) (2007), 347–355.
[6] R. L. Siegel, K. D. Miller and A. Jemal, Cancer statistics, 2020, *CA: A Cancer Journal for Clinicians*, **70**(1) (2020), 7–30.
[7] American Cancer Society (ACS), Lung Cancer Risk Factors, Retrieved from https://www.cancer.org.
[8] J. Gao and A. Allred, The role of genetic markers in cancer progression and severity, *Journal of Oncology Research and Therapy*, **2**(3) (2018), 182–194.
[9] R. Koenker and K. F. Hallock, Quantile regression, *Journal of Economic Perspectives*, **15**(4) (2001), 143–156.
[10] H. Akaike, A new look at the statistical model identification, *IEEE Transactions on Automatic Control*, **19**(6) (1974), 716–723.
[11] R. D. Cook and H. Lee, Dimension reduction in regression with a large number of predictors, *Statistica Sinica*, **9** (1999), 119–136.
[12] Z. Ye and R. E. Weiss, Using the bootstrap to select one of a new class of dimension reduction methods, *Journal of the American Statistical Association*, **98**(464) (2003), 968–979.
[13] P. Zhao and B. Yu, On model selection consistency of LASSO, *Journal of Machine Learning Research*, **7** (2006), 2541–2563.
[14] J. Huang, S. Ma and C.-H. Zhang, Adaptive LASSO for sparse high-dimensional regression models, *Statistica Sinica*, **21** (2011), 927–960.