

## Research Article

# HRLR-LOGISTIC: A Factor Selection Machine Learning Method Coupled with Binary Logistic Regression

Haoyan Xie,<sup>1</sup> Maryam Sadiq ,<sup>2</sup> Hai Huang ,<sup>3,4</sup> and Sughra Sarwar<sup>2</sup>

<sup>1</sup>The Philosophy Department, Nanjing University, Nanjing 210023, China

<sup>2</sup>Department of Statistics, University of Azad Jammu and Kashmir, Muzaffarabad 13100, Pakistan

<sup>3</sup>School of Marxism Studies, Xuchang University, Xuchang 461000, China

<sup>4</sup>Mobile Source Emission Aftertreatment Institute of Tongji University, Tongji University, Shanghai 200082, China

Correspondence should be addressed to Maryam Sadiq; [hussainulahmad@gmail.com](mailto:hussainulahmad@gmail.com) and Hai Huang; [xawangjuan01@sina.com](mailto:xawangjuan01@sina.com)

Received 17 February 2022; Accepted 5 May 2022; Published 21 July 2022

Academic Editor: Ardashir Mohammadzadeh

Copyright © 2022 Haoyan Xie et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The selection of influential predictor factors with maximum model accuracy is the main goal of the regression domain. The present study is conducted to integrate an innovative method, that is, “a hybrid of relaxed lasso and ridge regression,” with a logistic regression model in the context of dichotomous factors. The efficacy of the proposed approach is illustrated using both simulated and real-life data. The results suggested that HRLR-logistic selected the best subset compared to standard logistic, Lasso, and Ridge regression. Based on the Akaike information criterion (3065.85) and the Bayesian information criterion (3151.46), the proposed approach is proved to have the highest efficiency for cesarean section data. In addition, the study identified the elements that contribute to the cesarean section in Pakistan. It is evidenced that woman’s literacy level ( $\beta=0.5828$ ), place of delivery ( $\beta=0.8990$ ), availability of nurse as an assistant ( $\beta=0.7370$ ), and care during the first two days of delivery ( $\beta=0.7837$ ) are remarkable factors associated with cesarean section.

## 1. Introduction

Regression models play a vital role in establishing the association between the response and predictor factors to help forecast the outcome. The regression models have diverse applications in several scientific areas, including public health, genetics, clinical medicine, chemometrics, and bio-informatics. The most important phase of model building is the selection of significant factors, especially in the case of big data. Factor selection methods facilitate the user to choose the most influential factors by removing irrelevant or redundant ones to provide an optimal model with the highest efficiency. Due to advancements in research, big data is introduced in almost every field of science, which is difficult to model efficiently using traditional techniques. Consequently, factor selection has become a matter of interest to improve the accuracy of the predictive model by choosing the significant factors, especially in public health. The main targets of factor selection methods are to improve

interpretability, reduce noise, enhance model prediction performance, and accelerate modeling time (Aslam et al. [1]; Mehmood et al. [2]).

Regarding public health, an important subject is issues related to women’s health, e.g., cesarean section (CS) and its risk factors. Globally, rising CS rates have caused general populace specialists to analyze medical and nonmedical factors that have contributed to this rise (Ghosh [3]). Perinatal concerns are regarded as valid medical causes for cesarean section, including dystocia, previous cesarean section, fetal distress, breach births, postterm pregnancy, multiple pregnancies, and hypertensive illness (ElArdat et al. [4]). The mother’s age, socioeconomic status, schooling, career, and domicile have all been proven to be substantially connected to the style of delivery (Nilsen et al. [5]; Sadiq et al. [6]). Concerning the signs for performing CS, no standard characterization techniques exist, and they can be various or related. However, these difficulties, disconnecting the most well-known signs for CS method of delivery, are a key to

targeting preventive methodologies. Precisely perceiving signs related to maternal or fetal demise can assist with bringing down mortality. Understanding the effect of different elements on the choice with regards to the CS is fundamental. Hence, this study is expected to survey the greatness and elements related to the cesarean section in Pakistan using advanced and efficient factor selection methods. Bivariate and multivariate strategic relapse examinations were generally applied to survey factors adding to Cesarean delivery.

Logistic regression (LR) is a classification approach to examine the relationship among dichotomous factors and find the optimal model. Logistic regression assumes independence of observations, linearity of explanatory factors and log odds, and no or little multicollinearity among predictors. The logistic model can interpret regression estimates as an indicator of the importance of factors. Let  $Y$  represent binary response and let  $X$  be a  $(n * p)$  matrix of covariates, and then the LR model takes its form. Suppose that the response  $Y_i$  and the  $(n * p)$  matrix of predictors  $X$  are dichotomous, and then, the binary LR model is

$$\ln\left(\frac{Y}{1-Y}\right) = \alpha + \beta X + \epsilon, \quad (1)$$

where  $\alpha$  is the intercept,  $\beta$  is the  $P * 1$  vector of unknown regression estimates,  $\epsilon$  is the  $P * 1$  vector of residuals, and the mathematical expression  $(Y/1 - Y)$  denotes an odds ratio. The regression modeling is based on the estimation of unknown coefficients  $\beta$ 's. Several approaches have been introduced in the context of binary classification to estimate  $\beta$ 's and hence to select influential factors based on these coefficients estimates. The forward selection method and backward elimination method are the most popular traditional factor selection techniques for LR.

The advanced factor selection methods include least absolute shrinkage and selection operator (lasso) (Tibshirani [7]); elastic net regularization (Zou and Hastie [8]); relaxed lasso (Meinshausen [9]); and ridge regression (Hoerl and Kennard [10]). Recently, Pelawa Watagoda et al. [11] proposed an ML-based regularization and variable selection method called Hybrid of Relaxed Lasso regularization and Ridge regression (HRLR) for quantitative response in the context of a linear regression model. The study demonstrated that the proposed HRLR method provides an optimal model compared to lasso and relaxed lasso for high-dimensional data. The proposed technique is the combination of properties of relaxed lasso and ridge regression. A more efficient approach, namely, Hybrid of Relaxed Lasso and Ridge Regression coupled with Logistic model (HRLR-Logistic), is proposed as the optimal method for classification approach adopted from ML technique.

The formal statements of the problem are as follows:

- (1) to introduce a more efficient factor selection method for a dichotomous response;
- (2) to compare the proposed method with standard methods using simulated and real data set;
- (3) to determine the significant risk factors for cesarean section (CS) in Pakistan.

## 2. Materials and Methods

This study used three standard reference methods, namely, standard logistic regression model coupled with forward selection method, lasso regression, and ridge regression, to compare with machine learning-based proposed method, namely, HRLR-logistic.

**2.1. Logistic Regression Model.** The baseline logistic regression model has the form

$$\ln\left(\frac{Y}{1-Y}\right) = a + Xb, \quad (2)$$

where  $Y$  is a binary response, and  $X$  is a matrix of predictors. The parameters of the model are  $a$  and  $b$ , where  $a$  is the intercept from the regression equation,  $b$  is the set of estimates, and  $Y/1 - Y$  is the odds ratio.

The logistic regression model is used as a baseline classification method. The stepwise forward variable selection technique is the most commonly used classical approach in the context of logistic regression and hence is called standard logistic. It starts without considering any predictor variable in the model and gradually adds them in each subsequent step. At each phase, one factor that has been excluded from the model is reexamined for inclusion in the model. When an excluded factor is introduced to the model, a certain test statistic is calculated. Firstly, the most crucial variable is inserted, and then, the model was refitted using this variable. This approach is repeated until there are no more significant variables (Hosmer et al. [12]).

**2.2. Lasso Regression.** The lasso proposed by Tibshirani [7] is an appealing regularization approach for multiscale regression. The coefficients of lasso  $\beta^L = \sum_{j=1}^p |\beta_j|$  minimize

$$\sum_{i=1}^n (y_i - \hat{y}_i)^2 + \lambda \sum_{j=1}^p |\beta_j|. \quad (3)$$

There are various advantages and disadvantages of the lasso method. Firstly, it involves variable selection by setting specific coefficients to zero. Secondly, it employs systematic algorithmic computations. Regarding limitations, lasso's convergence is relatively low for  $p \gg n$  and it produces poor coefficient estimates when the data set comprises strongly correlated predictors (Tibshirani (1996b)).

**2.3. Ridge Regression.** The ridge regression coefficient estimates denoted by  $\beta_R = \sum_{j=1}^p \beta_j^2$  are given as

$$\beta_R = (X^t X + \lambda I)^{-1} (X^t y). \quad (4)$$

The  $\beta_R$  minimizes the expression as

$$\sum_{i=1}^n (y_i - \hat{y}_i)^2 + \lambda \sum_{j=1}^p \beta_j^2 = \text{RSS} + \lambda \sum_{j=1}^p \beta_j^2, \quad (5)$$

where  $\lambda > 0$  is the calibrating criterion. The shrinking penalty is  $\lambda \sum_{j=1}^p \beta_j^2$ . All  $p$  predictors are included in the final model employing ridge regression (Hoerl and Kennard [13]).

**2.4. Hybrid of Relaxed Lasso and Ridge Regression (HRLR).** A hybrid of Relaxed Lasso and Ridge regression (HRLR) is a new regularization and variable selection technique proposed by Pelawa Watagoda et al. [11]. This method used the characteristics of ridge regression and relaxed lasso.

HRLR-logistic estimator can be defined for  $\lambda_1 \in [0, \infty)$ ,  $\lambda_2 \in [0, \infty)$  and  $\Phi \in (0, 1]$  as

$$\hat{\beta}^{\lambda_1, \lambda_2, \Phi} = \operatorname{argmin}_{\beta} n^{-1} \sum_{i=1}^n (Y_i - X_i^T \beta \cdot 1_M)^2 + \Phi \lambda_1 |\beta_1| + \lambda_2 |\beta \cdot 1_{M_{\lambda_1}}|_2. \quad (6)$$

The  $M_{\lambda_1}$  is the characteristic function on the collection of variables  $M \subseteq \{1, \dots, p\}$  in case for every  $k \in \{1, \dots, p\}$  and

$$\beta \cdot 1_{M_{\lambda_1}} = \begin{cases} 0, & k \notin M_{\lambda_1}, \\ \beta_k, & k \in M_{\lambda_1}. \end{cases} \quad (7)$$

HRLR estimator is a two-step process: firstly, the ridge regression coefficients for each fixed  $\lambda_2$  are determined, and secondly, the relaxed lasso-type shrinkage is estimated along the solution paths for the relaxed lasso coefficient ( $\Phi$  and  $\lambda_1$ ). The general scheme of factor selection techniques used in this study is presented in Figure 1.

**2.5. Data Simulation.** The simulated data is generated from binomial distribution using the following probabilities of success  $p = 0.5, 0.6, 0.7, 0.8, 0.9$ . A dichotomous response having 100 predictors with 5000 observations is generated to compare the performance of standard-Logistic, lasso, and ridge with the proposed method HRLR. R-software is used for the generation of simulated data.

**2.6. Real Dataset.** The data originated from the Pakistan Demographic and Health Survey (PDHS) 2017-2018, conducted by the National Institute of Population Studies (NIPS) through the Ministry of National Health Services, Regulations and Coordination, Pakistan.

The Pakistan Demographic and Health Survey (PDHS) 2017-18 is the fourth in the international series of Demographic and Health Surveys (DHS). The National Institute of Population Studies (NIPS), a prime research institution in population and development specialty, accomplished the PDHS with technical cooperation from the International Classification of Functioning, Disability, and Health (ICF) as well as the Pakistan Bureau of Statistics (PBS) and financial aid from the United States Agency for International Development (USAID).

The 2017-18 PDHS had a broad goal of gathering high-quality data on fertility rates and preferences, contraception use, mother and infant health, neonatal mortality, immunization, dietary patterns of mothers and children, disability, relocation, women's empowerment, domestic abuse, HIV/

AIDS awareness, and other health-related issues. The delivery method is taken as the binary response variable with two categories; cesarean section (CS) and vaginal delivery having 1115 and 2230 observations, respectively, and 112 categorical predictors with complete information are initially included in this study.

### 3. Results

For model comparison, three standard regression models and a proposed approach are executed over a simulated and real data set.

**3.1. Model Comparison Based on Simulation Data.** The HRLR technique is introduced for continuous response in terms of linear regression. In this study, the HRLR approach is integrated with logistic regression to assess the efficiency of this proposed method in terms of binary response. An outcome variable following binomial distribution with success probabilities ranging from 0.3 to 0.9 is generated with 100 variables having 5000 observations. Table 1 presents the efficiency comparison of four methods based on AIC, BIC, and the number of significant variables. The standard-logistic method picked out the maximum number of predictors, and the HRLR approach selected the least number of influential variables.

Figure 2 shows the accuracy comparison of HRLR with standard-Logistic regression, lasso regression, and ridge regression based on AIC and BIC. The visual display demonstrated that the AIC measure of HRLR-logistic is  $<$  AIC of lasso  $<$  AIC of ridge regression  $<$  AIC of standard logistic. The results presented that HRLR-logistic is the optimal variable selection method for the simulated data set.

**3.2. Model Comparison Based on Real Data Set.** Initially, 112 predictors with a total of 3345 observations are included in this study. After the data cleaning process, 89 independent variables are left for further analysis. The basic assumptions of logistic regression were checked before model execution. The data consisted of a sufficiently large sample size for modeling logistic regression. To detect the independence of error terms, binned plot showed the average residual and fitted value for each bin in Figure 3 and exhibited the independence of error terms. The presence of multicollinearity is examined by the correlation plot presented in Figure 3. The strength and intensity of colors showed the strength of correlation among predictors. The plot exhibited that 18 predictors are highly correlated and, hence, eliminated from the data set. After removing highly correlated variables, 71 predictors are included in the final analysis of logistic modeling.

The upper and lower lines inside the boundary of Figure 3 represented  $\pm 0.10$  standard-error bounds, within which approximately 90% of the binned residuals are expected to fall. Residual's spread is represented by inner points.

The pattern of residuals shows independence as the points fall inside the 90% confidence bands. Also, the

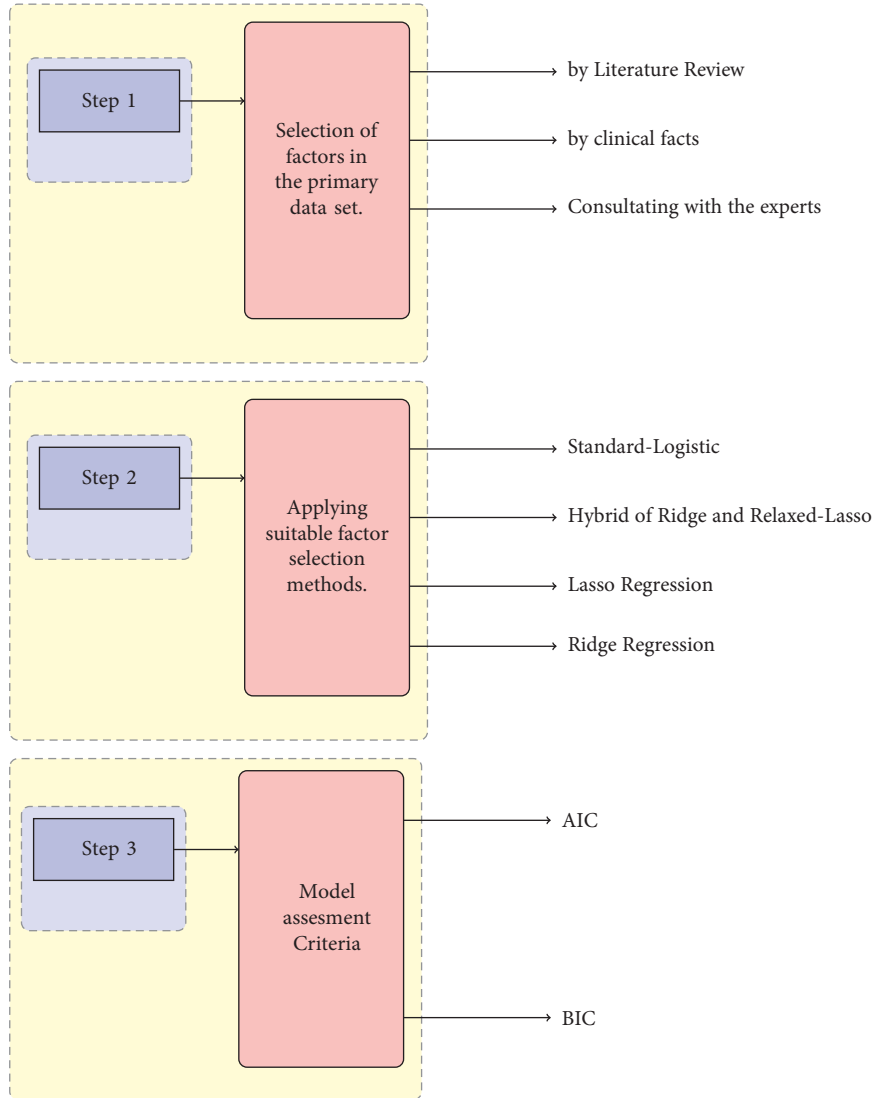


FIGURE 1: Factor selection procedure.

TABLE 1: Comparison of models over simulated data.

Methods	Number of variables	AIC	BIC
Standard-logistic	35	6257.29	6915.26
HRLR	18	6153.9	6277.74
Lasso	24	6161.14	6324.07
Ridge	26	6255.38	6907.1

residual's points depict the dissociated and arbitrary functioning of residuals. Figure 3 shows that the residual terms are distributed independently and randomly.

The correlations between the predictors are examined by using a correlation map and shown in Figure 4(a). The correlation map showed positive associations by red dots and negative relations by blue dots. The size of the dots reflects the magnitude of the correlation. The intensity of dots reflects the strength of association between predictors. A high correlation ( $>0.7$ ) between 14 predictors is observed, and the presence of multicollinearity is identified. A commonly suggested and convenient measure is to remove

highly correlated variables. After removing highly correlated predictors, Figure 4(b) represents the correlation map having uncorrelated and less correlated variables.

Three reference methods including standard-Logistic regression, Ridge Regression, Lasso Regression are employed to measure the accuracy comparison of the proposed HRLR method based on AIC, BIC, and the number of selected variables. Table 2 shows the performance comparison of all four models demonstrating the highest performance of the proposed HRLR method compared to the standard Logistic, Ridge, and Lasso Regression method. The results indicated that the HRLR approach selected the least number of variables with the highest performance efficiency and the Standard-logistic method picked out 25 influential predictors with the lowest performance. The Lasso and Ridge regression methods showed lower efficiency compared to the proposed HRLR algorithm and higher efficiency compared to Standard-logistic regression.

Based on AIC and BIC, the efficiency comparison presented in Figure 5 signaled the preference of the HRLR

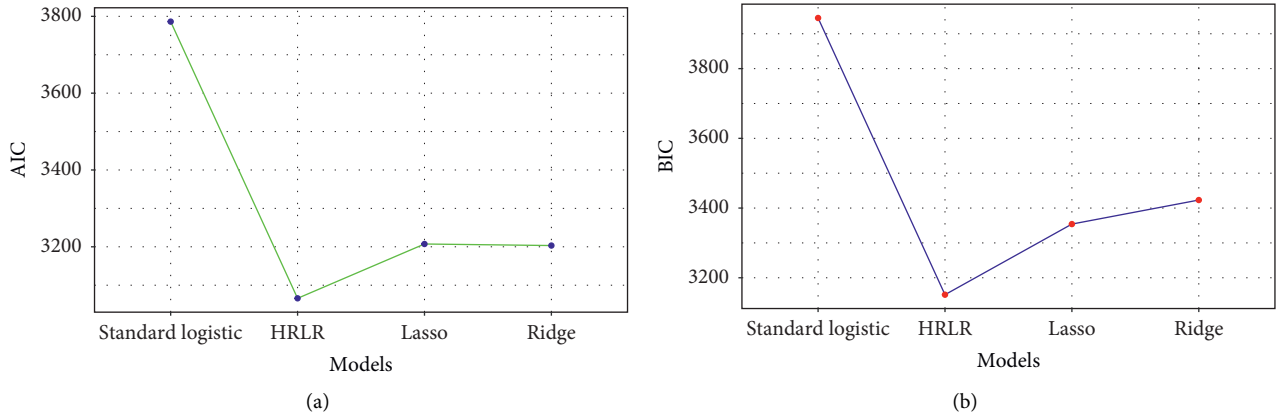


FIGURE 2: Comparison of models over simulated data. (a) AIC. (b) BIC.

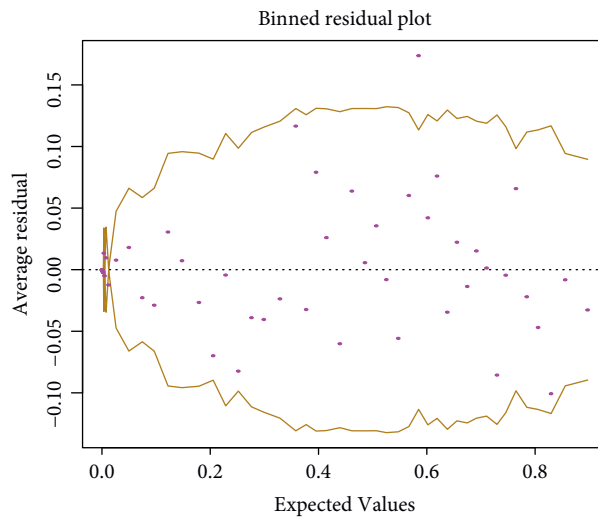


FIGURE 3: Residual plots.

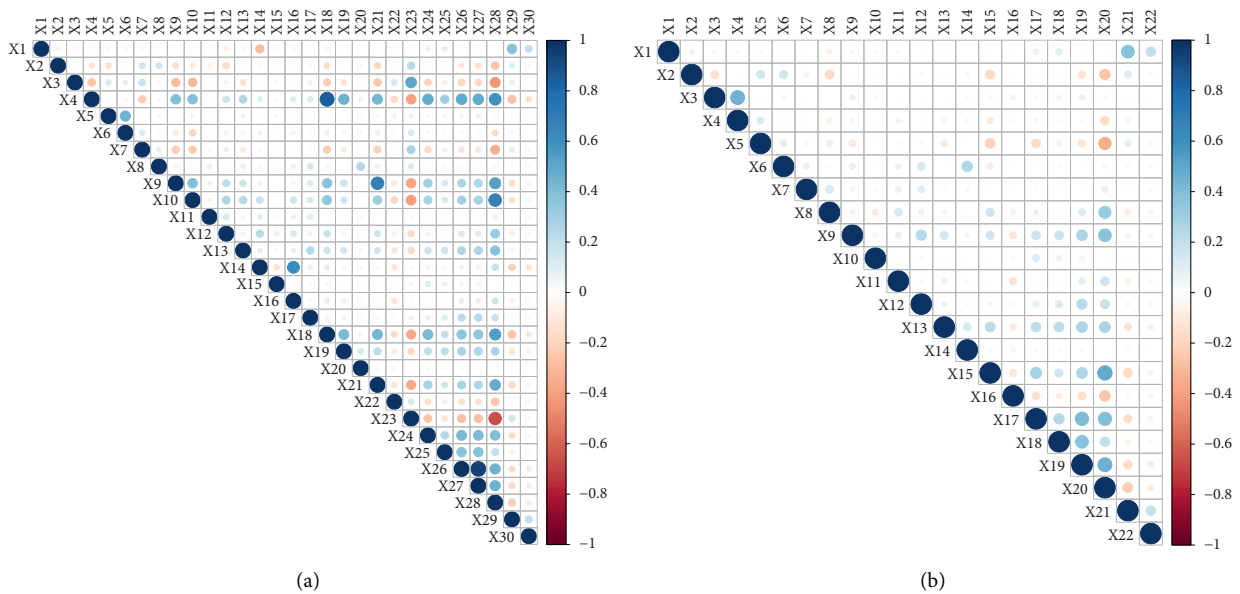


FIGURE 4: Correlation plots for predictors. (a) Before removal of multicollinearity. (b) After removal of multicollinearity.

TABLE 2: Comparison of models by using cesarean data.

Methods	Number of variables	AIC	BIC
Standard-logistic	25	3786.268	3945.264
Lasso	23	3207.284	3354.05
Ridge	35	3203.226	3423.374
HRLR	14	3065.851	3151.464

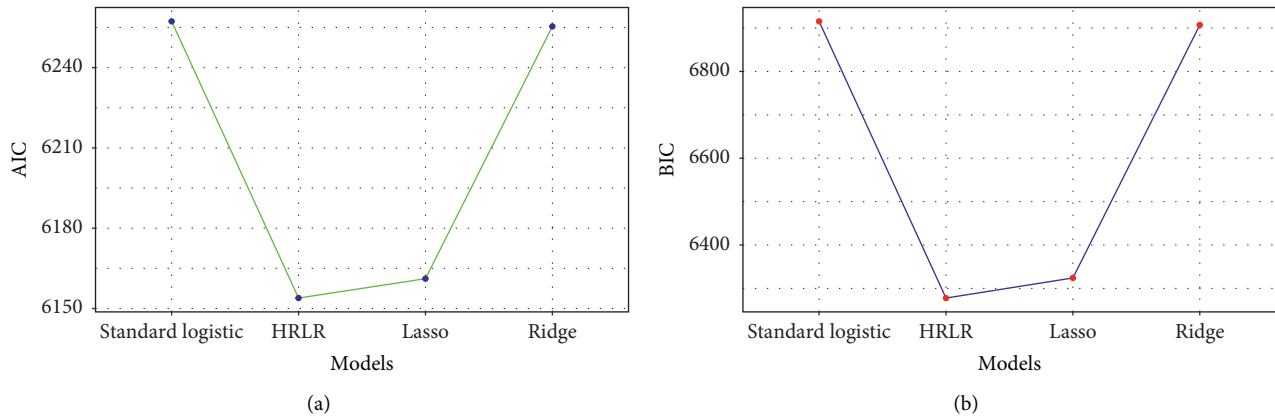


FIGURE 5: Comparison of models over cesarean section data. (a) AIC. (b) BIC.

method as the optimal variable selection method for the observed data set with a dichotomous response. The HRLR method is finally used for the selection of influential predictors of cesarean section delivery and presented in Table 3. The remarkable selected factors of the cesarean section included a source of drinking water, literacy, wealth index combined, frequency of watching television and using the Internet, age of respondents at first birth, marriage to the first birth interval, assistance nurse, number of antenatal visits during pregnancy, place of delivery, blood sampled taken during pregnancy, health provider measuring temperature during first two days, and whether ever been tested for hepatitis B or C. All 14 variables designate positive coefficients that mean that they will increase log-odds of CS delivery in Pakistan.

The log probability of CS increased by 0.12991 and 0.5828 units for every unit change in the source of drinking water and literacy, respectively. Similarly, one unit change in frequency of watching television and frequency of using the Internet caused 0.06901 and 0.02130 units change in log-odds of CS, respectively. One unit increase in wealth index, marriage to the first birth interval, place of delivery, and testing for Hepatitis B or C resulted in an increase of log-odds by 0.10781, 0.2357, 0.8990, and 0.25367 units, respectively.

Figure 6 represents the coefficient estimates of selected variables for the Standard-logistic, Lasso, ridge, and HRLR-Logistic model. The visual display showed the difference between the strength of association of CS with predictors. Only Standard-logistic and HRLR-logistic showed a significant association of CS with the source of drinking water. The literacy level of women showed a higher influence on CS regarding Standard-logistic while lasso and Ridge regression demonstrated a moderate relationship of women's education

TABLE 3: Regression coefficient estimates of finally fitted HRLR-Logistic selection method to select important factors of cesarean section.

Selected factor	Estimate
Source of drinking water	0.12991
Woman's literacy level	0.5828
Frequency of watching television	0.06901
Frequency of using Internet last month	0.02130
Wealth index	0.10781
Age of respondent at first birth	0.4756
Marriage to first birth interval	0.2357
Births in last three years	0.5936
Assistance: Nurse	0.7370
Number of antenatal visits during pregnancy	0.3729
Place of delivery	0.8990
During pregnancy: blood sample taken	0.6383
During first two days health provider took care	0.7837
Ever been tested for hepatitis B or C	0.25367

level with CS delivery of mode. Frequency of watching television, frequency of using Internet last month, marriage to the first birth interval, births in last three years, assisted by a nurse, number of antenatal visits during pregnancy, during first two days health provider took care, and whether ever been tested for Hepatitis B or C are the significant predictors identified by HRLR-logistic method only. The factors including wealth index, age of respondent at first birth, place of delivery, and a blood sample taken during pregnancy are identified as important by Lasso, ridge, and HRLR-Logistic regression, but not by Standard-logistic regression. The difference in identified factors by all four methods demonstrated that the selected predictors influence the efficiency of the corresponding method.

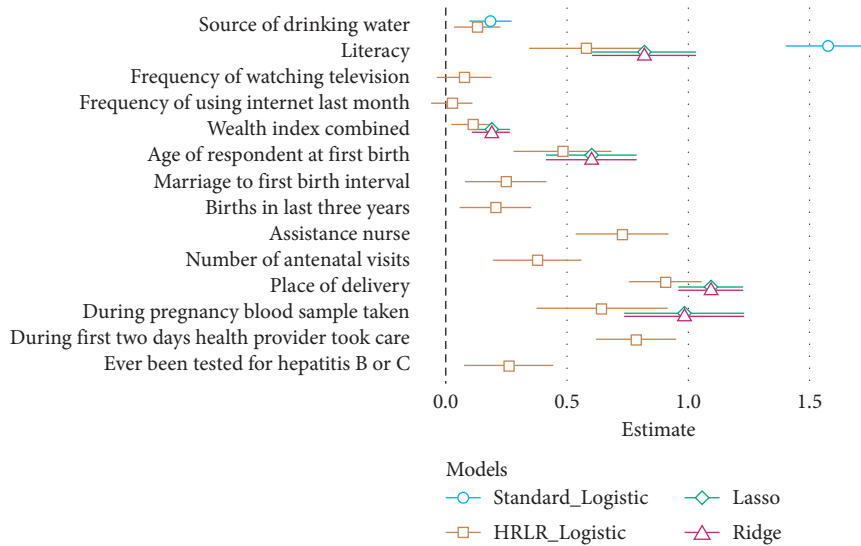


FIGURE 6: Regression estimates.

### 4. Discussion

The intrusion of CS into developing countries, where infant and maternal mortality and morbidity rates are high, is pivotal. However, Pakistan’s prevalence of CS exceeds WHO standards, implying that Pakistan is a sect of a global mania of nonmedical CS, and these may be more readily accepted in private health facilities in the country than in public ones.

The ambition of the present study included a finding of the most optimal variable selection statistical method as well as the determination of the factors contributing to CS in Pakistan. This was achieved by using the most recent machine learning technique proposed by a combination of relaxed lasso and ridge regression (HRLR-logistic). The performance of the models is tested by using simulated as well as real-life data of CS based on AIC and BIC. For comparison of models, the maternal data from the Pakistan Demographic and Health Surveys (PDHS) from 2018 to 2019 is utilized. The results suggested that the HRLR-logistic picked 14 highly significant variables in association with CS from a total of 71 predictors. The results evidenced that HRLR-logistic is the most efficient and effective variable selection technique as compared to the standard logistic, lasso, and Ridge method. These results were in good agreement with another study that reported HRLR, the best-fitted method for variable selection.

As reported in the current study, maternal age was the foremost commonly utilized variable in previous CS prediction models, with positive coefficients in all studies (Burke et al. [14]; Janssen et al. [15]; Souza et al. [16]). Medically, women in their thirties and beyond are more likely to have a CS. Fetal distress, stress, exhaustion, and managing the risks of mortality and morbidity for both mother and child are the most common causes of CS in advanced maternal age. The number of antenatal visits, the type of dai/traditional/nurse assistance, and the health provider during the first two days are all linked to determining the CS group in the present study. Several studies

have reported the correlation of CS to prenatal care, facilities, and antenatal visits (Sadiq et al. [17]). Factors like the age of mother at 1st birth, previous birth interval, and terminated pregnancy were reported to be linked with CS in the present analysis, and these findings were concurring with the results obtained from previous studies that evaluated the association between the CS ratio and terminated previous pregnancies, maternal age, and birth intervals (Edmonds et al. [18]).

A noteworthy relationship between CS and literacy, wealth index, and marriage to the birth interval was observed in the current study. Prior studies have shown that the CS rates are influenced by a parent’s educational level, wealth index, and birth interval (Yaya et al. [19]). Because education is correlated to women’s empowerment, women with higher education can take decisions about whether or not to have a C-section. High education, on the other hand, is not always linked to increased risk of a C-section probably because highly educated ladies are aware of the prospect of needless C-sections, because education gives awareness about health-promoting acts.

Furthermore, women who lived in the highest quantiles of wealth (richer and richest) had a higher probability of labor in a hospital than at home. In the present study, the factor being tested for hepatitis B or C was found to be highly significant in association with the CS group.

### 5. Conclusions

The present study proposed a more efficient machine learning variable selection technique, a hybrid of relaxed lasso, and ridge regression (HRLR) for dichotomous response in the context of the logistic model. The HRLR integrated with logistic regression is proved to be optimum compared to Standard-Logistic regression, Ridge Regression, and Lasso Regression for simulated and real data of CS. Hence, the HRLR approach is recommended as a more efficient variable selection modeling strategy for binary

variables. This technique is found to be more likely to remove highly correlated variables with greater efficiency. The analysis of the CS data set illustrated the fitness of the proposed method with real-world problems. Lastly, HRLR-logistic is the most efficient and appropriate variable selection technique.

### Data Availability

The data are available at <https://dhsprogram.com/data>.

### Additional Points

Factor subset selection is a matter of high concern and an essential step in the modeling approach across all scientific fields regarding big data analysis. Several factor selection methods including forward and backward elimination, lasso, ridge, and relaxed lasso are broadly implemented. This study used “A Hybrid of Relaxed Lasso and Ridge Regression (HRLR)”, which is developed by coupling the properties of relaxed lasso and ridge regression in the context of dichotomous factors. This method provided a factor selection method for the logistic model with higher performance. The practitioners may develop logistic models integrated with the HRLR method, by using the mathematical computations provided in this article.

### Conflicts of Interest

The authors declare that they have no conflicts of interest.

### References

- [1] M. Aslam, M. Sadiq, and T. Mehmood, “Assessment of maternal health services utilization in Pakistan: the role of socio-demographic characteristics,” *Asian Biomedicine*, vol. 14, no. 1, pp. 3–7, 2020.
- [2] T. Mehmood, M. Sadiq, and M. Aslam, “Filter-based factor selection methods in partial least squares regression,” *IEEE Access*, vol. 7, pp. 153499–153508, 2019.
- [3] S. Ghosh, *Increasing Trend in Caesarean Section Delivery in India: Role of Medicalisation of Maternal Health*, Institute for Social and Economic, Bangalore, India, 2010.
- [4] M. ElArdat, S. Izetbegovic, E. Mehmedbasic, and M. Duric, “Frequency of vaginal birth after cesarean section at clinic of gynecology and obstetrics in sarajevo,” *Medical Archives*, vol. 67, no. 6, p. 435, 2013.
- [5] C. Nilsen, T. Østbye, A. K. Daltveit, B. T. Mmbaga, and I. F. Sandøy, “Trends in and socio-demographic factors associated with caesarean section at a tanzanian referral hospital, 2000 to 2013,” *International Journal for Equity in Health*, vol. 13, no. 1, p. 87, 2014.
- [6] M. Sadiq, A. T. Abdulrahman, R. Alharbi, D. K. F. Alnagar, and S. M. Anwar, “Modeling the Ranked Antenatal Care Visits Using Optimized Partial Least Square Regression,” *Computational and Mathematical Methods in Medicine*, vol. 2022, Article ID 2868885, , 2022.
- [7] R. Tibshirani, “Regression shrinkage and selection via the lasso,” *Journal of the Royal Statistical Society: Series B*, vol. 58, no. 1, pp. 267–288, 1996.
- [8] H. Zou and T. Hastie, “Regularization and variable selection via the elastic net,” *Journal of the Royal Statistical Society: Series B*, vol. 67, no. 2, pp. 301–320, 2005.
- [9] N. Meinshausen, “Relaxed lasso,” *Computational Statistics & Data Analysis*, vol. 52, no. 1, pp. 374–393, 2007.
- [10] A. E. Hoerl and R. W. Kennard, “Ridge regression: biased estimation for nonorthogonal problems,” *Technometrics*, vol. 12, no. 1, pp. 55–67, 1970.
- [11] L. C. R. Pelawa Watagoda, A. T. Arnholt, and H. S. R. Arachchige Don, “Hrlr regression,” *Renaissance Manuscript Studies: Research in Mathematics & Statistics*, vol. 8, no. 1, Article ID 1921904, 2021.
- [12] D. W. Hosmer, S. Lemeshow, and R. X. Sturdivant, *Applied logistic regression*, Vol. 398, John Wiley & Sons, , Hoboken, 2013.
- [13] A. E. Hoerl and R. W. Kennard, “ridge regression: biased estimation for nonorthogonal problems,” *Technometrics*, vol. 42, no. 1, pp. 80–86, 2000.
- [14] N. Burke, G. Burke, F. Breathnach et al., “Prediction of cesarean delivery in the term nulliparous woman: results from the prospective, multicenter genesis study,” *American Journal of Obstetrics and Gynecology*, vol. 216, no. 6, pp. 598.e1–598.e11, 2017.
- [15] P. A. Janssen, J. J. C. Stienen, R. Brant, and G. E. Hanley, “A predictive model for cesarean among low-risk nulliparous women in spontaneous labor at hospital admission,” *Birthkit*, vol. 44, no. 1, pp. 21–28, 2017.
- [16] J. Souza, A. Betran, A. Dumont et al., “A global reference for caesarean section rates (c-model): a multicountry cross-sectional study,” *BJOG: An International Journal of Obstetrics and Gynaecology*, vol. 123, no. 3, pp. 427–436, 2016.
- [17] M. Sadiq, T. Mehmood, and M. Aslam, “Identifying the factors associated with cesarean section modeled with categorical correlation coefficients in partial least squares,” *PLoS One*, vol. 14, no. 7, Article ID e0219427, 2019.
- [18] J. K. Edmonds, M. Paul, and L. Sibley, “Determinants of place of birth decisions in uncomplicated childbirth in Bangladesh: an empirical study,” *Midwifery*, vol. 28, no. 5, pp. 554–560, 2012.
- [19] S. Yaya, O. A. Uthman, A. Amouzou, and G. Bishwajit, “Disparities in caesarean section prevalence and determinants across sub-saharan africa countries,” *Global health research and policy*, vol. 3, no. 1, p. 19, 2018.