International Journal of Mathematical Sciences and
Optimization: Theory and Applications
12(1), 2026, Pages 1 - 10
https://doi.org/10.5281/zenodo.20383827

Penalized Cox Regression and Its Application to Microarray Survival Data

F. A. Okolie [1*,2,3], E. E. E. Akarawak [3], O. O. Olufolabo [4]

1. Department of Mathematical Sciences, Ajayi Crowther University, Oyo State, Nigeria.
2. Department of Statistics, Olabisi Onabanjo University, Ago-Iwoye, Ogun State, Nigeria.
3. Department of Statistics University of Lagos, Akoka, Lagos State, Nigeria.
4. Department of Statistics, Yaba College of Technology, Lagos State, Nigeria.

* Corresponding author: fa.okolie@acu.edu.ng, eakarawak@unilag.edu.ng,
oluyomi.olufolabo@yabatech.edu.ng

Article Info
Received: 30 May 2023 Revised: 27 January 2026
Accepted: 04 February 2026 Available online: 15 March 2026

Abstract

Survival analysis provides a statistical framework for examining the relationship between time to event outcomes and explanatory variables. The Cox proportional hazards (Cox PH) model is widely used for this purpose, but its performance deteriorates in high-dimensional settings, where the number of predictors is large and strong collinearity is present. Penalized Cox regression methods have been developed to address these limitations by incorporating regularization into the estimation process. This study applies penalized Cox regression models to breast cancer microarray survival data obtained from the Gene Expression Omnibus dataset GSE20685, which contains gene expression measurements for 54,682 genes across 327 patients. The aim of this study is to fit and compare Ridge, LASSO, and Elastic Net Cox models with the classical Cox PH model. Model performance is evaluated using root mean square error (RMSE), with emphasis placed on test set results to assess predictive generalization. The results show that penalized Cox regression models consistently outperform the Cox PH model on the test set across all evaluated predictor dimensions. While the Cox PH model fails to produce reliable predictions when the number of predictors is large, the penalized models remain stable and effective. Among the penalized approaches, Ridge regression demonstrates the most robust performance in ultra-high-dimensional settings, whereas LASSO and Elastic Net provide competitive performance at lower and moderate dimensional levels through feature selection. These findings highlight the importance of regularization for survival modelling in high-dimensional genomic data and demonstrate that penalized Cox regression offers a more reliable alternative to the classical Cox PH model for microarray-based survival prediction.

Keywords: Cox Regression, Elastic net Regression, LASSO Regression, Penalized Cox Regression and Ridge Regression.
MSC2010: 60G55.

1 Introduction

Survival analysis provides a statistical framework for modelling the time until an event of interest occurs and for examining how covariates influence that timing. Although its early development was largely motivated by biomedical applications, particularly the study of patient mortality, survival analysis is now widely applied across diverse fields such as engineering, economics, social sciences and modern biostatistics. A major milestone in the development of survival modelling was the proportional hazards model proposed by Cox [1], which offers a semi-parametric approach for relating covariates to event risk without requiring explicit specification of the baseline hazard function. Owing to its flexibility and interpretability, the Cox proportional hazards (Cox PH) model has remained a cornerstone of survival analysis for several decades.

In contemporary biomedical research, particularly in oncology, the focus has increasingly shifted toward predicting patient outcomes using high-dimensional genomic data. Microarray technologies generate expression measurements for thousands of genes simultaneously, resulting in datasets where the number of predictors far exceeds the number of observed patients. Landmark studies in breast cancer genomics have demonstrated the strong prognostic value of gene expression profiles, but also highlighted the statistical challenges associated with analysing such data [2,3]. In these high-dimensional settings, classical Cox PH modelling becomes unreliable due to issues such as multicollinearity, unstable coefficient estimates, complex interaction structures and potential violations of modelling assumptions. As noted by Emmert-Streib and Dehmer [4] and Wang et al. [5], these challenges have motivated the development of alternative survival modelling approaches that are better suited to high-dimensional data. Examples include Bayesian Additive Regression Trees [6,7] and large-scale regularized Cox models [8], which are designed to maintain stability and predictive accuracy when traditional methods fail.

The limitations of the classical Cox model in high-dimensional contexts have long been recognised. Early methodological work by Verweij and Van Houwelingen [9] introduced penalized likelihood approaches to Cox regression as a way of stabilizing estimation in the presence of many covariates. Subsequent review studies, such as Van Wieringen et al. [10], systematically compared survival prediction methods for gene expression data and emphasized the superiority of penalized and regularized approaches over unpenalized Cox models in terms of prediction accuracy and robustness.

Among modern approaches, penalized Cox regression has emerged as a particularly effective and interpretable solution for genomic survival analysis. By introducing penalty terms such as the L_1 (LASSO), L_2 (Ridge), or elastic-net combinations into the Cox partial likelihood, penalized models are able to shrink regression coefficients, reduce overfitting and perform automatic variable selection. Theoretical and methodological developments in this area include the formulation of L_1-penalized Cox models for sparse estimation [11], efficient coordinate descent algorithms for computing regularization paths [12], and penalized Cox regression methods tailored specifically for high-dimensional, low-sample-size microarray data [13]. More recent studies continue to extend and refine penalized Cox methods, including grouped penalties that account for biological pathway structures [14], adaptations for left-truncated survival data [15], improved tuning parameter selection strategies [16], mixture-based penalized Cox models for heterogeneous patient populations [17], and broader evaluations of penalized estimators in Cox regression [18]. These developments underscore the central role of regularized Cox models in analysing complex, high-throughput biomedical datasets.

Motivated by these advances, this study applies penalized Cox regression techniques to breast cancer microarray survival data and evaluates their predictive performance. Specifically, Ridge, LASSO and Elastic Net penalties are implemented and compared with the classical Cox proportional hazards model. The objectives of the study are to demonstrate the practical advantages of penalized Cox regression in high-dimensional settings, to compare the predictive accuracy of different penalty structures, and to assess the conditions under which penalized approaches outperform the traditional Cox PH model.

2 Methodology

2.1 Cox Regression Model

A common mathematical approach for modelling survival data is the Cox proportional hazards (Cox PH) model. The model was originally introduced by Cox [1] and has since become one of the most widely used tools in survival analysis. The Cox PH model is semi-parametric in nature, as it specifies the effect of covariates on the hazard function while leaving the baseline hazard unspecified. This flexibility makes the model robust and allows it to produce results that are often comparable to those obtained from correctly specified parametric survival models.

Suppose we have a sample size of n from which to estimate the relationship between the survival time and the predictor levels $X_1, X_2, \cdots, X_p$ predictors. Due to censoring, for $i = 1, \cdots, n$, the i th datum in the sample is denoted by $(t_i, \delta_i, x_{i1}, x_{i2+}, \cdots, x_{ip})$, where δ_i is the censoring indicator and t_i is the censoring time if $\delta_i = 0$ or the survival time if $\delta_i = 1$ and $x_i = \{x_{i1}, x_{i2}, \cdots, x_{ip})\}'$ is the vector of the predictor level of p predictors for the ith sample. The Cox regression model demonstrates that the hazard function $\eta(t)$ which means the risk of death at time t for an individual with predictor profile, is given by:

$$\eta(t \mid X) \quad = \quad \lambda_0(t) \exp\left(\beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_p X_p\right) \tag{2.1}$$

where $\lambda_0(t)$ represents the baseline hazard function, and the linear predictor is given by:

$$\beta' X = \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_p X_p.$$

Here, $\beta = (\beta_1, \ldots, \beta_p)'$ is the column vector of regression coefficients, and the predictor variables are represented by the column vector $X = (X_1, \ldots, X_p)'$. The corresponding sample values for the ith individual are given by $x_i = (x_{i1}, x_{i2}, \ldots, x_{ip})'$.

We define the linear risk score function as $f(X) = \beta' X$. Using the available sample data, the Cox partial likelihood (Cox, 1972) is expressed as:

$$L(\beta) = \prod_{r \in D} \frac{\exp\left(\beta' x_r\right)}{\sum_{j \in R_r} \exp\left(\beta' x_j\right)} \tag{2.2}$$

where D represents the set of indices corresponding to events (such as deaths), while R_r denotes the set of indices for individuals who are still at risk just before time t_r (i.e., at $t_r - 0$).

Taking the logarithm of the likelihood function is more convenient to apply than the original likelihood function. Thus, taken the logarithm of the Cox partial likelihood, we have:

$$l(\beta) = \log L(\beta) = \sum_{r \in D} \left(\beta' x_r - \log \left(\sum_{j \in R_r} \exp\left(\beta' x_j\right) \right) \right) \tag{2.3}$$

We now follow the normal maximum likelihood estimation method to calculate unknown parameters.

2.2 Ridge Regression

The L_2- penalized regression, also known as the ridge regression [19] is the most popular form of penalized regression. It conserves all the parameters to construct prediction models. To add regularization term into the Cox regression model, the log partial likelihood function will be rewritten as follows;

$$l(\beta)=\sum_{r\in D}^{p}\left(\beta' x_r-\log\left(\sum_{j\in R_r}\exp\left(\beta' x_j\right)\right)\right)-\lambda\sum_{j=1}^{p}\beta_j^2 \tag{2.4}$$

where λ is a tuning parameter and p is the number of predictors.

2.3 LASSO Regression

The LASSO (Least Absolute Shrinkage and Selection Operator) regression [20], L_1-penalized regression was first proposed by Tibshirani (1996). Because of some constraints in LASSO's principles, it tends to convert some coefficients to zero finally. According to this special characteristic, the LASSO estimation is often applied in parameter shrinkage to build simpler models. The log partial likelihood function will be rewritten when the L_1-norm regularization is added into the Cox regression model as:

$$l(\beta)=\sum_{r\in D}^{p}\left(\beta' x_r-\log\left(\sum_{j\in R_r}\exp\left(\beta' x_j\right)\right)\right)-\lambda\sum_{j=1}^{p}|\beta_j| \tag{2.5}$$

Comparing equations (2.4) and (2.5), we see that the LASSO and ridge regression have similar formulations. The only difference is that the β_j^2 term in the ridge regression penalty equation (2.4) has been replaced by $|\beta_j|$ in the LASSO penalty (2.5). In statistical parlance, the LASSO uses an L_1 (pronounced "ell 1") penalty instead of an L_2 penalty. As with ridge regression, the LASSO shrinks the coefficient estimates towards zero. However, in the case of the LASSO, the L_1 penalty has the effect of forcing some of the coefficient estimates to be exactly equal to zero when the tuning parameter λ is sufficiently large. Hence, the LASSO performs variable selection. As a result, models generated from the LASSO are generally much easier to interpret than those produced by ridge regression. We say that the LASSO yields sparse models that is, sparse models that involve only a subset of the variables.

2.4 Elastic Net Regression

Elastic-Net regression [21] includes both the L_1 and L_2 norm regularization terms to the Cox regression model. Its log partial likelihood is given as

$$l(\beta)=\sum_{r\in D}^{p}\left(\beta' x_r-\log\left(\sum_{j\in R_r}\exp\left(\beta' x_j\right)\right)\right)-\sum_{j=1}^{p}\left(\lambda_1\beta_j^2+\lambda_2\left|\beta_j\right|\right) \tag{2.6}$$

where the λ_1 and λ_2 are corresponding tuning parameters of L_2 and L_1 penalties, respectively. This gives us the benefits of both LASSO and Ridge regression. It has been found to have predictive power better than LASSO, while still performing feature selection. We therefore get the best of worlds, performing feature selection with the feature-group selection of Ridge. Elastic-Net comes with the additional overhead of determining the two $\lambda s'$.

2.5 Criteria for Model Prediction Performance

Root Mean Square Error (RMSE)

Root mean squared error (RMSE) is the square root of the mean of the square of all of the error. The use of RMSE is very common, and it is considered an excellent general-purpose error metric for numerical predictions.

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \tag{2.7}$$

where y_i are the observations, y_i predicted values of the a variable y, and n the number of observations available for analysis. RMSE is a good measure of accuracy, but only to compare prediction errors of different models or model configurations for a particular variable and not between variables, as it is scale-dependent.

2.6 Data Summary

The dataset used in this study was obtained from the National Center for Biotechnology Information (NCBI) Gene Expression Omnibus (GEO). Specifically, the study utilizes the GSE20685 dataset, which contains gene expression profiles of breast cancer cases. The dataset consists of 54,682 genes and 327 samples, sourced from Taipei, Taiwan [22].

In this study, the dependent variable represents the waiting time until the event (death) occurs, while the covariates are the gene expression levels. To facilitate model training and evaluation, the dataset was partitioned into training (70%) and test (30%) sets.

3 RESULTS

Figure 1 is the cross-validated errors for the $\log\lambda$ - sequences examined. The top left plot for ridge regression, top right for lasso and bottom left for Elastic net regression. For each method $\log\lambda_{min}$ and $\log\lambda_{1se}$ are indicated by vertical lines. The left line indicates the value of $\log\lambda_{min}$ the right line indicates the value of $\log\lambda_{1se}$. Using lasso we would, in this case, get a model about 13 variables using $\log\lambda_{min}$ (≈ -5) and shrinks all coefficients to zero using $\log\lambda_{1se}$ and using elastic net we would, in this case, get a model with about 12 variables using $\log\lambda_{min}$ (≈ -4.2) and no variable using $\log\lambda_{1se}$.

Figure 2 is the trace plot for the λ-sequences examined. The top left plot for ridge regression, top right for LASSO and bottom left for Elastic net regression. As λ increases some of the LASSO coefficients and elastic net coefficients are shrunken to zero which is not the case for the ridge coefficlents. Hence, depending on λ LASSO and elastic net can produce a model with any number of covariates while the ridge model will always include all covariates, even though some of the can be very small.

Model performance on Train set

Table 1 below shows the performance comparison between the Cox Proportional Hazards (Cox PH) model and the penalized Cox models (Ridge, LASSO, and Elastic Net) reveals significant differences in predictive accuracy, as measured by Root Mean Square Error (RMSE). The standard Cox PH model exhibits substantially higher RMSE values across all available predictor counts, starting at 8.0777 with 30 predictors, increasing to 9.9099 at 100 predictors, and slightly improving to 9.3549 at 229 predictors. In contrast, the penalized Cox models consistently achieve much lower RMSE values, indicating better predictive performance and generalization. Ridge regression, which retains all predictors but applies shrinkage, starts with an RMSE of 1.0647 at 30 predictors and progressively improves as the predictor count increases, reaching 0.3304 at 54,627 predictors, demonstrating its strong ability to handle high-dimensional data effectively. LASSO and Elastic Net, which enforce sparsity by setting some coefficients to zero, perform better than Ridge in lower-dimensional settings, with RMSE values of 0.8390 and 0.8457 at 30 predictors, respectively, and further improvements at 229 predictors (0.5418 and 0.5491, respectively). However, unlike Ridge, their performance

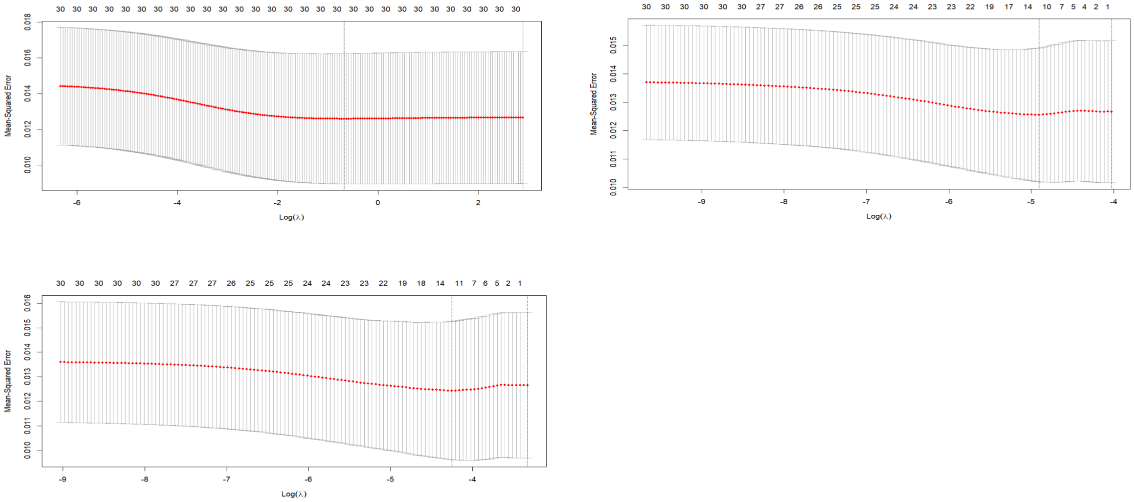

Figure 1: *Cross-validation plots for ridge, lasso, and elastic net regression, showing the relationship between the logarithm of the regularization parameter (λ) and the partial likelihood deviance. Red dots represent the mean deviance, while vertical bars indicate standard errors. The dotted vertical lines mark the optimal λ values that minimize deviance, highlighting the trade-off between model complexity and predictive accuracy.*

stabilizes at high dimensions, maintaining RMSE values around 0.5050–0.5098 at 30,000 predictors and 0.5045–0.5090 at 54,627 predictors, suggesting a limitation in handling extremely large feature spaces compared to Ridge. The lack of reported RMSE values for the Cox PH model beyond 229 predictors suggests that the model either failed to converge or exhibited excessive prediction error, further highlighting its inferiority to penalized Cox models in high-dimensional survival analysis.

Model Validation

Table 1: Model performance on train set

Model	Number of predictors in the model					
	30	100	229	230	30,000	54627
Ridge	1.0647	0.9720	0.8575	1.0461	0.4218	0.3304
LASSO	0.8390	0.7574	0.5418	0.7489	0.5050	0.5045
Elastic Net	0.8457	0.7633	0.5491	0.7579	0.5098	0.5090
Cox PH	8.0777	9.9099	9.3549	–	–	–

Table 2 presents the comparison between the Cox Proportional Hazards (Cox PH) model and penalized Cox models (Ridge, LASSO, and Elastic Net) revealing stark differences in their ability to generalize to new data. A well-validated model should demonstrate low RMSE across different predictor counts, ensuring stability and reliability in predictions. The Cox PH model consistently exhibits the highest RMSE values, starting at 8.0777 with 30 predictors, increasing to 9.9099 at 100 predictors, and slightly improving to 9.3549 at 229 predictors, which suggests severe poor generalization. Additionally, its performance degrades further as the number of predictors increases ($p > n$). In contrast, penalized Cox models exhibit significantly lower RMSE values, indicating better generalization and validation performance. Ridge regression, which shrinks coefficients without eliminating predictors, shows progressive RMSE improvement from 1.0647 at 30 predictors to

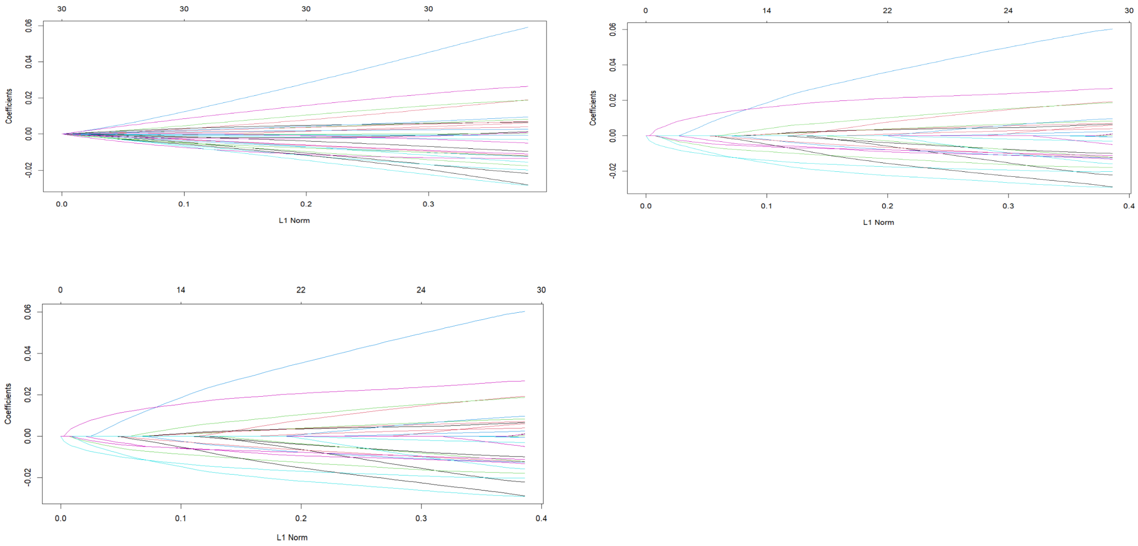

Figure 2: The trace plot for the $\log \lambda - sequences$ examined.

0.3304 at 54,627 predictors, demonstrating superior stability in high-dimensional settings. LASSO and Elastic Net, which improve interpretability by selecting relevant features, outperform Ridge in smaller models, achieving 0.8390 and 0.8457 RMSE at 30 predictors, respectively, and further reducing error at 229 predictors (0.5418 and 0.5491, respectively). However, their RMSE stabilizes at 0.5045–0.5098 for larger models, suggesting that feature selection may limit performance improvements in extremely high-dimensional settings. Since effective model validation requires a balance between bias and variance, the penalized models demonstrate better validation properties compared to Cox PH, which exhibits extreme variance and an inability to generalize well. The absence of Cox PH results beyond 229 predictors suggests validation failure due to excessive complexity or convergence issues, further reinforcing the superior validation performance of penalized Cox models for survival analysis, particularly when dealing with high-dimensional data.

Table 2: Model performance on test set

Model	Number of predictors in the model					
	30	100	229	230	30,000	54627
Ridge	1.5338	1.2664	1.3692	0.9436	1.6985	0.8182
LASSO	1.5033	1.4833	4.4887	2.1924	1.4988	0.8827
ELastic Net	1.5166	1.4018	4.6906	2.0551	1.5013	0.8731
COx PH	5.1916	5.7665	4.8990	–	–	–

Discussion

In this research, we have proposed penalized Cox regression in survival analysis and have compared three different methods—Ridge, LASSO, and Elastic Net to the traditional Cox Proportional Hazards (Cox PH) model. The Ridge method, a widely used penalization approach, shrinks regression coefficients toward zero but retains all predictors in the model. The LASSO method, a more recent alternative, not only shrinks coefficients but also performs variable selection by setting some coefficients to zero, making it particularly effective for sparse models. The Elastic Net method combines the benefits of both Ridge and LASSO, allowing for feature selection while still maintaining some level of coefficient shrinkage.

The test set results clearly show that the Cox PH model performs poorly on the test set across all predictor sizes for which it could be estimated. At 30, 100, and 229 predictors, the Cox PH model records substantially higher RMSE values, ranging from 4.8990 to 5.7665, compared to RMSE values close to or below 1.6 for all penalized Cox models. This indicates that even in relatively low-dimensional settings, the Cox PH model exhibits weak generalization performance when applied to this microarray dataset.

In contrast, the penalized Cox regression models demonstrate consistently superior predictive accuracy across all evaluated predictor dimensions. At lower predictor counts, Ridge, LASSO, and Elastic Net produce comparable RMSE values, all markedly lower than those of the Cox PH model. This suggests that regularization improves model stability and reduces overfitting even before the dimensionality exceeds the sample size.

As the number of predictors increases beyond the sample size, the advantages of penalization become more pronounced. The Cox PH model fails to produce valid results once the number of predictors exceeds 229, reflecting the instability of maximum likelihood estimation in high-dimensional settings. All penalized models, however, remain computationally stable and continue to yield reliable predictions.

Among the penalized approaches, Ridge regression exhibits the most stable performance at very high predictor dimensions. At 30,000 and 54,627 predictors, Ridge achieves the lowest RMSE values, indicating superior generalization in ultra-high-dimensional settings. This behavior is consistent with the structure of microarray data, where predictors are highly correlated and many genes contribute small but cumulative effects to survival outcomes. By shrinking coefficients without eliminating predictors, Ridge regression effectively balances bias and variance under extreme dimensionality.

LASSO and Elastic Net perform competitively at smaller and moderate predictor sizes but show relatively higher RMSE values as dimensionality increases. The sparsity induced by these methods, while beneficial for interpretability, may lead to the exclusion of weak yet collectively informative predictors in large genomic feature spaces. Elastic Net partially mitigates this limitation by combining L1 and L2 penalties, although its performance remains slightly inferior to Ridge regression in the highest-dimensional scenarios considered.

The test set results demonstrate that penalized Cox regression models substantially outperform the classical Cox PH model for survival prediction in microarray data. In particular, Ridge regression provides the most robust and reliable performance when the number of predictors is large relative to the sample size. These findings underscore the necessity of regularization for high-dimensional survival analysis and highlight the limitations of the Cox PH model in modern genomic applications.

Conclusion

In the analysis of larger studies of time to event data with large number of predictors, the usual standard technique such as the Cox model may not be appropriate due to violation of some assumptions such as collinearity of predictors and high dimensionality of the data. Alternative methods to Cox methods should be employed for better prediction.

References

[1] Cox, D. R. (1972). Regression models and life-tables. *Journal of the Royal Statistical Society: Series B (Methodological)*, 34(2), 187–202.

[2] Van't Veer, L. J., Dai, H., Van De Vijver, M. J., He, Y. D., Hart, A. A., Mao, M., & Friend, S. H. (2002). Gene expression profiling predicts clinical outcome of breast cancer. *Nature*, 415(6871), 530–536.

[3] Van De Vijver, M. J., He, Y. D., Van't Veer, L. J., Dai, H., Hart, A. A., Voskuil, D. W., & Bernards, R. (2002). A gene-expression signature as a predictor of survival in breast cancer. *New England Journal of Medicine*, 347(25), 1999–2009.

[4] Emmert-Streib, F., & Dehmer, M. (2019). Introduction to survival analysis in practice. *Machine Learning and Knowledge Extraction*, 1(3), 1013–1038.

[5] Wang, P., Li, Y., & Reddy, C. K. (2019). Machine learning for survival analysis: A survey. *ACM Computing Surveys (CSUR)*, 51(6), 1–36.

[6] Sparapani, R. A., Logan, B. R., McCulloch, R. E., & Laud, P. W. (2016). Nonparametric survival analysis using Bayesian additive regression trees (BART). *Statistics in Medicine*, 35(16), 2741–2753.

[7] Okolie, F. A., Fagbemigun, B. O., & Samson, O. J. (2025). Microarray Survival Analysis of Five Cancer Gene Sequences and its Application using Bayesian Additive Regression Trees (BART). *Sri Lankan Journal of Applied Statistics*, 26(3).

[8] Mittal, S., Madigan, D., Burd, R. S., & Suchard, M. A. (2014). High-dimensional, massive sample-size Cox proportional hazards regression for survival analysis. *Biostatistics*, 15(2), 207–221.

[9] Verweij, P. J., & Van Houwelingen, H. C. (1994). Penalized likelihood in Cox regression. *Statistics in Medicine*, 13(2–24), 2427–246.

[10] Van Wieringen, W. N., Kun, D., Hampel, R., & Boulesteix, A. L. (2009). Survival prediction using gene expression data: a review and comparison. *Computational Statistics & Data Analysis*, 53(5), 1590–1603.

[11] Goeman, J. J. (2010). L1 penalized estimation in the Cox proportional hazards model. *Biometrical Journal*, 52(1), 70–84.

[12] Simon, N., Friedman, J. H., Hastie, T., & Tibshirani, R. (2011). Regularization paths for Cox's proportional hazards model via coordinate descent. *Journal of Statistical Software*, 39, 1–13.

[13] Gui, J., & Li, H. (2005). Penalized Cox regression analysis in the high-dimensional and low-sample size settings, with applications to microarray gene expression data. *Bioinformatics*, 21(13), 3001–3008.

[14] Dang, X., Huang, S., & Qian, X. (2021). Penalized Cox's proportional hazards model for high-dimensional survival data with grouped predictors. *Statistics and Computing*, 31(6), 77.

[15] McGough, S. F., Incerti, D., Lyalina, S., Copping, R., Narasimhan, B., & Tibshirani, R. (2021). Penalized regression for left-truncated and right-censored survival data. *Statistics in Medicine*, 40(25), 5487–5500.

[16] Dai, B., & Breheny, P. (2024). Cross-validation approaches for penalized Cox regression. *Statistical Methods in Medical Research*, 33(4), 702–715.

[17] Peyraud, E., Jacques, J., Metzler, G., Faivre, I., & Dousse, M. (2024). Mixture of Cox regression models with L1-penalization for modeling patients' survival time after liver transplantation. *Statistics in Medicine*.

[18] Eslami, Z., Norouzirad, M., & Arashi, M. (2021). Penalized estimators in Cox regression model. *Andishe-ye Amari*, 25(1), 53–67.

[19] Hoerl, A. E., & Kennard, R. W. (1970). Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics*, 12(1), 55–67.

[20] Tibshirani, R. (1996). Regression shrinkage and selection via the LASSO. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 58(1), 267–288.

[21] Zou, H., & Hastie, T. (2005). Regularization and variable selection via the Elastic Net. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 67(2), 301–320.

[22] National Center for Biotechnology Information (2012). Gene Expression Omnibus dataset GSE20685: Breast cancer gene expression profiles. *NCBI Gene Expression Omnibus*. Available at: https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE20685