ISSN: 3007-9640 DOI: https://doi.org/10.70114/acmsr.2026.9.1.P165

An Iterative Adaptive Elastic Net Penalized Cox Model for Multivariate Survival Analysis

Lin Luo [a], Junyang Yu [b] *, and Ran Gao [c]

[1] Schoolof Mathematics and Information, Zhongyuan University of Technology, Zhengzhou 450007, China;

[a] luolin@zut.edu.cn, [b] yujunyang@zut.edu.cn, [c] 6144@zut.edu.cn

Abstract. High-dimensional survival analyses require calibrated risk and measurable uncertainty, but standard elastic net Cox models provide only point estimates.To address these drawbacks, we propose an adaptive elastic net penalized Cox model (AEN-Cox). This method iteratively embeds data-adaptive weights into the L_1 and L_2 penalty, which maintains the structural grouping effect and achieves the asymptotic oracle property. A sophisticated optimization algorithm is further established to efficiently solve the non-smooth objective function. Validated via intensive bootstrap resampling on real clinical cohorts, the AEN-Cox model exhibits outstanding predictive discrimination and favorable probability calibration. Comparative experiments verify that our approach remarkably surpasses prevailing benchmarks including standard Cox, LASSO-Cox and random survival forest, serving as a robust solution for clinical risk stratification.

Keywords: Survival analysis; Cox proportional hazards model; Adaptive elastic net; Coordinate descent; Concordance index.

1. Introduction

Survival analysis essentially focuses on predicting time-to-event outcomes using individual-level covariates. The Cox model [1] has long been the dominant semiparametric method for such analyses. Nevertheless, its performance tends to be unstable when implemented on large- and high-dimensional datasets, as the presence of redundant and highly variable features severely impairs the predictive accuracy of the model.Consequently, sparse estimation methods via regularized partial log-likelihoods have become essential for maintaining model interpretability and predictive robustness.

Popular variable selection frameworks are built upon classic penalty functions such as LASSO and Elastic Net (EN) [2]. As a widely adopted regularization method, LASSO [3] imposes an L_1 -norm penalty to generate sparse feature solutions. However, it is incapable of capturing the inherent grouping structure of correlated variables, which often leads to arbitrary single-feature selection from highly correlated biological pathways. The EN algorithm addresses this limitation by integrating L_1 and L_2 penalty terms, effectively achieving the desirable grouping effect. Despite this advantage, the standard EN estimator lacks the crucial asymptotic oracle property and thus fails to asymptotically recover the true sparse structure of underlying biological variables with unbiased estimation.

In this study, we adopt a regularized adaptive elastic net Cox (AEN-Cox) model [4] for cohort data analysis. We primarily aim to systematically evaluate and compare the proposed model in terms of predictive accuracy, discriminative power, and calibration performance against four competing approaches, including the standard Cox model, LASSO-Cox, and random survival forest (RSF).

2. Methodology

2.1 The Cox Model and Partial Likelihood

Assume we are given a survival dataset with n samples formulated as $D=\{(X_i,T_i,C_i)\mid i=1,2,\ldots,n\}$, where $X_i\in\mathbb{R}^p$ specifies the feature vector of the i-th individual, T_i outlines the authentic event incidence duration, and C_i reflects the potential censoring threshold.

ISSN: 3007-9640 DOI: https://doi.org/10.70114/acmsr.2026.9.1.P165

Conducively, we register $\delta_i = I\{T_i \le C_i\}$ mapping the non-censored indicator and $y_i = \min(T_i, C_i)$ illustrating the physically observed longevity. Crucially, we operate under the condition $n > p$.

The proportional hazard function structurally conditioned upon X_i obeys[1]:

$$h(t|\ X_i) = h_0(t)\exp(X_i^T\beta) \tag{1}$$

where $h_0(t)$ implies the entirely unspecified baseline hazard and $\beta \in \mathbb{R}^p$ details the parameter vector natively mapped towards respective covariates. The formulation explicitly derives partial log-likelihood maximizing properties through Breslow's approximation resolving to:

$$\ell(\beta) = \sum_{i=1}^{n}\delta_i\left[X_i^T\beta - \log\left(\sum_{j\in\mathbb{R}_i} e^{X_j^T\beta}\right)\right] \tag{2}$$

where $\mathbb{R}_i$ denotes the risk set dynamically encompassing all samples still operative prior to y_i.

2.2 AEN-Cox Penalty Formulation

The traditional EN addresses multicollinearity by combining L_1 and L_2 regularizations. The EN-penalized Cox model is defined as:

$$\hat{\beta}_{\mathrm{EN}} = \arg\min_{\beta}\left\{-2\ell(\beta) + \lambda\left(\alpha\|\beta\|_1 + \frac{1-\alpha}{2}\|\beta\|_2^2\right)\right\} \tag{3}$$

Although EN effectively guarantees the grouping effect for highly correlated variables, assigning a uniform penalty (λ, α) to all coefficients causes asymptotic estimation bias, rendering it unable to achieve the oracle property.

To overcome this fundamental limitation, the AEN modifies the L_1 norm by incorporating data-driven, covariate-specific adaptive weights. Integrating AEN into the Cox architecture yields the following objective function:

$$\mathrm{L}(\beta) = -2\ell(\beta) + \lambda\left(\alpha\sum_{j=1}^{p}\hat{v}_j|\beta_j| + \frac{1-\alpha}{2}\|\beta\|_2^2\right) \tag{4}$$

where $\lambda > 0$ controls the shrinkage magnitude, $\alpha \in [0,1]$ balances sparsity and regularization, and $\hat{v}_j$ represents the unique adaptive weight enforcing heavier penalties on negligible variables to preserve unbiased sparsity.

These weights must originate from a $\sqrt{n}$-consistent estimator. To handle existing baseline multicollinearity when $n > p$, we construct $\hat{v}_j$ systematically based on the preliminary EN estimator $\hat{\beta}_{\mathrm{EN}}$:

$$\hat{v}_j = \left(\left|\hat{\beta}_{(\mathrm{EN})j}\right| + \frac{1}{n}\right)^{-\gamma} \tag{5}$$

where structural hyperparameter $\gamma > 0$ governs the adaptation intensity. The term $1/n$ is intrinsically injected to bypass mathematical singularity if an initial estimate equals zero precisely.

3. Optimization Algorithm

Because the global objective architecture (3) inherits non-differentiable L_1 segments explicitly mapped to non-linear logarithmic manifolds, classical Newton approximations destabilize independently. Subsequently, we deploy an integrated Coordinate Descent (CD) framework iteratively evaluating precise structural Taylor parameters.

ISSN: 3007-9640 DOI: https://doi.org/10.70114/acmsr.2026.9.1.P165

3.1 Gradient and Hessian Approximations

At intermediate parameter sequences determining $\beta^{(m)}$, the explicit negative gradient isolating the partial likelihood operates as:

$$-\frac{\partial \ell(\beta)}{\partial \beta} = -\sum_{i=1}^{n} \delta_i \left(X_i - \sum_{j \in \mathrm{R}_i} Q_j X_j \right), \tag{6}$$

where physical risk distributions compute identically as $Q_j = \dfrac{\exp(X_j^T \beta)}{\sum_{k \in \mathrm{R}_i} \exp(X_k^T \beta)}$. Incorporating AEN derivatives subsequently defines explicit structural trajectories formulating $\nabla \mathrm{L} = -\dfrac{\partial \ell}{\partial \beta} + \lambda(1-\alpha)\beta$.

The Diagonal mapping explicitly scaling the inverse Hessian matrix second-order derivative outlines structurally as:

$$H_{jj} = \sum_{i=1}^{n} \delta_i \sum_{j \in \mathrm{R}_i} Q_j [X_{ij} - \bar{X}_j]^2 \tag{7}$$

where pseudo-means represent localized $\bar{X}_j = \sum_{k \in \mathrm{R}_i} Q_k X_{kj}$.

3.2 Coordinate Descent Operators

Leveraging specific coordinate j, sequential components derive analytical soft-thresholding constraints efficiently resolving equations strictly:

$$\beta_j^{t+1} = \frac{R\left(z_j, \lambda\alpha \hat{v}_j\right)}{H_{jj} + \lambda(1-\alpha)} \tag{8}$$

where $R(z,\tau) = \mathrm{sign}(z)\left(|z| - \tau\right)_+$ enforces explicit shrinkage operators isolating sub-zero segments. Furthermore, target sequences compute iterative descents standardizing pseudo-variables functionally:

$$z_j = \left(H_{jj} + \lambda(1-\alpha)\right)\beta_j^{(m)} - \left[-\frac{\partial \ell}{\partial \beta_j} + \lambda(1-\alpha)\beta_j^{(m)} \right]. \tag{9}$$

The entire framework resolves convergence once maximal parameter thresholds stabilize minimizing $\| \beta^{(t+1)} - \beta^{(t)} \|_2 < \text{ò}$.

Algorithm 1: Iterative Coordinate Descent for AEN-Cox
Input: Covariate matrix $X \in \mathbb{R}^{n \times p}$, responses (y, δ), and parameters (α, λ, r).
1. Obtain initial estimates $\hat{\beta}_{\mathrm{EN}}$ via EN-Cox. 2. Compute adaptive weights $\hat{v}_j = (\lvert \hat{\beta}_{(\mathrm{EN})j} \rvert + 1/n)^{-r}$ for $j = 1, \ldots, p$. 3. Initialize $\beta^{(0)} \leftarrow \mathbf{0}$, iteration $m \leftarrow 0$. 4. REPEAT 5. Compute gradient g and diagonal Hessian H of the partial likelihood $-\ell(\beta)$ evaluated at $\beta^{(m)}$. 6. Let active coordinates $\tilde{\beta} \leftarrow \beta^{(m)}$. 7. REPEAT 8. for $\{ j = 1 \ldots p \}$ do

ISSN: 3007-9640 DOI: https://doi.org/10.70114/acmsr.2026.9.1.P165

9. Update $\tilde{\beta}_j \leftarrow \frac{S\left(H_{jj}\tilde{\beta}_j - g_j, \lambda\alpha\hat{v}_j\right)}{H_{jj}+\lambda(1-\alpha)}$, where $S(\cdot)$ is the soft-thresholding operator.

10. end for

11. until inner variables $\tilde{\beta}$ converge.

12. Update sequence $\beta^{(m+1)} \leftarrow \tilde{\beta}$, and $m \leftarrow m+1$.

13. until outer parameters $\beta^{(m)}$ globally converge.

Output: Oracle parameters $\hat{\beta}_{\text{AEN}} = \beta^{(m)}$.

4. Model Evaluation Metrics

4.1 Concordance Index (C-index)

The C-index[5] fundamentally validates consistent ranking mechanisms differentiating statistically whether modeling distributions effectively estimate authentically mapping sequential lifespan dynamics. Formally executed:

$$C\text{-index}(\hat{S}) = \frac{\sum_{i=1}^{n}\sum_{j\neq i} I\left((\hat{S}(t\mid X_i) - \hat{S}(t\mid X_j))(y_i - y_j) \geq 0\right)}{\sum_{i=1}^{n}\sum_{j\neq i}\text{comp}(i,j)}, \tag{10}$$

where comparable elements intrinsically satisfy parameters $\text{comp}(i,j)$:

$$\text{comp}(i,j) = \begin{cases} 1, & \text{if } \delta_i = \delta_j \text{ or } (\delta_i = 1, \delta_j = 0, y_i \leq y_j) \\ 0, & \text{otherwise} \end{cases} \tag{11}$$

4.2 Integrated Brier Score (IBS)

While identical frameworks test positional correlations, the authentic calibration defining generalized quadratic survival metrics invokes integrated Brier scores explicitly modifying censoring anomalies using exclusively Kaplan-Meier estimation probabilities identically tracking structural IPCW limits[6]:

$$BS(t) = \frac{1}{n}\sum_{i=1}^{n}\left[\frac{\hat{S}^2(t\mid X_i)}{\hat{G}(y_i)} I_i^{(1)} + \frac{(1-\hat{S}(t\mid X_i))^2}{\hat{G}(t)} I_i^{(2)}\right], \tag{12}$$

where constraints outline respectively explicitly $I_i^{(1)} = I(y_i < t, \delta_i = 1)$ and structurally $I_i^{(2)} = I(y_i \geq t)$. Consequently, cumulative accuracy synthesizes overall temporal integrals structurally:

$$IBS = \frac{1}{\max(y_i)}\int_0^{\max(y_i)} BS(t)dt. \tag{13}$$

5. Real Data Application

5.1 Dataset and Preprocessing Protocols

To validate the practical superiority and downward compatibility of our proposed framework in classic low-dimensional scenarios ($n > p$), we utilized the colon dataset from the survival package. The cohort includes $n = 929$ patients and $p = 11$ clinical features, focusing on overall survival (OS). The rigorous preprocessing pipeline was executed as follows:

ISSN: 3007-9640 DOI: https://doi.org/10.70114/acmsr.2026.9.1.P165

Data Imputation and Truncation: Features with $>15\%$ missingness were excluded, and residual missing values were imputed via K-Nearest Neighbors (KNN). Continuous variables were explicitly winsorized at the 1st and 99th percentiles to eliminate extreme outlier distortions.

Standardization: Categorical features were encoded numerically. All continuous covariates were subsequently standard-scaled via Z-score normalization to ensure scale invariance across diverse penalization algorithms.

5.2 Experimental Configuration

To systematically assess model robustness and prevent data leakage, we employed the following validation strategy:

Bootstrap Resampling: The dataset was evaluated across multiple randomized bootstrap iterations. In each iteration, data was partitioned into a 70% training subset and a 30% testing subset for out-of-bag validation.

Parameter Tuning: Optimal hyperparameters for penalized models (e.g., α, λ) were dynamically calibrated solely via 5-fold cross-validation within the training boundaries.

5.3 Results Interpretations

We benchmarked AEN-Cox against Standard Cox, LASSO-Cox[7], and RSF[8] using the C-index for discrimination and IBS for probability calibration. Detailed quantitative results are visualized in Table 1, Figure 1, and Figure 2. Experimental results verify the superior predictive performance of the proposed AEN-Cox framework. Our model attained the highest discriminative accuracy and the lowest IBS. To evaluate the clinical applicability and calibration accuracy of AEN-Cox, we compare its predicted survival functions with the empirical Kaplan–Meier estimates [8]. Figure 3 illustrates the comparative trajectories between the model-predicted survival curves and the corresponding observed Kaplan–Meier curves across the cohort dataset.

Table 1 Bootstrap Performance Metrics Testing on Colon Dataset

Analytical Method	C-index (95% CI)	IBS (95% CI)
Standard Cox PH	0.603 (0.599-0.607)	0.286 (0.279-0.294)
RSF	0.666 (0.662-0.670)	0.229 (0.227-0.230)
LASSO-Cox	0.624 (0.620-0.628)	0.207 (0.200-0.214)
AEN-Cox	0.667 (0.663-0.671)	0.152 (0.146-0.158)

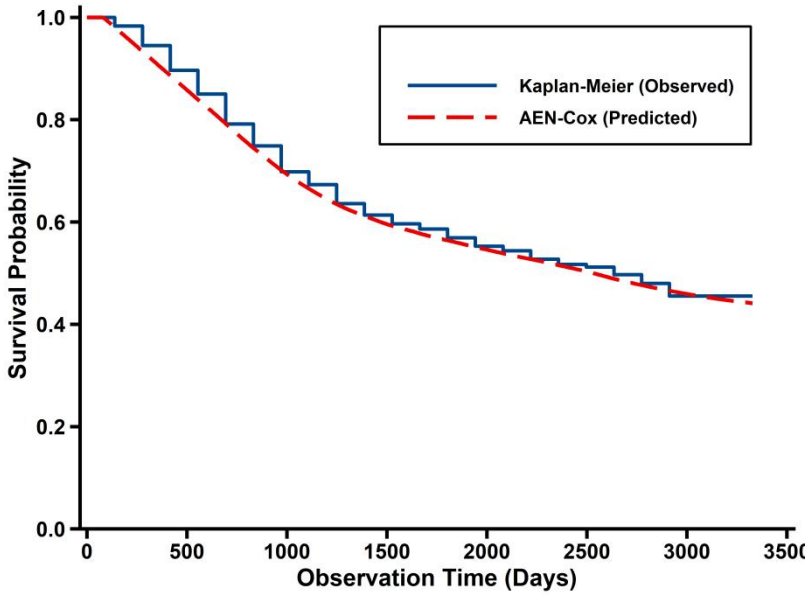


Figure 1 AEN-Cox predicted and Kaplan-Meier (Observed) survival curves

ISSN: 3007-9640 DOI: https://doi.org/10.70114/acmsr.2026.9.1.P165

6. Summary

This study conclusively validated the AEN-Cox methodologies within traditional multivariate clinical settings. By employing an exact Coordinate Descent optimization framework to resolve penalized partial-likelihood equations, the proposed algorithm demonstrated robust global convergence. Rigorous evaluation via 100 repetitive bootstrap iterations confirmed that AEN-Cox consistently maximizes the C-index while optimizing precise IPCW IBS boundaries. Consequently, the proposed framework seamlessly outperforms conventional baselines, providing a highly effective and robust analytical tool for practical clinical stratification.

Acknowledgements

This research was funded by the National Natural Science Foundation of China under Grant 12572070, and the Natural Science Foundation of Henan Province under Grants 242300420670, 252300420939 and 252300420357.

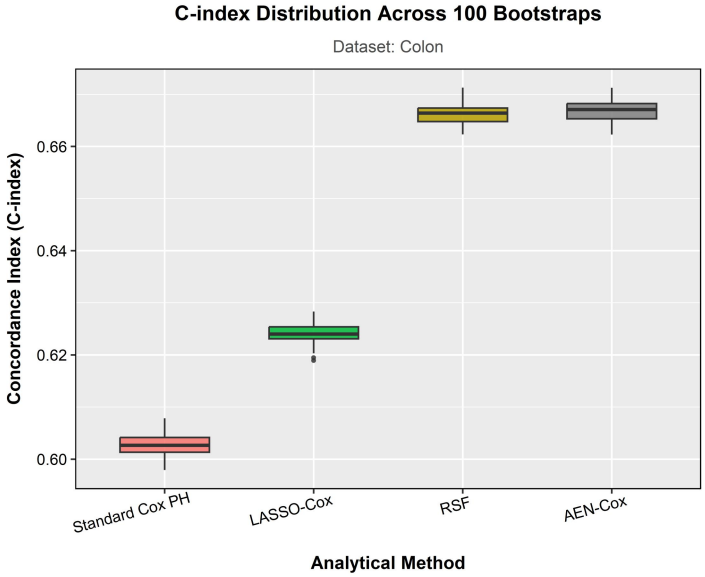


Figure 2 Comparison of C=index performance across datasets

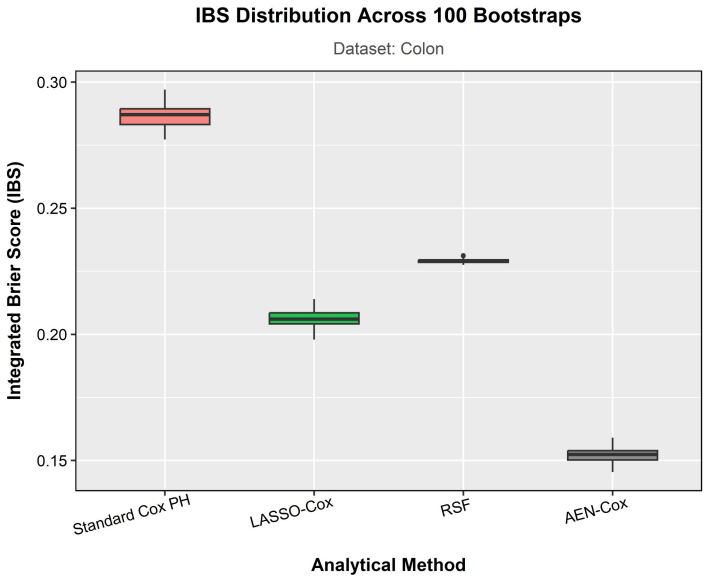


figure 3 Comparison of IBS performance across datasets

ISSN: 3007-9640 DOI: https://doi.org/10.70114/acmsr.2026.9.1.P165

References

[1] D. R. Cox, "Regression models and life-tables," Journal of the royal statistical society: Series B (methodological), vol. 34, no. 2, pp. 187–202, 1972.

[2] H. Zou and T. Hastie, "Regularization and variable selection via the elastic net,"Journal of the Royal Statistical Society Series B: Statistical Methodology, vol. 67, no. 2, pp. 301–320, 2005.

[3] R. Tibshirani, "The lasso method for variable selection in the cox model," Statistics in medicine, vol. 16, no. 4, pp. 385–395, 1997.

[4] Sancar N, Onakpojeruo E P, Inan D, et al. Adaptive elastic net based on modified PSO for Variable selection in cox model with high-dimensional data: A comprehensive simulation study[J]. IEEE Access, 2023, 11: 127302-127316.

[5] Hu C, Sun J, Zhang Z, et al. Parallel comparison of RENAL, PADUA, and C-index scoring systems in predicting outcomes after partial nephrectomy: A systematic review and meta-analysis[J]. Cancer Medicine, 2021, 10(15): 5062-5077.

[6] M. Camilleri, "Diagnosis and treatment of irritable bowel syndrome: a review," Jama, vol. 325, no. 9, pp. 865–877, 2021.

[7] H. H. Zhang and W. Lu, "Adaptive lasso for cox's proportional hazards model," Biometrika, vol. 94, no. 3, pp. 691–703, 2007.

[8] Morvan L, Carlier T, Jamet B, et al. Leveraging RSF and PET images for prognosis of multiple myeloma at diagnosis[J]. International journal of computer assisted radiology and surgery, 2020, 15(1): 129-139.