mathematics

MDPI

Article

A Bayesian Hierarchical Cox Model with Elastic Net Regularization for Improved Survival Prediction and Feature Selection

Bulus I. Doroh [1], Kazeem A. Dauda [2,*] and Rasheed K. Lamidi [1]

[1] Department of Mathematics and Statistics, Kwara State University, Malete, P.M.B. 1530, Ilorin 23431, Kwara State, Nigeria
[2] Department of Mathematics, University of Bergen, 5007 Bergen, Norway
* Correspondence: kazeem.dauda@uib.no or qdauda70@gmail.com

Abstract

In recent years, the growing availability of large-scale data across a wide range of disciplines has created new opportunities for developing models that improve the predictive accuracy of statistical models. Although techniques such as regularization and Bayesian hierarchical methods are commonly used for building predictive models, substantial challenges remain, particularly when dealing with high-dimensional datasets that contain considerable noise. In this study, we propose a Bayesian hierarchical model that employs a spike-and-slab hierarchical elastic net prior that regularizes the Cox Proportional Hazards (Cox-PH) model. The method combines Bayesian modeling with the regularized partial log-likelihood of the Cox-PH framework, incorporating an Elastic Net penalty to estimate the joint posterior distribution under a hierarchical elastic net prior. We compute this posterior using an Expectation–Maximization Cyclic Coordinate Descent Algorithm (EM-CCDA), which streamlines feature selection and enhances overall predictive performance. We evaluate the algorithm's performance through Monte Carlo simulations and apply it to three real-world datasets, comparing the results with those from established classical and Bayesian survival analysis approaches. The findings demonstrate notable gains in both feature selection and predictive accuracy, highlighting the model's strong ability to predict patient survival and identify relevant genes in real biological datasets.

Keywords: penalization function; elastic net; lasso; joint posterior function; hierarchical

MSC: 62N02

Check for updates

Academic Editors: Bogdan Oancea, Adrian Pană and Cătălina Liliana

Received: 22 December 2025
Revised: 20 February 2026
Accepted: 20 February 2026
Published: 25 February 2026

Copyright: © 2026 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license.

1. Introduction

The growing availability of large and complex datasets across various fields has increased the need for modern and robust statistical models that can effectively handle complex feature spaces while minimizing the impact of noise [1–3]. In the field of biomedical research, these challenges are particularly evident and also affect modern statistical and machine learning methods. Data from genomic, proteomic, and clinical studies are often large, complex, and highly correlated, yet sample sizes remain limited. As a result, traditional approaches frequently struggle with instability, overfitting, and variable selection. Thus, there is an increasing demand for advanced modeling approaches that can seamlessly accommodate the high-dimensional structure, perform effective regularization, and capture

the underlying biological association driving survival outcomes. In the context of survival analysis, the Cox Proportional Hazards (Cox-PH) model [4] is widely utilized due to its flexibility and interpretability, enabling the estimation of hazard ratios and the analysis of time-to-event data [5]. However, the model's performance is often unstable when applied to large-scale datasets, since such data often contain redundant or highly variable features that can compromise predictive accuracy [6–8]. Recent studies have focused on filling these gaps through the development of regularization and Bayesian approaches that enhance stability and sparsity in high-dimensional modeling. Penalized regression and Bayesian frameworks have gained significant attention in modern statistical research for variable selection in high-dimensional settings [9–11]. By applying an L1-penalty, these methods shrink many coefficients to zero, achieving both variable selection and model sparsity. Consequently, they enhance predictive performance and effectively address the issues of multicollinearity and overfitting commonly observed in multi-omic settings. While penalized regressions provide effective shrinkage and selection mechanisms, Bayesian hierarchical models extend this framework by incorporating prior knowledge and uncertainty quantification [12,13].

Identifying prognostic genes from high-dimensional data remains essential, and several Bayesian and hierarchical approaches have been successfully applied [14,15]. Notably, Bayesian approaches have become increasingly popular as flexible and effective tools for variable selection and modeling in high-dimensional survival data [16–20]. Several authors, including [21–27], introduced the Bayesian spike-and-slab LASSO method for high-dimensional linear modeling. Their approach is based on a spike-and-slab framework constructed from a mixture of double-exponential prior distributions. This prior structure, which forms the basis of Bayesian variable selection methods, has been shown to perform efficiently in numerous applications.

Earlier work by [28] introduced a Bayesian hierarchical Cox proportional hazards model known as the spike-and-slab Lasso Cox model, specifically designed for high-dimensional molecular data. Through extensive simulation studies, the authors demonstrated that this approach outperforms conventional LASSO Cox regression. Nevertheless, the model is constrained to the LASSO framework, which limits its flexibility and restricts its ability to capture more complex sparsity structures commonly observed in large-scale, high-dimensional datasets.

Beyond these contributions, the generalized spike-and-slab prior combined with the Expectation Propagation algorithm has been shown to effectively perform group feature selection in linear regression problems. This framework achieves state-of-the-art predictive performance and demonstrates strong potential for active learning applications [29]. Following these developments, ref. [30] developed spike-and-slab LASSO generalized linear models (GLMs) that deliver precise prognostic predictions and improve gene discovery in large-scale molecular datasets, thereby enhancing both the interpretability and robustness of cancer prognosis models.

In addition, ref. [16] introduced Bayesian regularization for Cox survival models using flexible baseline hazards, which provide robust estimates, indicating that B-splines are less dependent on regularization compared to piecewise specifications. However, these studies primarily focus on linear regression with Gaussian responses, and posterior computation has been noted to be particularly challenging due to the complexity involved in updating the discrete indicators that control model parameters. While these studies highlight the growing role of hierarchical priors in high-dimensional modeling, most existing Bayesian

Cox methods still face challenges related to posterior instability, limited scalability, and weak integration of structured priors in high-dimensional survival settings.

To address these gaps, this study proposes a Bayesian hierarchical Cox Proportional Hazards (Cox-PH) model that incorporates spike-and-slab hierarchical elastic net prior regularization, a hybrid approach that combines the strength of L_1 (lasso) and L_2 (ridge) penalties. The purpose of this penalty function is to simultaneously perform variable selection and control multicollinearity, thereby enhancing model stability, adaptive shrinkage, and predictive accuracy in high-dimensional survival data. By incorporating the Elastic Net penalty within a Bayesian hierarchical setting, the model balances variable selection and coefficient shrinkage, allowing both strongly and moderately correlated predictors to be retained with appropriate regularization strength. Essentially, the inclusion of the spike-and-slab hierarchical elastic net prior helps to provide feature selection by differentiating between significant and irrelevant covariates while adaptively shrinking small effects toward zero. This prior structure promotes sparsity in the coefficient space and helps to capture underlying signals in complex biomedical data. Posterior inference and parameter estimation are performed using the Expectation–Maximization (EM) and cyclic coordinate descent (CCD) algorithms, which iteratively optimize the model parameters to facilitate stable and computationally efficient convergence [31,32]. Specifically, our study differs from [28] in two main aspects. First, with respect to the regularization procedure, we employ the elastic net, whereas [28] used the Lasso. Second, in terms of prior specification, we adopt a hierarchical elastic net prior, while [28] relied on a Laplace prior.

This paper is structured as follows: Section 2 presents the methodology; Section 3 reviews related methods; Section 4 outlines the formulation of the proposed framework; Section 5 describes the coordinate descent algorithm and the E- and M-steps for the proposed spike-and-slab elastic net Cox–PH model; Section 6 reports the results; and Section 7 concludes this paper.

2. Methodology

The Cox proportional hazard model is a classical method of analyzing survival data. However, it faces challenges in the presence of high-dimensional data due to the singularity matrix involved, which can seldom be invertible. To address this issue, a penalty function is introduced to handle the problem of sparsity and convergence. In the Bayesian concept, the spike and slab formulation provides a probabilistic construct for sparse recovery, inducing sparsity through a penalty function integrated into constrained optimization.

This study introduces the Cox proportional hazard model as censored survival outcomes, where a pair of response variables $Y = (t_i, d_i)$ are observed for each patient. The censoring indicator d_i is 1 when the observed survival time t_i is uncensored and 0 when censored. The true survival time for patient i is represented as T_i. If $d_i = 1$, then $T_i = t_i$, and if $d_i = 0$, then $T_i > t_i$. Past studies indicate that Cox proportional hazard models are popular for identifying associations between survival responses and explanatory variables X, assuming that the hazard function for survival time T is of the form

$$h(t \mid X) = h_0(t) \exp(X\beta) \tag{1}$$

where $h_0(t)$ is the baseline hazard function and X and β are vectors of explanatory variables and coefficients, respectively. The prognostic index is represented by $X\beta$. Classical Cox models estimate β by maximizing the partial log-likelihood (pl):

$$\mathrm{pl}(\beta) = \sum_{i=1}^{n} \left[d_i \log \left(\frac{\exp(X_i\beta)}{\sum_{l \in R(t_i)} \exp(X_l\beta)} \right) \right] \tag{2}$$

where $R(t_i)$ is the risk set at time t_i and l represents each individual inside the risk set. This study penalized the partial likelihood of the Cox proportional hazards (Cox–PH) model using the Elastic-net penalty function, in contrast to the already established penalized Lasso [28]. We start by showing how the penalized partial likelihood of the Cox proportional hazards (Cox–PH) model is formulated. Let the Lasso penalty $p(\beta)$ be defined as

$$p(\beta) = \lambda \sum_{j=1}^{J} |\beta_j|, \tag{3}$$

and, thus, the penalized partial likelihood of the Cox–PH model is expressed as

$$ppl(\beta) = pl(\beta) - \lambda \sum_{j=1}^{J} |\beta_j|. \tag{4}$$

The parameter λ regulates the strength and size of all independent joint posteriors, controlling the degree of regularization and the selection of variables in the model. The Lasso penalty is represented as $\sum_{j=1}^{J} |\beta_j|$, with various ways of penalizing $pl(\beta)$.

3. Overview of the Posterior Development

This section reviews established Bayesian approaches to variable selection within the Cox proportional hazards framework. We begin by formulating the joint posterior distribution for the Lasso Cox–PH model, which combines the likelihood function with appropriate prior distributions to estimate regression coefficients under penalization. This formulation follows the framework introduced by [28] and serves as the foundation for the modified spike-and-slab Elastic-Net Cox–PH model proposed in this study.

Joint Posterior for the Spike-and-Slab Lasso Cox–PH Model (Sslasso-Coxph)

Within the Bayesian paradigm, the posterior distribution is obtained by combining the likelihood with the prior distributions. Specifically, the joint posterior distribution for the spike-and-slab Lasso Cox–PH model (sslasso-coxph) by [28,33] is defined as

$$\log p(\beta, \gamma, \theta \mid t, d) = \log p(t, d \mid \beta, h_0) + \sum_{j=1}^{J} \log p(\beta_j \mid S_j) + \sum_{j=1}^{J} \log p(\gamma_j \mid \theta) + \log p(\theta) \tag{5}$$

Here, the regression coefficients β are assumed to follow a double-exponential (Laplace) prior, h_0 is the baseline hazard function, the indicator variables γ (known to be a latent variable) and $\gamma \in \{0, 1\}$ follow independent Bernoulli distributions, and the hyperparameter θ is assigned a uniform prior. The corresponding log joint posterior distribution can be written as

$$\log p(\beta, \gamma, \theta \mid t, d) = \log p(\beta) + \sum_{j=1}^{J} \left(-\frac{1}{2s} |\beta_j| + \gamma_j \log(\theta) + (1 - \gamma_j) \log(1 - \theta) \right). \tag{6}$$

This formulation was originally developed by [28]; however, in the present work, we extend this framework by adopting a spike-and-slab hierarchical elastic net prior for Cox–PH models.

4. Bayesian Hierarchical Modeling for the Elastic Net Cox Proportional Hazards Model (Sselastic-Net-Coxph)

This study adopts a Bayesian hierarchical framework to penalize the Cox proportional hazards (Cox–PH) model using the elastic net penalty while incorporating a spike-and-slab mixture prior. The spike-and-slab formulation is employed as a prior distribution on the regression coefficients, and the corresponding joint posterior distribution is formulated as described below.

4.1. Penalized Cox–PH Model Using Elastic-Net Penalty Function

The partial likelihood of the Cox proportional hazards model is penalized using the elastic net penalty, as presented in Equation (7). The elastic net penalty combines both L_1 (lasso) and L_2 (ridge) regularization terms and is defined as

$$\text{pen}_\lambda(\beta) = \lambda_1 \sum_{j=1}^{J} |\beta_j| + \lambda_2 \sum_{j=1}^{J} \beta_j^2. \tag{7}$$

Accordingly, the penalized partial likelihood (ppl) for the Cox–PH model with an elastic net penalty is expressed as

$$ppl(\beta) = \text{pl}(\beta) - \left(\lambda_1 \sum_{j=1}^{J} |\beta_j| + \lambda_2 \sum_{j=1}^{J} \beta_j^2 \right). \tag{8}$$

Here, λ_1 and λ_2 denote the regularization (tuning) parameters governing the relative contribution of the L_1 and L_2 penalties. Both control the amount and type of shrinkage applied to the regression coefficients. The terms $|\beta_j|$ and β_j^2 correspond to the lasso and ridge components, respectively.

4.2. Spike-and-Slab Elastic Net Cox–PH Model

The Spike-and-Slab Elastic Net Cox–PH Model is a Bayesian hierarchical model that builds on the elastic net prior framework. For a more convenient representation, the elastic net prior in Equation (7) is characterized in the literature [34–36] by the standardized hierarchical elastic net prior distribution given in Equation (9).

$$p(\beta \mid \lambda_1, \lambda_2) \propto \exp\left(-\frac{\lambda_2}{2} \sum_{j=1}^{J} \beta_j^2 - \frac{\lambda_1}{2} \sum_{j=1}^{J} |\beta_j| \right). \tag{9}$$

For easier computation and follow-up, we introduce global scale parameters s_1 and s_2, where $s_1 = \frac{1}{\lambda_1}$ and $s_2 = \frac{1}{\lambda_2}$, respectively. Thus, a reparameterized version is obtained and presented in Equation (10).

$$p(\beta \mid s_1, s_2) \propto \exp\left(-\frac{1}{2s_2} \sum_{j=1}^{J} \beta_j^2 - \frac{1}{2s_1} \sum_{j=1}^{J} |\beta_j| \right). \tag{10}$$

The scale parameters s_1 and s_2 control the amount of shrinkage; smaller scales induce stronger shrinkage and force the estimates of β_j toward 0. Specifically, if s_1 decreases, then $\frac{1}{s_1}$ increases, leading to a stronger L_1 penalty; similarly, if s_2 decreases, then $\frac{1}{s_2}$ increases, resulting in a stronger L_2 penalty.

Next, we show the development of the spike-and-slab elastic net Cox model using the quantity in Equation (10):

$$p(\beta_j) = (1-\gamma_j)\,\exp\left(-\frac{1}{2s_{2,0}}\,\beta_j^2 - \frac{1}{2s_{1,0}}\,|\beta_j|\right) + \gamma_j\,\exp\left(-\frac{1}{2s_{2,1}}\,\beta_j^2 - \frac{1}{2s_{1,1}}\,|\beta_j|\right). \quad (11)$$

Equation (11) can also be written in a hierarchical form by conditioning on the latent indicator variable γ_j, similar to the formulation in [28]. Rather than expressing the prior as a mixture of spike and slab elastic net components, we represent it using a single conditional prior:

$$\beta_j \mid \gamma_j \propto \exp\left(-\frac{1}{2s_{2j}}\beta_j^2 - \frac{1}{2s_{1j}}|\beta_j|\right), \quad (12)$$

where the scale parameters depend on the value of the indicator variable, defined as

$$s_{1j} = (1-\gamma_j)s_{1,0} + \gamma_j s_{1,1}, \qquad s_{2j} = (1-\gamma_j)s_{2,0} + \gamma_j s_{2,1}.$$

This representation is equivalent to the mixture form in Equation (11), but it is easier to work with in a hierarchical setting and simplifies the computational implementation, similar to [28]. The scales are chosen such that $s_{1,1} > s_{1,0} > 0$ and $s_{2,1} > s_{2,0} > 0$, ensuring that the slab component imposes weaker shrinkage, whereas the spike component enforces stronger shrinkage toward zero.

The slab component permits the regression coefficients to assume relatively large values, whereas the spike component aggressively shrinks coefficients toward zero, promoting sparsity. The latent indicators $\gamma_j \in \{0,1\}$ represent model inclusion variables and are assumed to follow a Bernoulli distribution with an unknown inclusion probability parameter θ_j, defined as follows:

$$\gamma_j \mid \theta \sim \text{Bernoulli}(\theta), \qquad p(\gamma_j \mid \theta) = \theta^{\gamma_j}(1-\theta)^{1-\gamma_j}. \quad (13)$$

The inclusion probability parameter θ is assumed to follow a uniform distribution, $\theta \sim U(0,1)$.

4.3. Joint Posterior for Spike-and-Slab Elastic Net Cox–PH Model

In the Bayesian school of thought, the posterior is considered as the product of the log-likelihood of a model and the given prior distribution. We jointly use $P(\cdot \mid t,d)$, which represents the posterior, and $\prod(\cdot)$, which stands for the prior. Thus, the log joint posterior of the Spike-and-Slab Elastic Net Cox–PH Model $P(\beta,\gamma,\theta \mid t,d)$ is

$$\log p(\boldsymbol{\beta},\boldsymbol{\gamma},\theta \mid t,d) = \log p(t,d \mid \boldsymbol{\beta}) + \sum_{j=1}^{J} \log p(\beta_j \mid \gamma_j) + \sum_{j=1}^{J} \log p(\gamma_j \mid \theta) + \log p(\theta). \quad (14)$$

where $\log p(t,d \mid \boldsymbol{\beta})$ is proportional to the partial log-likelihood $pl(\beta)$ defined in Equation (2). Therefore, the log joint posterior for the spike-and-slab elastic net Cox–PH model can be fully expressed as

$$\log p(\boldsymbol{\beta},\boldsymbol{\gamma},\theta \mid t,d) = pl(\beta) - \sum_{j=1}^{J}\left(\frac{1}{2s_{2j}}\beta_j^2 + \frac{1}{2s_{1j}}|\beta_j|\right) + \sum_{j=1}^{J}\left[\gamma_j \log\theta + (1-\gamma_j)\log(1-\theta)\right] + \log p(\theta). \quad (15)$$

5. Procedure for EM–Coordinate Descent Algorithm

The EM–coordinate descent algorithm considers the indicator variables γ_j as latent quantities and estimates the parameters $(\boldsymbol{\beta}, \theta)$ by integrating over their posterior distributions. In this formulation, the indicator variables represent whether each coefficient belongs to the spike or slab component. The EM algorithm is based on the log joint posterior density, and the parameter estimates are updated iteratively through the E-step and M-step.

5.1. E-Step

In the E-step, we compute the expectation of the log joint posterior density with respect to the conditional posterior distributions of the latent indicators γ_j. The conditional posterior probability that the jth coefficient belongs to the slab component is given by

$$p_j = P(\gamma_j = 1 \mid \beta_j, \theta, t, d) = \frac{p(\beta_j \mid \gamma_j = 1)\, P(\gamma_j = 1 \mid \theta)}{p(\beta_j \mid \gamma_j = 0)\, P(\gamma_j = 0 \mid \theta) + p(\beta_j \mid \gamma_j = 1)\, P(\gamma_j = 1 \mid \theta)}. \tag{16}$$

where $P(\gamma_j = 1 \mid \theta) = \theta$ and $P(\gamma_j = 0 \mid \theta) = 1 - \theta$.

For the proposed spike-and-slab elastic-net model, the conditional prior for each coefficient β_j, depending on whether $\gamma_j = 1$ (slab, $p(\beta_j \mid \gamma_j = 1)$) or $\gamma_j = 0$ (spike, $p(\beta_j \mid \gamma_j = 0)$), is defined as

$$p(\beta_j \mid \gamma_j) \propto \exp\left(-\frac{1}{2s_{2j}} \beta_j^2 - \frac{1}{2s_{1j}} |\beta_j| \right), \tag{17}$$

where the scale parameters depend on the indicator variable as

$$s_{1j} = (1 - \gamma_j) s_{1,0} + \gamma_j s_{1,1}, \qquad s_{2j} = (1 - \gamma_j) s_{2,0} + \gamma_j s_{2,1},$$

with $s_{1,1} > s_{1,0}$ and $s_{2,1} > s_{2,0}$ so that the slab component imposes weaker shrinkage than the spike component.

Therefore, the conditional posterior expectations of the inverse scale parameters are obtained as

$$E(s_{1j}^{-1} \mid \beta_j) = \frac{1 - p_j}{s_{1,0}} + \frac{p_j}{s_{1,1}}, \tag{18}$$

$$E(s_{2j}^{-1} \mid \beta_j) = \frac{1 - p_j}{s_{2,0}} + \frac{p_j}{s_{2,1}}. \tag{19}$$

These expectations provide adaptive shrinkage weights for the L_1 and L_2 penalties, respectively.

5.2. M–Step

In the M-step, $(\boldsymbol{\beta}, \theta)$ are updated by maximizing the expected log joint posterior, with the latent indicators γ_j and the inverses of the scale parameters, s_{1j}^{-1} and s_{2j}^{-1}, substituted by their conditional posterior expectations, as given in Equations (18) and (19). From the log joint posterior density (Equation (15)), it can be seen that $\boldsymbol{\beta}$ and θ can be updated separately because the coefficients $\boldsymbol{\beta}$ are only involved in the prior penalty terms, whereas the probability parameter θ is only involved in $\sum_j \gamma_j \log\theta + (1 - \gamma_j)\log(1 - \theta)$.

Therefore, the coefficients $\boldsymbol{\beta}$ are updated by maximizing the expression

$$Q_1(\boldsymbol{\beta}) = pl(\boldsymbol{\beta}) - \sum_{j=1}^{J}\left(\frac{1}{2}\,E[s_{2j}^{-1}]\,\beta_j^2 + \frac{1}{2}\,E[s_{1j}^{-1}]\,|\beta_j|\right), \tag{20}$$

Given the adaptive scale parameters, the term

$$\sum_{j=1}^{J}\left(\frac{1}{2}E[s_{2j}^{-1}]\beta_j^2 + \frac{1}{2}E[s_{1j}^{-1}]|\beta_j|\right)$$

serves as an elastic net penalty with coefficient-specific penalty factors. Thus, the coefficients can be updated by maximizing $Q_1(\boldsymbol{\beta})$ using the cyclic coordinate descent algorithm, allowing some coefficients to be shrunk toward zero while others remain relatively large.

The probability parameter θ is updated by maximizing

$$Q_2(\theta) = \sum_{j=1}^{J}\big(\gamma_j \log\theta + (1-\gamma_j)\log(1-\theta)\big), \tag{21}$$

while $\theta = \frac{1}{J}\sum_{j=1}^{J} p_j$.

Therefore, the model uses cyclic coordinate descent to update β_j by maximizing $Q_1(\beta)$, and θ is updated by maximizing $Q_2(\theta)$. In summary, the EM coordinate descent algorithm for fitting the spike-and-slab elastic net Cox model is presented in Algorithm 1. The convergence is assessed using the relative change in the partial log-likelihood, as given in Equation (22), with tolerance $\varepsilon = 10^{-5}$.

$$\frac{|d^{(t)} - d^{(t-1)}|}{0.1 + |d^{(t)}|} < \varepsilon, \tag{22}$$

where $d^{(t)} = pl(\boldsymbol{\beta}^{(t)})$.

5.3. Selecting Optimal Scale Values

This study aims to select optimal scale values through a hierarchical spike-and-slab framework using two sets of scale parameters, $(s_{1,0}, s_{1,1})$ for the L_1 component and $(s_{2,0}, s_{2,1})$ the L_2 component. The objective is to determine appropriate scale values that control the strength of shrinkage in the proposed spike-and-slab elastic net Cox–PH model.

Instead of focusing on a single model, the slab scales $s_{1,1}$ and $s_{2,1}$ are fixed at relatively large values to induce weak or minimal shrinkage, while the spike scales $s_{1,0}$ and $s_{2,0}$ are varied over sequences of smaller values such that

$$0 < s_{1,0}^{(1)} < s_{1,0}^{(2)} < \cdots < s_{1,0}^{(L)} < s_{1,1}, \qquad 0 < s_{2,0}^{(1)} < s_{2,0}^{(2)} < \cdots < s_{2,0}^{(L)} < s_{2,1}.$$

As the spike scales are increased, more coefficients are allowed to shift from the spike to the slab component. This process incorporates additional non-zero effects into the model, similar to what is seen in spike-and-slab lasso models [28].

Algorithm 1: EM Coordinate Descent Algorithm for the Spike-and-Slab Elastic Net Cox–PH Model

Input : Initial values of $\boldsymbol{\beta}$ and θ

Output: Estimated coefficients $\hat{\boldsymbol{\beta}}$ and inclusion probability $\hat{\theta}$

Initialize $\boldsymbol{\beta}^{(0)} = \mathbf{0}$, $\theta^{(0)} = 0.5$, and set $t = 0$;

while *convergence criterion not satisfied* **and** $t < t_{\max}$ **do**

  $t \leftarrow t + 1$;

  E-step: Update posterior inclusion probabilities p_j using Equation (16);

  Compute the conditional posterior expectations of the inverse scales using Equations (18) and (19);

$$E[s_{1j}^{-1} \mid \beta_j^{(t-1)}] = \frac{1 - p_j}{s_{1,0}} + \frac{p_j}{s_{1,1}}, \quad E[s_{2j}^{-1} \mid \beta_j^{(t-1)}] = \frac{1 - p_j}{s_{2,0}} + \frac{p_j}{s_{2,1}}$$

  M-step:;

    Update $\boldsymbol{\beta}^{(t)}$ using cyclic coordinate descent to maximize $Q_1(\boldsymbol{\beta})$ in Equation (20);

    Update the probability parameter $\theta^{(t)} = \frac{1}{J}\sum_{j=1}^{J} p_j$;

  Compute convergence criterion: $\dfrac{|pl(\boldsymbol{\beta}^{(t)}) - pl(\boldsymbol{\beta}^{(t-1)})|}{0.1 + |pl(\boldsymbol{\beta}^{(t)})|} < \varepsilon$;

end

Output: Final estimates $\hat{\boldsymbol{\beta}} = \boldsymbol{\beta}^{(t)}$ and $\hat{\theta} = \theta^{(t)}$

This approach is similar to the elastic net regularization path, where the `glmnet` package fits models across a range of tuning parameters, allowing the selection of an optimal model based on predictive performance [10,32]. In the present study, the adaptive scale parameters play a role similar to the tuning parameters by controlling the amount of shrinkage applied to each coefficient.

For predictive assessment, a pre-validation strategy (a variant of cross-validation) is employed. The data are partitioned into K subsets, and the model is fitted using $K - 1$ subsets while the remaining subset is used for validation. The prognostic index for the Kth subset is calculated as $\hat{\eta}_K = X_K \hat{\beta}^{(-K)}$. This represents the pre-validated prognostic index, which is based on models created without including the Kth subset.

This procedure is repeated across all subsets to obtain cross-validated performance measures. Additionally, cross-validated partial likelihood (CVPL) is used as an alternative validation criterion, defined as

$$\text{CVPL} = \sum_{K=1}^{K} \left[\text{PL}_{(-K)}(\hat{\beta}) - \text{PL}_{(-K)}(\hat{\beta}_{(-K)}) \right], \tag{23}$$

where $\hat{\beta}_{(-K)}$ denotes the parameter estimate obtained by excluding the Kth subset. The CVPL compares the partial likelihood evaluated using all observations with that obtained when one subset is omitted, providing an additional measure of predictive accuracy. This strategy allows our proposed spike-and-slab elastic net Cox model to perform simultaneous variable selection and coefficient stabilization, particularly in the presence of correlated predictors. The final code and wrapper for reproducing the results of the sselastic-net-coxph will be made publicly available on GitHub (https://github.com/Dydx1989 (accessed on 10 February 2026)) to ensure reproducibility.

6. Results

6.1. Simulation Procedure

In this section, simulation studies are conducted to evaluate the performance of the penalization methods, with particular emphasis on the proposed hierarchical spike-and-slab elastic net Cox model. The penalty functions considered in the study share common parameters, namely, the tuning parameters and the regression coefficients $\boldsymbol{\beta}$. The simulation procedure is designed to accommodate all the penalization methods under comparison.

Five simulation scenarios with varying levels of complexity (overlapping groups, group sizes, non-overlapping structures, number of non-zero groups, and correlation coefficients r) are considered. For each dataset, $n = 500$ observations are generated with a survival response consisting of observed survival times t_i and censoring indicators d_i. A vector of continuous predictors $X_i = (X_{i1}, \ldots, X_{iJ})$, with $J = 200, 1000$, is also generated.

The predictors are generated in groups. Specifically, sub-vectors $X_i(50k + 1), \ldots, X_i(50k + 50)$ are drawn from a multivariate normal distribution $N_{50}(0, \Sigma)$, where Σ is a covariance matrix with correlation coefficient r among predictors within each group. The predictors are correlated within groups but assumed to be independent between groups.

The "true" survival times T_i are generated from an exponential distribution as $T_i \sim exp\left(\sum_{j=1}^{J} X_{ij}\beta_j\right)$. The censoring times C_i for each individual are also generated from an exponential distribution $C_i \sim \exp(\phi_i)$, where ϕ_i is randomly selected from a standard normal distribution. The observed censored time t_i is set as the minimum between the "true" survival time and the censoring time $t_i = \min(T_i, C_i)$, while the censoring indicator d_i is set to 1 if $C > T$ and 0 otherwise. This data-generating mechanism follows standard procedures used in survival analysis simulation studies [28,33].

Table 1 presents the true non-zero variables used in the simulation settings. In Scenario 1, consisting of 500 samples and 200 predictors, and Scenario 2, consisting of 500 samples with 1000 predictors, the relevant features were successfully identified across the simulated replicates.

Table 1. Best Variables Selected.

Scenario	Variable 1	Variable 2	Variable 3	Variable 4
Scenario 1 ($n = 500, J = 200$)	X_5	X_{55}	X_{105}	X_{155}
Coefficient Values	0.6454	−0.6331	0.6294	−0.6361
Scenario 2 ($n = 500, J = 1000$)	X_5	X_{225}	X_{505}	X_{755}
Coefficient Values	0.6371	−0.6771	0.6500	−0.6269

Table 2 shows the simulation results for five different penalization methods applied to the Cox proportional hazards model across six separate scenarios. The evaluation used 10-fold cross-validation—a widely adopted approach in penalized Cox regression studies [37,38]—along with system time (Time), the concordance index (C-Index), deviance, and mean absolute error (MAE). The methods being compared include the Bayesian Cox model with a double-exponential prior (d-e bcoxph), the Bayesian Cox model with a Student-t prior (t-prior bcoxph), the Lasso-penalized Cox model (Lasso coxph), the Spike-and-Slab Lasso Cox model (sslasso-coxph), and the Spike-and-Slab Elastic Net Cox model (sselastic-net coxph).

In Scenario 1, we consider 500 observations with 200 predictors, the sselastic-net coxph model demonstrates superior performance, achieving the highest C-Index of 0.721, indicating strong predictive accuracy. Its deviance 2259.279 offers a better model fit compared to

the other methods, while the MAE of 0.7826 reflects a reliable predictive accuracy. However, the Lasso coxph model exhibits faster computation, providing a lower C-Index (0.696) and a higher deviance (2301.044), indicating reduced accuracy and poor fit. The t-prior bcoxph model, with a C-Index of 0.605 and a deviance of 3010.550, performs notably worse, as reflected by its high MAE (1.5519), suggesting overfitting and weak generalization.

In Scenario 2, we maintained the number of observations and predictors, where the sselastic-net coxph model continued to perform consistently well, with a C-Index of 0.721 and a deviance of 2269.679, maintaining a strong balance between predictive accuracy and model fit. Its system time (0.045) and MAE (0.7769) also demonstrated efficiency and stability across replications. In contrast, the Lasso coxph model demonstrated a slight decline in predictive strength (C-Index = 0.686) and higher deviance (2306.598), while the t-prior bcoxph again recorded poor predictive and fitting metrics.

Table 2. Evaluation of Proposed and Existing Methods Across Six Simulated Scenarios Using 10-Fold Cross-Validation.

Simulation	Method	10-Fold CV Time (s)	C-Index	Deviance	MAE
Scenario 1 ($n = 500$, $J = 200$)	d-e bcoxph	0.745	0.569	2484.278	0.650
	t-prior bcoxph	0.979	0.605	3010.550	1.552
	Lasso coxph	0.014	0.696	2301.044	**0.478**
	sslasso-coxph	0.035	0.707	2281.189	0.652
	sselastic-net coxph	0.040	**0.721**	**2259.279**	0.783
Scenario 2 ($n = 500$, $J = 200$)	d-e bcoxph	0.909	0.602	2855.699	1.407
	t-prior bcoxph	1.059	0.588	3164.973	1.551
	Lasso coxph	0.014	0.686	2306.598	**0.478**
	sslasso-coxph	0.042	0.698	2292.370	0.656
	sselastic-net coxph	0.045	**0.721**	**2269.679**	0.777
Scenario 3 ($n = 500$, $J = 200$)	d-e bcoxph	0.931	0.610	2949.448	1.411
	t-prior bcoxph	1.019	0.593	2986.948	1.555
	Lasso coxph	0.016	0.700	2292.258	**0.510**
	sslasso-coxph	0.047	0.703	2287.087	0.659
	sselastic-net coxph	0.043	**0.719**	**2277.174**	0.754
Scenario 4 ($n = 500$, $J = 1000$)	d-e bcoxph	4.241	0.592	2845.307	1.299
	t-prior bcoxph	5.280	0.577	5899.464	1.429
	Lasso coxph	0.033	0.697	2278.758	**0.510**
	sslasso-coxph	0.080	0.714	2248.719	0.659
	sselastic-net coxph	1.138	**0.891**	**2143.213**	0.754
Scenario 5 ($n = 500$, $J = 1000$)	d-e bcoxph	4.348	0.601	5570.253	1.299
	t-prior bcoxph	5.298	0.588	6084.526	1.429
	Lasso coxph	0.030	0.697	2281.190	**0.510**
	sslasso-coxph	0.081	0.713	**2246.745**	0.659
	sselastic-net coxph	1.089	**0.876**	2283.666	0.754
Scenario 6 ($n = 500$, $J = 1000$)	d-e bcoxph	4.337	0.598	5340.907	1.299
	t-prior bcoxph	5.239	0.580	6385.465	1.429
	Lasso coxph	0.031	0.689	2284.913	**0.510**
	sslasso-coxph	0.086	0.712	2249.343	0.659
	sselastic-net coxph	1.106	**0.890**	**1958.092**	0.754

For the sslasso-coxph model, the spike scale S_0 varied across scenarios as follows: $0.045, 0.05, 0.02$ for Scenarios 1–3 and $0.029, 0.034, 0.042$ for Scenarios 4–6, while the slab scale $S_1 = 0.5$ was fixed for all scenarios. For the sselastic-net coxph model, $s_{1,0}$ varied across scenarios, as reported above: $s_{1,1} = 0.5$, $s_{2,0} = 0.01$, and $s_{2,1} = 1.0$ for all scenarios.

In Scenario 3, we obtained the same dimension of 500 observations and 200 predictors, and the sselastic-net coxph model outperformed all methods, with a C-Index of 0.719,

deviance of 2277.174, and MAE of 0.7543. The consistent performance across multiple scenarios highlights the robustness and adaptability of the model. The Lasso coxph model followed but remained less accurate, with a C-Index of 0.700 and a higher deviance of 2292.258, reinforcing its limitations in predictive power.

The number of predictors was increased to 1000 in scenario 4. The sselastic-net coxph model sustained its dominance, achieving a higher C-Index of 0.891, a lower deviance of 2143.213, and an MAE of 0.7543. Despite the high dimensionality, its system time of 1.138 indicated strong computational efficiency. The Lasso coxph model, while faster, showed inferior results (C-Index = 0.697; deviance = 2278.758), suggesting reduced adaptability to complex, high-dimensional data. The d-e bcoxph and t-prior bcoxph models exhibited both high deviance and longer computation times, underscoring their limited suitability in such settings. In Scenario 5, 500 observations and 1000 predictors were used, and the sselastic-net coxph model maintained a robust performance, achieving a C-Index of 0.876, a deviance of 2283.666, and an MAE of 0.7543. Its system time (1.089) further illustrated efficiency relative to model complexity. Although the Lasso coxph model ran faster, it delivered lower accuracy (C-Index = 0.697) and higher deviance (2281.190). The t-prior and d-e bcoxph models continued to underperform, marked by elevated deviance and MAE values.

Finally, in Scenario 6, which retained the same sample size and predictor count, the sselastic-net coxph model achieved its best performance, with the highest C-Index of 0.890, a lower deviance of (1958.092), and a stable MAE of 0.7543. This confirms its robustness and predictive reliability even in high-dimensional environments. Although the Lasso coxph model computed faster, it again showed weaker predictive ability (C-Index = 0.689; deviance = 2284.913), while the remaining methods continued to exhibit sub-optimal fit and higher prediction errors.

Table 3 shows the results of six simulation scenarios using the spike-and-slab elastic net Cox model for survival prediction. The scenarios differ in the number of observations and predictors. Evaluation metrics include system time $(Time\alpha)$, C-Index, deviance, and mean absolute error (MAE).

Table 3. Outcomes of Six Simulation Scenarios Using the Spike-and-Slab Elastic Net Cox Model for Survival Prediction.

Simulation	Method	$Time^{\alpha}$ (s)	C-Index	Deviance	MAE
Scenario 1 ($n = 500, m = 200$)	sselastic-net coxph	0.745	0.721	2259.279	0.478
Scenario 2 ($n = 500, m = 200$)	sselastic-net coxph	0.909	0.721	2269.679	0.478
Scenario 3 ($n = 500, m = 200$)	sselastic-net coxph	0.043	0.719	2277.174	0.510
Scenario 4 ($n = 500, m = 1000$)	sselastic-net coxph	1.138	0.891	2143.213	0.510
Scenario 5 ($n = 500, m = 1000$)	sselastic-net coxph	1.089	0.876	2246.745	0.510
Scenario 6 ($n = 500, m = 1000$)	sselastic-net coxph	1.106	0.891	1958.092	0.510

In Scenario 1, 500 observations and 200 predictors were used, and the model achieved a C-Index of 0.721, a deviance of 2259.279, and an MAE of 0.478470. These results offer a good balance between accuracy and model fit, with a short computation time of 0.745, showing high efficiency for moderate datasets. In Scenario 2, we utilized the same observations and predictors in scenario 1. Thus, the outcomes were almost the same, with a C-Index of 0.721 and an MAE of 0.478470, with a slight increase in deviance to 2269.679. This consistency highlights the robustness and reliability of the model in repeated simulations. In Scenario 3, we used 500 observations and 200 predictors. The C-Index revealed a slight drop to 0.719,

while the deviance scaled to 2277.174 and the MAE increased to 0.5100032, thus showing a small decline in accuracy, but the model continued to perform well and provide reliable predictions. In Scenario 4, we scaled the number of predictors to 1000, which signified an improvement in performance. The C-Index rose to 0.891 and showed high discriminative ability. The deviance reduced to its lowest value of 2143.213, indicating a strong fit to the model. MAE remained at 0.5100032, revealing that the model maintained stable accuracy even with more predictors. In Scenario 5, we used 500 observations and 1000 predictors, where the C-Index decreased slightly to 0.876, while the deviance increased to 2246.745. MAE remained at 0.5100032, showing that the model maintained strong accuracy despite the added complexity. In Scenario 6, the C-Index of 0.891 was the same as in Scenario 4. The deviance was reduced to 1958.092, indicating the best data fit. The MAE remained stable at 0.5100032, showing that the model consistently improved accurate predictions even under the most complex scenario.

The estimated regression coefficients from the Bayesian hierarchical elastic-net Cox proportional hazards model (sselastic-net-coxph) for Scenarios 1 and 4 are presented visually and described in Figure 1. Figure 1 presents the simulation results for Scenario 1, comparing Bayesian Cox proportional hazards models fitted with hierarchical elastic net and double-exponential (Laplace) priors. The figure highlights differences in coefficient estimation, shrinkage behavior, and posterior uncertainty under the two prior specifications.

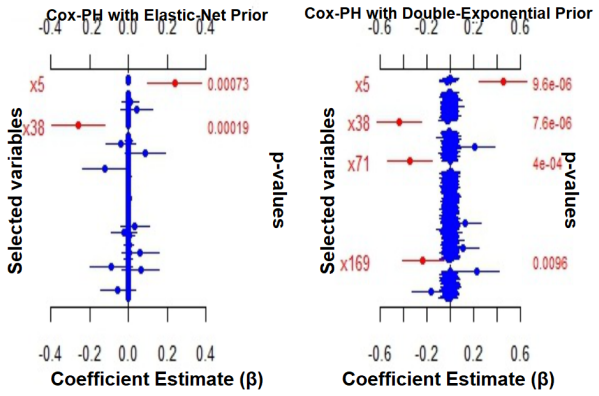


Figure 1. Estimated regression coefficients from the Bayesian Cox proportional hazards model under hierarchical elastic net and double-exponential (Laplace) priors for Scenario 1. Red points indicate posterior mean estimates, while blue lines represent 95% credible intervals.

In this scenario, the simulated dataset consisted of $n = 500$ observations and $m = 200$ predictors. The results illustrate how prior choice affects coefficient shrinkage and variable selection in the Bayesian Cox model. Among the 200 predictors, only a small subset was identified as relevant, with two predictors selected under the hierarchical elastic net prior and four under the Laplace priors. The hierarchical elastic net prior induces stronger shrinkage, pulling many coefficient estimates toward zero and promoting sparsity. In contrast, the Laplace prior exhibits heavier tails, allowing greater flexibility by retaining larger coefficient estimates for predictors associated with stronger signals. These differences are reflected in the relative dispersion of the posterior means and credible intervals. The spike-and-slab hyperparameters, $s_\lambda = 0.045$ and $s_1 = 0.5$, control the degree of sparsity by regulating the probability of variable inclusion and the strength of penalization. Overall, both priors effectively recover the underlying signal while suppressing noise. However, the hierarchical elastic net and Laplace priors favor more aggressive shrinkage. In contrast,

other priors permit greater variability in the estimated coefficients, illustrating the trade-off between sparsity and flexibility in Bayesian survival modeling.

Figure 2 illustrates the impact of hierarchical elastic net and Laplace priors on variable selection and coefficient shrinkage in a high-dimensional Bayesian Cox proportional hazards (CoxPH) model. The dataset for Scenario 4 consists of $n = 500$ observations and $p = 1000$ predictors ($p \gg n$). This scenario demonstrates the effect of prior choice on coefficient shrinkage and variable selection. Under the hierarchical elastic net prior, six predictors were selected, whereas the Laplace prior retained nine predictors. The selected predictors and their corresponding p-values are shown in the figure. Red points and lines indicate the posterior mean estimates of the regression coefficients, while blue lines represent 95% credible intervals, illustrating estimation uncertainty. The hierarchical elastic net prior enforces strong shrinkage, pulling most coefficients toward zero and retaining only the most relevant variables. In contrast, the Laplace prior almost selects the same variables with the hierarchical elastic net, but allows additional predictors with moderate effects to be included. The spike-and-slab hyperparameters, $s_\lambda = 0.034$ and $s_1 = 0.5$, regulate the level of sparsity and the probability of variable inclusion, balancing the retention of relevant predictors against the exclusion of noise. Overall, both priors successfully identify important predictors, with the hierarchical elastic net prior yielding a more compact model and the Laplace prior capturing additional potential effects. These results highlight how prior selection influences sparsity, coefficient flexibility, and model stability in Bayesian survival analysis.

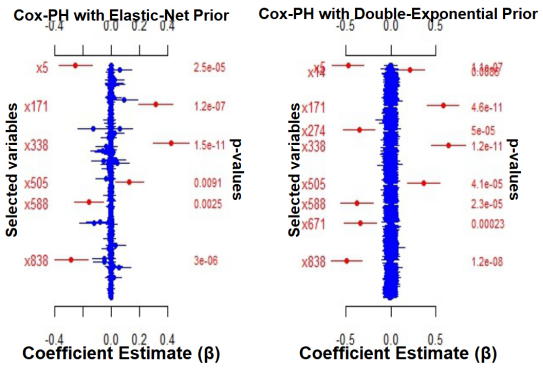


Figure 2. Estimated regression coefficients from the Bayesian Cox proportional hazards model under double-exponential and Student-t priors for Scenario 4. Red points indicate posterior mean estimates and blue lines represent 95% credible intervals.

6.2. Real-Life Studies

This section focused on the practical application of the proposed Spike-and-Slab for the Elastic-Net COX-PH, with the existing models using three (3) different microarray gene expression data sets. The three data sets include GSE14333 data [39] (Data 1), Dutch breast cancer data [40] (Data 2), and Diffuse large-B-cell lymphoma data [41] (Data 3). A detailed description of the datasets is presented in Table 4.

Table 4. Characteristics of the Used Clinical and High-Dimensional Datasets.

Name	Observations	Features	Censoring Rate
GSE14333 [39]	226	16,383	25
Dutch Breast Cancer Data (DBCD) [40]	295	4919	73
Diffuse Large-B-Cell Lymphoma data (DLBCL) [41]	240	7402	42
Diffuse Large-B-Cell Lymphoma data (DLBCL) [41]	240	7402	42

Table 4 summarizes the main characteristics of the three datasets, GSE14333, DBCD, and DLBCL, used in this analysis. All three have a high-dimensional structure in which the number of predictors greatly exceeds the number of observations ($p \gg n$). This difference creates serious challenges for classical survival models, including risks of multicollinearity and overfitting. The GSE14333 dataset contains 226 samples and 16,383 gene expression features, representing a typical high-dimensional colorectal cancer cohort. The DBCD dataset includes 295 observations with 4919 predictors, offering a more complex setting due to its larger sample size and substantial feature space. The DLBCL dataset consists of 240 samples and 7402 features, providing an intermediate scenario in both dimensionality and sample characteristics. Together, these datasets cover a wide range of feature dimensions and sample sizes, offering a strong basis for assessing the scalability, robustness, and predictive performance of Bayesian hierarchical Cox models. The distinct $p > n$ structure in all datasets highlight the importance of using effective shrinkage priors such as Gaussian, Laplace, and Spike-and-Slab variants to achieve reliable variable selection and improve generalization in high-dimensional genomic survival analysis.

6.3. Application to Real Life Data Sets

In this section, we introduce an alternative survival model using Bayesian regularization through a hierarchical modeling approach. This approach aims to identify and select relevant predictors by applying penalty functions, such as Lasso and elastic net, which induce shrinkage in the penalized partial log-likelihood of the Cox model for survival prediction. Specifically, hierarchical modeling with the spike-and-slab double-exponential prior, along with Lasso, spike-and-slab Lasso, and elastic net Cox models, were utilized to analyze real-world data.

Table 5 presents the number of selected genes in GSE14333, with a summarized outcome for the three methods applied in this study, providing the corresponding features selected and the proportion of features chosen by each method. Based on the outcomes, Lasso Coxph selected the fewest features. For the model fit, as shown by AIC values, the ssElastic model performs best with the AIC (1024.265), followed by ssLasso with AIC (1372.611), while the Lasso model has the highest AIC (1499.291), suggesting a poor fit. For accuracy, measured by MAE, the ssLasso model presents the lowest MAE (0.0016), showing superior predictive performance, while the ssElastic (0.0038) and Lasso (0.0039) models show a slightly lower performance. Generally, the ssElastic model achieves the best balance between fit and predictive accuracy.

Table 5. Number of Selected Genes from GSE14333 Survival Data.

Method	Genes Selected	Proportion of Genes Selected	AIC	MAE
Lasso CoxPH	10	0.00095	1499.291	0.0039
ssLasso CoxPH	108	0.00642	1372.611	0.0016
ssElastic-Net CoxPH	94	0.00559	1024.265	0.0038

Table 6 shows the performance of three survival analysis methods on the Dutch Breast Cancer Data (DBCD) dataset, evaluated using metrics such as the number of selected genes, the proportion of selected genes, the AIC, and the MAE. The Lasso Cox model selected 16 genes, showcasing a more inclusive selection, while the spike-and-slab Lasso Cox model (ssLasso) picked 8 genes, selecting the most relevant predictors. The spike-and-slab elastic net Cox model (ssElastic) selected only 1 gene, indicating an extremely selective approach. The proportions of selected genes are 0.0033 for Lasso, 0.0016 for ssLasso, and 0.0002 for

ssElastic. Concerning the model fit, ssLasso provide the best AIC (730.5997), while Lasso and ssElastic had higher AIC values of 802.7552 and 806.567, respectively. In terms of predictive accuracy, ssElastic performed best with a lower MAE (0.0013), compared to 0.0026 for Lasso and 0.0027 for ssLasso. Generally, ssElastic provided the best predictive accuracy, while ssLasso offered the best balance between model fit and feature selection.

Table 6. Number of selected genes from DBCD high-dimensional survival data.

Method	Genes Selected	Proportion of Genes Selected	AIC	MAE
Lasso Coxph	16	0.0033	802.7552	0.0026
ssLasso Coxph	8	0.0016	730.5997	0.0027
ssElastic Coxph	1	0.0002	806.567	0.0013

Table 7 presents the results of survival analysis methods applied to the Diffuse Large B-Cell Lymphoma (DLBCL) dataset. The Lasso Cox model identified 29 genes, indicating an extensive selection process. In contrast, the ssLasso model selected 12 genes, while the ssElastic model selected 11 genes, showing a more precise feature selection approach. The proportion of selected genes for the Lasso model was 0.0039, providing moderate selection, while ssLasso and ssElastic models had proportions of 0.0016 and 0.0014, respectively, indicating a more focused selection. In terms of model performance, the Lasso Cox model had an AIC of 616.7418, suggesting space for improvement. The ssLasso model achieved a best fit with an AIC of 565.6131, while the ssElastic model had an AIC of 589.9531. Relating to prediction accuracy, the Lasso model had an MAE of 0.0017, the ssLasso model had a slightly lower MAE of 0.0015, and the ssElastic model had an MAE of 0.0016, revealing that all models performed well, with the ssLasso model showing the best predictive performance.

Table 7. Number of selected genes from DLBCL high-dimensional survival data.

Method	Genes Selected	Proportion of Genes Selected	AIC	MAE
Lasso Coxph	29	0.0039	616.7418	0.0017
ssLasso Coxph	12	0.0016	565.6131	0.0015
ssElastic Coxph	11	0.0014	589.9531	0.0016

Table 8 presents the performance of the Lasso-CoxPH, Spike-and-Slab Lasso (ssLasso), and Spike-and-Slab Elastic Net (ssElastic) models across three datasets: GSE14333, DBCD, and DLBCL. The table reveals the number of features each model selected with their ranking proportion, indicating the effectiveness of feature identification across datasets. The Lasso-CoxPH model selected a small number of features (10, 16, and 29), demontrating aggressive shrinkage. While this provides a sparse and simple model, it may miss some moderate but relevant effects, excluding the potentially predictive genes. The ssLasso CoxPH model identified more features (108 in GSE14333, 81 in DBCD, and 12 in DLBCL) and achieved the highest ranking proportion (1.0). This reveals its ability to adaptively shrink coefficients, effectively separating between noise and signals. Its flexibility allows it to retain informative predictors that the regular Lasso might exclude, enhancing predictive performance in high-dimensional settings. The ssElastic CoxPH model provided a balance between sparsity and stability, selecting a few numbers of features (94, 1, and 11), with a ranking proportion of 0.7. Its performance provides the advantage of incorporating ridge-type regularization to capture correlated predictors, enhancing interpretability while maintaining computational efficiency. Generally, Table 8 reveals that while Lasso-CoxPH

induces stronger sparsity, the Spike-and-Slab models, particularly ssLasso, offer more adaptive and stable feature selection, making them more reliable for high-dimensional genomic survival analysis with complex correlations and mild genetic effects.

Table 8. Summarized proportion and ranking on real life data.

Method	GSE14333	DBCD	DLBCL	Poportion of Rank
Lasso Coxph	10	16	29	0.3
ssLasso Coxph	108	81	12	1.0
ssElastic Coxph	94	1	11	0.7

Figure 3 presents a Cox coefficient plot for genes selected from the GSE14333 colorectal cancer dataset using the Lasso-CoxPH, Spike-and-Slab Lasso, and Spike-and-Slab Elastic Net models, indicating differences in variable selection and clinical interpretability.

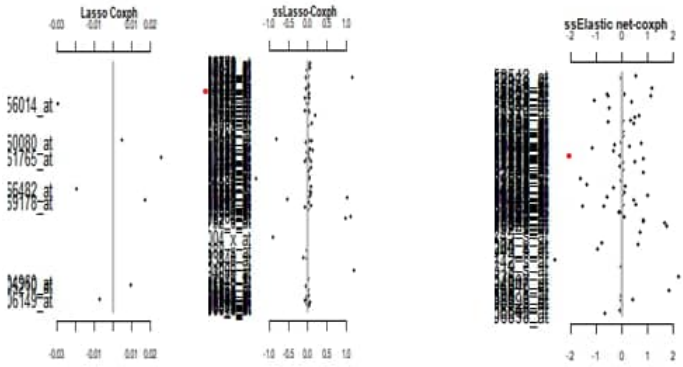


Figure 3. Cox coefficient estimates plot for genes selected in GSE14333. The red dot in each panel indicates a highly significant gene.

Figure 3 presents the Cox coefficient plots for genes selected from the GSE14333 colorectal cancer dataset, which contains 16,830 gene expression features, using three regularized models: Lasso-CoxPH, Spike-and-Slab Lasso, and Spike-and-Slab Elastic Net. The Lasso-CoxPH model provided a highly sparse representation, identifying only a few dominant genes by inducing strong global shrinkage across all coefficients. While effective in excluding noise, this approach tends to over-shrink moderately informative genes, resulting in potential loss of biological signals relevant to patient survival. The Spike-and-Slab Lasso model mitigates this constraint through adaptive shrinkage, capturing a large but more relevant number of genes. The selected genes that indicate both positive and negative coefficients represent risk-associated and protective genetic effects. This flexibility allowed the model to better differentiate between genes with negligible and substantial contributions to the hazard function. The Spike-and-Slab Elastic Net further improved selection stability by integrating both Lasso and ridge-type penalties. It identifies a slightly larger but more consistent number of predictive genes, effectively handling high intergene correlations typical of microarray data. The retain genes in this model demonstrated clearer biological interpretability, reflecating coherent risk and protective gene cluster relevant to colorectal cancer prognosis. Essentially, Figure 3 highlights that while all models achieve variable selection, the Spike-and-Slab Elastic Net model offers the most balanced performance, preserving sparsity while retaining biologically meaningful and stable predictors, making it particularly suitable for high-dimensional genomic survival modeling.

Figure 4 illustrates the Cox coefficient estimates for genes selected from the DBCD dataset using the Lasso-CoxPH, Spike-and-Slab Lasso, and Spike-and-Slab Elastic Net

models, revealing how each method balances sparsity and stability and identifies relevant predictors.

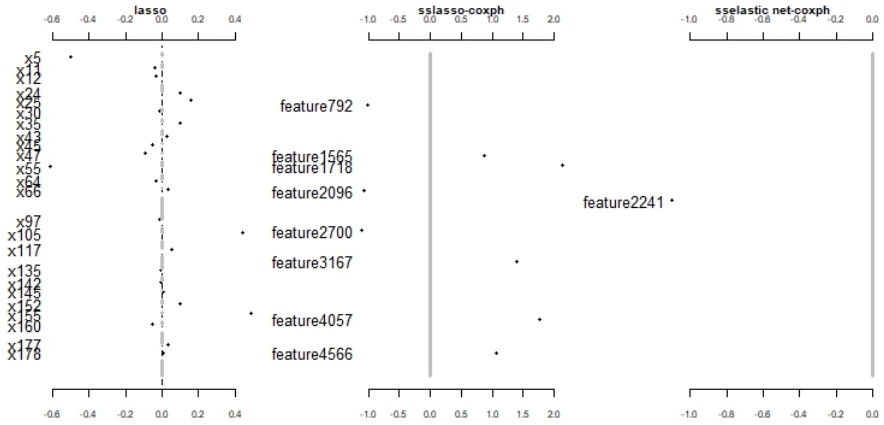


Figure 4. Cox coefficient estimates plot for the genes selected in DBCD Data.

Figure 4 revealed the estimated Cox coefficients of genes selected from the DBCD dataset, containing 4919 gene expression features, through the Lasso-CoxPH, Spike-and-Slab Lasso (ssLasso), and Spike-and-Slab Elastic Net (ssElastic) models. The Lasso-CoxPH model identified a few genes, which indicates its strong shrinkage effect, which induces high sparsity. Although this helps exclude noise predictors, it can also underestimate moderately important genes and overlook relevant biological signals linked to patient survival. The ssLasso model relaxed this constraint through adaptive shrinkage, offering a more balanced selection of genes. Its coefficients include both positive and negative values genes with positive log-hazard coefficients indicating higher risk and those with negative values indicating protective effects. This adaptive behavior allows the model to better distinguish between irrelevant and significant predictors. The ssElastic model further optimized the selection process by adding ridge-type regularization, causing it to capture correlated or jointly informative genes that may influence survival together. The selected genes in this model formed clear clusters of strong positive and negative effects, demonstrating its robustness in identifying biologically significant gene groups while maintaining model stability. Essentially, Figure 4 reveals that while all three models perform variable selection, the ssElastic model provides the best balance between sparsity, stability, and interpretability. It effectively retains influential prognostic genes while filtering out irrelevant ones, making it the most reliable method for large-scale genomics in the DBCD dataset.

Figure 5 displays the Cox coefficient estimates for genes selected from the DLBCL dataset using the Lasso-CoxPH, Spike-and-Slab Lasso, and Spike-and-Slab Elastic Net models, highlighting how each method balances sparsity and stability.

Figure 5 presents the Cox coefficient plot for the features selected from the 7399-gene DLBCL dataset using three penalized models: Lasso-CoxPH, Spike-and-Slab Lasso, and Spike-and-Slab Elastic Net. The figure illustrates both the direction and magnitude of the effects associated with the selected genes. The Lasso-CoxPH model selected a small number of genes through uniform shrinkage, effectively excluding weak predictors but often over-shrinking moderately strong effects. In contrast, the Spike-and-Slab Lasso model achieved more adaptive shrinkage, balancing sparsity and flexibility. It retained a moderate number of influential genes, separating irrelevant predictors (shrunk near zero) from truly significant ones with nonzero posterior means. These selected genes indicate

both positive and negative log-hazard coefficients, indicating features associated with increased and decreased risk, respectively. The Spike-and-Slab Elastic Net model further enhanced feature recovery by combining the Lasso sparsity with ridge-type regularization. This approach identified a slightly larger yet more stable set of genes, effectively handling correlated predictors, a common characteristic of gene expression data. The selected genes in this model showed distinct clusters of high positive coefficients (strong risk factors) and negative coefficients (protective factors), suggesting biologically meaningful differentiation in DLBCL prognosis. Generally, Figure 5 demonstrates that while all models achieve variable selection, the Spike-and-Slab Elastic Net provides the most balanced trade-off between sparsity, stability, and predictive interpretability, making it more reliable for high-dimensional genomic survival analysis.

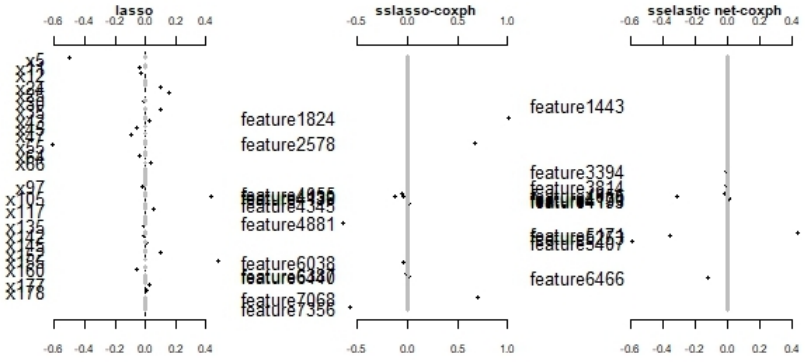


Figure 5. Cox coefficient estimates plot for the selected genes in DLBCL Data.

7. Conclusions

This study proposes a survival prediction approach designed to more effectively address high-dimensional data. The model integrates Bayesian variable selection with spike-and-slab priors and an EM-based estimation procedure, improving predictive accuracy while enabling clear identification of relevant predictors. Simulation results and empirical analyses indicate that the proposed model demonstrates favorable predictive performance and model stability compared with existing penalized Cox approaches, particularly in the high-dimensional scenarios evaluated in this study. Although motivated by high-dimensional applications, the method can also be applied to smaller datasets without compromising performance.

The primary methodological contribution is a spike-and-slab elastic-net Cox model that employs a spike-and-slab hierarchical elastic net prior. This prior permits varying levels of shrinkage across regression coefficients, thereby limiting the influence of irrelevant predictors and stabilizing coefficient estimates. For model estimation, the EM algorithm is adapted to the penalized partial likelihood of the Cox model, with joint posterior expectations computed efficiently through cyclic coordinate descent within each EM iteration. During the E-step, posterior expectations of the indicator variables (γ_i) are obtained for each coefficient, while, in the M-step, the regression coefficients are updated using standard coordinate descent procedures. This framework is particularly relevant for biomedical and clinical applications, where accurate risk stratification and reliable variable selection are critical for understanding disease progression and supporting evidence-based decision-making.

Overall, the proposed Bayesian hierarchical modeling framework for the Elastic Net Cox proportional hazards model (sselastic-net-coxph) provides a flexible, interpretable, and

computationally tractable approach for modeling complex survival data, while maintaining adaptability for extension to other structured prior formulations and inference methods.

7.1. Strengths, Limitations, and Future Perspectives

7.1.1. Strengths

The development of our proposed Spike-and-Slab Elastic Net Cox–PH model demonstrates several advanced methodological and practical strengths. The proposed model integrates Bayesian hierarchical spike-and-slab modeling with elastic net regularization, enabling simultaneous variable selection and coefficient stabilization. This is particularly advantageous in high-dimensional survival settings with correlated predictors, where standard LASSO may select only one variable from a correlated group, while ridge regression does not perform variable selection extensively.

The use of the EM-coordinate descent algorithm in this study provides a computationally efficient alternative to full Bayesian MCMC estimation. Estimating the posterior inclusion probabilities through the EM framework and updating regression coefficients via adaptive penalized coordinate descent, our proposed methodology achieves scalability while retaining probabilistic interpretation through the inclusion probabilities p_j. Moreover, varying the expected inverse scales $E(s_{1j}^{-1})$ and $E(s_{2j}^{-1})$ across predictors enhances model flexibility and improves stability in the presence of heterogeneous effect sizes. The use of cross-validated partial likelihood provides a comprehensive evaluation of both discrimination and calibration performance.

7.1.2. Limitations

Although the proposed Spike-and-Slab Elastic Net Cox–PH model demonstrated several improvements over existing methods, certain limitations should be acknowledged. The present methodology has the following limitations, which can be addressed in future work: (i) The EM algorithm was used to compute the estimates; faster computational methods could also be considered. (ii) The model evaluation in the present work primarily relied on the concordance index (C-index), deviance, and mean absolute error (MAE). Prediction accuracy measures such as the Integrated Brier Score and time-dependent AUC would also be appropriate. (iii) The cross-validation strategy, while effective for tuning scale parameters, may be sensitive to sample size and fold allocation, particularly in small datasets. This could affect the reproducibility of different random splits.

7.1.3. Future Perspectives

This study can be extended in several directions for future research, particularly for biomedical and clinical applications. First, integrating uncertainty quantification through approximate Bayesian inference methods, such as variational Bayes or Laplace approximation, would enhance interpretability while preserving computational efficiency. Second, incorporating structured priors, such as group spike-and-slab penalties or network-based regularization, could further improve performance when predictors exhibit known biological or functional relationships. Finally, future research may focus on extending the framework to accommodate more complex survival settings, including competing risks, structured priors, and scalable Bayesian inference techniques.

Author Contributions: Conceptualization, K.A.D.; Methodology, B.I.D. and K.A.D.; Software, K.A.D.; Validation, R.K.L.; Formal analysis, B.I.D.; Investigation, B.I.D. and K.A.D.; Resources, K.A.D.; Data curation, B.I.D. and R.K.L.; Writing—original draft, B.I.D. and K.A.D.; Writing—review & editing, R.K.L.; Visualization, R.K.L.; Supervision, K.A.D. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: The original contributions presented in this study are included in the article. Further inquiries can be directed to the corresponding author.

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

The following abbreviations are used in this manuscript:

Correlation-Based-IBMA	Correlation-Based approach and Iterative Bayesian Model Averaging (IBMA)
Cox-PH	Cox proportional hazard
EM	Expectation–Maximization
CCD	Cyclic Coordinate Descent
EM-CCDA	Expectation–Maximization Cyclic Coordinate Descent Algorithm
pl	Partial Log-likelihood
ppl	Penalized Partial Log-likelihood
CVPL	Cross-validated Partial Likelihood
C-Index	Concordance Index
MAE	Mean Absolute Error
DBCD	Dutch Breast Cancer Data
DLBCL	Diffuse Large-B-Cell Lymphoma Data
AIC	Akaike Information Criterion

References

1. Maturo, F.; Rambaud, S.C.; Ventre, V. Advances in statistical learning from high-dimensional data. *Qual. Quant.* **2025**, *59*, 1933–1937. [CrossRef]
2. Dinov, I.D. Methodological challenges and analytic opportunities for modeling and interpreting Big Healthcare Data. *Gigascience* **2016**, *5*, 12. [CrossRef] [PubMed]
3. Dauda, K.A.; Olorede, K.O.; Banjoko, A.W.; Yahya, W.B.; Ayipo, Y.O. Genetic Diagnosis, Classification, and Risk Prediction in Cancer Using Next-Generation Sequencing in Oncology. In *Computational Approaches in Biomaterials and Biomedical Engineering Applications*; CRC Press: Boca Raton, FL, USA, 2024; pp. 107–122.
4. Cox, D.R. Regression models and life-tables. *J. R. Stat. Soc. Ser. B Methodol.* **1972**, *34*, 187–202. [CrossRef]
5. Zhao, L.; Dong, Q.; Luo, C.; Wu, Y.; Bu, D.; Qi, X.; Luo, Y.; Zhao, Y. DeepOmix: A scalable and interpretable multi-omics deep learning framework and application in cancer survival analysis. *Comput. Struct. Biotechnol. J.* **2021**, *19*, 2719–2725. [CrossRef]
6. Piironen, J.; Vehtari, A. Projective inference in high-dimensional problems: Prediction and feature selection. *Bayesian Anal.* **2020**, *15*, 709–733. [CrossRef]
7. Dauda, K.A. Optimal tuning of Random Survival Forest hyperparameter with an application to liver disease. *Malays. J. Med. Sci.* **2022**, *29*, 67–76. [CrossRef]
8. Haredasht, F.N.; Dauda, K.A.; Vens, C. Exploiting Censored Information in Self-Training for Time-to-Event Prediction. *IEEE Access* **2023**, *11*, 96831–96840. [CrossRef]
9. Hastie, T.; Tibshirani, R.; Friedman, J. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, 2nd ed.; Springer: Berlin/Heidelberg, Germany, 2009.
10. Hastie, T.; Tibshirani, R.; Wainwright, M. *Statistical Learning with Sparsity: The Lasso and Generalizations*; Chapman and Hall/CRC: Boca Raton, FL, USA, 2015.

11. Tibshirani, R. Regression shrinkage and selection via the lasso. *J. R. Stat. Soc. Ser. B Methodol.* **1996**, *58*, 267–288. [CrossRef]
12. Boluki, S.; Esfahani, M.S.; Qian, X.; Dougherty, E.R. Incorporating biological prior knowledge for Bayesian learning via maximal knowledge-driven information priors. *BMC Bioinform.* **2017**, *18*, 552. [CrossRef]
13. Mallick, H.; Yi, N. Bayesian methods for high-dimensional linear models. *J. Biom. Biostat.* **2013**, *1*, 005. Available online: https://api.semanticscholar.org/CorpusID:23361307 (accessed on 18 October 2025).
14. Samorodnitsky, S.; Hoadley, K.A.; Lock, E.F. A hierarchical spike-and-slab model for pan-cancer survival using pan-omic data. *BMC Bioinform.* **2021**, *23*, 235. [CrossRef]
15. Dauda, K.A.; Olorede, K.O.; Aderoju, S.A. A novel hybrid dimension reduction technique for efficient selection of bio-marker genes and prediction of heart failure status of patients. *Sci. Afr.* **2021**, *12*, e00778. [CrossRef]
16. Lázaro, E.; Armero, C.; Alvares, D. Bayesian regularization for flexible baseline hazard functions in Cox survival models. *Biom. J.* **2020**, *63*, 26–37. [CrossRef] [PubMed]
17. Bonato, V.; Baladandayuthapani, V.; Broom, B.M.; Sulman, E.P.; Aldape, K.D.; Do, K. Bayesian ensemble methods for survival data with genomic covariates. *Bioinformatics* **2011**, *27*, 359–367. [CrossRef] [PubMed]
18. Ma, Z.; Teschendorff, A.E. A variational Bayes beta mixture model for feature selection in DNA methylation studies. *J. Bioinform. Comput. Biol.* **2013**, *11*, 1350005. [CrossRef] [PubMed]
19. Dauda, K.A.; Adeniyi, E.J.; Lamidi, R.K.; Wahab, O.T. Exploring Flexible Penalization of Bayesian Survival Analysis Using Beta Process Prior for Baseline Hazard. *Computation* **2025**, *13*, 21. [CrossRef]
20. Dauda, K.A.; Lamidi, R.K. Consistency and Stability in Feature Selection for High-Dimensional Microarray Survival Data in Diffuse Large B-Cell Lymphoma Cancer. *Data* **2025**, *10*, 26. [CrossRef]
21. George, E.I.; McCulloch, R.E. Variable selection via Gibbs sampling. *J. Am. Stat. Assoc.* **1993**, *88*, 881–889. [CrossRef]
22. George, E.I.; McCulloch, R.E. Approaches for Bayesian variable selection. *Stat. Sin.* **1997**, *7*, 339–373.
23. Rocková, V.; George, E.I. EMVS: The EM approach to Bayesian variable selection. *J. Am. Stat. Assoc.* **2014**, *109*, 828–846. [CrossRef]
24. Rocková, V.; George, E.I. The spike-and-slab lasso. *J. Am. Stat. Assoc.* **2016**, *111*, 431–444. [CrossRef]
25. Chipman, H.A. Bayesian variable selection with related predictors. *Can. J. Stat.* **1996**, *24*, 17–36. [CrossRef]
26. Ishwaran, H.; Rao, J.S. Spike and slab variable selection: Frequentist and Bayesian strategies. *Ann. Stat.* **2005**, *33*, 730–773. [CrossRef]
27. Nikooienejad, A.; Wang, W.; Johnson, V.E. Bayesian variable selection for binary outcomes in high-dimensional genomic studies using non-local priors. *Bioinformatics* **2016**, *32*, 1338–1345. [CrossRef]
28. Tang, Z.; Shen, Y.; Zhang, X.; Yi, N. The spike-and-slab lasso Cox model for survival prediction and associated genes detection. *Bioinformatics* **2017**, *33*, 2799–2807. [CrossRef]
29. Hernández-Lobato, D.; Hernández-Lobato, J.M.; Dupont, P. Generalized Spike-and-Slab Priors for Bayesian Group Feature Selection Using Expectation Propagation. *J. Mach. Learn. Res.* **2013**, *14*, 1891–1945. Available online: http://jmlr.org/papers/v14/hernandez-lobato13a.html (accessed on 15 July 2025).
30. Tang, Z.; Shen, Y.; Li, Y.; Zhang, X.; Wen, J.; Qian, C.; Zhuang, W.; Shi, X.; Yi, N. Group spike-and-slab lasso generalized linear models for disease prediction and associated genes detection by incorporating pathway information. *Bioinformatics* **2017**, *34*, 901–910. [CrossRef]
31. Wu, T.T.; Lange, K. Coordinate descent algorithms for lasso penalized regression. *Ann. Appl. Stat.* **2008**, *2*, 224–244. [CrossRef]
32. Friedman, J.; Hastie, T.; Tibshirani, R. Regularization Paths for Generalized Linear Models via Coordinate Descent. *arXiv* **2010**, arXiv:1001.0736. [CrossRef]
33. Tang, Z.; Lei, S.; Zhang, X.; Yi, Z.; Guo, B.; Chen, J.Y.; Shen, Y.; Yi, N. Gsslasso Cox: A Bayesian hierarchical model for predicting survival and detecting associated genes by incorporating pathway information. *BMC Bioinform.* **2019**, *20*, 94. [CrossRef]
34. Hans, C. Elastic net regression modeling with the orthant normal prior. *J. Am. Stat. Assoc.* **2011**, *106*, 1383–1393. [CrossRef]
35. Zou, H.; Hastie, T. Regularization and variable selection via the elastic net. *J. R. Stat. Soc. Ser. B Methodol.* **2005**, *67*, 301–320. [CrossRef]
36. Li, Q.; Lin, N. The Bayesian elastic net. *Bayesian Anal.* **2010**, *5*, 151–170. [CrossRef]
37. Simon, N.; Friedman, J.; Hastie, T.; Tibshirani, R. Regularization paths for Cox's proportional hazards model via coordinate descent. *J. Stat. Softw.* **2011**, *39*, 1–13. [CrossRef]
38. Dai, B.; Breheny, P. Cross-validation approaches for penalized Cox regression. *Stat. Methods Med. Res.* **2024**, *33*, 702–715. [CrossRef]

39. Jorissen, R.N.; Gibbs, P.; Christie, M.; Prakash, S.; Lipton, L.; Desai, J.; Kerr, D.; Aaltonen, L.A.; Arango, D.; Kruhøffer, M.; et al. Metastasis-associated gene expression changes predict poor outcomes in patients with Dukes stage B and C colorectal cancer. *Clin. Cancer Res.* **2009**, *15*, 7642–7651. [CrossRef] [PubMed]
40. van Houwelingen, H.C.; Bruinsma, T.; Hart, A.A.M.; van't Veer, L.J.; Wessels, L.F.A. Cross-validated Cox regression on microarray gene expression data. *Stat. Med.* **2006**, *25*, 3201–3216. Available online: https://api.semanticscholar.org/CorpusID:36336638 (accessed on 11 June 2025). [CrossRef] [PubMed]
41. Lenz, G.; Wright, G.; Dave, S.; Xiao, W.; Powell, J.; Zhao, H.; Xu, W.; Tan, B.; Goldschmidt, N.; Iqbal, J.; et al. Stromal gene signatures in large-b-cell lymphomas. *N. Engl. J. Med.* **2008**, *359*, 2313–2323. [CrossRef] [PubMed]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.