

Particle Swarm Optimization for Penalize Cox Models in Long-Term Prediction of Breast Cancer Data

Ehab Abbas Fadhil^{1,*}, Basad Al-Sarray¹

¹University of Baghdad, College of Science, Computer Science Department, Iraq

*Corresponding Author: Ehab Abbas Fadhil

DOI: <https://doi.org/10.52866/ijcsm.2023.04.04.017>

Receive July 2023; Accepted October 2023; Available online November 2023

ABSTRACT: The particle swarm optimization (PSO) algorithm was applied to penalize the Cox model for predicting long-term outcomes in breast cancer patients. This study utilized data from 198 breast cancer survivors, including their age, estrogen receptor status, tumor size, grade, and expression levels of 76 genes. The aim was to identify a feature subset that could accurately predict patient survival while mitigating overfitting and model complexity. PSO was used to search for optimal model parameters. The algorithm was designed to minimize a penalized partial likelihood function, which balances the trade-off between model accuracy and complexity. The values of the objective function were compared with that of other feature selection techniques, including the least absolute shrinkage and selection operator (LASSO) and elastic regression, and found to outperform them in predictive accuracy and feature selection. Results demonstrated that PSO-PCOX with cross-validation for regularization parameters achieved higher prediction accuracy than models trained with other feature selection methods. The PSO algorithm identified a subset of features that were consistently selected across multiple iterations, indicating their importance in predicting patient survival. Overall, this study showcases the potential of PSO-based feature selection in enhancing the accuracy and interpretability of Cox regression models for predicting long-term outcomes in breast cancer patients.

Keywords: PSO, Penalized Cox models, Partial likelihood function, Cox proportional hazard model.

1. INTRODUCTION

Over the last several years, the development of big data applications has gained significant importance. In fact, many government systems across various sectors increasingly rely on vast datasets. However, traditional data platforms and methodologies are less efficient in handling big data environment [1]. The primary objective of these real-world applications is to improve the accuracy of predicting the occurrence of specific events. A significant challenge in time-to-event data analysis is censoring, where important events are missed due to time constraints or become unobservable within the observation period [2, 3]. As the number of features grows, the problem of feature selection, often referred to as an NP-hard combinatorial problem, becomes increasingly complex.

The feature selection procedure employs machine learning (ML) algorithms to curate a high-quality feature set [4]. This process involves eliminating irrelevant, redundant, or undetectable features while uncovering correlation coefficients in the dataset to enhance accuracy.

Various evolutionary algorithms, such as genetic algorithms [5, 6], differential evolution [7, 8], particle swarm optimization (PSO) [9, 10], and evolutionary computation techniques, have been effectively applied to feature selection [11, 12]. The complexity of feature selection primarily arises from the challenge of reducing dimensions within the vast search space [13]. Selecting subsets with relevant features leads to improved accuracy and enhanced ML performance.

Penalized Cox models, also known as regularized Cox proportional hazards (CPH) models, are a type of survival analysis (SA) models used to predict the time until a specific event, such as death due to a disease or disease progression. The CPH model is a widely used SA method; however, it can suffer from overfitting when dealing with a large number of covariates [14]. To address this issue, penalized Cox models have been developed. These introduce a penalty term into the likelihood function, promoting the shrinkage of the coefficients toward zero. This approach yields more stable and interpretable results. The most commonly penalized Cox models are the least absolute shrinkage and selection operator (LASSO) Cox and ridge Cox models. The LASSO Cox model introduces an ℓ_1 penalty term to the likelihood function, encouraging sparse solutions by setting some coefficients to zero. In contrast, the ridge Cox model adds an ℓ_2 penalty term, promoting the shrinkage of all coefficients toward zero without setting any coefficients too close to zero. Elastic net penalized Cox models are another type of SA model; these combine elements of both LASSO and ridge Cox models.

PSO is a widely used optimization algorithm [15] that can be leveraged to find the optimal parameter set for a given objective function. It operates as a population-based algorithm, which means that it utilizes a population of selected solutions, known as particles, to identify the optimal solution.

This research proposes the PSO-PCOX method, which uses PSO to penalize the Cox model in order to optimize the positions of the particles. In each iteration of the PSO loop, the feature set corresponding to each particle's position is extracted, and a penalized Cox model is fitted using the feature set and regularization parameter λ . This approach provides a more robust and effective solution for analyzing time-to-event data, improving the accuracy of predictions, reducing the risk of overfitting, and shrinking features.

The motivation is to leverage the power of PSO to address challenges related to model complexity, overfitting, and parameter tuning in CPH models. This approach holds the potential to enhance accuracy and reliability in predicting long-term outcomes in breast cancer patients, which is crucial for making informed medical decisions and designing personalized treatment strategies.

The first study on widely used penalized Cox models was conducted by **Robert Tibshirani (1997)**, offering a novel method for CPH model feature selection and reduction [16]. In another study titled **Sparsity and Smoothness via the Fused LASSO**, the utilization of fused LASSO, a generalization suitable for scenarios where features may be effectively resolved, was proposed to encourage sparsity and smoothness [17]. A study called **Adaptive LASSO for the CPH Model** introduced the adaptive LASSO estimator, utilizing a penalized log partial likelihood with adaptively weighted ℓ_1 penalty on regression coefficients [18]. Moreover, **Regularization for the CPH Model with NP-Dimensionality** demonstrated reliable oracle characteristics of non-concave penalized algorithms for NP-dimensional data with censoring within the CPH model framework [19]. **Network Regularized High-Dimensional Cox Regression for Analysis of Genomic Data** recommended employing a network-based regularization method for high-dimensional Cox regression to incorporate previous knowledge of the variable's network structure [20]. Further, **Fast LASSO Method for Cox Model to UK Biobank** offered an approach for fitting a CPH model by maximizing the LASSO partial likelihood function using genetic data and the batch screening iterative LASSO (BASIL) [21]. **Parameters Optimization of Elastic Net Using PSO Algorithm** introduced the PSO-ENSVM model, a hybrid feature selection, optimization, and classification approach [22]. Additionally, **Adaptive LASSO Logistic Regression Based On Particle Swarm Optimization for Alzheimer's Disease Early Diagnosis** presented a new technique for early diagnosis of Alzheimer's disease. PSO may reduce the computation time for the subsequent stage by eliminating redundant features in the first stage, and adaptive LASSO primarily improves logistic regression, wherein adaptive weights can penalize various feature coefficients to ensure that feature selection is compatible with the ℓ_1 penalty [23].

The rest of this paper is arranged as follows: The second part concentrates on the survival analysis methodology used in this study. The third part introduces the optimization algorithm that includes PSO. The fourth part states the proposed method. The fifth part introduces the experimental results and discusses them further. The last part is the conclusion of this paper.

2. SURVIVAL ANALYSIS

SA, a subfield of statistics, focuses on analyzing time-to-event data, also known as survival times. Data from prospective group studies or data gathered from clinical trials are two types of data collected prospectively in time that are typically analyzed using SA methods [24]. The final event of interest should be adequately characterized to ensure a clear understanding of the times under consideration. (In the example mentioned, this could be a death resulting from cancer.) The time taken to progress from the origin to the endpoint can then be determined. One of the main challenges in SA is the presence of scenarios where the outcomes of an event become difficult to observe beyond a certain point in time. This is known as censoring. Right censoring is the most common type and, therefore, the simplest to analyze. It occurs when someone is followed from the origin (τ_0) up to a later time point (τ_c), because the person has not experienced the relevant event. Left censoring is another form of censorship; it indicates that the event occurred before the censoring time but could have occurred at any time in the past. Interval censoring is another option, wherein a person is only known to have experienced the event between two time periods, but the precise time of the event remains unrecorded. Censoring occurs in various real-life situations where complete information about an event is unavailable. This phenomenon is particularly relevant in SA, which deals with time-to-event data. The following are some illustrative examples of censoring in different contexts:

- **Medical Research:** In clinical trials, participants may drop out or be lost to follow-up before the event of interest (death due to a disease or disease progression) occurs.
- **Customer Churn:** In business, companies may track how long customers remain active before churning (ending their subscription or relationship).
- **Failure Analysis:** Engineers may study the time until failure for a product (electronic devices or machinery).
- **Epidemiology:** Researchers analyzing the time to recover from an illness might encounter left censoring if patients are recruited only after a certain duration of illness has passed.

Censoring has a crucial impact on data analysis and model development, as ignoring censoring can lead to biased estimates of survival probabilities and hazard rates. Failing to account for censoring can result in the development of

inaccurate predictive models, affecting their reliability in making future predictions. Events and censorship are depicted in Fig. 1.

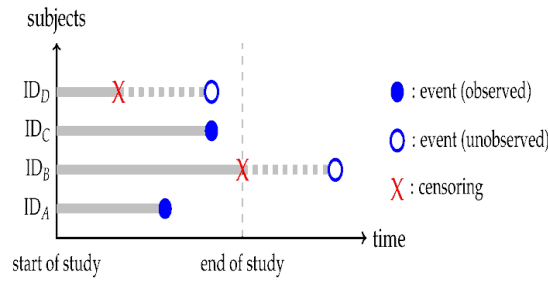


FIGURE 1. Event and censoring times [25].

2.1 COX PROPORTIONAL HAZARD (CPH) MODEL

The CPH model is a semi-parametric regression model used in SA to model the time to event of interest. It was introduced by Sir David Cox in 1972 and has since become widely used and highly influential in medical research, epidemiology, and other fields. The model assumes that the hazard rate, which represents the likelihood of experiencing the event of interest at a given time, is proportional to a set of covariates. In other words, regardless of the time elapsed since the study's inception, the model predicts that the likelihood of encountering the event of interest varies by a constant factor for each unit increase in a covariate. This proportionality assumption makes the model flexible and allows it to accommodate time-varying factors [26].

The estimation of the CPH model is done using maximum likelihood techniques, which involve finding the values of the regression coefficients that maximize the likelihood of the observed data. This model can be used to assess the survival probability or risk rate for a group of individuals with a specific set of covariate values and to compare the survival probability between different groups based on their covariate values. The CPH model finds many applications in medical research, such as predicting the risk of cancer recurrence or mortality in patients with a particular disease, and can also be applied in other fields where time-to-event data are collected. The hazard function is calculated by

$$\hat{h}(\tau|\kappa) = \hat{h}_0(\tau) e^{\beta_i \kappa} \quad (1.1)$$

Where:

- $\hat{h}(\tau|\kappa)$ is the hazard function for an individual with covariate values (κ) at time (τ).
- $\hat{h}_0(\tau)$ is the baseline hazard function that represents the hazard rate for an individual with all covariate values set to 0.
- β_i are the regression coefficients that represent the result of each covariate on the hazard rate.
- κ_1, κ_2 to κ_p are the values of the p covariates for the individual.

Under the CPH model, the hazard ratio between two subjects with covariates (τ, τ') can be written as

$$\mathcal{HR} = \mathcal{HR}(\kappa, \kappa') = \frac{\hat{h}(\tau|\kappa)}{\hat{h}(\tau|\kappa')} = e^{[\beta_i(\kappa - \kappa')]} \quad (1.2)$$

To define the distribution of the survival time, the cumulative distribution function can consistently be used:

$$\mathcal{P}(\mathcal{T} \leq \tau|\kappa) = 1 - e^{(-\int_0^\tau \hat{h}_0(\tau) e^{\beta_i \kappa} ds)} \quad (1.3)$$

One of the advantages of the Cox model is that despite the baseline function being non-parametric and the model being semi-parametric, the parameter β can be estimated without calculating the baseline function. The Cox model for the survival function can be expressed using Formula (1.4):

$$\zeta(\tau) = e^{-H_0(\tau) e^{\beta \kappa}} = \zeta_0(\tau) e^{\beta \kappa} \quad (1.4)$$

- $\mathcal{H}_0(\tau)$ is accumulative $\hat{h}_0(\tau)$.
- $\zeta(\tau)$ perform the baseline survival function.

The Cox model uses the log partial likelihood $\mathcal{L}(\beta)$ function [27], which only depends on the item of interest and is thus free of extraneous parameters.

$$\mathcal{L}(\beta) = \sum_{i: \mathcal{O}_i=1} \beta_i \kappa - \log \left(\sum_{j: \mathcal{T}_j \geq \mathcal{T}_i} e^{(\beta_i \kappa)} \right) \quad (1.5)$$

The CPH model is widely used in SA, for example, in predicting cancer survival. By applying this model, researchers can estimate the effect of each covariate on the hazard of death while assuming that hazard ratios remain constant over time. This allows them to identify significant factors affecting survival and quantify their impact.

2.2 REGULARIZATION OF COX MODELS

In statistics and ML, regularized Cox models serve as a survival model used to model time-to-event data. These models modify the likelihood function by adding a penalty term to the CPH model. The penalty term discourages the model from fitting noise in the data and encourages it to identify the most important predictors of the survival outcome. In many real-world scenarios, high-dimensional data are collected, because methods for gathering and analyzing data are ever evolving. Sometimes, the number of variables in the assumed data is nearly equal to or even greater than the number of instances. This presents a challenge in developing accurate predictive models that utilize all available features, as models may suffer from overfitting issues [28, 29]. The regularized Cox regression method is applied in Formula (1.6):

$$\hat{\beta} = \arg \min_{\beta} \mathcal{L}(\beta) + \lambda * P(\beta) \quad (1.6)$$

Here, $P(\beta)$ represents a sparsity norm, and λ is the regularization factor.

Table 1. Penalty Term Formulation

Method	Penalty Equations	N. of Eq.
Lasso-Cox [30].	$\arg \min_{\beta} \mathcal{L}(\beta) + \lambda * \beta $	(1.7)
Ridge-Cox [31, 32]	$\arg \min_{\beta} \mathcal{L}(\beta) + \frac{\lambda}{2} * \beta_2^2$	(1.8)
EN-Cox [33]	$\arg \min_{\beta} \mathcal{L}(\beta) + \lambda * (\beta + \beta_2^2)$	(1.9)
General-Cox [21]	$\arg \min_{\beta} \mathcal{L}(\beta) + \lambda * \sum \omega \beta $	(1.10)

Regularization techniques are used to prevent overfitting and improve the generalization of predictive models. For example, consider a telecommunications company, wherein analysts predict customer churn. They collect features such as usage patterns, contract length, and customer demographics. By employing techniques such as LASSO, ridge, or elastic net regularization, they can prevent the model from becoming overly complex and overfitting the training data. This ensures that only the most relevant features are retained, leading to a more interpretable and accurate churn prediction model.

Regularization is a vital tool when dealing with high-dimensional data and overfitting in CPH models. It helps control overfitting, aids in feature selection, improves model stability, enhances generalization, and mitigates multicollinearity, thereby ensuring that the model focuses on relevant patterns in the data. This leads to more accurate and interpretable survival predictions, enabling researchers and practitioners to make informed decisions based on reliable insights from complex datasets.

2.3 CROSS-VALIDATION (CV)

Cross-validation (CV) is a popular technique in ML for evaluating a model's performance. This method splits the dataset into K-folds of a similar size. After training on K-1 folds, the model is then tested on the remaining fold. In the K-time procedure, each fold is utilized exactly once as the validation data [34].

One of the benefits of CV is that it reduces the impact of random variations in the data, leading to a more accurate evaluation of the model's performance. It also ensures that all samples are used for both training and validation. CV is commonly used in ML to fine-tune hyperparameters, select models, and assess the efficacy of various methods on a given dataset. Moreover, this technique helps evaluate a model's generalization and identifies potential overfitting issues.

The pseudocode for performing k-fold cross-validation is as follows:

- i. Split the dataset into K subsets.
- ii. For each fold I in K:
 - a. Set aside subset I as the validation set.
 - b. Combine the remaining K-1 subsets to form the training set.
 - c. Train the model on the training set.
 - d. Test the trained model on the validation set and record the assessment metrics of interest.
 - e. Calculate the average of the evaluation metrics across all folds.

CV is crucial for assessing model performance and selecting the best model hyperparameters. The application of credit risk assessment is an example of CV. If a bank is to develop a predictive model for credit risk assessment, they would like to determine whether an applicant is likely to default on a loan. By employing CV, the bank can repeatedly split their dataset into training and validation sets. This allows them to evaluate how well their model generalizes to new data and provides opportunities to fine-tune hyperparameters such as regularization strength or feature selection thresholds. CV aids in selecting the best-performing model for accurate credit risk prediction.

3. PARTICLE SWARM OPTIMIZATION

The PSO algorithm is a metaheuristic algorithm inspired by the social behavior of fish communities and bird flocks. It is used to find the best solution to optimization problems by emulating the social behavior of a swarm of particles [35]. The algorithm generates a population of particles, each representing a potential solution to that problem. In each iteration, each particle's position and velocity are adjusted based on its own best position and the global best position found by the entire swarm. This movement of particles in the search space, along with tracking the best solution each particle discovers, allows the algorithm to explore the solution space [36].

Algorithm1: Particle Swarm Optimization Algorithm

```

Step 1: Set the population of particles.
    For each particle in the population:
        Initialize position and velocity.
        Set the initial personal best position as the current position.
        Update the global best position if the current position has a better fitness value.
    End for.
Step 2: Set the maximum number of iterations or a desired fitness value to achieve.
Step 3: While the end condition does not happen, do the following:
    For each particle in the population:
        Evaluate the fitness value of the particle's current position using the objective function.
        Update the private best position.
        Update the global best position.
        Update the velocity and position.
    For each dimension of the particle's position and velocity, do the following:
        Generate random numbers r1 and r2.
        Update the velocity component:
             $velocity[d] = inertia * velocity[d] + c1 * r1 * (pbest[d] - position[d]) + c2 * r2 * (gbest[d] - position[d])$ .
        Update the position component:
             $position[d] = position[d] + velocity[d]$ .
    End for.
Step 4: Check the end condition.
Step 5: Return to the global optimal position solution to the optimization problem.

```

4. THE PROPOSER METHOD

Improving the penalized Cox model through the PSO algorithm helps optimize the selection of features and model parameters in SA. The penalized Cox model uses regularization to select a subset of features and prevent overfitting, while PSO optimizes model parameters and feature selection by emulating the societal behavior of a swarm of particles, achieving a balance between exploration and exploitation to discover optimal solutions in complex spaces.

This section illustrates how PSO-PCOX is applied to breast cancer data, as shown in Fig. 2. The process involves loading and preprocessing the survival dataset, splitting it into training and test sets using 5 K-fold CV, fitting the penalized Cox model with the training set, initializing the PSO swarm, and performing PSO to optimize the positions of the particles. In each iteration of the PSO loop, the feature set corresponding to each particle's position is extracted, and a penalized Cox model is fitted using the feature set and the regularization parameter lambda. Survival function prediction is performed on the test set, followed by the calculation of the concordance index (C-index) of the model. The fitness function is then penalized based on the number of feature selections. After evaluating the fitting of the particle, the positions of the particles are updated, both individually and globally. The PSO loop continues until the stopping requirements are satisfied. This paper aims to improve the penalized Cox model through PSO with CV in order to optimize the model's performance and reduce the risk of overfitting by selecting the most relevant features and tuning the regularization parameter.

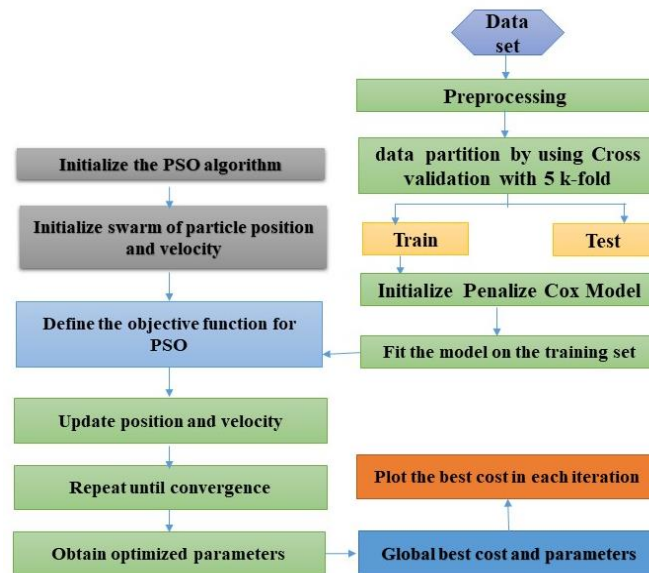


FIGURE 2. Flow chart of PSO-PCOX

Algorithm 2: Pseudo code of the proposed method (PSO-PCOX)

Step 1: Split the dataset into k non-overlapping subsets.

Step 2: for fold_idx in range (k):

a. Set aside fold_idx as the validation set.

b. To generate the training set, combine the remaining k-1 subset.

c. Use the PSO algorithm to define the Objective Function. The Concordance index (C-index) of the Cox Model with the ℓ_1 or ℓ_2 penalty term should be determined using this objective function.

d. Apply PSO to optimize the Cox Model coefficients using the specified Objective Function.

e. Train the regularized Cox model with the optimal coefficients on the training set.

f. Record the evaluation metric(s) of interest for the current fold.

Step 3: Calculate the average of the evaluation metric(s) across all k folds.

In the context of SA, the objective function of PSO combines two penalty terms: ℓ_1 (LASSO) and ℓ_2 (ridge) regularization.

For ℓ_1 penalty, the objective function becomes Objective Function = $NLL + \lambda * \sum |\beta|$. Here, NLL is negative log likelihood, β represents the coefficients of the Cox model, and λ controls the strength of regularization. The absolute values of coefficients ($\sum |\beta|$) are summed up, penalizing large coefficient values. LASSO tends to drive some coefficients to exactly zero, effectively performing feature selection and producing a sparse model.

For ℓ_2 penalty, the objective function becomes Objective Function = $NLL + \lambda * \sum \beta^2$. In ℓ_2 regularization, the squared coefficients ($\sum \beta^2$) are summed up. Ridge regularization does not force coefficients to exactly zero but rather shrinks them toward zero, leading to a more moderate reduction of coefficients and addressing multicollinearity.

Effective optimization of hyperparameters ensures that PSO-PCOX strikes a balance between model complexity, regularization, and predictive performance. These hyperparameters include the number of particles, maximum number of iterations, inertia weight, acceleration coefficients, and penalty parameter (λ).

The PSO-PCOX model addresses the challenges by offering the following:

- Flexibility in Hazard Estimation: By combining PSO with a penalized Cox model, the PSO-PCOX model can capture non-proportional hazards more effectively than traditional Cox models. The optimization process helps identify complex relationships between covariates and survival outcomes.
- Regularization for Improved Performance: The ℓ_1 or ℓ_2 penalty term incorporated through regularization aids in controlling model complexity, which is particularly beneficial when dealing with high-dimensional data.
- Feature Selection and Interpretability: The PSO-PCOX model naturally performs feature selection through ℓ_1 regularization, automatically identifying relevant predictors while excluding less important ones.
- Optimal Parameter Tuning: PSO optimizes the penalty parameter (λ) in the penalized Cox model, ensuring that the right amount of regularization is applied.

The potential real-world applications of the PSO-PCOX model extend beyond clinical settings and can be applied in various domains, including clinical prognosis and treatment personalization, disease outcomes and progression, clinical trial design and patient recruitment, healthcare resource allocation, financial risk assessment in insurance, time-to-event analysis in business and engineering, customer churn prediction in business, public health and epidemiology,

environmental monitoring and disaster response, and genomic and proteomic analysis. One common metric for SA is the C-index, which is used to evaluate the predictive accuracy of a survival model, especially in the context of SA. It provides insight into how well the model's predicted survival probabilities align with the actual outcomes. The C-index measures the ability of a survival model to correctly rank the relative risks of pairs of individuals in terms of their event times. It assesses the consistency between the predicted survival probabilities and the observed outcomes. In other words, the C-index quantifies the proportion of pairs of subjects whose predicted survival times are correctly ordered in relation to their actual survival times. Its definition is the rate of concordant pairs to all similar pairs. It acts as a rank-order statistic for predictions against actual outcomes. Assuming the equivalent instance pair (i, j) by (τ_i) and (τ_j) are the observed times, the $\zeta(\tau_i)$ and $\zeta(\tau_j)$ are the forecast survival times.

Then, the concordance probability can be given by

$$C - index = PR(\hat{\tau}_i < \hat{\tau}_j \mid \tau_i < \tau_j) \quad (1.11)$$

5. EXPERIMENTAL RESULTS

This section presents the results of the proposed method (PSO-PCOX). The hardware and software used to test the PSO-PCOX model had the following components: Intel Core i5-1135G7 Processor with 64-bit architecture, 8 GB RAM, Windows 11 operating system, and the Python 3.7 programming language.

To illustrate the use of penalized Cox models, breast cancer survival data for 198 patients was employed, including their age, estrogen receptor status, tumor size, grade, and the expression levels of 76 genes [37].

The following key parameters were considered to set up the PSO optimizer: 10 particles, a search space dimension of five, and options for updating the velocity and position of each particle in the swarm ($c1 = 0.5$, $c2 = 0.3$, $w = 0.9$, and K-fold = 5).

Table (2) illustrates the results of models used in the proposed method. The LASSO Cox model selected the most important 20 features and achieved the highest C-index result with CV.

Table 2. Penalized the Cox models

Models	Partial Log-Likelihood	Partial Log-Likelihood With cross-validation	C-index	C-index With cross-validation	Number of features after cross-validation
Lasso-Cox	-19460561.89665379	-3389580.2250737506	95.50817610062893	87.705936	20
Ridge-Cox	-19457574.655256554	-3369297.332312067	95.54381551362684	86.733949	37
Elastic-Cox	-19460457.074499007	-3374589.002166087	95.53249475890985	84.136190	21

The penalized Cox model with CV is a successful and popular SA method that provides valuable information about factors affecting the time to event of interest and is a powerful tool for outcome prediction. Fig. 3 shows the variation of log partial likelihood with different values of alpha.

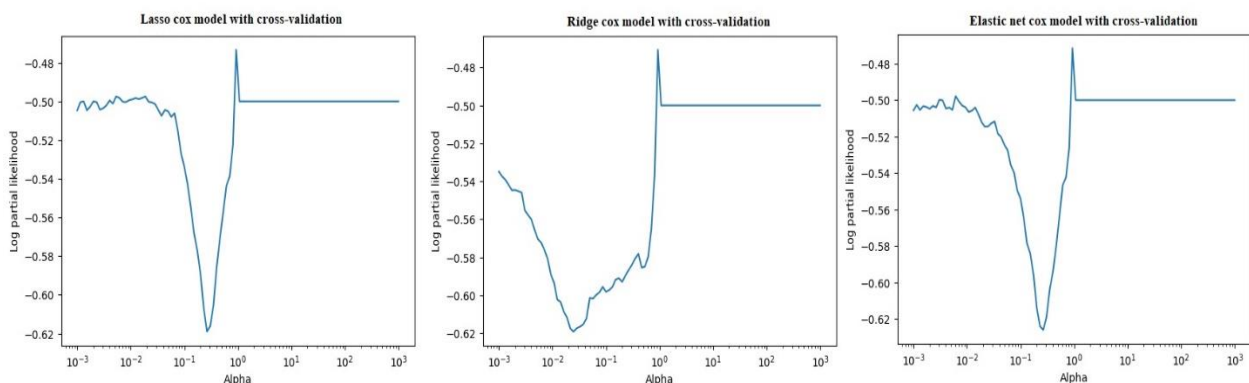


FIGURE 3. Penalize the Cox model with Cross-Validation

The mean C-index across all folds for each alpha may also be used for visualizing the results. According to the 0.5 C-index of a perfectly random model, an alpha value to the right results in a wide range and sets all coefficients to zero, as shown in Fig. 4. When alpha becomes too small, the model's performance again resembles that of a random model, because too many features have been included.

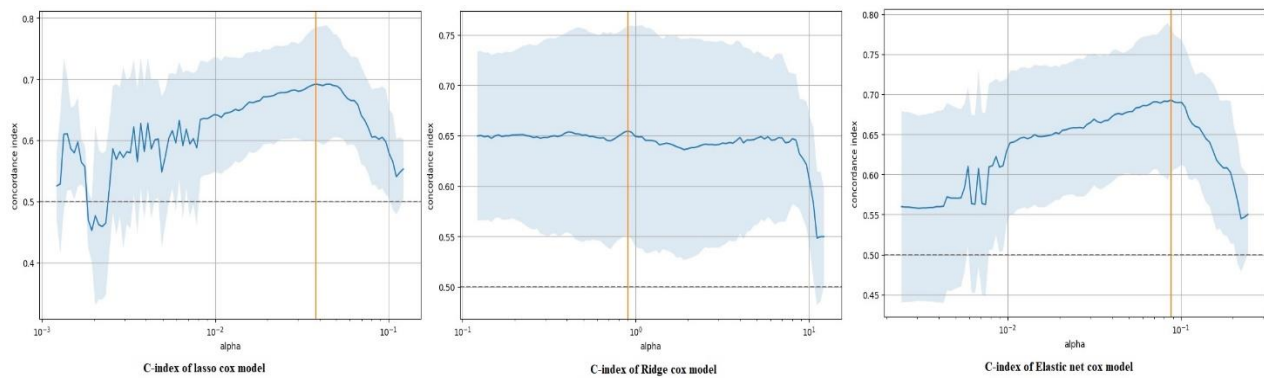


FIGURE 4. Concordance index

Table (3) presents the results of the proposed PSO-PCOX method. The PSO-LASSO Cox model achieved the highest C-index with CV, and PSO with all penalization types presented highest C-index results and lowest partial log likelihood values. Fig. 5 shows the best cost in each iteration based on the C-index.

Table 3. Results of PSO-PCOX

Models	Partial Log-Likelihood	C-index
PSO-Lasso Cox	-3471984.2966590463	96.889256
PSO-Ridge Cox	-4105575.5977999666	89.733482
PSO-Elastic Cox	-4587280.185964314	92.733909

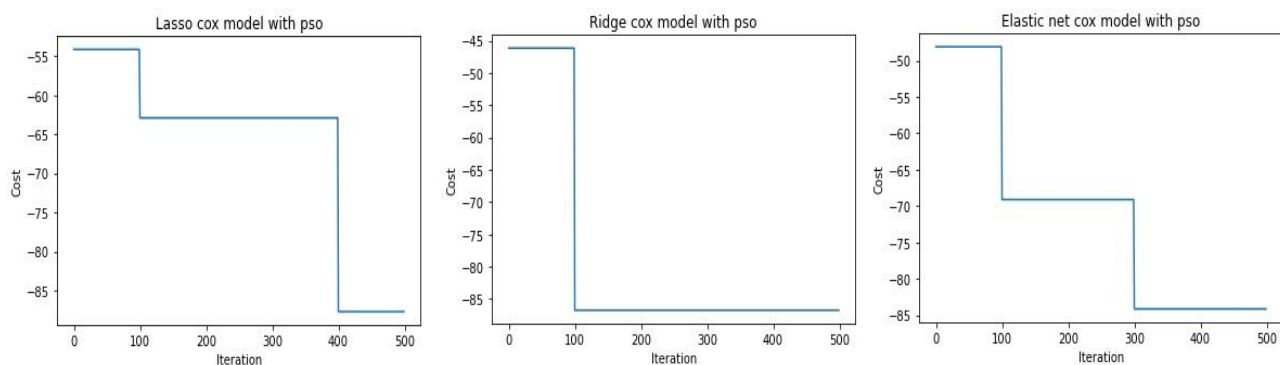


FIGURE 5. Best cost in each iteration for PSO-PCOX

6. CONCLUSIONS AND FUTURE WORK

In conclusion, the proposed PSO-PCOX model optimizes penalized Cox models using PSO, enhancing the positions of the particles. In each iteration of the PSO loop, the feature set corresponding to each particle's position is extracted, and a penalized Cox model is fitted using the feature set and the regularization parameter lambda. This approach offers a more robust and effective solution for analyzing time-to-event data, improving the accuracy of predictions, reducing the risk of overfitting, and enhancing predictive performance through feature selection and regularization adjustment. Using a dataset of breast cancer survivors' clinical and gene expression features imparts specificity to the research setting and proves the method's relevance. Future studies may explore the application of other optimization approaches, such as genetic algorithm or DE, with penalized Cox models to expand and improve the research.

Funding

None

ACKNOWLEDGEMENT

The first author would like to thank the reviewers for providing useful suggestions, allowing for the improved presentation of this paper.

CONFLICTS OF INTEREST

None

REFERENCES

- [1] A. Oussous, F.-Z. Benjelloun, A. A. Lahcen, and S. Belfkih, "Big Data technologies: A survey," *Journal of King Saud University-Computer and Information Sciences*, vol. 30, no. 4, p. 431, 2018. <https://doi.org/10.1016/j.jksuci.2017.06.001>.
- [2] K. -M. Leung, R. M. Elashoff, and A. A. Afifi, "Censoring issues in survival analysis," *Annual review of public health*, vol. 18, no. 1, p. 83, 1997. <https://doi.org/10.1146/annurev.publhealth.18.1.83>.
- [3] G. James, D. Witten, T. Hastie, and R. Tibshirani, *An introduction to statistical learning*. Springer, p. 607, 2013. <https://doi.org/10.1007/978-1-0716-1418-1>.
- [4] M. Qaraad, S. Amjad, H. Fathi, and I. I. Manhrawy, "Feature selection techniques for cancer classification applied to microarray data: a survey," in *2019 international conference on intelligent systems and advanced computing sciences (ISACS)*, IEEE, p. 1, 2019. <https://doi.org/10.1109/ISACS48493.2019.9068865>.
- [5] K. Deb, "Genetic algorithm in search and optimization: the technique and applications," in *Proceedings of International Workshop on Soft Computing and Intelligent Systems,(ISI, Calcutta, India)*, 1998: Proceedings of International Workshop on Soft Computing and Intelligent, p. 58, 1998. <https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.33.5371&rep=rep1&type=pdf>.
- [6] N. A. K. Hussein and B. Al-Sarray, "Deep Learning and Machine Learning via a Genetic Algorithm to Classify Breast Cancer DNA Data," *Iraqi Journal of Science*, p. 3153, 2022. <https://doi.org/10.24996/ijcs.2022.63.7.36>.
- [7] S. Rahnamayan, H. R. Tizhoosh, and M. M. Salama, "Opposition-based differential evolution," *IEEE Transactions on Evolutionary Computation*, vol. 12, no. 1, p. 64, 2008. <https://doi.org/10.1109/TEVC.2007.894200>.
- [8] R. D. Al-Dabbagh, "Differential evolution with adaptive repository of strategies and parameter control schemes," in 2015 IEEE Congress on Evolutionary Computation (CEC), IEEE, p. 361, 2015. <https://doi.org/10.1109/CEC.2015.7256914>.
- [9] J. J. Liang, A. K. Qin, P. N. Suganthan, and S. Baskar, "Comprehensive learning particle swarm optimizer for global optimization of multimodal functions," *IEEE transactions on evolutionary computation*, vol. 10, no. 3, p. 281, 2006. <https://doi.org/10.1109/TEVC.2005.857610>.
- [10] N. H. Salman and S. N. Mohammed, "Image Segmentation Using PSO-Enhanced K-Means Clustering and Region Growing Algorithms," *Iraqi Journal of Science*, p. 4988, 2021. <https://doi.org/10.24996/ijcs.2021.62.12.35>.
- [11] N. Ansari, "A Survey on Feature Selection Techniques using Evolutionary Algorithms," *Iraqi Journal of Science*, p. 2796, 2021. <https://doi.org/10.24996/ijcs.2021.62.8.32>.
- [12] O. Al-Janabee and B. Al-Sarray, "Evaluation Algorithms Based on Fuzzy C-means for the Data Clustering of Cancer Gene Expression," *Iraqi Journal for Computer Science and Mathematics*, vol. 3, no. 2, p. 27, 2022. <https://doi.org/10.52866/ijcsm.2022.02.01.004>.
- [13] I. A. Gheyas and L. S. Smith, "Feature subset selection in large dimensionality domains," *Pattern recognition*, vol. 43, no. 1, p. 5, 2010. <https://doi.org/10.1016/j.patcog.2009.06.009>.
- [14] S. M. Redha and A. T. A. Hadia, "Estimate the Survival Function By Using The Genetic Algorithm," *journal of Economics And Administrative Sciences*, vol. 26, no. 122, p. 440, 2020. Available on <https://www.iasj.net/iasj/download/0979743ab12f89f7>.
- [15] J. Kennedy and R. Eberhart, "Particle swarm optimization", vol. 4: IEEE, p. 1942, 1995. <https://doi.org/10.1109/ICNN.1995.488968>.
- [16] R. Tibshirani, "The lasso method for variable selection in the Cox model," *Statistics in medicine*, vol. 16, no. 4, p. 385, 1997. [https://doi.org/10.1002/\(SICI\)1097-0258\(19970228\)16:4<385::AID-SIM380>3.0.CO;2-3](https://doi.org/10.1002/(SICI)1097-0258(19970228)16:4<385::AID-SIM380>3.0.CO;2-3).
- [17] R. Tibshirani, M. Saunders, S. Rosset, J. Zhu, and K. Knight, "Sparsity and smoothness via the fused lasso," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, vol. 67, no. 1, p. 91, 2005. <https://doi.org/10.1111/j.1467-9868.2005.00490.x>.
- [18] H. H. Zhang and W. Lu, "Adaptive Lasso for Cox's proportional hazards model," *Biometrika*, vol. 94, no. 3, p. 691, 2007. <https://doi.org/10.1093/biomet/asm037>.
- [19] J. Bradic, J. Fan, and J. Jiang, "Regularization for Cox's proportional hazards model with NP-dimensionality," *Annals of statistics*, vol. 39, no. 6, p. 3092, 2011. <https://doi.org/10.1214/11-AOS911>.

- [20] H. Sun, W. Lin, R. Feng, and H. Li, "Network-regularized high-dimensional Cox regression for analysis of genomic data," *Statistica Sinica*, vol. 24, no. 3, p. 1433, 2014. <https://doi.org/10.5705/ss.2012.317>.
- [21] R. Li *et al.*, "Fast Lasso method for large-scale and ultrahigh-dimensional Cox model with applications to UK Biobank," *Biostatistics*, vol. 23, no. 2, p. 522, 2022. <https://doi.org/10.1093/biostatistics/kxaa038>.
- [22] M. Qaraad, S. Amjad, P. El-Kafrawy, H. Fathi, and I. I. Manhrawy, "Parameters optimization of elastic NET for high dimensional data using PSO algorithm," in *2020 International Conference on Intelligent Systems and Computer Vision (ISCV)*, IEEE, p. 1, 2020. <https://doi.org/10.1109/ISCV49265.2020.9204218>.
- [23] X. Cui *et al.*, "Adaptive LASSO logistic regression based on particle swarm optimization for Alzheimer's disease early diagnosis," *Chemometrics and Intelligent Laboratory Systems*, vol. 215, p. 104316, 2021. <https://doi.org/10.1016/j.chemolab.2021.104316>.
- [24] C. Kartsonaki, "Survival analysis," *Diagnostic Histopathology*, vol. 22, no. 7, p. 263, 2016. <https://doi.org/10.1016/j.mpdhp.2016.06.005>.
- [25] F. Emmert-Streib and M. Dehmer, "Introduction to survival analysis in practice," *Machine Learning and Knowledge Extraction*, vol. 1, no. 3, p. 1013, 2019. <https://doi.org/10.3390/make1030058>.
- [26] D. R. Cox, "Regression models and life-tables," *Journal of the Royal Statistical Society: Series B (Methodological)*, vol. 34, no. 2, p. 187, 1972. <https://doi.org/10.1111/j.2517-6161.1972.tb00899.x>.
- [27] Cox, D. R. "Partial Likelihood." *Biometrika* 62, no. 2, p. 269, 1975. <https://doi.org/10.2307/2335362>.
- [28] H. van Houwelingen and H. Putter, *Dynamic prediction in clinical survival analysis*. CRC Press, p. 250, 2011. <https://doi.org/10.1201/b11311>.
- [29] T. Hastie, R. Tibshirani, J. H. Friedman, and J. H. Friedman, *The elements of statistical learning: data mining, inference, and prediction*. Springer, Vol. 2, p. 1, 2009. <https://doi.org/10.1007/978-0-387-21606-5>.
- [30] T. Robert. "Regression shrinkage and selection via the lasso," *Journal of the Royal Statistical Society. Series B (Methodological)*, Vol. 58, No.1, p. 267, 1996. <https://doi.org/10.1111/j.2517-6161.1996.tb02080.x>.
- [31] A. E. Hoerl and R. W. Kennard, "Ridge regression: Biased estimation for nonorthogonal problems," *Technometrics*, vol. 12, no. 1, p. 55, 1970. <https://doi.org/10.1080/00401706.2020.1742207>.
- [32] P. J. Verweij and H. C. Van Houwelingen, "Penalized likelihood in Cox regression," *Statistics in medicine*, vol. 13, no. 23-24, p. 2427, 1994. <https://doi.org/10.1002/sim.4780132307>.
- [33] 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, p. 301, 2005. <https://doi.org/10.1111/j.1467-9868.2005.00503.x>.
- [34] M. M. Rahma and A. D. Salman, "Heart Disease Classification–Based on the Best Machine Learning Model," *Iraqi Journal of Science*, p. 3966, 2022. <https://doi.org/10.24996/ijs.2022.63.9.28>.
- [35] D. Wang, D. Tan, and L. Liu, "Particle swarm optimization algorithm: an overview," *Soft computing*, vol. 22, no. 2, p. 387, 2018. <https://doi.org/10.1007/s00500-016-2474-6>.
- [36] Y. Zhang, S. Wang, and G. Ji, "A comprehensive survey on particle swarm optimization algorithm and its applications," *Mathematical problems in engineering*, vol. 2015, p. 1, 2015. <https://doi.org/10.1155/2015/931256>.
- [37] C. Desmedt *et al.*, "Strong time dependence of the 76-gene prognostic signature for node-negative breast cancer patients in the TRANSBIG multicenter independent validation series," *Clinical cancer research*, vol. 13, no. 11, p. 3207, 2007. <https://doi.org/10.1158/1078-0432.CCR-06-2765>.