Mathematical Biology and Bioinformatics
2023. V. 18. № 2. P. 282-293. doi: 10.17537/2023.18.282

========== **DATA MINING** ==========

An Investigational Modeling Approach for Improving Gene Selection using Regularized Cox Regression Model

Ghada Yousif Ismail Abdallh[1] and Zakariya Yahya Algamal[2,3*]

[1]*Department of Mechanical Technology, Technical Institute in Mosul, Northern Technical University, Mosul, Iraq*
[2]*Department of Statistics and Informatics, University of Mosul, Mosul, Iraq*
[3]*College of Engineering, University of Warith Al-Anbiyaa, Karbala, Iraq*

Abstract: By producing the required proteins, the process of gene expression establishes the physical properties of living things. Gene expression from DNA or RNA may be recorded using a variety of approaches. Regression analysis has evolved in prominence in the area of genetic research recently. Several of the genes in high dimensional gene expression information for statistical inference may not be related to their illnesses, which is one of the major problems. The ability of gene selection to enhance the outcomes of several techniques has been demonstrated. For censored survival data, the Cox proportional hazards regression model is the most widely used model. In order to identify important genes and achieve high classification accuracy, a new technique for selecting the tuning parameter is suggested in this study using an optimization algorithm. According to experimental findings, the suggested strategy performs much better than the two rival methods in terms of the area under the curve and the number of chosen genes. This study provides a comprehensive assessment of the latest work on performance evaluation of regression analysis in gene selection. In addition to its performance analysis, this research conducts a thorough assessment of the numerous efforts done on various extended models based on gene selection in recent years.

Key words: *Cox regression model; penalized method; gene selection; crow search algorithm.*

INTRODUCTION

Gene selection is a method used to reduce the quantity of duplicated, under-expression, or uninformative genes in a gene expression dataset, such as a DNA microarray. Gene choice depends on the Feature Selection approach based on regression analysis, which is ideally suited for applications requiring thousands of characteristics. Identifying the relevant and expressing genes and getting rid of the redundant genes from the original space are the two major goals of using gene selection techniques. The model's performance should decline as the number of genes increases, and overfitting might undermine the generalization. Focusing on relevance, redundancy, and complementarity is necessary to achieve meaningful results. When a gene provides critical details about a certain class, either alone or in combination with some other genes, it is deemed significant. The feature subset can be divided into very relevant, mildly relevant, and unimportant in technical terms. Strongly relevant, weakly relevant and non-redundant features are where one will find the majority of the significant attributes.

* zakariya_yahyaa56@outlook.com

The problem of analyzing time-to-event data arises in a number of applied fields, such as medicine, biology, public health, and epidemiology [1, 2]. Nowadays, high dimensional gene expression data are increasingly used for modeling various clinical outcomes to facilitate disease diagnosis, disease prognosis, and prediction of treatment outcome [3].

Regression modeling is a standard practice to study jointly the effects of multiple predictors on a response. The Cox proportional hazards model is ubiquitous in the analysis of time-to-event data. When the number of predictors is large, building a Cox proportional hazards model including all of them is undesirable because it has low prediction accuracy and is hard to interpret [4, 5]. For these reasons, variable selection has become an important focus in Cox proportional hazards modeling.

Penalized methods are very effective variable selection methods. These methods combine the Cox proportional hazards model with a penalty to perform variable selection and estimation simultaneously. With deferent penalties, several Cox proportional hazards models can be applied, among which are, LASSO, which is called the least absolute shrinkage and selection operator [6], smoothly clipped absolute deviation (SCAD) [7], elastic net [8], and adaptive LASSO [9]. Unquestionably, LASSO is considered one of the most popular procedures in the class of penalized methods. However, LASSO has a limitation: It applies the same amount of the penalty to all variables. Thus, it is an inconsistent variable selection method [9–11].

In general, methods for classifying the genomic information such as the signal-to-noise ratio (SNR) method, the partial least squares method, the Pearson correlation coefficient method and the t-test statistic method [12], typically use a set of criteria pertaining to the correlation extent to prioritize and select key genes. DNA microarray computer modeling also employs autonomous principal components analysis. Researcher conducted a rigorous and thorough examination of various key algorithms [13] in order to outfit the system with the optimal setting of classification, genetic identification, and cross-validation procedures.

Comparing BagBoosting [14] against a number of well-known class classification techniques for microarray data demonstrates the estimates of the future. Moreover, it offers to consolidate the classification performance of the rules for trustworthy predictions after discovering a variety of noteworthy and diverse rules using high-dimensional profile data. Low-ranked characteristics are present in the identified rules that were occasionally required for classifiers to reach 100% accuracy. Three alternative supervised machine learning algorithms, including boosted decision trees and C4.5 decision trees [15], have been the subject of some prior research on the categorization. They have examined the classification and prediction performances of various techniques using classification tasks on seven freely accessible neoplastic genetic analyses.

The relevant attributes are preserved with the use of feature selection. Most often, feature selection is used with high-dimensional statistical data. In disciplines like RNA sequencing and DNA microarray, in which there are too many characteristics and not enough samples, feature selection is quite helpful [16]. To extract the relevant selected features from the initial feature space, the main goal of feature selection, which has recently gained popularity. Techniques for feature selection help in managing the dimensionality, better interpreting the feature space, increasing prediction accuracy, and shortening the length of the modeling training phase. The result of feature selection, which aids in the prediction step is the optimum quantity of attributes that are pertinent to the specified classification model.

The process of turning the initial feature region into a prominence region, which might be a straight or non-linear collection of the initial space of features, is known as feature extraction. The main disadvantage is that it drastically changes the original feature space, which ultimately makes the data unintelligible. The transition is typically pricey as well. The representative sample is inadequate, and the gene expression data is extremely dimensional on the other extreme. The data's high degree of dimensionality is a result of the thousands of values produced for each gene in a genome. A gene's thousands of molecules may be analyzed in a specific sample with the use of sophisticated technology like microarray. Nevertheless, the drawback of microarray is its high cost [17].

The notions of unmonitored, supervising, and semi-supervised feature selection are essential since information about gene expression is often unmarked, labelled, or semi-labelled. Unsupervised learning seems to have no prior knowledge of the capabilities; yet, based on distribution, variability, and specificity, it verifies gene selection. Meaningful classification methods and details about the functionality are part of the labelled data. Then, depending on the relevancy and significance score of the designated traits, gene selection will be carried out. Combining a little quantity of unlabeled data with labelled data and vice versa adds extra information in the semi-supervised or semi-unsupervised model. The significance of features or gene selection for a better outcome is discussed in this work [18]. The remaining portions discuss the background and progression of feature selection, the overall concept in feature selection, a thorough analysis of numerous works on gene selection in the literature, unresolved problems, and possible future study directions related to the gene expression data. The generalized representation of gene selection is depicted in Figure 1.

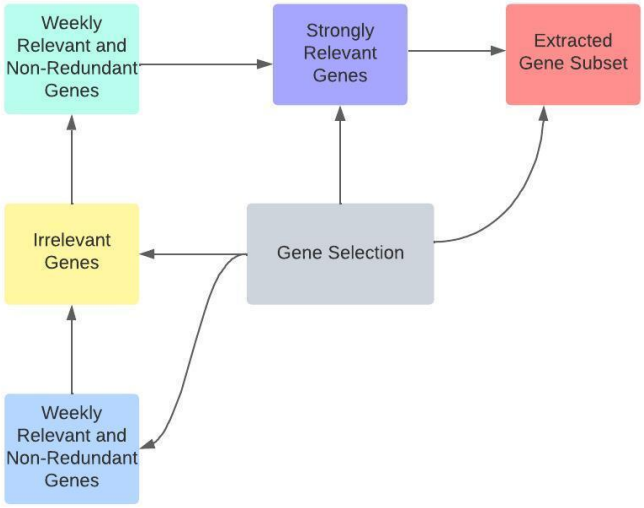

Fig 1. Simpler depiction of gene selection.

To increase the power of informative gene selection, in the present study, an adaptive Cox proportional hazards model is proposed. More specifically, a new weight inside LASSO is proposed, which can correctly reduce the estimation error. This weight will reflect the importance amount of each gene. Experimentally, comparisons between our proposed gene selection method and other competitor methods are performed. The experimental results prove that the proposed method is very effective for selecting the relevant genes with high prediction accuracy.

PANELIZED COX PROPORTIONAL HAZARDS MODEL

Survival analysis is the statistical branch studying time-to-event data, or, more precisely, the time elapsing from a well-defined initiating event to some particular endpoint. The Cox proportional hazards regression model is one of the most popular and useful models in survival analysis [19].

Consider an analysis with time-to-event outcome, we denote the observed triplet as $\{(t_i, \delta_i, x_i) : i = 1, ..., n\}$ where t_i is the survival time if $\delta_i = 1$ and censored time if $\delta_i = 0$, and $x_i = (x_{i1}, ...x_{ip})$ is a p-dimensional explanatory variables. Under the proportional hazards framework, the Cox proportional hazards model (CPHM) can be defined as

$$h(t_i \mid x_i) = h_0(t_i) \exp(\beta^T x_i), \quad (1)$$

where $h_0(t_i)$ is the baseline hazard function, and $\beta = (\beta_1, ..., \beta_p)^T$ is a ($p \times 1$)-vector of unknown regression coefficients. Assuming that the subjects are statistically independent of each other, the joint probability of all realized events is the following partial likelihood

$$L(\beta) = \prod_{i=1}^{n} \frac{\exp\left(\beta^T x_i\right)}{\sum_{j \in R_i} \exp\left(\beta^T x_j\right)}, \tag{2}$$

where R_i is the set of subjects that are at risk just before time t_i .

The estimation of the regression parameters of Eq. (1) is commonly carried out by minimizing the partial log likelihood function (Eq. (2)) as

$$\hat{\beta}_{CPHM} = \arg\min_{\beta} \left(\log L(\beta)\right) = -\sum_{i=1}^{n} \left[\beta^T x_i - \log \sum_{j \in R_i} \exp(\beta^T x_j) \right]. \tag{3}$$

Panelized Cox proportional hazards model (PCPHM) adds a nonnegative penalty term to Eq. (1), such that the size of variable coefficients can be controlled. Several penalty terms have been discussed in the literature considering the Cox proportional hazards model [20-37]. The LASSO method, proposed by Tibshirani [6], is one of the popular penalty terms. The LASSO performs variable selection and estimation simultaneously by constraining the log-likelihood function of variable coefficients. Generally, the PCPHM is defined as

$$PCPHM = \sum_{i=1}^{n} \left[\beta^T x_i - \log \sum_{j \in R_i} \exp(\beta^T x_j) \right] - \lambda P(\beta), \tag{4}$$

where $\lambda P(\beta)$ is the penalty term that regularized the estimates. The penalty term depends on the positive tuning parameter, $\lambda > 0$, which controls the tradeoff between fitting the data to the model and the effect of the regularization. In other words, it controls the amount of shrinkage. For the $\lambda = 0$, we obtain the CPHM solution in Eq. (3). In contrast, for large values of λ, the influence of the penalty term on the coefficient estimates increases.

Without loss of generality, it is assumed that the explanatory variables are standardized, $\sum_{i=1}^{n} x_{ij} = 0$ and $(n^{-1}) \sum_{i=1}^{n} x^2_{ij} = 1, \quad \forall j \in \{1, 2, ..., p\}$. The estimation of the vector β using LASSO is obtained by minimizing Eq. (4) as [31–33]

$$\hat{\beta}^{LASSO}_{Cox} = \arg\min_{\beta} \left[-\sum_{i=1}^{n} \left[\beta^T x_i - \log \sum_{j \in R_i} \exp(\beta^T x_j) \right] + \lambda \sum_{j=1}^{p} |\beta_j| \right]. \tag{5}$$

Equation (5) can be efficiently solved by the coordinate descent algorithm [41].

The LASSO has an advantage in that it is computationally feasible in high-dimensional data.

CROW SEARCH ALGORITHM

The crow search algorithm is one of the most recent evolutionary algorithms inspired from the social behavior of the crow. This algorithm was introduced in 2016 by Askarzadeh [42]. A novel kind of optimization using swarm intelligence entity is crow search algorithm was developed by imitating the clever behavior of crows in concealing and scavenging for food. The method features a straightforward structure, a small number of process variables, and uncomplicated implementations. The crow is a remarkably intelligent bird that can recall human faces and alert its species of impending danger. The capacity of crows to conceal feed and recollect the placements of the concealed sustenance represents one of the most blatant examples of their cunning. They will accompany one another simultaneously in search of a better food supply, but if a crow discovers that other crows are following it, it will attempt to move the location of its food to prevent food stealing.

The following are the crow search optimizer's capabilities:

Setting the input values for the crow search optimizer's variables, which primarily encompass the population levels, repetitions, flying step size, and sensitivity probabilities.

The storage matrices and individual crows are then initialized, and crows are then created in the search area where every one of which represents symbolizes a potential solution to a specific issue. It is presumed that the initial memory matrix represents the beginning location because the initial population lacks experience.

The fitness function's efficiency of each crow is then assessed by creating a new location for every crow in the search area. This is accomplished by presuming that a crow would follow another crow arbitrarily in an effort to locate where it will discover its hidden food.

It may be capable of determining a position in two situations relying on the crow's position update. The first crow is unaware of the ones that really are trailing it. Following, if it discovers the one after, it will assign the pursued crow a random place.

In CSA, the idea is motivated from the storing process of the excess food in hiding places then restoring it in the necessary time. It is known that the crow is very intelligent bird that observes the others hide their food and steal it once they leave. After committing the theft, it hides to avoid being a victim in the future. It is assumed that a flock of n_c crows, the crow number i has position at iteration t is x_i^t. The hiding place of the food followed by crow i is memorized. Crow moves in the search plane and tries to find the best food source, which is defined as M_i^t. The searching approach in CSA has two probable scenarios; the first one is that the owner crow j of food source M_j^t does not know the thief crow i follows it therefore the thief crow reaches to the hide place of owner crow. The updating process of the crow thief position is done by

$$x_i^{t+1} = x_i^t + \tau \times fl \times (M_j^t - x_i^t), \quad i = 1, 2, ...n_c, \tag{6}$$

where fl is the flight length and τ is a random number in the interval $[0,1]$.

The second scenario is that the owner crow j knows that the thief crow i follows it therefore, the owner crow will deceive crow i by going to any another position of search space. The position of crow i is updated by a random position. In CSA, the scenario is determined by the following expression:

$$x_i^{t+1} = \begin{cases} x_i^t + \tau \times fl \times (M_j^t - x_i^t), & \text{if } \theta \geq \text{AP} \\ \text{random position}, & \text{otherwise} \end{cases} \tag{7}$$

where θ a random number in the interval $[0,1]$ and AP is the probability of awareness.

THE PROPOSED METHOD

In the context of gene expression data problems, the goal of gene selection is to improve prediction performance, to provide faster and more cost-effective genes, and to achieve a better knowledge of the underlying problem. High dimensionality can negatively influence the performance of the Cox proportional hazards regression model by increasing the risk of overfitting and lengthening the computational time. Therefore, removing irrelevant and noisy genes from the original microarray gene expression data is essential for applying Cox proportional hazards regression model to analyze the microarray gene expression data.

A crucial part of variable selection and model estimation using the penalized methods is the selection of the tuning parameter. In practice the tuning parameter, λ, has to be chosen by a data driven procedure [11, 43, 44]. This can be achieved by using cross-validation or generalized cross-validation. These criteria can be defined as

$$\mathrm{CV}_{(\lambda)} = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{y_i - \hat{y}_{i(\lambda)}}{1 - s_{ii}}\right)^2, \tag{8}$$

$$\mathrm{GCV}_{(\lambda)} = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{y_i - \hat{y}_{i(\lambda)}}{1 - tr\mathbf{S}/n}\right)^2, \tag{9}$$

where $\hat{y}_{i(\lambda)}$ is the fitted values and s_{ii} is the i-th diagonal element of the hat matrix, $\mathbf{S}$, of the selected predictors, where $\mathbf{S} = \mathbf{X}\,[\mathbf{X}^T\mathbf{X} + \boldsymbol{\Sigma}_{(\lambda)}(\hat{\beta}_{Cox}^{LASSO})]^{-1}\mathbf{X}^T$.

It is worth mentioning that CV and GCV method is greatly dependent on the fold assignment process, which leads to large variability in selecting the shrinkage parameter value and, consequently, will negatively affect the prediction performance of the penalized Cox regression using LASSO. This is happen because repeating the observations assignment to folds might result in significantly different values of λ [45, 46].

For LASSO penalty, we have one parameter, α. This tuning parameter treated as a position in BA. Specifically, our improving of penalized regression model is depending on giving a wide search space for the α values with less time. Consequently, the steps of our proposed improving are as:

Step 1: The number of agents is $nc = 30$ and the maximum number of iterations is $t_{\max}$ =100.

Step 2: The positions of each patient are randomly specified. For the λ, the position is randomly generated from uniform distribution $U(0,50)$.

Step 3: The fitness function is defined as

$$\text{fitness} = \min\left[\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2\right], \tag{10}$$

where the fitness is calculated for the testing dataset.

Step 4: The positions of the agent are updated using Eq. (7).

Step 5: Steps 3 and 4 are repeated until a $t_{\max}$ is reached.

REAL APPLICATION

"To evaluate the performance of the proposed method, three real gene datasets were used. A brief introduction and summary of the used datasets are given in Table 1. The first dataset is the Diffuse large B-cell lymphoma dataset (DLBCL) [47]. There are 240 lymphoma patients' samples. Each patient's data consists 7399 gene expression measurements, and its survival time, including censored or not.

The second dataset is the Dutch breast cancer dataset (DBC) [48]. In this dataset, there was 295 breast cancer patients' information collected in this dataset. Each patient's data consist 4919 gene expression measurements.

The third dataset is the Lung cancer dataset (LC) [49]. This dataset contains 86 lung cancer patients' information including 7129 gene expression measurements, survival time and whether the survival time is censored.

Table 1. The details of the three used real microarray datasets

Dataset	Sample	Gene	Censored
DLBCL	240	7399	102
DBC	295	4919	207
LC	86	7129	62

To demonstrate the usefulness of the proposed method, comparative experiments with the CV and GCV are conducted. To do so, each gene expression dataset is randomly partitioned into the training dataset and the test dataset, where 70% of the sample is selected for training dataset and the rest 30% are selected for testing dataset. For a fair comparison and for alleviating the effect of the data partition, all the used methods are evaluated, for their classification performance metrics using 10 folds cross validation, averaged over 100 partitioned times. Depending on the training dataset, the tuning parameter value, λ, for each used method was fixed as $0 \le \lambda \le 50$. To assess how well the model predicts the outcome, the idea of time-dependent receiver-operator characteristics (ROC) curves for censored data and area under the curve (AUC) as our criteria. The real application results are summarized in Tables 2–4.

The Table 2 and the Figure 2 show the average results of different used methods applied to the three real datasets. Obviously, the number of genes selected by CV is much larger than that of GCV and the proposed method. Between the other two methods, the proposed method selected the least subset of genes. For example, in LC dataset, the proposed method selected 21 gens out of 7129 genes comparing to 61 and 75 selected genes by GCV and CV, respectively.

Table 2. The selected genes results

Dataset	CV	GCV	Proposed
DLBCL	122	94	54
DBC	87	44	32
LC	75	61	21

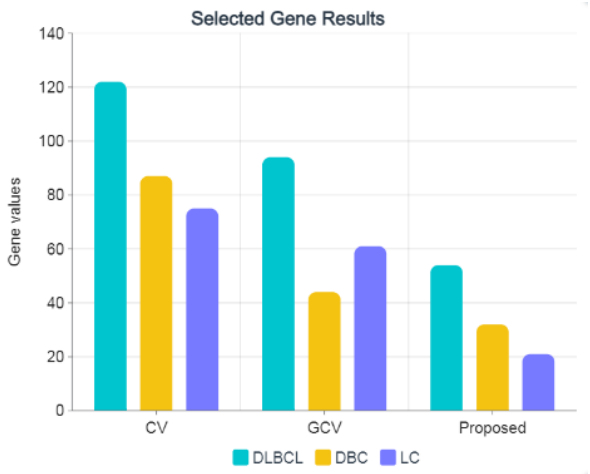

Fig.2. Selected gene results.

In order to test the prediction accuracy of the different used methods, their average values of AUC for both the training and testing dataset are given in Tables 3 and 4, respectively. In the observation of Table 3 and figure 3, in terms of AUC, the proposed method achieved a maximum accuracy of 95.8%, 96.7% and 97.1% for DLBCL, DBC, and LC datasets, respectively. Furthermore, it is clear from the results that the proposed method outperformed the GCV for all datasets. This improvement in AUC is mainly due to the proposed method ability in taking into account the new weight. Moreover, the proposed method improved the classification accuracy compared to CV. The improvements were 8.5%, 7.7%, and 7.3% for the DLBCL, DBC, and LC datasets, respectively.

Table 3. The AUC results for the training dataset

Dataset	CV	GCV	Proposed
DLBCL	0.871	0.912	0.958
DBC	0.884	0.924	0.967
LC	0.901	0.937	0.971

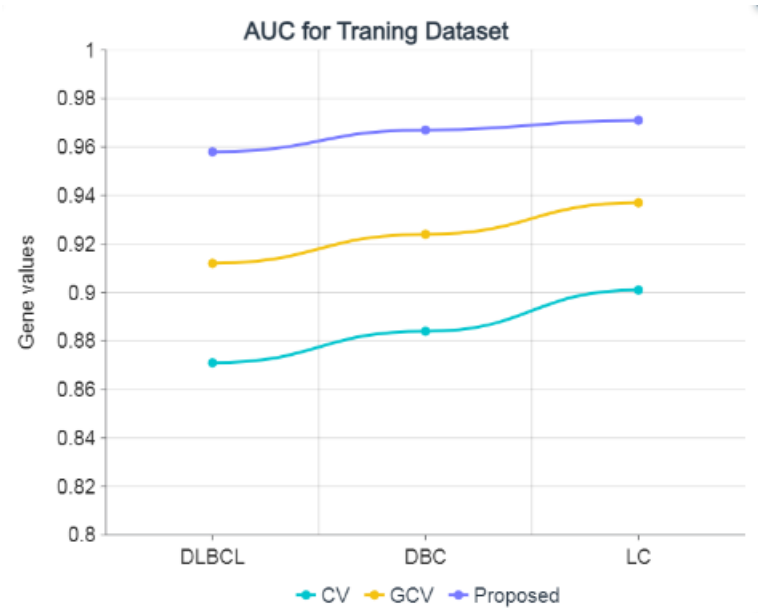

Fig 3. Results of AUC for the training dataset.

It can also be seen from Table 4 and Figure 4 that the proposed method has the best results in terms of the AUC for the testing dataset. The proposed method has the largest AUC of 93.8%, 94.9%, and 95.6% for the DLBCL, DBC, and LC datasets, respectively. This indicated that the proposed method significantly succeeded in identifying the patients who are in fact having the cancer with a probability of greater than 0.93.

Table 4. The AUC results for the testing dataset

Dataset	CV	GCV	Proposed
DLBCL	0.854	0.905	0.938
DBC	0.814	0.911	0.949
LC	0.882	0.923	0.956

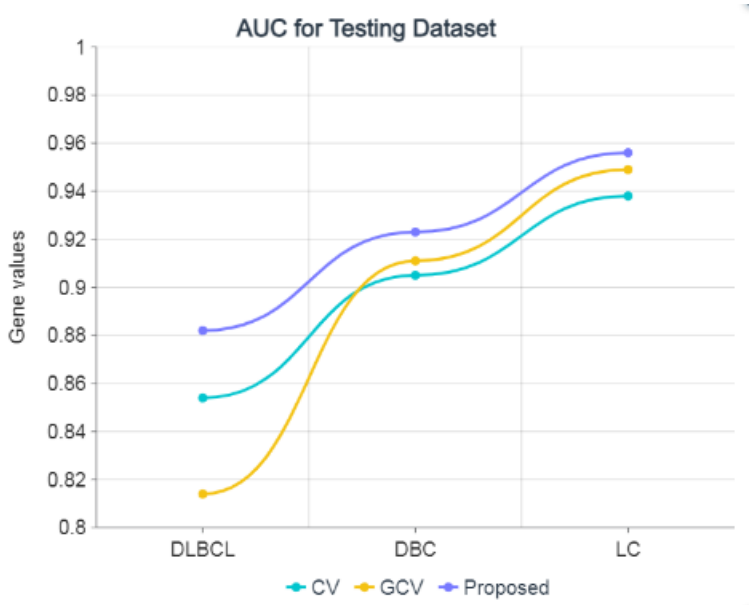

Fig. 4. Results of AUC for the testing dataset.

CONCLUSION

Regression modelling is frequently used to evaluate how numerous factors interact to produce a result. In the examination of time-to-event data, the Cox proportional hazards model is frequently used. Building a Cox proportional hazards model with every predictor is undesirable when there are many predictors since it has poor prediction accuracy and is challenging to grasp. These factors have led to a significant shift in the importance of variable selection in Cox proportional hazards modelling. In reality, the traditional Cox regression algorithms only consider choosing one biomarker, disregarding the high association between genes. Even though network-based Cox regression techniques circumvent these issues, the life science field uses these network-based methods less frequently. This work introduces the penalized Cox proportional hazards regression model, which combines the Cox proportional hazards regression model with the LASSO, to identify the significant genes in gene expression data. A unique approach to selecting tuning parameters, motivated by natural processes, was proposed. The crow search approach was used for this particular endeavor. Comparisons with existing methods and experimental analysis of our proposed strategy involve higher predictive power of the recommended technique was shown by the AUC. In fact, we demonstrated that, given an insignificant number of variables, all the approaches under consideration could choose the changed genes in various simulated conditions. On the other hand, the analysis demonstrated that it is possible to find gene signatures with a precise prognosis capability by integrating network information with tuning parameters into Cox regression algorithms.

The results of this study have several significant implications for current and future clinical practice. First, in order to condense the size of the feature space to a reasonable scale, a study based on a quick screening technique might be developed. To incorporate biological data into statistical screening analysis and give a more certain picture of the gene-regulatory networks, various screening approaches could actually be merged. Secondly, adding clinical data and data from other omics to the screening process may also result in research that is more thorough and avoid the shortcomings of the current approaches. Additionally, a proteome-scale map of the human binary interactome can be compared to alternative network maps to provide a deeper knowledge of the relationships between genotype and phenotype, allowing for a more accurate biomarkers investigation. Finally, in order to turn this methodological framework into a useful tool, it will be important to create a user-friendly interface.

REFERENCES

1. Cockeran M., Meintanis S.G., Allison J.S. Goodness-of-fit tests in the Cox proportional hazards model. *Communications in Statistics – Simulation and Computation*. 2019. V. 50. No. 12. P. 4132–4143. doi: 10.1007/978-1-4612-0103-8_18
2. Emura T., Chen Y.H., Chen, H.Y. Survival prediction based on compound covariate under Cox proportional hazard models. *PLoS One*. 2012. V. 7. No. 10. Article No. e47627. doi: 10.1371/journal.pone.0047627
3. Huang J., Liu L., Liu Y., Zhao X. Group selection in the Cox model with a diverging number of covariates. *Statistica Sinica*. 2014. P. 1787–1810. doi: 10.5705/ss.2013.061
4. Karabey U., Tutkun N.A. Model selection criterion in survival analysis. *AIP Conference Proceedings*. 2017. V. 1863. No. 1. Article No. 120003. doi: 10.1063/1.4992296
5. Leng C., Zhang H.H. Model selection in nonparametric hazard regression. *Journal of Nonparametric Statistics*. 2006. V. 18. No. 7–8. P. 417–429. doi: 10.1080/10485250601027042
6. Tibshirani R. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*. 1996. V. 58. No. 1. P. 267–288. doi: 10.1111/j.2517-6161.1996.tb02080.x
7. Fan J., Li R. Variable selection via nonconcave penalized likelihood and its oracle properties. *Journal of the American Statistical Association*. 2001. V. 96. No. 456. P. 1348–1360. doi: 10.1198/016214501753382273
8. Zou H., Hastie T. Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*. 2005. V. 67. No. 2. P. 301–320. doi: 10.1111/j.1467-9868.2005.00503.x
9. Zou H. The adaptive lasso and its oracle properties. *Journal of the American Statistical Association*. 2006. V. 101. No. 476. P. 1418–1429. doi: 10.1198/016214506000000735
10. Algamal Z.Y., Lee M.H. Penalized logistic regression with the adaptive LASSO for gene selection in high-dimensional cancer classification. *Expert Systems with Applications*. 2015. V. 42. No. 23. P. 9326–9332. doi: 10.1016/j.eswa.2015.08.016
11. Algamal Z.Y., Lee M.H. Regularized logistic regression with adjusted adaptive elastic net for gene selection in high dimensional cancer classification. *Computers in Biology and Medicine*. 2015. V. 67. P. 136–145. doi: 10.1016/j.compbiomed.2015.10.008
12. Golub T.R., Slonim D.K., Tamayo P., Huard C., Gaasenbeek M., Mesirov J.P., Coller H., Loh M.L., Downing J.R., Caligiuri M.A., et al. Molecular classification of cancer: class discovery and class prediction by gene expression monitoring. *Science*. 1999. V. 286. No. 5439. P. 531–537. doi: 10.1126/science.286.5439.531
13. Nguyen D.V., Rocke D.M. Tumor classification by partial least squares using microarray gene expression data. *Bioinformatics*. 2002. V. 18. No. 1. P. 39–50. doi: 10.1093/bioinformatics/18.1.39
14. Xiong M., Jin L., Li W., Boerwinkle E. Computational methods for gene expression-based tumor classification. *Biotechniques*. 2000. V. 29. No. 6. P.1264–1270. doi: 10.2144/00296bc02
15. Baldi P., Long A.D. A Bayesian framework for the analysis of microarray expression data: regularized *t*-test and statistical inferences of gene changes. *Bioinformatics*. 2001. V. 17. No. 6. P. 509–519. doi: 10.1093/bioinformatics/17.6.509
16. Shibly F.H.A., Kumar R.L. Image Processing for Automatic Cell Nucleus Segmentation Using Super pixel and Clustering Methods on Histopathological Images. *Tamjeed Journal of Healthcare Engineering and Science Technology*. 2023. V. 1. No. 1. P. 54–63.
17. Statnikov A., Aliferis C.F., Tsamardinos I., Hardin D., Levy S. A comprehensive evaluation of multicategory classification methods for microarray gene expression cancer diagnosis. *Bioinformatics*. 2005. V. 21. No. 5. P. 631–643. doi: 10.1093/bioinformatics/bti033

18. Liu Y. Detect key gene information in classification of microarray data. *EURASIP Journal on Advances in Signal Processing*. 2008. Article No. 612397 (2008). doi: 10.1155/2008/612397
19. Cox D.R. Regression models and life-tables. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*. 1972. V. 34. No. 2. P. 187–202. doi: 10.1111/j.2517-6161.1972.tb00899.x
20. Du P., Ma S., Liang H. Penalized Variable Selection Procedure for Cox Models with Semiparametric Relative Risk. *Ann. Stat*. 2010. V. 38. No. 4. P. 2092–2117. doi: 10.1214/09-AOS780
21. Fu Z., Parikh C.R., Zhou B. Penalized variable selection in competing risks regression. *Lifetime Data Anal*. 2017. V. 23. No. 3. P. 353–376. doi: 10.1007/s10985-016-9362-3
22. Gui J., Li H. Penalized Cox regression analysis in the high-dimensional and low-sample size settings, with applications to microarray gene expression data. *Bioinformatics*. 2005. V. 21. No. 13. P. 3001–3008. doi: 10.1093/bioinformatics/bti422
23. Hossain S., Ahmed S.E. Penalized and Shrinkage Estimation in the Cox Proportional Hazards Model. *Communications in Statistics – Theory and Methods*. 2014. V. 43. No. 5. P. 1026–1040. doi: 10.1080/03610926.2013.826368
24. Hou W., Song L., Wang X. Penalized Empirical Likelihood via Bridge Estimator in Cox's Proportional Hazard Model. *Communications in Statistics – Theory and Methods*. 2013. V. 43. No. 2. P. 426–440. doi: 10.1080/03610926.2012.657325
25. Kauermann G. Penalized spline smoothing in multivariable survival models with varying coefficients. *Computational Statistics & Data Analysis*. 2005. V. 49. No. 1. P. 169–186. doi: 10.1016/j.csda.2004.05.006
26. Lin C.Y., Halabi S. A Simple Method for Deriving the Confidence Regions for the Penalized Cox's Model via the Minimand Perturbation. *Commun. Stat. Theory Methods*. 2017. V. 46. No. 10. P. 4791–4808. doi: 10.1080/03610926.2015.1085568
27. Park E., Ha, I.D. Penalized variable selection for accelerated failure time models. *Communications for Statistical Applications and Methods*. 2018. V. 25. No. 6. P. 591–604. doi: 10.1002/sim.8023
28. Shi Y., Xu D., Cao Y., Jiao Y. Variable Selection via Generalized SELO-Penalized Cox Regression Models. *Journal of Systems Science and Complexity*. 2019. V. 32. No. 2. P. 709–736. doi: 10.1007/s11424-018-7276-8
29. Suchting R., Hebert E.T., Ma P., Kendzor D.E., Businelle M.S. Using Elastic Net Penalized Cox Proportional Hazards Regression to Identify Predictors of Imminent Smoking Lapse. *Nicotine and Tobacco Research*. 2019. V. 21. No. 2. P. 173–179. doi: 10.1093/ntr/ntx201
30. Wang D., Wu T. T., Zhao Y. Penalized empirical likelihood for the sparse Cox regression model. *Journal of Statistical Planning and Inference*. 2019. V. 201. P. 71–85. doi: 10.1016/j.jspi.2018.12.001
31. Wu T.T., Gong H., Clarke E.M. A Transcriptome Analysis by Lasso Penalized Cox Regression for Pancreatic Cancer Survival. *Journal of Bioinformatics and Computational Biology*. 2012. V. 09. No. Supp01. P. 63–73. doi: 10.1142/S0219720011005744
32. Huang H.H., Liang Y. Hybrid L1/2+2 method for gene selection in the Cox proportional hazards model. *Comput. Methods Programs Biomed*. 2018. V. 164. P. 65–73. doi: 10.1016/j.cmpb.2018.06.004
33. Huang J., Sun T., Ying Z., Yu Y., Zhang C.H. Oracle Inequalities for the Lasso in the Cox Model. *Ann. Stat*. 2013. V. 41. No. 3. P. 1142–1165. doi: 10.1214/13-AOS1098
34. Jiang H.K., Liang Y. The L1/2 regularization network Cox model for analysis of genomic data. *Comput. Biol. Med*. 2018. V. 100. P. 203–208. doi: 10.1016/j.compbiomed.2018.07.009

35. Li Y., Dicker L., Zhao S.D. The Dantzig Selector for Censored Linear Regression Models. *Stat. Sin*. 2014. V. 24. No. 1. P. 251–2568. doi: 10.5705/ss.2011.220
36. Liu C., Liang Y., Luan X.Z., Leung K.S., Chan T.M., Xu Z.B., Zhang H. The L1/2 regularization method for variable selection in the Cox model. *Applied Soft. Computing*. 2014. V. 14. P. 498–503. doi: 10.1016/j.asoc.2013.09.006
37. Zhang H.H., Lu W. Adaptive Lasso for Cox's Proportional Hazards Model. *Biometrika*. 2007. V. 94. No. 3. P. 691–703. doi: 10.1093/biomet/asm037
38. Bradic J., Fan J., Jiang J. Regularization for Cox's proportional hazards model with NP-dimensionality. *Annals of Statistics*. 2011. V. 39. No. 6. P. 3092–3120. doi: 10.1214/11-AOS911
39. Goeman J.J. L1 penalized estimation in the Cox proportional hazards model. *Biom. J.* 2010. V. 52. No. 1. P. 70–84. doi: 10.1002/bimj.200900028
40. Tibshirani R. The lasso method for variable selection in the Cox model. *Statistics in Medicine*. 1997. V. 16, No. 4. P. 385–395. doi: 10.1002/(SICI)1097-0258(19970228)16:4<385::AID-SIM380>3.0.CO;2-3
41. Simon N., Friedman J., Hastie T., Tibshirani R. Regularization paths for Cox's proportional hazards model via coordinate descent. *Journal of Statistical Software*. 2011. V. 39. No. 5. P. 1–13. doi: 10.18637/jss. v039.i05
42. Askarzadeh A. A novel metaheuristic method for solving constrained engineering optimization problems: crow search algorithm. *Computers & Structures*. 2016. V. 169. P. 1–12. doi: 10.1016/j.compstruc.2016.03.001
43. Kawano S. Selection of tuning parameters in bridge regression models via Bayesian information criterion. *Statistical Papers*. 2014. V. 55. No. 4. P. 1207–1223. doi: 10.1007/s00362-013-0561-7
44. Algamal Z.Y., Lee M.H. Penalized logistic regression with the adaptive LASSO for gene selection in high-dimensional cancer classification. *Expert Systems with Applications*. 2015. V. 42. No. 23. P. 9326–9332. doi: 10.1016/j.eswa.2015.08.016
45. Algamal Z.Y. Shrinkage parameter selection via modified cross-validation approach for ridge regression model. *Communications in Statistics-Simulation and Computation*. 2020. V. 49. No. 7. P. 1922–1930. doi: 10.1080/03610918.2018.1508704
46. Algamal Z.Y. A new method for choosing the biasing parameter in ridge estimator for generalized linear model. *Chemometrics and Intelligent Laboratory Systems*. 2018. V. 183. P. 96–101. doi: 10.1016/j.chemolab.2018.10.014
47. Rosenwald A., Wright G., Chan W.C., Connors J.M., Campo E., Fisher R.I., Gascoyne R.D., Muller-Hermelink H.K., Smeland, E.B., Giltnane, J.M., et al. The use of molecular profiling to predict survival after chemotherapy for diffuse large-B-cell lymphoma. *New England Journal of Medicine*. 2002. V. 346. No. 25. P. 1937–1947. doi: 10.1056/NEJMoa012914
48. Van Houwelingen H.C., Bruinsma T., Hart A.A., Van't Veer L.J., Wessels L.F. Cross-validated Cox regression on microarray gene expression data. *Statistics in Medicine*. 2006. V. 25. No. 18. P. 3201–3216. doi: 10.1002/sim.2353
49. Beer D.G., Kardia S.L., Huang C.C., Giordano T.J., Levin A.M., Misek D.E., Lin L., Chen G., Gharib T.G., Thomas D.G., et al. Gene-expression profiles predict survival of patients with lung adenocarcinoma. *Nature Medicine*. 2002. V. 8. No. 8. P. 816–824. doi: 10.1038/nm733

Received 01.02.2023.
Revised 27.05.2023.
Published 20.07.2023.