



Research article

Survival prediction model for right-censored data based on improved composite quantile regression neural network

Xiwen Qin, Dongmei Yin, Xiaogang Dong*, Dongxue Chen and Shuang Zhang

School of Mathematics and Statistics, Changchun University of Technology, Changchun 130012, China

* Correspondence: Email: dongxiaogang@ccut.edu.cn.

Abstract: With the development of the field of survival analysis, statistical inference of right-censored data is of great importance for the study of medical diagnosis. In this study, a right-censored data survival prediction model based on an improved composite quantile regression neural network framework, called rcICQRNN, is proposed. It incorporates composite quantile regression with the loss function of a multi-hidden layer feedforward neural network, combined with an inverse probability weighting method for survival prediction. Meanwhile, the hyperparameters involved in the neural network are adjusted using the WOA algorithm, integer encoding and One-Hot encoding are implemented to encode the classification features, and the BWOA variable selection method for high-dimensional data is proposed. The rcICQRNN algorithm was tested on a simulated dataset and two real breast cancer datasets, and the performance of the model was evaluated by three evaluation metrics. The results show that the rcICQRNN-5 model is more suitable for analyzing simulated datasets. The One-Hot encoding of the WOA-rcICQRNN-30 model is more applicable to the NKI70 data. The model results are optimal for $k = 15$ after feature selection for the METABRIC dataset. Finally, we implemented the method for cross-dataset validation. On the whole, the Cindex results using One-Hot encoding data are more stable, making the proposed rcICQRNN prediction model flexible enough to assist in medical decision making. It has practical applications in areas such as biomedicine, insurance actuarial and financial economics.

Keywords: right-censored data; composite quantile regression; neural network; whale optimization algorithm; feature selection

1. Introduction

Prognostic modeling in medical survival analysis is used to analyze and understand disease processes, to predict the behavior of new patients in the context of available data, and to explore the interaction between disease factors. Statistical methods and machine learning-based methods are two common types of approaches to construct prognostic models [1]. Both statistical and machine learning methods can predict survival time and estimate the probability of survival. However, the former focuses more on characterizing the statistical properties of time-to-event distributions and parameter estimates by estimating survival curves, while the latter mainly predicts the occurrence of specific time-to-events by combining the power of traditional survival analysis methods with various machine learning techniques. Kaplan-Meier estimation is an asymptotic maximum of the survival function under incomplete data likelihood estimation, which provides a solid theoretical foundation for dealing with right-censored data [2]. The problem of sparse estimation of censored median regression models is an important issue in the analysis of high-dimensional survival data. A sparse and robust median estimate can be obtained by minimizing the weighted minimum absolute bias loss of the inverse censored probability under an adaptive lasso penalty [3]. Giussani and Bonetti explored multivariate survival techniques for binary right-censored data analysis and investigated a new parametric bivariate vulnerability model [4]. For consideration of the correlation between two survival times, the Marshall-Olkin bivariate exponential distribution (MOBVE) was used to model two joint distributions. Yu investigated the problem of great likelihood estimation and data fitting diagnostics associated with uniform distributions and applied it to a cancer research dataset [5].

When investigating regression models, different loss functions are usually considered. The most commonly utilized one is the squared loss, i.e., the least squares estimation. However, it is sensitive to heavy-tailed distributions and outliers and is not robust. The quantile regression model [6] can not only handle heteroskedasticity and outliers, but also construct confidence intervals for parameters using empirical likelihood inference methods without estimating the asymptotic variance. Zou and Yuan proposed the first composite quantile regression method by considering the loss of multiple quantile points simultaneously, which greatly expanded the theoreticality and stability of the estimation results [7]. In recent years, many scholars have done a lot of work based on the composite quantile regression method, such as Shim et al. proposed a new nonparametric regression method, the composite support vector quantile regression (CSVQR) [8]. Bang et al. proposed the weighted composite quantile regression (WCQR) [9], and in that same year, Bang proposed composite kernel quantile regression (CKQR) [10]. Xu et al. proposed a new composite quantile regression neural network model (CQRNN) [11]. In 2021, a weighted composite quantile regression model for linear models [12] was used for the problem of right-censored data with random missing censored indicators, and an adaptive penalty procedure was applied to discuss the variable selection problem in the model.

Neural networks have an excellent ability to handle censored data in survival analysis and can identify the importance of extremely complex interactions in the data. Therefore, it can be used to improve traditional survival analysis techniques [13]. Katzman et al. introduced DeepSurv, a state-of-the-art Cox proportional risk deep neural network survival method for modeling the interaction between patient covariates and treatment effects in order to provide personalized treatment recommendations [14]. Wang et al. proposed a novel multi-task based neural network SurvNet [15]. Anika and Olivier constructed a multimodal neural network model to develop an unsupervised encoder for personalized treatment of cancer patients [16]. Currently, there are an increasing number of models

for cancer survival prediction [17–20], and developing a model that is both accurate and interpretable is to remain a challenge.

For datasets with censored data, the selection of estimation methods is essential due to the incompleteness of the data, which happens to make excellent use of the quantile nature. In this paper, we use a composite quantile regression approach to deal with censored data. Moreover, many deep learning algorithms have been adapted to tackle such censored data and to deal with other challenging problems that arise in realistic data. In this study, a survival prediction model based on an improved composite quantile regression neural network framework is proposed. It extends the deep neural network framework to composite quantile regression, combined with an inverse probability weighting method for survival prediction. The whale optimization algorithm (WOA) is used for hyperparameter tuning, and the binary whale optimization algorithm (BWOA) is applied for variable selection of high-dimensional covariate data. Through extensive tests on simulated datasets and three real datasets of NKI70, METABRIC, and TCC, the results show that the proposed rcICQRNN method has high prediction accuracy for right-censored data.

The purpose of this study is to help healthcare professionals make sound decisions by establishing a more accurate method of predicting cancer patients. Cancer patients can benefit from accurate prognosis prediction, and physicians can benefit by making timely judgments and suggesting appropriate treatment options. For healthcare professionals, cancer prognosis analysis is an important part of their profession, especially when dealing with patients with short survival times, which need to be supported by more robust models. Deep learning can greatly facilitate cancer treatment and has the potential to bring it closer to accurate prognostic outcomes. This paper employs a state-of-the-art neural network architecture optimized for cancer survival analysis and validated on open-source and publicly accessible multiple data, and also shows that the present method can enable predictions from other domain datasets. The application of composite quantile regression captures the overall picture of the survival distribution, enhances the robustness, fit, predictive power and nonlinear processing of neural networks, and fills the gap between composite quantile regression and neural network methods in the field of survival analysis. Compared with state-of-the-art models, this paper is able to enhance the predictive power of survival analysis and facilitate the development of deep learning techniques in personalized clinical decision making.

2. Materials and methods

2.1. Materials

2.1.1. NKI70 dataset

In this paper, the proposed rcICQRNN method and DeepQuantreg [21] are compared with two real breast cancer datasets. The first is the NKI70 dataset, which is from the R package *penalized*. This dataset is described in detail on the website <https://rdrr.io/cran/penalized/man/nki70.html>. The NKI70 dataset contains 144 patients with lymph node positive breast cancer. In an earlier study, metastasis-free survival, 5 clinical risk factors, and gene expression measurements for 70 genes were found to be predictive of metastasis-free survival in 144 patients. Among these patients, 96 were censored samples, with a censoring ratio of approximately 67%. Four variables in the dataset were of type categorical: tumor diameter (two levels), number of affected lymph nodes (two levels), estrogen receptor status

(two levels), and tumor grade (three ordered levels). We implemented integer encoding and One-Hot encoding to encode categorical features, and the number of covariates after One-Hot encoding was 80. We subsequently compared the performance of integer encoding and One-Hot encoding in rcICQRNN for survival prediction of breast cancer data.

2.1.2. METABRIC dataset

The most important part of the clinical decision-making process for cancer patients is the accurate estimation of prognosis and survival. Breast cancer patients with the similar disease stage and the similar clinical features can have different treatment responses and overall survival due to the fact that cancer is associated with genetic abnormalities. The Molecular Taxonomy of Breast Cancer International Consortium (METABRIC) database is a Canadian-UK project containing targeted sequencing data from 1980 major breast cancer samples. Clinical and genomic data can be downloaded from cBioPortal. For this study, the METABRIC dataset was obtained from the Kaggle website at <https://www.kaggle.com/raghadalharbi/breast-cancer-gene-expression-profiles-metabric/download>. The METABRIC dataset included 31 clinical attributes, m-RNA level z scores for 331 genes, and 175 gene mutations in 1904 breast cancer patients. After processing missing values, the data included 1902 patients with 518 features. Again, integer encoding and One-Hot encoding were performed for 18 categorical features. Due to the large number of variables in the dataset, direct survival prediction gives poor results. Therefore, a variable selection method, called BWOA, is proposed in this paper. The number of variables after integer encoding was 433 and the number of variables after One-Hot encoding was 241, both with a censoring ratio of 66%.

2.1.3. Telco Customer Churn dataset (TCC)

To validate the performance of this study on different domain datasets, this paper analyzes the characteristics that influence telecom customer churn and the prediction of customer churn probability. The data was obtained from 7034 telecommunication customers. It consists of 50% female customers, 50% male customers, 18 characteristics, and a censoring ratio of 73%. This data set may be found at <https://www.kaggle.com/datasets/blastchar/telco-customer-churn>. The data set of ‘gender’, ‘Partner’ and other 15 non-numeric features in the data set were converted to numeric form.

The selected features corresponding to each dataset are given in Table 1.

Table 1. Reduced feature set using BWOA.

dat	Available features		BWOA features	
	Integer	One-Hot	Integer	One-Hot
NKI70	75	80	-	-
METABRIC	518	578	433	241
TCC	18	45	-	-

2.2. *rcICQRNN* method

Taylor and Cannon considered a quantile regression neural network (QRNN) method by

combining the excellent properties of neural networks and quantile regression (QR) [22,23]. In this section, the censored composite quantile regression neural network (rcICQRNN) model is proposed by adding a deep neural network and an inverse probability weighting structure to the composite quantile regression method. For random variables T , the τ^{th} quantile of T is defined as:

$$Q_T(\tau) = \inf\{t: Pr(T \leq t) \geq \tau\}, \tau \in (0,1), \quad (1)$$

$Q_T(\tau)$ is a function of τ , it can completely describe the distribution of random variable T . Given the covariate $X = (x_1, \dots, x_n)$, The τ^{th} conditional quantile of T can be defined as:

$$Q_{T|X}(\tau) = \inf\{t: Pr(T \leq t|X) \geq \tau\}, \tau \in (0,1). \quad (2)$$

The log linear quantile regression model is defined as follows:

$$Q_\tau(T|X) = \exp(X^T \beta(\tau)), \tau \in (0,1), \quad (3)$$

where $\beta(\tau) = (\beta_1(\tau), \dots, \beta_n(\tau))^T$ is the regression coefficient. Since the original test function is not defined at the origin, it cannot be differentiable everywhere. Therefore, the Huber function [24] is used to smooth the test function in this paper.

$$\begin{aligned} \rho_\tau(u) &= h(u)(\tau - I(u \leq 0)) \\ &= \begin{cases} h(u)\tau u > 0 \\ h(u)(\tau - 1) u \leq 0 \end{cases}, \\ h(u) &= \begin{cases} \frac{u^2}{\xi} & 0 \leq |u| \leq \xi \\ |u| - \frac{\xi}{2} & |u| > \xi \end{cases}, \end{aligned} \quad (4)$$

where $I(\cdot)$ is the indicator function. In traditional quantile regression, the model will be trained separately for each quantile, and the training objective is to minimize the average loss function $Q\rho_\tau$ for the τ^{th} quantile:

$$\min Q\rho_\tau = \frac{1}{n} \sum_{i=1}^n \rho_\tau(\log(t_i) - X_i^T \beta). \quad (5)$$

The composite quantile regression is extended to different levels of τ while considering the multiple quantile regression model. In general, take an equidistant $\tau_k = \frac{k}{K+1}$, $k = 1, 2, \dots, K$, and define the following objective function $CQ\rho_{\tau_k}$:

$$\min CQ\rho_{\tau_k} = \frac{1}{K} \sum_{k=1}^K Q\rho_{\tau_k} = \frac{1}{nK} \sum_{k=1}^K \sum_{i=1}^n \rho_{\tau_k}(\log(t_i) - X_i^T \beta). \quad (6)$$

Suppose a deep feedforward network with two hidden layers is considered. The overall structure is: an input layer, consisting of m input variables $x_i, i = 1, \dots, m$. In the middle are two hidden layers, consisting of n and p neural units, respectively. The final layer is the output layer. $w_{in}^{(h)}, w_{nl}^{(h)}, b_n^{(h)}, b_l^{(h)}$ is the weight and bias of the hidden layer, and $w_l^{(o)}, b^{(o)}$ is the weight and bias of the output layer. Then, the output of the first hidden layer is:

$$g_n = f_1^{(h)} \left(\sum_{i=1}^m x_i w_{in}^{(h)} + b_n^{(h)} \right) n = 1, 2, \dots, N. \quad (7)$$

The output of the second hidden layer is:

$$h_l = f_2^{(h)} \left(\sum_{n=1}^N g_n w_{nl}^{(h)} + b_l^{(h)} \right) \quad l = 1, 2, \dots, L, \quad (8)$$

where $f_1^{(h)}$ and $f_2^{(h)}$ are the activation functions of the hidden layer. Common activation functions include sigmoid, relu, etc. [21]. Finally, the output layer of the network is a single node with linear activation function. The estimation of the τ^{th} conditional quantile of the i^{th} object is Eq (9).

$$\log \hat{Q}_i^{(\tau)} = \sum_{l=1}^L h_{l,i} w_l^{(o)} + b^{(o)}. \quad (9)$$

Therefore, the objective of CQRNN is to minimize the empirical loss function as Eq (10).

$$CQRNN(\theta) = \frac{1}{K} \sum_{k=1}^K Q \rho_{\tau_k} = \frac{1}{nK} \sum_{k=1}^K \sum_{i=1}^n \rho_{\tau_k} (\log(t_i) - \log \hat{Q}_i^{(\tau)}), \quad (10)$$

where θ includes all coefficients, i.e., weights and bias, thus trained and estimated.

Now consider the data with right-censored loss and analyze mainly the time of death of the target subject, or other known points of interest that occurred over a period of time. The Cox proportional risk model is a widely used statistical method for modeling risk rates. For overcoming the complexity of risk ratio interpretation, a composite quantile regression neural network is introduced to the censored data, and the interpretation is more intuitive. First, T is defined as the failure time. T_0 is the maximum follow-up study time. X_i is the p dimension prediction variable that may affect the failure time. $C_i \leq T_0, \{C_i; i = 1, \dots, n\}$ indicates the censoring time, and the distribution of C_i depends on X_i . For the right-censored dataset $(Y_i, \delta_i, C_i), i = 1, \dots, n$, there are independent observations, where Y_i is the observed failure time, $Y_i = \min\{T_i, C_i\}$, δ_i are the censored indicator variables, i.e,

$$\delta_i = I(T_i \leq C_i) \begin{cases} 1 & \text{uncensored} \\ 0 & \text{censored} \end{cases} \quad (11)$$

Combined with the characteristics of right-censored data and assuming the conditional independence between T_i and C_i of a given x , define a weight function:

$$\omega_i = \frac{\delta_i}{\hat{G}(Y_i)}, \quad (12)$$

where $\hat{G}(Y_i)$ is the Kaplan-Meier estimation based on censored distribution of observed data $Y_i, I(\delta = 0)$. This weight function can be shown to be equivalent to a jump in the Kaplan-Meier estimator of the logarithmic distribution function of the failure time [21,25]. Similar to the ordinary censored quantile regression, in this study, the loss function of rcICQRNN is shown in Eq (13).

$$rcICQRNN(\theta) = \frac{1}{nK} \sum_{k=1}^K \sum_{i=1}^n \omega_i \cdot \rho_{\tau_k} (\log(t_i) - \log \hat{Q}_i^{(\tau)}). \quad (13)$$

In summary, the inverse censored probability weighting method is combined with the composite quantile regression neural network method considering the characteristics of the right censored data, and the neural network methods used include single hidden layer feedforward neural networks and deep feedforward networks. For the hyperparameters in the neural network, the Whale optimization algorithm in the next section is used for parameter optimization. Also, to prevent overfitting, the dropout layer is added to the neural network in this paper.

2.3. Whale optimization algorithm (WOA)

The whale optimization algorithm is a novel population intelligence optimization method proposed by Mirjalili and Lewis at Griffith University, Australia, in 2016 [26]. Its thought is derived from the unique special predatory behavior of humpback whales in the ocean, and the algorithm achieves the optimization search purpose through the process of whale encirclement and bubble attack on the prey. The algorithm mainly consists of three search mechanisms, using contraction and spiral mechanism to achieve the local search of the algorithm, and using stochastic learning strategy to achieve the global search of the algorithm. The most important feature is to simulate the hunting behavior of humpback whales with random individuals or optimal individuals, and to simulate the bubble net attack mechanism of humpback whales with spirals.

In recent years, machine learning methods have developed rapidly and have greater advantages in some aspects compared with traditional statistical methods, and the integration of the two is increasingly being studied, which is a trend in the development of statistical research at this stage. Therefore, after studying and deeply researching statistical methods and machine learning methods, this paper integrates the characteristics of different methods, combines the composite quantile regression and neural network methods, and uses WOA optimization parameters to propose a new method with higher comprehensive utility.

2.4. Binary WOA (BWOA)

In this paper, a feature selection method based on an optimization algorithm is proposed, inspired by the feature selection strategy in [27]. The process of selecting the smallest subset of survival predictions from the most informative features. A binary version of the WOA is utilized to perform this task by constructing the input feature matrix most relevant to the model.

The original WOA algorithm is a continuous algorithm, which cannot solve the binary problem directly. For solving this problem, we propose a binary version of WOA. In BWOA, set the upper bound (ub) to 1, the lower bound (lb) to 0, and the threshold to 0.5. The whale individual is initialized by $X = lb + (ub - lb) * rand$ to 1 if X is greater than 0.5, and to 0 if X is less than or equal to 0.5. In the discrete binary space, the position update implies a change between “0” and “1” are mutually changed. This study sets the following conversion function to update the whale position in binary space.

$$\vec{X}(t+1) = \begin{cases} 1 & \vec{X}(t) > 0.5 \\ 0 & \vec{X}(t) \leq 0.5 \end{cases}. \quad (14)$$

Note that the updated individual position is obtained, and it is judged whether it satisfies the upper and lower bound requirements. If it is less than the lower bound, it is made equal to the lower bound, and if it is greater than the upper bound, it is made the upper bound.

2.5. Algorithm implementation

The rcICQRNN algorithm requires defining weights and performance indicator functions, and then splitting the input data for representation. In the main function, train, test, layer, node, n_epoch, bsize, activation, optimizer, dropout, k are used as parameters. In this study, the results were compared for hidden layers of 1, 2 and 3 and k of 5, 10, 15, 19, 30 and 50. In the base model, node

is set to 50, n_epoch is set to 200, bsize is set to 64, activation is set to sigmoid, optimizer is set to Adam, and dropout is set to 0.2. The newly defined function is used as the loss function of the rcICQRNN algorithm to construct the prediction model, and the output is the result of the evaluation index MMSE, C-index, and QL.

The operating system used in this paper is Windows 10. The running software is Python version 3.8.8 and the running tool is Jupyter Notebook. Scikit-learn version is Sklearn 0.24.1. The processor is Intel(R) Core(TM) i7-10700 CPU @ 2.90 GHz 2.90 GHz and the RAM is 16.0 GB.

Due to the specificity of right-censored data, this paper proposes a survival analysis method that extends neural networks to composite quantile regression by effectively integrating the loss functions of neural networks and composite quantile regression. To optimize the hyperparameters of the survival prediction model, the WOA algorithm was used to fit the model to the training set with the best parameters and to evaluate it on the test set. The prediction results of the model proposed in this paper and the quantile regression neural network method under different quantile are compared, and the performance of the model is evaluated comprehensively by quantitative analysis of evaluation indexes to find the best prediction model. The general framework diagram of this paper is shown in Figure 1.

2.6. Performance metrics

When evaluating the prediction model, the main measure is the difference between the predicted value \hat{y}_i and the observed value y_i . In this paper, three important metrics were used to evaluate survival prediction models using the consistency index (C-index) [28], modified mean squared error (MMSE) [21], and quantile loss (QL) [21].

In the field of survival analysis, one of the most commonly used metrics to assess model prediction accuracy is the C-index, which is an extension of the area under the subject operating characteristic curve (ROC) to censored time-to-event data. The C-index represents the fraction of patient pairs that show agreement between predictions and outcomes, i.e., patients are considered to have greater predicted survival times if longer-lived patients is consistent with the outcome, as shown in Eq (15).

$$C = \frac{\sum_{i,j} \delta_i \times I(\eta_i > \eta_j) \times I(t_i < t_j)}{\sum_{i,j} \delta_i \times I(t_i < t_j)}, \quad (15)$$

where i and j are the patient pairs in the sample and $I(\cdot)$ denotes the indicator function. If $\eta_i > \eta_j$, then $I(\eta_i > \eta_j) = 1$, otherwise $I(\eta_i > \eta_j) = 0$. η_i and η_j are the predicted risks of the patient pairs i and j . Where patient pairs with greater predicted risk have shorter survival rates, indicating the agreement between the model's risk prediction and the true survival outcome. The C-index measures the overall performance of the survival prediction model within an interval [0,1]. The larger the C-index, the higher the prediction accuracy.

The MMSE is defined as the mean of the sum of squares of the residuals between the observed event times and the times predicted by the quantile estimates in the censored condition. Since no true event times are observed for the censored sample, the MMSE is calculated only at the observed event times. The MMSE can be defined as Eq (16).

$$MMSE = \frac{\sum_{i=1}^n [\delta_i \cdot \{log(t_i) - log(\hat{Q}_i^{(r)})\}]^2}{\sum_{i=1}^n \delta_i}. \quad (16)$$

QL is the defined quantile loss and is another effective way to assess the difference between the

predicted and true quantile. QL is defined as Eq (17).

$$QL = \frac{1}{n} \sum_{i=1}^n \left[w_i \cdot \rho_\tau(\log(t_i \wedge u) - \log(\hat{Q}_i^{(\tau)})) \right], \quad (17)$$

where u is a constant, slightly smaller than the true follow-up time.

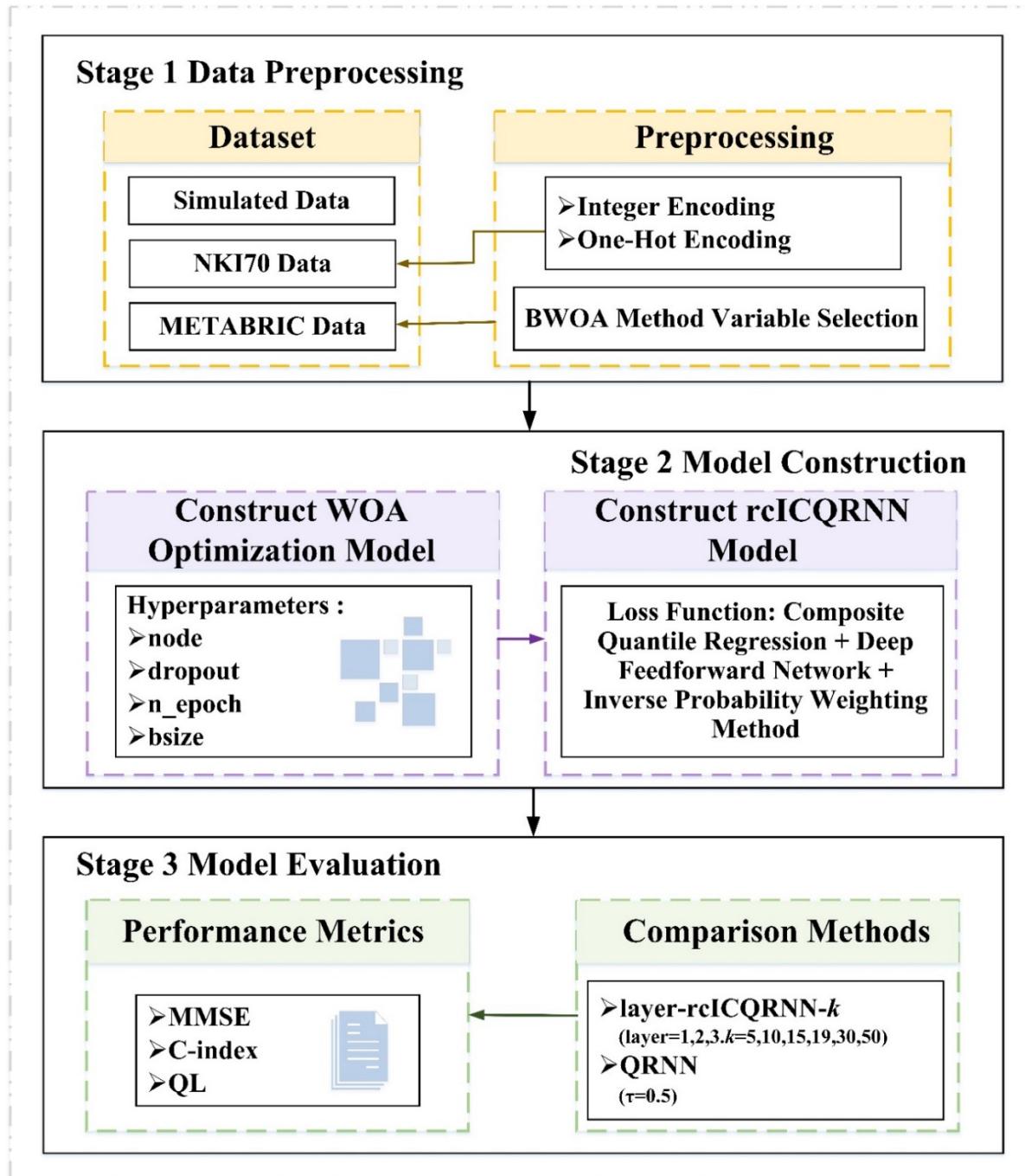


Figure 1. Framework diagram of the proposed rcICQRNN model.

3. Results and discussions

In this study, the above method was applied to multiple data sets to predict survival time. To avoid overfitting phenomenon, the dataset was divided into two parts: 75% for data model training and 25% for evaluating prediction accuracy of rcICQRNN-k, Integer-rcICQRNN-k, One-Hot-rcICQRNN-k, GS-rcICQRNN-k, WOA-rcICQRNN-k, QRNN, COX proportional hazard model.

3.1. Predictive performance on simulated datasets

In this study, numerical simulation experiments are used to evaluate the performance of the proposed estimation method on a simulated data set. The proposed inverse probability weighted composite quantile regression neural network model method is then applied to the real data. An observational data with 10 covariates was generated with 4000 training samples and 1000 test samples. Each covariate was derived from a uniform distribution and patients' survival times were generated by an exponential Cox model [29], where the log-risk function was set as a Gaussian function [14].

Table 2. Evaluation index results of all models on the simulated data set.

layer	Models	Cindex	MMSE	QL
layer = 1	rcICQRNN-5	0.6467	2.3290	0.2889
	rcICQRNN-10	0.6483	2.3111	0.2872
	rcICQRNN-15	0.6474	2.3107	0.2874
	rcICQRNN-19	0.6478	2.3122	0.2878
	rcICQRNN-30	0.6485	2.3051	0.2866
	rcICQRNN-50	0.6484	2.3218	0.2882
	QRNN	0.6489	2.3054	0.2871
layer = 2	rcICQRNN-5	0.6531	2.0199	0.2778
	rcICQRNN-10	0.6528	2.0158	0.2774
	rcICQRNN-15	0.6524	2.0275	0.2789
	rcICQRNN-19	0.6520	2.0219	0.2783
	rcICQRNN-30	0.6524	2.0196	0.2781
	rcICQRNN-50	0.6531	2.0268	0.2782
	QRNN	0.6524	2.0158	0.2778
layer = 3	rcICQRNN-5	0.6330	2.1673	0.2881
	rcICQRNN-10	0.6304	2.1695	0.2888
	rcICQRNN-15	0.6254	2.1689	0.2897
	rcICQRNN-19	0.6288	2.1777	0.2888
	rcICQRNN-30	0.6286	2.1887	0.2902
	rcICQRNN-50	0.6312	2.1690	0.2886
	QRNN	0.6308	2.1668	0.2885

Table 2 shows the prediction results of 21 survival prediction models run 1 time using different hidden layers. The best results for Cindex, MMSE and QL are shown in bold. rcICQRNN-5 denotes the composite quantile regression neural network with $k = 5$, and the rest also denotes the model at

different k values, with which the real data set in the next section is consistent. Scatter plots of actual versus predicted survival times are in the supplemental material. Among all models with a single hidden layer, the QRNN model has a slightly better Cindex and QL of 0.6489 and 0.2871, respectively. rcICQRNN-30 is only 0.0004 lower than it in terms of Cindex. While the MMSE index is optimal for the composite quantile regression neural network model when $k = 30$. Among the models combining composite quantile regression with feedforward neural networks with two hidden layers, both rcICQRNN-5 and rcICQRNN-50 show optimal properties on Cindex, 0.0003 higher than the next best model rcICQRNN-10, and significantly better than the single hidden layer model by 0.0051. The MMSE and QL of the rcICQRNN-10 model are 2.0158 and 0.2774, respectively, which are better than the rest of the models. In all models with three hidden layers, the performance is worse than that of the model with two hidden layers. In general, the rcICQRNN-5 model with two hidden layers is more suitable for analyzing the simulated data set.

3.2. Predictive performance on real datasets

Breast cancer is the deadliest type of cancer in women worldwide, and there is evidence that 50–60% of breast cancer cases are detected at a late stage, thus the survival rate of breast cancer patients is low, and early diagnosis and care will improve the survival rate of breast cancer patients [30]. In the present study, two real breast cancer datasets were selected and the proposed model was applied to the prediction of survival time. We ran the different methods for each dataset 50 times separately and took the average results. In this section, the grid search method (GS) used is the one in the article by Jia et al. [21], where the individual parameters are adjusted in the training set using 5-fold cross validation. The GS optimized parameters result in discrete values in the range, and the WOA optimized parameters result in continuous values in the range.

In Figure 2, we plot the Kaplan-Meier survival curves for the overall data. The characteristics of patient survival over time are depicted. The average age at diagnosis or disease examination in the NKI70 dataset was 44 years, and the average survival time of patients was 7.35 months. The average age at diagnosis or disease examination in the METABRIC dataset was 61 years, and the average survival time of patients was 126.55 months. 2757 clients in the TCC dataset had a duration of use of more than 40 months.

Figure 3 shows the histograms of the performance results of different models for the NKI70 dataset. Parameter results for the GS and WOA optimization models after integer encoding and One-Hot encoding are shown in the Supplementary Material. The three plots on the left side of Figure 3 show the results for the integer encoding and the right side shows the results for the One-Hot encoding. The results show that the model performance of One-Hot encoding is better than the model performance of integer encoding. The performance after WOA optimized parameters is significantly better than the model performance of GS optimized parameters. The rcICQRNN model proposed in this paper significantly outperforms the rcICQRNN with two hidden layers than the rcICQRNN with one hidden layer and three hidden layers due to other evaluation algorithms. In integer encoding, the performance of WOA optimized parameters is significantly better than that of GS optimized parameters when layer = 1 the average Cindex of the WOA-rcICQRNN-30 model is 0.8194, which is 0.0870, 0.0987, 0.0271, 0.0897, 0.0867 higher than rcICQRNN-30, QRNN, GS-rcICQRNN-30, GS-QRNN, and WOA-QRNN, respectively. The average Cindex of the WOA-rcICQRNN-50 model when layer = 2 is 0.8761, which is 0.1414, 0.1873, 0.0818, 0.1724, 0.1849 higher than rcICQRNN-50,

QRNN, GS-rcICQRNN-50, GS-QRNN, and WOA-QRNN, respectively. The QL of the WOA-rcICQRNN-50 model is also the lowest value among all models. The average Cindex of the WOA-rcICQRNN-15 model when layer = 3 is 0.8699, which is 0.1132, 0.1886, 0.0473, 0.1852, and 0.1818 higher than that of rcICQRNN-15, QRNN, GS-rcICQRNN-15, GS-QRNN, and WOA-QRNN, respectively. In addition, the MMSE of QRNN is significantly higher than the other models.

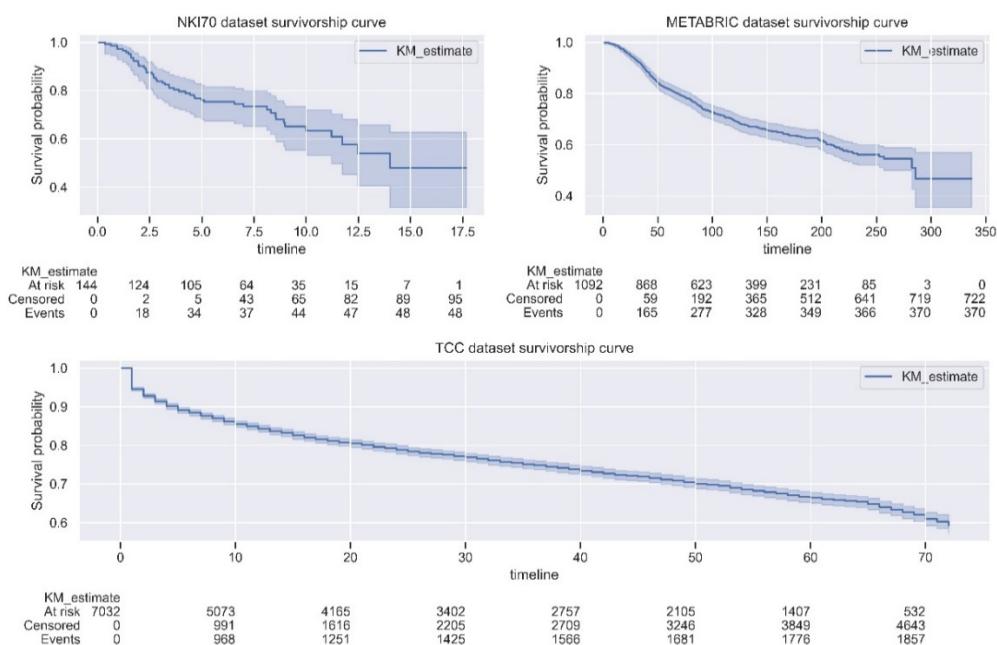


Figure 2. Survival curves for different datasets.

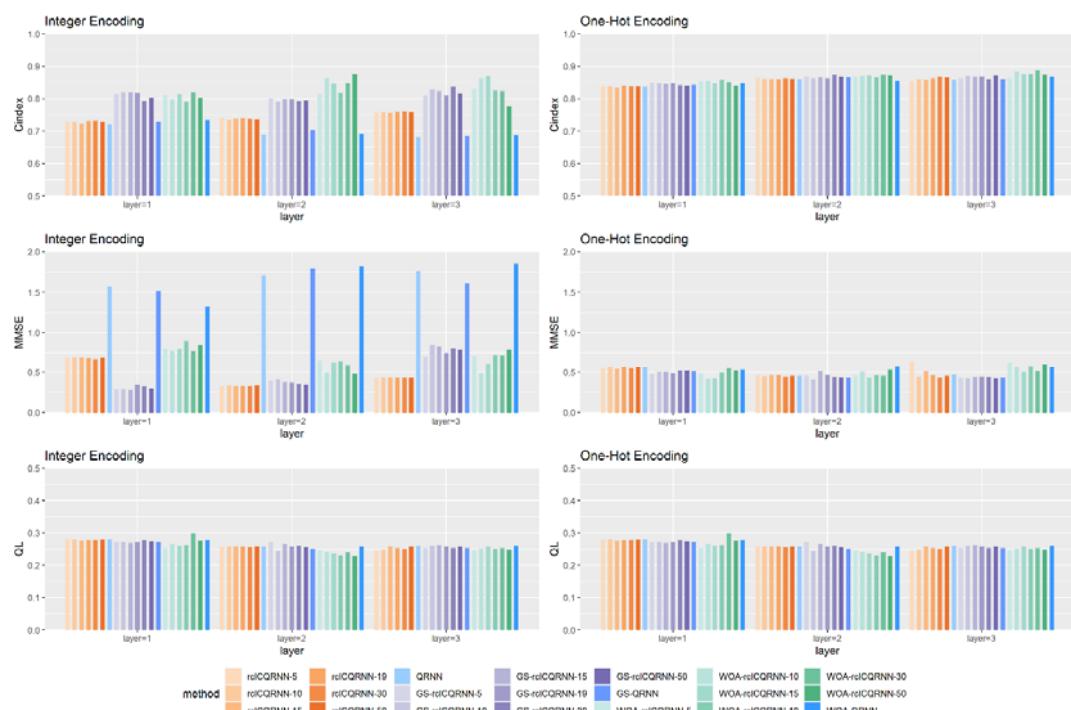


Figure 3. Performance results of different models for the NKI70 dataset.

The One-Hot encoding all had a Cindex above 0.8. If the Cindex is used as the criterion, where WOA-rcICQRNN-19 performs the best among all rcICQRNN models with a single hidden layer, the Cindex is 0.8582, the MMSE is 0.4991 and the QL is 0.2615. The WOA-rcICQRNN-30 model outperformed all rcICQRNN models in both the double hidden layer and the triple hidden layer with a Cindex of 0.8736 and an MMSE of 0.2615. The WOA-rcICQRNN-30 model also outperforms all rcICQRNN models in both the double-hidden layer and the triple-hidden layer with a Cindex of 0.8736, an MMSE of 0.4419, and a QL of 0.2588. In the triple-hidden layer, it has a Cindex of 0.8864, an MMSE of 0.4583, and a QL of 0.2534.

This paper also compares the Cindex results of 50 runs of different models to compare the prediction accuracy of the proposed algorithm with the existing state-of-the-art algorithms. Figure 4 shows the box line plots of the Cindex results. The results show that the span of the 50 times Cindex results of the model is smaller in the Integer encoding dataset. The box line plot of the QRNN model is significantly lower than the other models, and the mean and standard deviation of the dual hidden layer WOA-rcICQRNN-50 model are better than the other models, respectively.

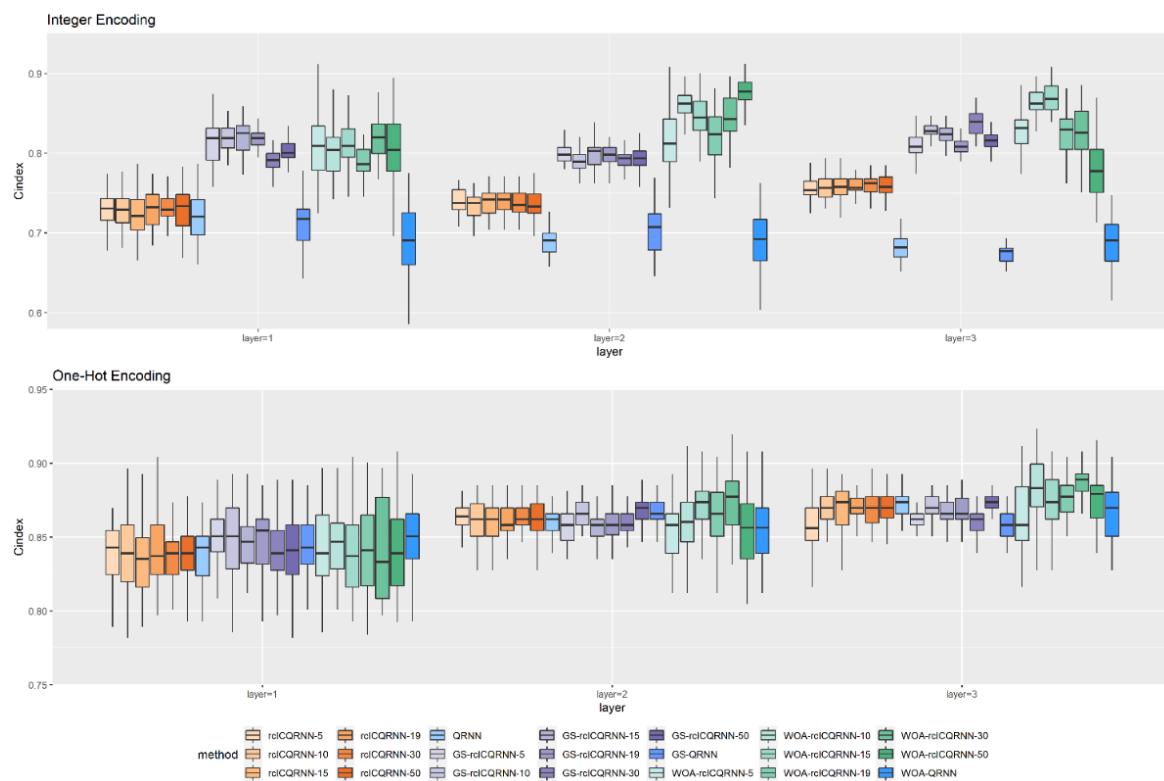


Figure 4. Cindex box plot of different models for the NKI70 dataset.

The span of the Cindex results of the single hidden layer model is larger in the One-Hot encoding dataset. Standard deviation is larger. Although the difference in the Cindex results between the different models was small, the superiority of the WOA-rcICQRNN-30 model could still be seen, further illustrating our method as the best breast cancer survival prediction model.

The survival prediction results of the METABRIC dataset before feature selection are shown in Table 3, and the values in parentheses are the standard deviations of the results of 50 runs. The results show that the Cindex of different models is around 0.6 and the MMSE is around 0.8, and the accuracy

of the prediction model is low and the error is large. Therefore, considering the dimensionality of the dataset, the BWOA algorithm was used for variable selection of the data, and variables that slightly affect survival time were eliminated.

Table 3. Original survival prediction results for the METABRIC dataset.

layer	Models	Cindex	MMSE	QL
layer = 1	Integer-rcICQRNN-5	0.6012(0.012)	0.8225(0.0427)	0.3472(0.007)
	Integer-rcICQRNN-10	0.6008(0.0115)	0.8286(0.0432)	0.3485(0.006)
	Integer-rcICQRNN-15	0.6016(0.0117)	0.8285(0.0392)	0.3488(0.0066)
	Integer-rcICQRNN-19	0.6014(0.0141)	0.8362(0.0484)	0.3473(0.0074)
	Integer-rcICQRNN-30	0.6037(0.0132)	0.8214(0.0511)	0.347(0.0067)
	Integer-rcICQRNN-50	0.6035(0.0113)	0.8235(0.0434)	0.3462(0.007)
	One-Hot-rcICQRNN-5	0.602(0.0117)	0.7909(0.0463)	0.3474(0.0062)
	One-Hot-rcICQRNN-10	0.6031(0.0127)	0.7898(0.0407)	0.3473(0.0064)
	One-Hot-rcICQRNN-15	0.6017(0.0112)	0.7968(0.0408)	0.3475(0.0061)
	One-Hot-rcICQRNN-19	0.6014(0.0124)	0.7962(0.0398)	0.348(0.0065)
layer = 2	One-Hot-rcICQRNN-30	0.6019(0.0122)	0.8079(0.0458)	0.3473(0.0052)
	One-Hot-rcICQRNN-50	0.5996(0.0106)	0.8034(0.0465)	0.3478(0.0066)
	Integer-rcICQRNN-5	0.6123(0.0093)	0.8862(0.0457)	0.323(0.0039)
	Integer-rcICQRNN-10	0.6123(0.0078)	0.8918(0.0497)	0.3227(0.0043)
	Integer-rcICQRNN-15	0.6139(0.0091)	0.3234(0.0036)	0.3234(0.0036)
	Integer-rcICQRNN-19	0.614(0.0078)	0.8839(0.0464)	0.3228(0.0039)
	Integer-rcICQRNN-30	0.6152(0.0068)	0.8817(0.0379)	0.3214(0.0036)
	Integer-rcICQRNN-50	0.6124(0.0087)	0.889(0.0469)	0.3228(0.0041)
	One-Hot-rcICQRNN-5	0.6052(0.0075)	0.8763(0.0465)	0.326(0.0041)
	One-Hot-rcICQRNN-10	0.6043(0.0085)	0.8882(0.0398)	0.3256(0.0045)
layer = 3	One-Hot-rcICQRNN-15	0.6061(0.0091)	0.8765(0.0409)	0.325(0.0045)
	One-Hot-rcICQRNN-19	0.6044(0.0074)	0.8801(0.0389)	0.3265(0.0051)
	One-Hot-rcICQRNN-30	0.6058(0.0089)	0.8838(0.045)	0.3254(0.0038)
	One-Hot-rcICQRNN-50	0.6061(0.0101)	0.8775(0.0421)	0.3258(0.0044)
	Integer-rcICQRNN-5	0.6141(0.0104)	0.8984(0.0586)	0.3222(0.0049)
	Integer-rcICQRNN-10	0.6154(0.0096)	0.9007(0.0589)	0.3214(0.0048)
	Integer-rcICQRNN-15	0.6128(0.0091)	0.9082(0.0583)	0.3233(0.005)
	Integer-rcICQRNN-19	0.6145(0.0097)	0.8987(0.056)	0.3219(0.0048)
	Integer-rcICQRNN-30	0.6139(0.01)	0.9053(0.0619)	0.3221(0.0048)
	Integer-rcICQRNN-50	0.6141(0.0085)	0.8996(0.0481)	0.3213(0.0052)

Table 4. One-Hot encoding prediction results for feature selection of METABRIC dataset.

layer	Models	Cindex	MMSE	QL
layer = 1	One-Hot-FS-GS-rcICQRNN-5	0.8546(0.0077)	0.4745(0.0205)	0.2726(0.0038)
	One-Hot-FS-GS-rcICQRNN-10	0.8524(0.0082)	0.468(0.0228)	0.2724(0.0035)
	One-Hot-FS-GS-rcICQRNN-15	0.8556(0.0075)	0.4715(0.0212)	0.2714(0.0029)
	One-Hot-FS-GS-rcICQRNN-19	0.8581(0.0092)	0.4692(0.0255)	0.2702(0.0032)
	One-Hot-FS-GS-rcICQRNN-30	0.8556(0.0077)	0.4693(0.0188)	0.2715(0.0042)
	One-Hot-FS-GS-rcICQRNN-50	0.8569(0.0066)	0.4713(0.0182)	0.2716(0.0031)
	One-Hot-FS-GS-QRNN	0.8573(0.0081)	0.4659(0.0187)	0.2702(0.0039)
	One-Hot-FS-WOA-rcICQRNN-5	0.8573(0.0081)	0.4851(0.0236)	0.2733(0.0041)
	One-Hot-FS-WOA-rcICQRNN-10	0.8562(0.0085)	0.4995(0.0263)	0.2738(0.0036)
	One-Hot-FS-WOA-rcICQRNN-15	0.8522(0.0055)	0.4949(0.0158)	0.2756(0.0028)
layer = 2	One-Hot-FS-WOA-rcICQRNN-19	0.8607(0.005)	0.4794(0.015)	0.2721(0.0026)
	One-Hot-FS-WOA-rcICQRNN-30	0.8619(0.007)	0.4745(0.016)	0.2713(0.0025)
	One-Hot-FS-WOA-rcICQRNN-50	0.8614(0.0086)	0.4847(0.0215)	0.272(0.0034)
	One-Hot-FS-WOA-QRNN	0.8573(0.0074)	0.4891(0.0162)	0.273(0.0032)
	One-Hot-FS-GS-rcICQRNN-5	0.858(0.0053)	0.4874(0.0308)	0.2605(0.0031)
	One-Hot-FS-GS-rcICQRNN-10	0.8548(0.0057)	0.5218(0.0197)	0.2628(0.0029)
	One-Hot-FS-GS-rcICQRNN-15	0.85(0.0055)	0.5731(0.0282)	0.2652(0.003)
	One-Hot-FS-GS-rcICQRNN-19	0.8562(0.0063)	0.5108(0.0215)	0.2625(0.0027)
	One-Hot-FS-GS-rcICQRNN-30	0.8524(0.0058)	0.555(0.02)	0.2641(0.0027)
	One-Hot-FS-GS-rcICQRNN-50	0.8572(0.0067)	0.5011(0.0274)	0.2605(0.0024)
layer = 3	One-Hot-FS-GS-QRNN	0.8581(0.0068)	0.5141(0.0309)	0.2633(0.0039)
	One-Hot-FS-WOA-rcICQRNN-5	0.8476(0.0049)	0.5646(0.0192)	0.2648(0.0026)
	One-Hot-FS-WOA-rcICQRNN-10	0.858(0.0065)	0.5314(0.0268)	0.262(0.004)
	One-Hot-FS-WOA-rcICQRNN-15	0.8609(0.0083)	0.5161(0.0331)	0.2604(0.0083)
	One-Hot-FS-WOA-rcICQRNN-19	0.8598(0.0083)	0.5006(0.0375)	0.2616(0.0037)
	One-Hot-FS-WOA-rcICQRNN-30	0.8591(0.0082)	0.5154(0.0271)	0.2635(0.0041)
	One-Hot-FS-WOA-rcICQRNN-50	0.8593(0.0062)	0.5215(0.023)	0.2613(0.0028)
	One-Hot-FS-WOA-QRNN	0.8512(0.0082)	0.5084(0.0241)	0.2653(0.0041)
	One-Hot-FS-GS-rcICQRNN-5	0.8507(0.005)	0.5531(0.0211)	0.2654(0.0037)
	One-Hot-FS-GS-rcICQRNN-10	0.8466(0.0052)	0.5931(0.0275)	0.2672(0.0041)
	One-Hot-FS-GS-rcICQRNN-15	0.8326(0.0064)	0.6542(0.0352)	0.2725(0.0047)
	One-Hot-FS-GS-rcICQRNN-19	0.8483(0.0063)	0.5617(0.0285)	0.2653(0.0034)
	One-Hot-FS-GS-rcICQRNN-30	0.8405(0.0061)	0.6174(0.0347)	0.2721(0.0056)
	One-Hot-FS-GS-rcICQRNN-50	0.8297(0.0055)	0.6568(0.0251)	0.2719(0.0032)
	One-Hot-FS-GS-QRNN	0.8483(0.0051)	0.5805(0.0252)	0.267(0.0052)
	One-Hot-FS-WOA-rcICQRNN-5	0.8506(0.0065)	0.5683(0.0299)	0.2661(0.0043)
	One-Hot-FS-WOA-rcICQRNN-10	0.8533(0.0065)	0.5314(0.0327)	0.2662(0.0079)
	One-Hot-FS-WOA-rcICQRNN-15	0.8544(0.0054)	0.4943(0.0209)	0.2623(0.0039)
	One-Hot-FS-WOA-rcICQRNN-19	0.85(0.0078)	0.5853(0.0372)	0.2668(0.0047)
	One-Hot-FS-WOA-rcICQRNN-30	0.8484(0.0088)	0.5562(0.0356)	0.2665(0.0044)
	One-Hot-FS-WOA-rcICQRNN-50	0.8529(0.0069)	0.5152(0.0263)	0.2631(0.004)
	One-Hot-FS-WOA-QRNN	0.8541(0.008)	0.5074(0.0346)	0.2638(0.0042)

We found a significant increase in accuracy after subjecting the data to variable selection. The same GS and WOA algorithms were chosen to optimize the parameters in order for the model to achieve more effective predictive power. Since the previous datasets show that One-Hot encoding has better results than integer encoding, although some of the models in Table 3 have better performance after integer encoding, the performance of each model after feature selection still shows that One-Hot encoding is optimal. Therefore, in Table 4, only the performance metrics of each model with One-Hot encoding are shown. The performance of each model after feature selection and the performance results of each model after integer encoding optimization are in the supplementary material.

In Table 4, when layer = 1, the average Cindex of the WOA-rcICQRNN-30 model is 0.8619, which is 0.0063, 0.0046, 0.0046 higher than GS-rcICQRNN-30, GS-QRNN, and WOA-QRNN, respectively. When layer = 2, the average Cindex of the WOA-rcICQRNN-15 model is 0.8609, which is 0.8609, 0.0028, and 0.0097 higher than GS-rcICQRNN-15, GS-QRNN, and WOA-QRNN, respectively. the QL of the WOA-rcICQRNN-15 model is also the lowest among all models lowest value. When layer = 3, again the average Cindex of the WOA-rcICQRNN-15 model is 0.8544, which is better than the other models, and is 0.0218, 0.0061, and 0.0003 higher than GS-rcICQRNN-15, GS-QRNN, and WOA-QRNN, respectively. And the MMSE and QL of the WOA-rcICQRNN-15 model were 0.4943 and 0.2623, respectively, the lowest values among all models.

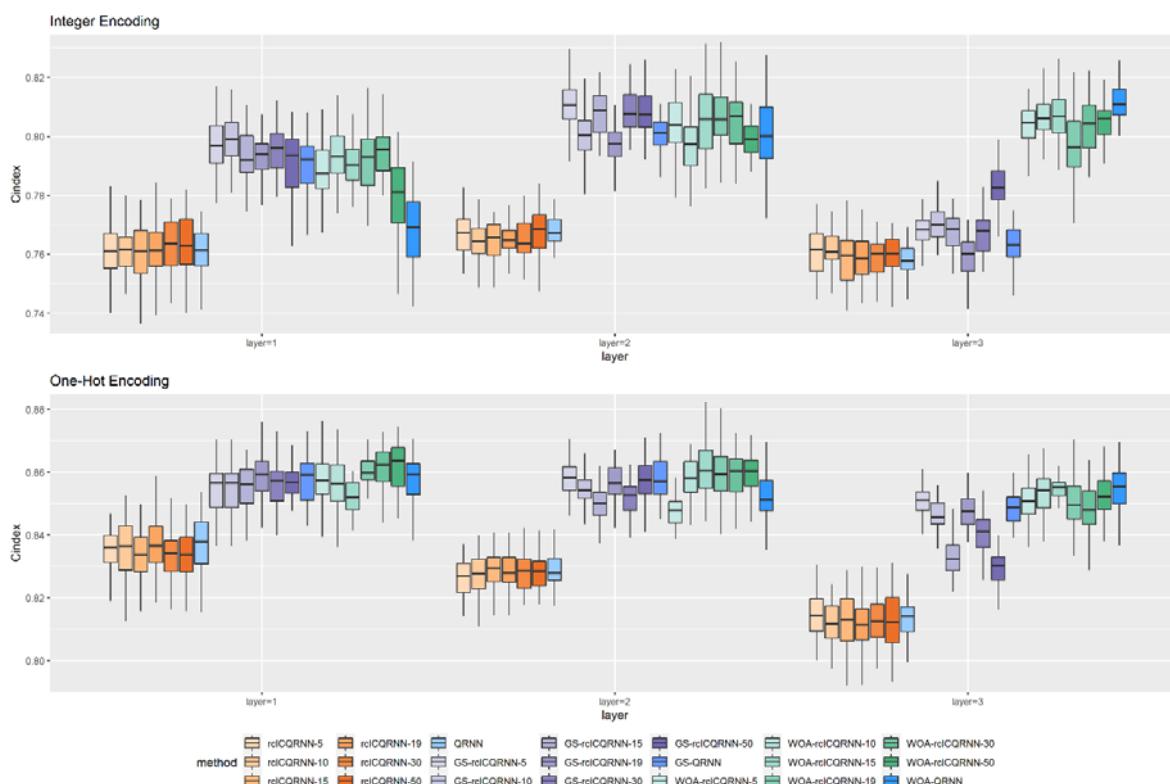


Figure 5. Cindex box plot of different models after feature selection for METABRIC dataset.

Figure 5 represents the Cindex box plots of different models after feature selection for the METABRIC dataset. Parameter results for the integer encoding and One-Hot encoding datasets after optimization of parameters by GS and WOA algorithms are in the supplementary material. It can be seen that the Cindex performance of the model for the feature-selected dataset is much improved after

the parameters are optimized using the intelligent optimization algorithm. The Cindex results for One-Hot encoding data are more stable at this time. Except for the single-hidden layer and double-hidden layer models in integer encoding, all other models optimized by WOA can show better prediction accuracy, making the proposed rcICQRNN more suitable for predicting right-censored data.

The survival prediction results for the TCC dataset are shown in Table 5. In this section, the results of the data after integer encoding and One-Hot encoding are further described. The results show that for each layer, One-Hot encoding helps to improve the prediction accuracy. When layer = 3, the rcICQRNN_19 model has a Cindex of 0.9813, MMSE of 0.1265, and QL of 0.0732 for the TCC dataset, which is better than the integer encoding model. However, the QRNN technique produced the least significant results.

Table 5. Survival prediction results after encoding the TCC dataset.

layer	Models	Integer encoding			One-Hot encoding		
		Cindex	MMSE	QL	Cindex	MMSE	QL
layer = 1	rcICQRNN-5	0.9720	0.1971	0.0834	0.9740	0.1971	0.0834
	rcICQRNN-10	0.9742	0.2342	0.0911	0.9750	0.2222	0.0852
	rcICQRNN-15	0.9753	0.2139	0.0853	0.9719	0.1685	0.0789
	rcICQRNN-19	0.9723	0.2034	0.0857	0.9732	0.2018	0.0839
	rcICQRNN-30	0.9738	0.2243	0.0855	0.9754	0.2372	0.0895
	rcICQRNN-50	0.9738	0.2243	0.0855	0.9717	0.2073	0.0876
	QRNN	0.9672	0.0941	0.0970	0.9685	0.1572	0.0852
layer = 2	rcICQRNN-5	0.9791	0.1304	0.0682	0.9806	0.1351	0.0721
	rcICQRNN-10	0.9805	0.1463	0.0717	0.979	0.1446	0.0716
	rcICQRNN-15	0.9801	0.1443	0.0697	0.9771	0.1147	0.0647
	rcICQRNN-19	0.9795	0.1227	0.0646	0.9797	0.1171	0.0643
	rcICQRNN-30	0.9780	0.1155	0.0642	0.9788	0.1501	0.0768
	rcICQRNN-50	0.9796	0.1019	0.0605	0.9804	0.1513	0.0723
	QRNN	0.9707	0.0577	0.0731	0.9722	0.0621	0.0634
layer = 3	rcICQRNN-5	0.9810	0.0981	0.0595	0.9792	0.0910	0.0626
	rcICQRNN-10	0.9802	0.0989	0.0646	0.9789	0.0961	0.0632
	rcICQRNN-15	0.9797	0.1082	0.0647	0.9793	0.0823	0.0578
	rcICQRNN-19	0.9804	0.1348	0.0684	0.9813	0.1265	0.0732
	rcICQRNN-30	0.9707	0.0823	0.0569	0.9797	0.1124	0.0659
	rcICQRNN-50	0.9801	0.0780	0.0592	0.9804	0.1039	0.0637
	QRNN	0.9694	0.0873	0.0901	0.9724	0.0597	0.0642

The obtained results were compared with QRNN, COX proportional hazard models (in terms of Cindex and MMSE) based on the same data set. A summary of the obtained results is presented in Table 6 below. It can be observed that the method proposed in this study has a higher precision and accuracy than the other techniques applied in the listed studies.

Table 6. A comparison of the results obtained from the proposed technique and other techniques.

data	Models	Cindex	MMSE
simulated	rcICQRNN-5	0.65	2.02
	QRNN	0.65	2.02
	COX	0.52	13.15
NKI70 Integer	rcICQRNN-30	0.76	0.43
	QRNN	0.72	1.57
	COX	0.80	0.86
NKI70 One-Hot	rcICQRNN-30	0.87	0.44
	QRNN	0.86	0.46
	COX	0.81	0.97
METABRIC Integer	rcICQRNN-10	0.62	0.90
	COX	0.75	0.65
METABRIC One-Hot	rcICQRNN-15	0.61	0.88
	COX	0.80	2.14
METABRIC FS Integer	rcICQRNN-50	0.77	0.74
	QRNN	0.77	0.73
	COX	0.77	1.12
METABRIC FS One-Hot	rcICQRNN-19	0.84	0.52
	QRNN	0.84	0.52
	COX	0.78	3.31
TCC Integer	rcICQRNN-5	0.98	0.10
	QRNN	0.97	0.06
	COX	0.93	1.25
TCC One-Hot	rcICQRNN-19	0.98	0.13
	QRNN	0.97	0.06
	COX	0.92	1.32

4. Conclusions

Machine learning has achieved great success in the field of predictive modeling dealing with survival data containing right censored data, taking advantage of his modeling of nonlinear relationships [1]. Currently, many approaches use the patient's survival status as the output of a neural network and analyze the relationship between features and survival time to the point of improving model predictive power [31,32]. The traditional accelerated time-to-failure (AFT) model, which directly relates the logarithm of the time-to-event to a covariate or predictor variable, has become a popular alternative. However, AFT models are usually modeled parametrically because the logarithm of time-to-event is linearly associated with predicted values with a specific error distribution. Therefore, the quantile-based approach, due to its nonparametric nature and defined from a cumulative distribution function or survival function, allows estimating the entire spectrum of quantile based on covariate values, providing the overall shape of the survival distribution, while allowing statistical inference for specific percentile of interest if necessary, actually offering some advantages over AFT-based and hazard-based models. Combined with the ability of neural networks to capture nonlinearities, more accurate predictions can be achieved.

This study focuses on right censored type data, including data with failure time as the response

variable and covariates affecting the failure time, and conducts an in-depth study for the prediction of failure time to estimate the nonparametric effects of covariates on the response variable. A hybrid algorithm of composite quantile regression and deep neural network is applied to the censored data to simulate the quantile function of failure time and to investigate the covariate effects in different quantile. In this study, a survival prediction model based on an improved composite quantile regression neural network framework combined with an inverse probability weighting method is proposed for survival prediction. WOA was subsequently used for hyperparameter tuning, and BWOA was applied for variable selection of high-dimensional covariate data. Evaluation of survival prediction models using three important metrics: C-index, MMSE, and QL. For the categorical variables in the real dataset, integer encoding and One-Hot encoding are implemented to encode the categorical features. Through extensive tests on simulated datasets and three real datasets of NKI70, METABRIC, and TCC, the results show that the proposed rcICQRNN method has high prediction accuracy for right-censored data. In the simulated dataset, the composite quantile regression deep neural network rcICQRNN-5 model with two hidden layers exhibits optimal properties on Cindex, which is significantly better than the single hidden layer model by 0.0051. Overall, the two-hidden layer rcICQRNN-5 model is more suitable for analyzing the simulated dataset. the One-Hot encoding WOA-rcICQRNN-30 model is more suitable for NKI70 data, and in the three hidden layers, the WOA-rcICQRNN-30 model has a Cindex of 0.8864, an MMSE of 0.4583, and a QL of 0.2534. The WOA-rcICQRNN-15 model is more used in the METABRIC dataset. The Cindex performance of the model for the feature-selected dataset was much improved after optimizing the parameters using an intelligent optimization algorithm, when the C-index results for the One-Hot encoding data were more stable, making the proposed rcICQRNN more suitable for predicting right-censored data and providing useful insights for better diagnosis and implementation of specific treatments for breast cancer patients.

This study combines composite quantile regression and neural network techniques to provide a new framework to enhance cancer prognostic survival analysis and prediction by fusing multidimensional features. The advantages of the proposed rcICQRNN model compared with previous methods are as follows. First, it is theoretically and numerically demonstrated that the rcICQRNN model utilizes the advantages of both composite quantile regression and feedforward networks. It can not only flexibly exploit the properties of neural networks to explore the nonlinear relationships between variables, but also can use the composite quantile regression features to improve the estimation efficiency and prediction ability. Second, it integrates the use of multiple regression quartiles. In practical applications, it always has advantages in terms of robustness and accuracy of predictions. The superiority of the rcICQRNN model is illustrated by simulation studies and three practical applications. In practice, $k = 1930$ seems to be a favorable choice for the rcICQRNN model.

Despite the contributions of this study, there are still some limitations that need to be considered in future research. We should acknowledge that rcICQRNN models, while successfully extracting potentially informative features from the data and avoiding manual intervention to generate features, come at the cost of the lack of transparency and interpretability of these models. Attention mechanisms can be extended to this model in the future, not only to rank the importance of input variable features, but also to demonstrate the dynamics of feature importance over time.

Another future work of this study is to fuse composite quantile regression neural networks with different penalty functions to develop an accurate and reliable prediction model that more effectively targets high-dimensional genomic features for survival analysis. It helps physicians to understand the impact of patients' genetic characteristics on breast cancer to effectively predict human survival.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (Grant No. 12026430), Education Department Project of Jilin Province (Grant No. JJKH20210716KJ), and Science and Technology Department Project of Jilin Province (Grant No. 20200403182SF). The authors thank the editors for their comments and suggestions.

Conflict of interest

The author declares that there is no conflict of interest to report regarding the present study.

References

1. P. Wang, Y. Li, C. K. Reddy, Machine learning for survival analysis, *ACM Comput. Surv.*, **51** (2019), 1–36. <https://doi.org/10.1145/3214306>
2. E. L. Kaplan, P. Meier, Nonparametric estimation from incomplete observations, *J. Am. Stat. Assoc.*, **53** (1958), 457–481. <https://doi.org/10.2307/2281868>
3. J. H. Shows, W. Lu, H. Z. Hao, Sparse estimation and inference for censored median regression, *J. Stat. Plann. Inference*, **140** (2010), 1903–1917. <https://doi.org/10.1016/j.jspi.2010.01.043>
4. A. Giussani, M. Bonetti, Marshall—Olkin frailty survival models for bivariate right-censored failure time data, *J. Appl. Stat.*, **46** (2019), 2945–2961. <https://doi.org/10.1080/02664763.2019.1624694>
5. Q. Yu, The MLE of the uniform distribution with right-censored data, *Lifetime Data Anal.*, **27** (2021), 1–17. <https://doi.org/10.1007/s10985-021-09528-2>
6. R. Koenker, G. W. Bassett, Regression quantiles, *Econometrica*, **46** (1978), 33–50. <https://doi.org/10.2307/1913643>
7. H. Zou, M. Yuan, Composite quantile regression and the oracle model selection theory, *Ann. Stat.*, **36** (2008), 1108–1126. <https://doi.org/10.1214/07-AOS507>
8. J. Shim, C. Hwang, K. Seok, Composite support vector quantile regression estimation, *Comput. Stat.*, **29** (2014), 1651–1665. <https://doi.org/10.1007/s00180-014-0511-4>
9. S. Bang, H. Cho, M. Jhun, Adaptive lasso penalised censored composite quantile regression, *Int. J. Data Min. Bioinf.*, **15** (2016), 22–46. <https://doi.org/10.1504/IJDMB.2016.076015>
10. S. Bang, S. H. Eo, M. Jhun, H. J. Cho, Composite kernel quantile regression, *Commun. Stat. Simul. Comput.*, **46** (2016), 2228–2240. <https://doi.org/10.1080/03610918.2015.1039133>
11. Q. Xu, K. Deng, C. Jiang, F. Sun, X. Huang, Composite quantile regression neural network with applications, *Expert Syst. Appl.*, **76** (2017), 129–139. <https://doi.org/10.1016/j.eswa.2017.01.054>
12. J. Wang, W. Jiang, F. Xu, W. Fu, Weighted composite quantile regression with censoring indicators missing at random, *Commun. Stat. Theory Methods*, **50** (2019), 1–18. <https://doi.org/10.1080/03610926.2019.1678638>
13. L. M. De, P. M. Ravdin, Survival analysis of censored data: Neural network analysis detection of complex interactions between variables, *Breast Cancer Res. Treat.*, **32** (1994), 113–118. <https://doi.org/10.1007/BF00666212>

14. J. L. Katzman, U. Shaham, A. Cloninger, J. Bates, Y. Kluger, DeepSurv: Personalized treatment recommender system using a Cox proportional hazards deep neural network, *BMC Med. Res. Method.*, **18**(2018), 24. <https://doi.org/10.1186/s12874-018-0482-1>
15. C. Anika, G. Olivier, Deep learning with multimodal representation for pancancer prongosis prediction, *Bioinformatics*, **35** (2019), i446–i454. <https://doi.org/10.1093/bioinformatics/btz342>
16. J. Wang, N. Chen, J. Guo, X. Xu, Z. Yi, SurvNet: A novel deep neural network for lung cancer survival analysis with missing values, *Front. Oncol.*, **10** (2021), 588990–588990. <https://doi.org/10.3389/FONC.2020.588990>
17. J. H. Oh, W. Choi, E. Ko, M. Kang, A. Tannenbaum, J. O. Deasy, PathCNN: Interpretable convolutional neural networks for survival prediction and pathway analysis applied to glioblastoma, *Bioinformatics*, **37** (2021), i443–i450. <https://doi.org/10.1093/BIOINFORMA/TICS/BTAB285>
18. B. Ma, G. Yan, B. Chai, X. Hou, XGBLC: An improved survival prediction model based on XGBoost, *Bioinformatics*, **38** (2021), 410–418. <https://doi.org/10.1093/bioinformatics/btab675>.
19. N. Arya, S. Saha, Multi-modal advanced deep learning architectures for breast cancer survival prediction, *Knowl. Based Syst.*, **221** (2021), 106965. <https://doi.org/10.1016/J.KNOSYS.2021.106965>
20. S. M. Zahra, M. Alexa, A two-stage modeling approach for breast cancer survivability prediction, *Int. J. Med. Inf.*, **149** (2021), 104438. <https://doi.org/10.1016/J.IJMEDINF.2021.104438>
21. Y. Jia, J. H. Jeong, Deep learning for quantile regression under right censoring: DeepQuantreg, *Comput. Stat. Data Anal.*, **165** (2022), 107323. <https://doi.org/10.1016/J.CSDA.2021.107323>
22. J. W. Taylor, A quantile regression neural network approach to estimating the conditional density of multiperiod returns, *J. Forecasting*, **19** (2000), 299–311. [https://doi.org/10.1002/1099-131X\(200007\)19:4<299::AID-FOR775>3.0.CO;2-V](https://doi.org/10.1002/1099-131X(200007)19:4<299::AID-FOR775>3.0.CO;2-V)
23. A. J. Cannon, Quantile regression neural networks: Implementation in r and application to precipitation downscaling, *Comput. Geosci.*, **37** (2011), 1277–1284. <https://doi.org/10.1016/j.cageo.2010.07.005>
24. P. J. Huber, Robust Regression: Asymptotics, Conjectures and Monte Carlo, *Ann. Stat.*, **1** (1973), 799–821. <https://doi.org/10.1214/aos/1176342503>
25. H. Jian, S. Ma, H. Xie, Least absolute deviations estimation for the accelerated failure time model, *Stat. Sin.*, **17** (2007), 1533–1548. <https://www.jstor.org/stable/24307687>
26. S. Mirjalili, A. Lewis, The whale optimization algorithm, *Adv. Eng. Software*, **95** (2016), 51–67. <https://doi.org/10.1016/j.advengsoft.2016.01.008>
27. W. Zheng, X. Peng, D. Lu, D. Zhang, Y. Liu, Z. Lin, et al, Composite quantile regression extreme learning machine with feature selection for short-term wind speed forecasting: A new approach, *Energy Convers. Manage.*, **151** (2017), 737–752. <https://doi.org/10.1016/j.enconman.2017.09.029>
28. F. E. Harrell, K. L. Lee, D. B. Mark, Multivariable prognostic models: Issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors, *Stat. Med.*, **15** (1996), 361–687. [https://doi.org/10.1002/\(SICI\)1097-0258\(19960229\)15:4<361::AID-SIM168>3.0.CO;2-4](https://doi.org/10.1002/(SICI)1097-0258(19960229)15:4<361::AID-SIM168>3.0.CO;2-4)
29. P. C. Austin, Generating survival times to simulate Cox proportional hazards models with time-varying covariates, *Stat. Med.*, **31**(2012), 3946–3958. <https://doi.org/10.1002/sim.5452>

30. T. Hanaa, A. Mostafa, E. Nawal, S. Hanaa, A novel deep autoencoder based survival analysis approach for microarray dataset, *PeerJ Comput. Sci.*, **7** (2021), e492–e492. <https://doi.org/10.7717/PEERJ-CS.492>
31. E. Biganzoli, P. Boracchi, L. Mariani, E. Marubini, Feed forward neural networks for the analysis of censored survival data: A partial logistic regression approach, *Stat. Med.*, **17** (1998), 1169–1186. [https://doi.org/10.1002/\(SICI\)1097-0258\(19980530\)17:10<1169::AID-SIM796>3.0.CO;2-D](https://doi.org/10.1002/(SICI)1097-0258(19980530)17:10<1169::AID-SIM796>3.0.CO;2-D)
32. P. J. G. Lisboa, H. Wong, P. Harris, R. Swindell, A Bayesian neural network approach for modelling censored data with an application to prognosis after surgery for breast cancer, *Artif. Intell. Med.*, **28** (2003), 1–25. [https://doi.org/10.1016/S0933-3657\(03\)00033-2](https://doi.org/10.1016/S0933-3657(03)00033-2)



AIMS Press

©2022 the Author(s), licensee AIMS Press. This is an open access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0>)