Expert Systems With Applications 227 (2023) 120218

Contents lists available at ScienceDirect

Expert Systems With Applications

journal homepage: www.elsevier.com/locate/eswa

CoxNAM: An interpretable deep survival analysis model

Liangchen Xu, Chonghui Guo [*]

Institute of Systems Engineering, Dalian University of Technology, Dalian 116024, China

ARTICLE INFO

Keywords:
Survival analysis
Neural additive model
Interpretable AI
Cox proportion hazards model

ABSTRACT

Survival analysis is widely used in medicine, engineering, economics and other fields as an effective method to model the relation between the time of an event of interest occurring and related features. However, traditional survival analysis models lack the ability to capture nonlinearity. In addition, most nonlinear survival analysis models, especially deep learning-based methods, lack interpretability, which limits the practical application of these models. For these gaps, we proposed an interpretable deep survival analysis model named CoxNAM. This model is based on the Cox proportion hazards model and uses neural additive model to predict the hazard function. We also used the backpropagation algorithm to train the model based on the corresponding loss function. When performing a survival analysis, we can obtain the survival functions, shape functions of features, and the importance of related features while predicting the probability of the occurrence of the event of interest. We conducted numerical experiments on two synthetic datasets and one public breast cancer dataset to verify the performance of the model, at the same time, we compared the interpretability with the SHAP framework on the two synthetic datasets and the results demonstrated the effectiveness of the proposed model's interpretation. We also applied the model for prognostic analysis of gastric cancer patients to illustrate its application. The experimental results indicate that the proposed model performs better on C-index than the classic statistical survival analysis model (i.e., Cox proportional hazards model) and machine learning-based survival analysis models (i.e., random survival forest and DeepSurv), and it can also provide the importance of features related to the time of the occurrence of events of interest and the effect of the feature values on the results. The proposed method shows promising performance and realistic interpretability. The model can potentially be extended to survival analysis problems in multiple domains for relevant decision-making.

1. Introduction

Survival analysis has been applied in many fields, such as medical, engineering, and economics, to model the time until an event of interest occurs. The objective of survival analysis is to model the time of the distribution of events of interest as a continuous function of time. The survival function and hazard function are two fundamental functions utilized in survival analysis, which are described in Eq. (1) and Eq. (2).

Survival function $S\left(t\,|\boldsymbol{x}_i\right)$: the probability of surviving observation i with feature vector $\boldsymbol{x}_i$ up to time t, i.e.,

$$S\left(t\,|\boldsymbol{x}_i\right) = Pr\left(T > t\,|\; \boldsymbol{x}_i\right), \tag{1}$$

where T represents time of the event of interest occurring.

Hazard function $\lambda\left(t\,|\boldsymbol{x}_i\right)$: the probability of an event of interest occurring at an extra infinitesimal amount of time δ given that no event occurred before time t, i.e.,

$$\lambda\left(t\,|\boldsymbol{x}_i\right) = \lim_{\delta\to 0}\frac{Pr\left(t \le T < t+\delta\,|T \ge t, \boldsymbol{x}_i\;\right)}{\delta}. \tag{2}$$

Many survival analysis models have been developed and used for survival analysis, which can be roughly divided into two categories (Wang, Li & Reddy, 2019): statistical methods and machine learning-based methods.

Statistical survival analysis methods can be subdivided into three categories depending on the assumptions made and the way parameters are used in the models: (i) nonparametric methods, (ii) parametric methods, and (iii) semiparametric methods. Regarding nonparametric methods, the survival function can be obtained by the Kaplan–Meier (KM) method (Kaplan & Meier, 1958), Nelson–Aalen (NA) method (Andersen, Borgan, Gill, & Keiding, 2012), or Life-Table (LT) method (Cutler & Ederer, 1958). Nonparametric methods are more efficient when no suitable theoretical distributions are known, but they are difficult to interpret and may yield inaccurate estimates. The parametric methods are more efficient when the time to the event of interest follows a specific distribution that can be specified in terms of some parameters. Parametric methods mainly include accelerated failure time (AFT) (Wei, 1992). When the distribution assumption is violated, which

[*] Corresponding author.
E-mail addresses: xuliangchen@yeah.net (L. Xu), dlutguo@dlut.edu.cn (C. Guo).

https://doi.org/10.1016/j.eswa.2023.120218
Received 7 September 2022; Received in revised form 17 April 2023; Accepted 17 April 2023
Available online 29 April 2023

is common in practical applications, the parametric methods may be inconsistent and can yield suboptimal results. The semiparametric methods do not require knowledge of the underlying distribution of survival times, which are widely used in survival analysis. Of available semiparametric methods, the Cox proportional hazards (CPH) model (Cox, 1972) is the most popular one and has been widely used in many fields (Grant, Hickey, & Head, 2019). Many expansion studies based on the Cox proportional hazards model have attracted the attention of many scholars (Tibshirani, 1997). The main problem with vanilla form of these CPH-based models is the assumption of linear relation between related features and event time. Although adding extra terms allows these models to capture nonlinearity. However, these terms are often added ad hoc and rely heavily on the expertise of the users.

One solution for this gap is to transform the survival analysis problem into a binary classification problem commonly encountered in machine learning and solve it with advanced nonlinear machine learning models, and many studies (Sim et al., 2020; Wang, Wang et al., 2019; de Lima Lemos, Silva, & Tabak, 2022) have used advanced machine learning techniques to simply treat the survival analysis problem as a binary classification problem. Moncada-Torres, van Maaren, Hendriks, Siesling, and Geleijnse (2021) used machine learning techniques, including extreme gradient boosting, to predict survival. Hu et al. (2022) classified the multiomics data of gastric cancer using a deep feature selection method, which is a binary classification method that aims to select related features for survival analysis. Gao et al. (2020) presented a mortality risk prediction ensemble model for COVID-19 (MRPMC) based on four machine learning methods. Smith and Alvarez (2021) also used some popular machine learning algorithms to predict the mortality of COVID-19 patients and used Shapley values to identify mortality factors. Dolatsara, Chen, Evans, Gupta, and Megahed (2020) proposed a two-stage machine learning framework to predict heart transplantation survival probabilities over time with a monotonic probability constraint, which uses multiple models to predict the mortality risk over multiple time periods. Sim et al. (2020) also used a hybrid data mining approach to predict breast cancer survival in three different time periods. However, the large number of time periods limits the application of these models. Although the binary classification can provide predictions of a specific time point, it may lose the interpretability and flexibility provided by modeling the event probabilities as a function of time. In addition, some observations are not followed completely to their event time in survival analysis data, and this phenomenon is referred to as censored data. Since converting survival analysis to a binary classification problem requires removal of censored data, such an operation may lose some information contained in the censored data, which may also provide an important reference for clinical decision-making. These censored data also contain information that provides a lower bound on the event time (Vale-Silva & Rohr, 2021) that at least at a certain point the event has not yet occurred, but the binary classification model ignores it. However, the survival analysis models can handle these censored data.

Another solution is to combine the machine learning model with the survival analysis model, which will take into account the characteristics of both models. A number of models have been developed to relax the assumptions of linear relation accepted in the CPH model. An important class of these nonlinear survival analysis models is deep survival analysis models, such as DeepSurv (Katzman et al., 2018), DeepHit (Lee, Zame, Yoon, & Van Der Schaar, 2018), Dynamic-DeepHit (Lee, Yoon, & Van Der Schaar, 2019), and Deep Survival Machines (Nagpal, Li, & Dubrawski, 2021). Béjar, Pérez, Vilalta, Álvarez-Napagao, and Garcia-Gasulla (2022) use the nonlinear survival analysis algorithm DeepSurv to predict the sick leave duration and obtain better performance. Nezhad, Sadati, Yang, and Zhu (2019) proposed a survival analysis approach based on deep learning and active learning for precision treatment recommendations. Most of these models use a deep neural network to model interactions between an observation's features and the time of an event of interest occurring to obtain a relative risk function (Kvamme, Borgan, & Scheel, 2019). Another important nonlinear model is called random survival forests (RSF) (Ishwaran, Kogalur, Blackstone, & Lauer, 2008). RSF is a type of random forest method with new survival splitting rules for the analysis of right-censored survival data. These models can all capture the nonlinear relation between features and targets. With the application of advanced machine learning techniques, such as deep learning, to survival analysis (Behrad & Abadeh, 2022), the model performance has been improved, but the interpretability has been lost compared to traditional survival analysis models because many powerful and efficient machine learning models, especially deep learning models, are treated as black-box models (Pouyanfar et al., 2018). This term indicates that we do not know the mechanism by which the model provides the corresponding results. We have no way to understand why we obtained the corresponding results, which limits the application of the model in practice.

Therefore, how the machine learning models obtain the corresponding results has received more attention (Naeem, Alshammari, & Ullah, 2022; Ullah, Moon, Naeem, & Jabbar, 2022). There are roughly two types of methods, interpretable machine learning methods and explainable machine learning methods. Interpretable machine learning methods can directly provide insights into how predictive results are obtained, for example, decision tree and logistic regression. Explainable machine learning methods cannot directly provide interpretability by themselves and rely on other methods to explain, for example, SHAP methods (Austin, 2012). The SHAP framework provides explanation methods for many unexplainable machine learning models. SHAP is model agnostic and many studies have used it to explain the results (Chen et al., 2021; Janssen et al., 2022). However, the SHAP framework belongs to the pos-hoc explainable methods. Post-hoc explanation may not truly reflect the real situation at the time the model predicted the outcome.

For these gaps, we aim to develop a intrinsically interpretable model that directly gives the details of the model's predictions, so we proposed an interpretable deep survival analysis model. In our work, we followed the idea of combining machine learning methodology with the survival analysis model and extend the CPH model with neural additive models (NAM) (Agarwal et al., 2021), which is a type of interpretable machine learning method with neural nets. Therefore, we named the proposed model CoxNAM. The fusion of NAM and the CPH model offers two advantages: the fused model can capture the nonlinear relation because NAM is a nonlinear model and the shape functions of the features can be obtained to explain the importance of the features. The architecture of NAM is based on GAM. GAM can be written as Eq. (3).

$$g\left(\boldsymbol{E}\left(y\left(\boldsymbol{x}\right)\right)\right) \;=\; g_1\left(\boldsymbol{x}^1\right)+g_2\left(\boldsymbol{x}^2\right)+\cdots+g_M\left(\boldsymbol{x}^M\right), \tag{3}$$

where $\boldsymbol{x}=\left(\boldsymbol{x}^1,\ \boldsymbol{x}^2,\ldots,\boldsymbol{x}^M\right)^T$ is the feature vector related to the event, y is the target variable, $g_i(\bullet)$ is a univariate shape function, and $g(\bullet)$ is a link function relating y to the related features.

As an extension of GAM, the shape functions of NAM are neural networks. Fig. 1 shows that NAM consists of M subnetworks to process the M-dimensional dataset $\boldsymbol{x}$, and a single feature is fed to one subnetwork. We can use backpropagation to jointly train subnetworks and learn corresponding shape functions. The values of subnetworks $g_m\left(\boldsymbol{x}^m\right), m=1,2,\ldots,M$, can be regarded as impacts of the corresponding features. Based on this information, we can obtain important influencing factors to explain the predicted results.

CoxNAM can be regarded as an extended CPH model by applying NAM instead of the linear combination of features and is one of interpretable methods. Each feature is fed into a neural network, and the results are then summed to obtain the hazard function value, which is brought into CPH to obtain the final result. This method has the characteristics of both CPH and NAM, and can measuring the importance of related features while modeling the nonlinear hazard function.

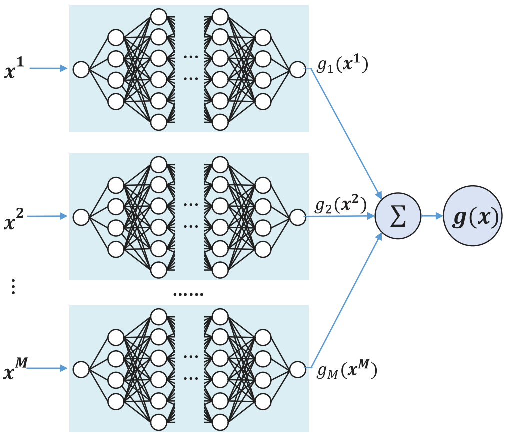


Fig. 1. An illustration of NAM.

A few models have also addressed nonlinear and interpretable problems from the same perspective, such as SurvNAM (Kvamme et al., 2019) and SurvLIME (Kovalev, Utkin, & Kasimov, 2020). SurvNAM and SurvLIME all adopt the idea of approximating the black-box model to pursue interpretability. Our model may be similar to SurvNAM in structure. However, SurNAM is an interpretable model learned by fitting RSF, which explains the black box model. Based on the survival analysis data, we used the loss function to directly train the model on the dataset so that useful information can be directly learned from the data, and important influencing factors can be given at the same time as the survival analysis.

The remainder of this paper is organized as follows. Section 2 introduces the details of the proposed CoxNAM model and the datasets used in the research and their processing. Section 3 presents the numerical results of evaluating the proposed CoxNAM model and other compared models on two synthetic datasets and two public real-world datasets. Section 4 presents the discussion about the proposed method. We conclude the paper and suggest future research in Section 5.

2. Methodology and materials

2.1. An overview of the proposed CoxNAM framework

Suppose that there is an available survival analysis dataset $D = \{\boldsymbol{x}_i, T_i, E_i\}_{i=1}^{N}$, which includes N samples. Each sample includes M features (that is, $\boldsymbol{x}_i \in \boldsymbol{R}^M$), one time of the event of interest occurring variable T_i, and one indicator variable $E_i \in \{0, 1\}$. When $E_i = 1$, it indicates that the event of interest was observed (the uncensored observation). When $E_i = 0$, it indicates that the event of interest was not observed (the censored observation). Note that we only focus on right-censored data, where the survival time is less than or equal to the true survival time.

The basic idea of survival analysis is to characterize the relation function $\boldsymbol{\Phi}$ between the features of the data $\boldsymbol{x}_i$ and the time to the event of interest occurring T_i when the data can observe the outcome ($E_i = 1$), that is, $T_i = \boldsymbol{\Phi}\left(\boldsymbol{x}_i \,|E_i = 1\right)$. For better survival analysis and interpretability of the results, we proposed the CoxNAM. CoxNAM is an extension of the CPH model. For relation function $\boldsymbol{\Phi}$, we replaced the linear regression of the predicted risk function in CPH model with the predicted value of the NAM model. We also designed the corresponding loss function and trained the model based on the backpropagation algorithm. The model has the functions of survival analysis and important factors analysis. For a new example with features $\boldsymbol{x}_{new}$, the model with input $\boldsymbol{x}_{new}$ produces the corresponding output $S\left(t\,|\boldsymbol{x}_{new}\right) = Pr\left\{\left[\boldsymbol{\Phi}\left(\boldsymbol{x}_{new}\right) > t\right] |\boldsymbol{x}_{new}\right\}$, and we can also calculate the importance of the features $Importance_{x^i}$ and the influence of the corresponding feature value according to features' shape functions. Fig. 2 presents a schematic illustration of the CoxNAM. The left part of the scheme illustrates the input data, the middle part of the scheme illustrates the CoxNAM model, and the right part of the scheme illustrates the output, which includes survival functions, feature importance and shape functions. The CoxNAM is a configurable interpretable deep survival framework, and we can use different NAM networks by setting different hyperparameters of the network.

2.2. Construction of CoxNAM

As shown in Fig. 2, CoxNAM is roughly divided into two parts: the NAM portion and the CPH model frame portion. The NAM portion is used to predict the relative hazard value, and the CPH model frame portion is used to obtain the final hazard value. Therefore, we can first build a NAM network model based on the survival data, then build a CPH framework based on the prediction results of NAM, and obtain the final hazard value. The specific process is outlines as Algorithm 1.

Algorithm 1 The algorithm for construction of CoxNAM

```
Input: survival data features set $\boldsymbol{x}$, the number of features $M$, the
    number of layers of the neural network hidden_size, dropout rate
    dropout_rate, feature dropout feature_dropout.
Output: hazard function $\lambda(t|\boldsymbol{x})$, importance of features
    {# Build a NAM model based on $\boldsymbol{x}$ and $M$}
for i in range($M$) do
    Build a network $DNN_i(hidden\_size, dropout\_rate, feature\_dropout)$

    return $DNN_i$
end for
for $\boldsymbol{x}^i$ in $\boldsymbol{x}$ do
    Calculate the shape function of features $S_i = DNN_i\left(\boldsymbol{x}^i\right)$
    return $S_i$
end for
Calculate the log-risk function $\hat{h}(\boldsymbol{x}) = \sum_{i=1}^{M} S_i$
return $\hat{h}(\boldsymbol{x}), S_i$
    {# Build the CPH framework to compute the final hazard value
    based on $\hat{h}(\boldsymbol{x})$}
Calculate the final hazard value $\lambda(\boldsymbol{x}) = h_0(t)exp\left(\hat{h}(\boldsymbol{x})\right)$
return $\lambda(\boldsymbol{x})$
    {#Calculate the importance of features}
Calculate the importance of the features and the influence of the
    feature value based to the shape function of the feature $S_i$
```

2.3. The estimation of model parameters

After the model was built, the model parameters need to be solved. The parameter solution of CoxNAM is divided into two steps: parametric fitting and nonparametric baseline hazard function $h_0(t)$ estimation. We can separately estimate these two components of the model. First, the NAM model parameters of the hazard function was solved by minimizing the negative log partial likelihood. Second, the nonparametric baseline hazard function $h_0(t)$ was estimated based on the parametric results.

2.3.1. Parametric parts fitting

According to the assumption of the hazard function of the CPH, we can abstract hazard function into a general form ignoring the linear regression assumption, as shown in Eq. (4).

$$\lambda\left(t\,|\boldsymbol{x}_i\right) = h_0(t)exp\left(h\left(\boldsymbol{x}_i\right)\right), \tag{4}$$

where $exp\left(h\left(\boldsymbol{x}_i\right)\right)$ is the relative hazard and the features that influence it, and $h\left(\boldsymbol{x}_i\right)$ is the log-risk function that we need to build to provide accurate log-risk prediction.

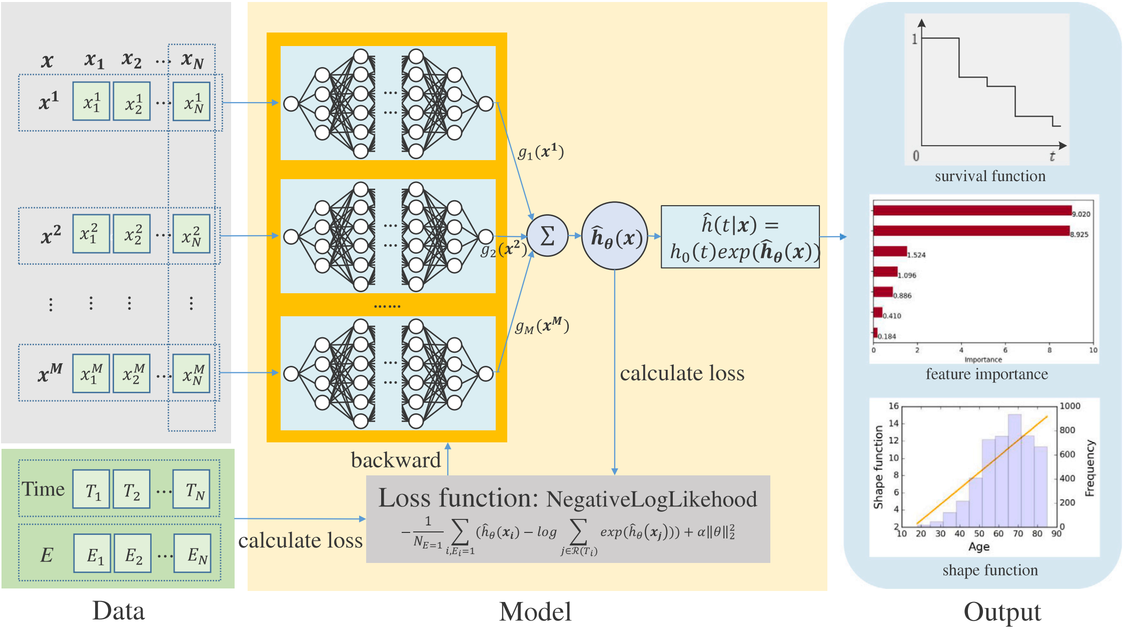


Fig. 2. A schematic illustration of the CoxNAM.

As shown in Eq. (4), the specific form of the $h(x_i)$ is $x_i\beta$, which is a linear function. Since a linear function cannot describe the nonlinear relation between the features and the target, we will replace it with a nonlinear function. In this research, we chose an interpretable nonlinear model called NAM as the log-risk function. We input the features x_i into NAM and obtain the corresponding log-risk value under the NAM parameters θ. Therefore we need to design a loss function to evaluate our model parameters and obtain the optimal model parameters θ^*. θ is the unknown parameters in the log-risk function $h(x_i)$ that we can use the maximum likelihood method to estimate.

For parameter estimation of the CPH model, we can optimize the Cox partial likelihood by tuning the weights β. The partial likelihood is the product of the probability at each event time T_i that the event has occurred to individual i given the set of individuals still at risk at time T_i. The Cox partial likelihood is defined as Eq. (5).

$$l(\beta) = \prod_{i=1,E=1}^{N} \frac{\hat{\lambda}(t|x_i)}{\sum_{j\in\mathcal{R}(T_i)} \hat{\lambda}(t|x_j)} = \prod_{i=1,E=1}^{N} \frac{exp(\hat{h}_\beta(x_i))}{\sum_{j\in\mathcal{R}(T_i)} exp(\hat{h}_\beta(x_j))}, \tag{5}$$

where $\mathcal{R}(T_i)$ is the set of individuals still at risk at time T_i.

Usually, the loss function of a machine learning model will have an additional regularization term to constrain the parameters of the model, mainly including two types: $\mathcal{L}_1$ or $\mathcal{L}_2$ regularization. $\mathcal{L}_1$ regularization is modeled with Lasso regression, which will generate a sparse weight matrix, so it is suitable for feature selection; $\mathcal{L}_2$ regularization uses Ridge regression modeling to prevent model overfitting (Yang, Lim, Qu, Li, & Ni, 2023). We choose $\mathcal{L}_2$ regularization to prevent model overfitting. We can train the NAM network by setting the average negative log partial likelihood of Eq. (5) with $\mathcal{L}_2$ regularization as the loss function, as shown in Eq. (6).

$$l(\theta) = -\frac{1}{N_{E=1}} \sum_{i,\ E_i=1} \left(\hat{h}_\theta(x_i) - log \sum_{j\in\mathcal{R}(T_i)} exp(\hat{h}_\theta(x_j))\right) + \alpha\|\theta\|_2^2, \tag{6}$$

where $N_{E=1}$ is the number of uncensored observations and α is the $\mathcal{L}_2$ regularization parameter.

Using the loss function $l(\theta)$ to train the NAM model through the backpropagation algorithm, the optimal parameters of the parameter part θ^* can be obtained.

2.3.2. Nonparametric baseline hazard function estimation

We can use the maximum likelihood method derived by Kalbfleisch and Prentice (1973) to estimate the baseline hazard function $h_0(t)$. Suppose that there are r distinct death times. When arranged in increasing order, these times are $t_1 < t_2 < \cdots < t_r$. The estimated baseline hazard function at time t_j is given as Eq. (7).

$$\hat{h}_0(t_j) = 1 - \hat{\xi}_j, \tag{7}$$

where $\hat{\xi}_j$ is the solution of Eq. (8).

$$\sum_{l\in D(t_j)} \frac{exp(\hat{h}_\theta(x_l))}{1 - \hat{\xi}_j^{exp(\hat{h}_\theta(x_l))}} = \sum_{l\in\mathcal{R}(t_j)} exp(\hat{h}_\theta(x_l)), \tag{8}$$

where $D(t_j)$ is the set of observations that die at the jth ordered death time t_j, $\mathcal{R}(t_j)$ is the set of observations that are at risk at time t_j, and $\hat{h}_\theta(x)$ is the predicted value of the hazard function of the observation with feature vector x.

Eq. (8) can be solved to yield the following Eq. (9).

$$\hat{\xi}_j = \left(1 - \frac{exp(\hat{h}_\theta(x_j))}{\sum_{l\in\mathcal{R}(t_j)} exp(\hat{h}_\theta(x_l))}\right)^{exp(-\hat{h}_\theta(x_j))}. \tag{9}$$

The quantity $\hat{\xi}_j$ can be regarded as an estimate of the probability that an individual survives through the interval from t_j to t_{j+1}. The baseline survival function can then be estimated as Eq. (10).

$$\hat{S}_0(t) = \prod_{j=1}^{k} \hat{\xi}_j, \tag{10}$$

for $t_k \le t < t_{k+1}, k = 1, 2, \ldots, r-1$. The number of intervals is r. The survival function is a step function, and the estimated value of the baseline survival function is unity for $t < t_1$ and zero for $t > t_r$.

2.4. The calculation of feature importance

The shape function of a feature indicates the degree of its influence on the predicted results. The shape function is taken as an absolute value, and the positive and negative effects are eliminated. Then, the average is weighted to represent the average importance of the feature. Therefore, we use the weighted average of one feature's values of its shape function to define its importance. A particular feature x^i has J_i

different values V_{ij}, and continuous variables can also be regarded as including a finite number of values. Each V_{ij} corresponds to a specific shape function value S_{ij}. The specific calculation formula is noted as Eq. (11).

$$Importance_{\boldsymbol{x}^i} = \frac{\sum_{j=1}^{J_i} N\left(V_{ij}\right) \bullet \left|S_{ij}\right|}{\sum_{j=1}^{J_i} N\left(V_{ij}\right)}, \tag{11}$$

where $N\left(V_{ij}\right)$ represents the number of the jth value of the ith feature.

2.5. Datasets and preprocessing

We used four datasets, including two synthetic datasets and two public real-world datasets, to evaluate the CoxNAM model. Using the synthetic datasets, we can design the form of the hazard function in advance and further test the fitting ability of the model to the hazard function according to the results. One synthetic dataset has a linear hazard function, and the other has a nonlinear hazard function. Using a public real-world dataset, namely Rotterdam & German Breast Cancer Study Group (Katzman et al., 2018), we can test the performance of the model on real data and compare it with other models. To further verify the effect of the model in real-world applications, we select a large-scale real-world gastric cancer dataset for verification.

2.5.1. Synthetic datasets

Two scenarios, including linear and nonlinear hazard functions, are used to study the CoxNAM on synthetic data. The objective is to evaluate the performance of the models when we perform two controlled experiments. In addition, since we can control the relation between features and targets in advance, we can compare the performance of our method and other methods in the interpretation of results in this section. We choose the most popular SHAP frameworks for comparison. The interpretation results of the proposed model are compared with those of DeepSurv+SHAP to verify which interpretation method is more in line with the previous setting.

We generated 6000 observations according to the exponential CPH model. Each observation $\boldsymbol{x}$ with $d = 10$ covariates, and each variable was randomly generated according to a uniform distribution on [−1,1). Then, we generated the time of occurrence T as a function of their covariates using the exponential CPH model (Austin, 2012) shown in Eq. (12).

$$T \sim exp\left(\lambda(t;\boldsymbol{x})\right) = exp\left(\lambda_0 \bullet \exp\left(h(\boldsymbol{x})\right)\right). \tag{12}$$

To verify that CoxNAM discerns the relevant covariates from the noise, we let the hazard function depend on two of the ten covariates in both synthetic dataset experiments with linear and nonlinear hazard functions. Here, 50% of the observations set served as censored data, that is, $E = 0$, and 50% of the observations have an observed event, $E = 1$. Next, we will conduct experiments to evaluate the linear and nonlinear fitting ability of the model.

(1) Linear experiment

To verify the linear fitting ability of the model, we assumed that the data have a linear hazard function, and the specific risk function assumption is described as Eq. (13).

$$h(\boldsymbol{x}) = x^0 + 2x^1. \tag{13}$$

We generated linear survival synthetic data according to such a hazard function, trained the proposed and compared models, and tested and evaluated the performance of these models.

(2) Nonlinear experiment

To verify the nonlinear fitting ability of the model, we assumed that the data have a Gaussian hazard function, and the specific risk function assumption is noted as Eq. (14).

$$h(\boldsymbol{x}) = log\left(\lambda_{max}\right) exp\left(-\frac{\left(x^0\right)^2 + \left(x^1\right)^2}{2r^2}\right). \tag{14}$$

In specific experiments, we set $\lambda_{max} = 5.0$, and the scale factor of $r = 0.5$. We generated nonlinear survival synthetic data according to such a hazard function, trained the proposed and compared models, and tested and evaluated the performance of these models.

2.5.2. Rotterdam & German breast cancer study group

To easily compare with existing models and verify the application of the model on real datasets, we used one public real-world dataset for model verification. The Rotterdam & German Breast Cancer Study Group (Rotterdam & GBSG) includes two datasets focused on breast cancer, namely, the Rotterdam tumor bank (Foekens et al., 2000) and the German Breast Cancer Study Group (GBSG) (Sauerbrei et al., 2000). The Rotterdam tumor bank dataset contains 1546 node-positive breast cancer patients, and approximately 90% of patients have an observed death time. The GBSG dataset contains 686 breast cancer patients, and 56% are censored.

2.5.3. Dataset for prognosis of gastric cancer

The latest data show that gastric cancer is the 5th most prevalent type of malignancy with 1.089 million new cases and an age-standardized incidence rate of 15.8/100,000 in men and 7.0/100,000 in women. Gastric cancer is the 4th most common cause of mortality with 769,000 new deaths and an overall mortality rate of 7.7/100,000 (Society of Gastric Cancer of China Anti-Cancer Association secretariat, 2022). Gastric cancer poses a huge threat to our health, and research on prognosis management and clinical decision-making for gastric cancer patients is of great significance.

The data used in this work are gastric cancer data acquired from the Surveillance, Epidemiology, and End Results (SEER) Program of the National Cancer Institute, which provides information on cancer statistics in an effort to reduce the cancer burden among the U.S. population. SEER is supported by the Surveillance Research Program (SRP) in NCI’s Division of Cancer Control and Population Sciences (DCCPS). The dataset can be requested online from the SEER website (http://www.seer.cancer.gov). We used the latest Incidence-SEER research Data, 17 Registries, Nov 2021 Sub (2000–2019) database through SEER*Stat 8.4.0. We obtained the relevant data from gastric cancer patients by setting “Site recode ICD-O-3/WHO 2008” to “Stomach”. Then, we selected features related to the prognosis of gastric cancer patients and extracted relevant data from the database.

(1) Selection of research cohorts

(a) To study the survival status of gastric cancer patients in recent years, gastric cancer patients diagnosed in 2009–2019 were selected as the research cohorts by excluding the data of gastric cancer patients before 2009.
(b) ‘Sequence number’ = 0 was selected to exclusively focus on gastric cancer patients and exclude the interference of other cancers.
(c) Excluding the interference of few gastric cancer patients younger than 18 years old, we focused on the survival status of adult gastric cancer patients.
(d) Patients with a survival time of 0 were excluded.

(2) Consolidation and transformation of features

Given that the data span a 10-year period, some features have different representations in different time periods. To unify the dataset, different representations in different time periods need to be merged. Details are shown in Table 1.

In addition, based on the opinions of clinical experts, we further process the original to form a new index, such as the lymph node ratio (LNR), that is, the ratio of regional nodes positive (1988+) and regional nodes examined (1988+). The lymph node index is a clinically important factor affecting cancer patients.

(3) Determination of E_i

Patients are censored or not based on the combination of ‘Survival months’ and ‘Vital status recode’ in the SEER database. ‘Survival months’ refers to patient survival time recorded by month. ‘Vital status

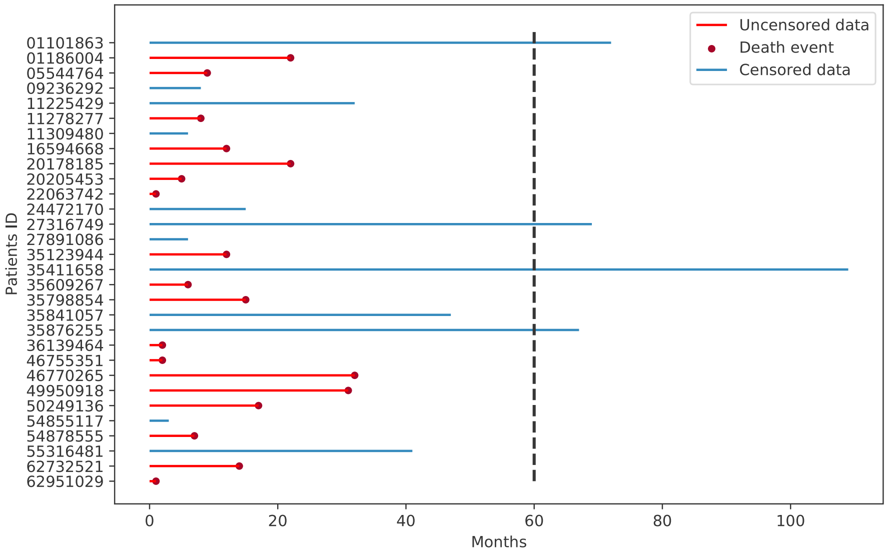


Fig. 3. Lifespans of 30 randomly selected gastric cancer patients.

Table 1
Consolidation of features.

Feature	Different expressions	Time period
Grade	Grade (thru 2017) Grade Clinical (2018+)	2009–2017 2018–2019
Tumor Size	CS tumor size (2004–2015) Tumor Size Summary (2016+)	2009–2015 2016–2019

recode' refers to identification of the patient's status, i.e., survival or death. A 'Vital status recode' equals to 1 indicates survival, and a 'Vital status recode' equal to 0 indicates death. According to the definition of censoring, we proposed a criterion for determining whether the value is censored. The details are as Eq. (15).

$$E_i = \begin{cases} 0, & \text{if ('Survival months'} \le 60\,\&\,\text{'Vital status recorde'} = 1) \\ & \quad\text{or ('Survival months'} > 60) \\ 1, & \text{if ('Survival months'} \le 60\,\&\,\text{'Vital status recorde'} = 0) \end{cases} \tag{15}$$

(a) If 'Survival months' is less than or equal to 60 and 'Vital status recode'=1, the patient has no observed death during the study period, and it is judged as censored data, that is, $E_i = 0$.
(b) If the 'Survival months' is greater than 60, the patient outcome cannot be observed during the study period, and it is judged as censored data, that is, $E_i = 0$.
(c) If 'Survival months' is less than or equal to 60 and 'Vital status recode'=0, the death outcome can be observed during the study period, that is, $E_i = 1$.

Fig. 3 presents the three cases of patient events described above. Thirty gastric cancer patients were randomly selected and arranged by ID, and their lifetimes were visualized. Patients 01101863, 27316749, 35411658, and 35876255 are surviving patients to the right of the 60-month survival baseline and still alive at the end of the observation period, and their survival time exceeds 60 months. Surviving patients to the left of the 60-month survival baseline, for example, patient 11225429, are lost to follow-up. Dead patients, for example, patient 01186004, die during the observation period, and their specific survival times are known.

Finally, the specific features and their descriptions are shown in Table 2.

After preprocessing, we obtained 15,506 data records, of which 7,715 are censored and 7,791 are observed outcomes. Thus, the censoring rate is approximately 50%.

3. Results

3.1. Evaluation metric for survival analysis

Common used evaluation metrics in the field of machine learning are not suitable for evaluating the performance of survival analysis models because censored observations exist in survival data. The concordance index (C-index) is the most popular evaluation metric and is calculated as the ratio of the number of pairs correctly ordered by the model to the total number of comparable pairs. A pair is not comparable if the two observations are both censored or the earliest time in the pair is censored. The C-index is defined as Eq. (16).

$$\frac{1}{N}\sum_{i,E=1}\sum_{j,T_i<T_j} I\left[S\left(\hat{T}_i \,\middle|\, \mathbf{x}_i\right) < S\left(\hat{T}_j \,\middle|\, \mathbf{x}_j\right)\right], \tag{16}$$

where N refers to the number of all comparable pairs, $S(\bullet)$ is the survival function, $I[\bullet]$ is the counting function, the outcome is the number of satisfied conditions, and $\hat{T}_i$ is the predictive value for the time of the event of interest occurring.

According to the Eq. (16), the value of C-index ranges from 0 to 1. The closer the value of C-index is to 1, the better the model performance is. The model is the same as random guessing when the value of C-index is 0.5 and worse if the value of C-index is less than 0.5.

3.2. Numerical results for all experiments on the indicator C-index

To verify the performance of the model, the model is compared with CPH, DeepSurv and RSF. We performed all the experiments on a PC equipped with Intel Core i5-6500 HQ CPUs at 3.20 GHz and 16 GB RAM, running in the PyCharm Community 2019.2.5 environment. All algorithms are implemented in Python language, and the compared models are implemented using relevant Python packages. The specific Python packages are shown in Table 3, and the specific hyperparameters can be found in appendix Table A.1.

We used 10-fold cross validation procedure to get the value of indicator C-index in all experiments. To further test that the performance of the models were statistically significant, we performed a Student's t test. The results for all experiments on the indicator C-index as shown in Table 4 and Fig. 4.

Table 2
Prognostic elements of gastric cancer survivability.

Feature	Description
ID	This is a dummy number to uniquely identify a patient.
Sex	The sex of each patient.
Age	Patient's age at diagnosis with all patients older than 85 years of age grouped together.
Year of diagnosis	The year the tumor was first diagnosed by a recognized medical practitioner.
Race	The race of each patient.
Origin	Identifies patients with Spanish/Hispanic surname or of Spanish origin.
Primary Site	The primary site code is provided in ICD-O-3.
Grade	Category based on the tumor appearance.
Tumor Size	Size in mm.
Lymph Node Ratio	The number of positive regional nodes as a percentage of the total regional nodes examined.
Diagnostic Confirmation	This data item records the best method used to confirm the presence of the cancer being reported.
Histology	Histology recode broad groupings.
Stage	Created from SEER Combined Summary Stage 2000 (2004–2017) & Derived Summary Stage 2018 (2018+).
Surg Prim Site	A surgical therapy that removes and/or destroys tissues of the primary site.
Scope Reg LN Sur	This data item records the number of regional lymph nodes examined in conjunction with surgery performed as part of the first course of treatment at all facilities.
Surg Oth Reg/Dis	A surgical therapy that removes and/or destroys tissues of the primary site.
Survival months	Patient survival time recorded by month.
Event	Event indicator of censored or uncensored.

Table 3
Python packages used for compared models' implementation.

Models	Python package
CPH	lifelines
DeepSurv	DeepSurv.pytorch
RSF	scikit-survival

From Table 4 and Fig. 4, we can see that the traditional statistical survival analysis model CPH performs well in linear experiment, but performs poorly in nonlinear experiment while machine learning-based survival analysis models (RSF, DeepSurv, and CoxNAM) performed well, which exposes the CPH's insufficient ability to capture nonlinearities. Furthermore, CoxNAM performs better in all experiments. Although there is no significant difference from the results of DeepSurv in the "Simulated Linear" and "Rotterdam & GBSG" experiments, the performance of CoxNAM is not worse than that of DeepSurv. In addition, CoxNAM can also provide interpretation which DeepSurv can only obtain by spending extra time and using other tools (such as SHAP), which leads to the irreplaceability and superiority of CoxNAM.

We also reported the average running time of one round for all experiments as shown in Table 5. From the Table 5 we can see that CoxNAM takes roughly the same time as DeepSurv. Although CoxNAM is time-consuming in some experiments, it is still comparable to DeepSurv. In addition, CoxNAM can also provide interpretation which DeepSurv can only obtain by spending extra time and using other tools. So even if there is no significant difference in C-index and the time may be longer with DeepSurv, the interpretable CoxNAM may be more likely to be selected and applied.

3.3. Results about the importance and the shape functions of related features

The analysis about the importance of features is also an important task of this research. In this section we analyzed the relevant important features in the four experiments. It should be emphasized that two synthetic datasets, including linear and nonlinear hazard functions, were used to estimate the interpretability of the models. Since we can control the relation between features and targets in advance, we can compare the performance of our method and other methods in the interpretation of results in this section. We choose the most popular SHAP frameworks for comparison. The interpretation results of the proposed model are compared with those of DeepSurv+SHAP to verify which interpretation method is more in line with the previous setting. Since accurate calculations by SHAP would take several hours, it should be noted that we used approximate calculation method. According to the simplified algorithm provided by SHAP, we clustered the samples into 10 categories and performed approximate calculations.

Fig. 5 shows the importance of features in linear experiments, Fig. 5(a) is the result obtained by CoxNAM, and Fig. 5(b) is the result obtained by DeepSurv+SHAP. From Fig. 5(a), we can see that x^1 and x^0 are the most important features. The importance of other features is close to 0, and the importance of x^1 is approximately two times that of x^0. These findings are consistent with our previous experimental settings. These results demonstrate that our model can effectively describe the importance of features to the target in linear experiments. Comparing Figs. 5(a) and 5(b), we can find that both interpretation methods capture important features. But the ratio of x^1 to x^0 in Fig. 5(a) is about 1.929, which is closer to the previous assumption (the ratio of x^1 to x^0 is 2) than the ratio of x^1 to x^0 in Fig. 5(b)(1.839). Therefore, the interpretation of the proposed CoxNAM model is more accurate in linear experiment. In addition, DeepSurv+SHAP spent an additional 205.31 s for the explanation of the result in this experiment.

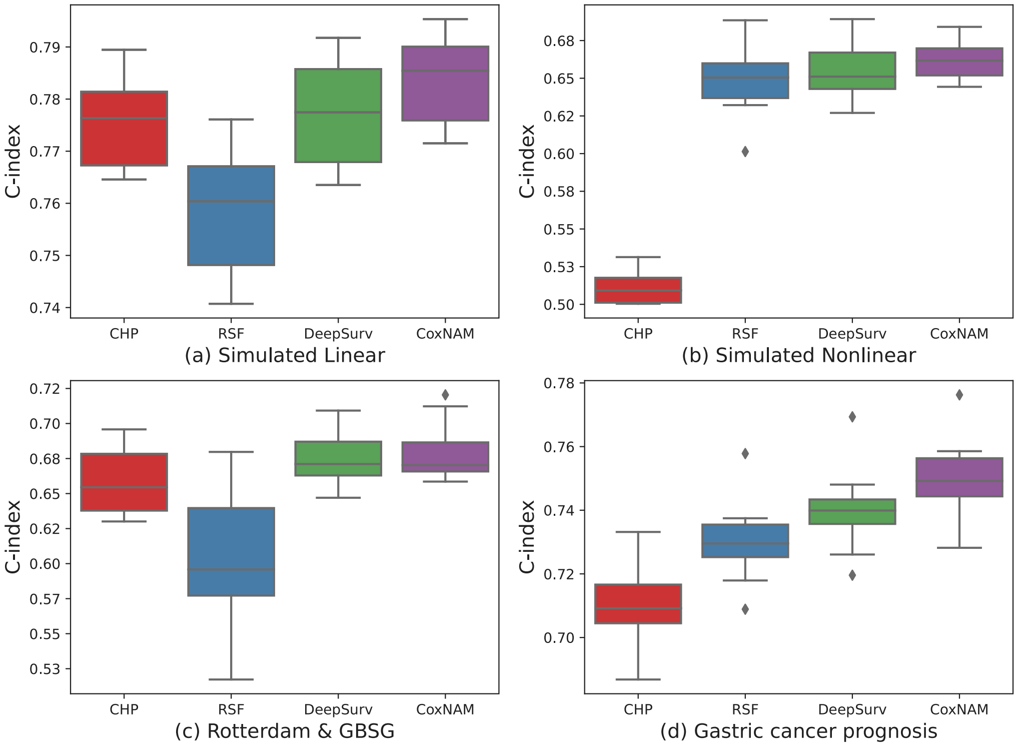
Fig. 6 shows the shape functions of the linear experiment. Fig. 6(a) shows the shape functions of the 10 features of the linear experiment in the same coordinate system. We can see from the figure that x^0 and x^1 play an important role in the prediction of the risk function. In addition, the shape function value of x^1 is approximately two times that of x^0, but the shape function value of $x^2 \sim x^9$ is approximately 0, which demonstrates that these features are not very important for the risk function. These findings are consistent with the previous assumptions. Figs. 6(b) and 6(c) shows the shape functions of x^0 and x^1 and the number of samples supported. The figure shows that the values of the features are uniformly distributed in the $[-1,1)$ interval. The shape functions of x^0 and x^1 are roughly linear and increase with increasing feature values. These results are consistent with the previous assumptions.

Fig. 7 shows the importance of features in nonlinear experiment, Fig. 7(a) is the result obtained by CoxNAM, and Fig. 7(b) is the result obtained by DeepSurv+SHAP. From Fig. 7(a), we can see that x^0 and x^1 are the most important features. The importance of other features is close to 0, and the importance of x^0 is very close to that of x^1. These findings are consistent with our previous experimental settings.

Table 4
Numerical results for all experiments on the indicator C-index (STD)

Experiment	CPH	RSF	DeepSurv	CoxNAM
Simulated Linear	0.7756(0.0084)***	0.7582(0.0115)**	0.7770(0.0097)***	0.7835(0.0080)
Simulated Nonlinear	0.5109(0.0101)***	0.6479(0.0217)*	0.6557(0.0204)	0.6627(0.0128)
Rotterdam & GBSG	0.6585(0.0227)**	0.6054(0.0468)**	0.6757(0.0204)	0.6801(0.0200)
Gastric cancer prognosis	0.7102(0.0121)***	0.7382(0.0123)***	0.7401(0.0126)***	0.7503(0.0121)

*Indicates the difference between CoxNAM and marked model is statistically significant at 0.05 significance level;
**Indicates the difference between CoxNAM and marked model is statistically significant at 0.01 significance level;
***Indicates the difference between CoxNAM and marked model is statistically significant at 0.001 significance level.



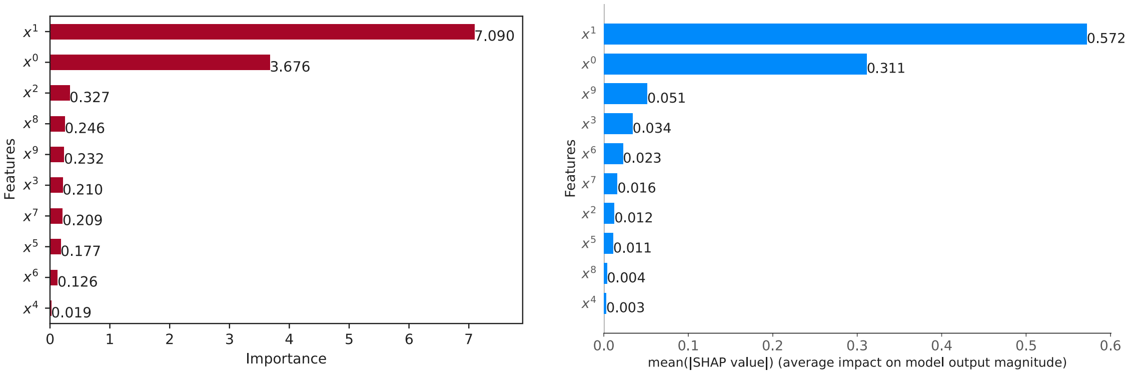
(a) Simulated Linear

(b) Simulated Nonlinear

(c) Rotterdam & GBSG

(d) Gastric cancer prognosis

Fig. 4. Boxplots of performance for proposed and compared models with 10-fold cross validation.



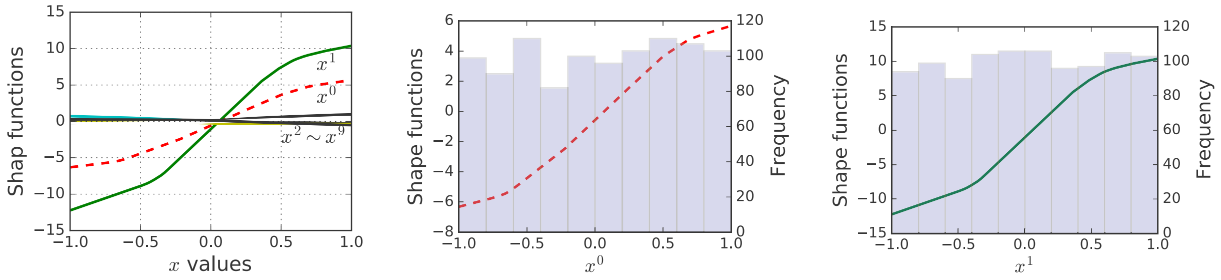
(a) CoxNAM

(b) DeepSurv+SHAP

Fig. 5. The importance of features in linear experiment.

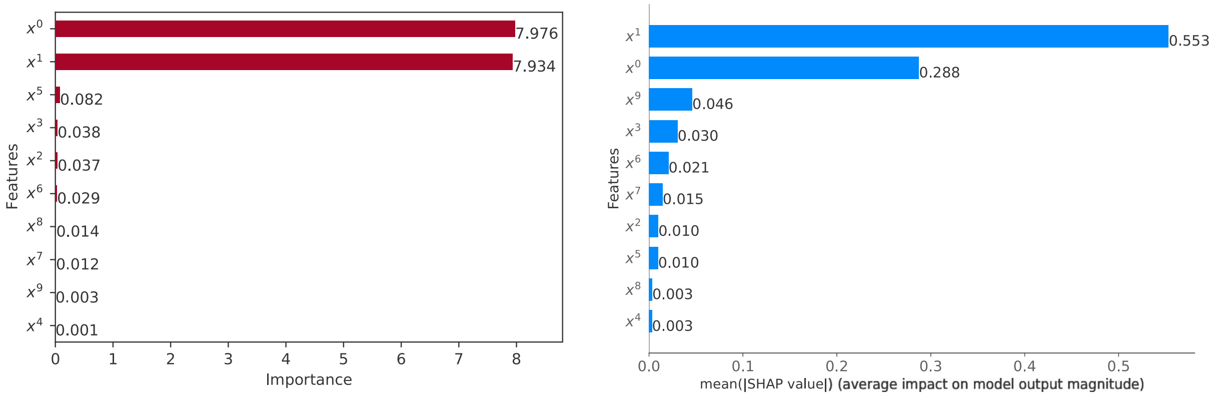
These results demonstrate that our model can also effectively describe the importance of features to the target in nonlinear experiments. Comparing Figs. 7(a) and 7(b), we can find that both interpretation methods also capture important features. But the ratio of x^1 to x^0 in Fig. 7(a) is about 1.005, which is closer to the previous assumption (the ratio of x^1 to x^0 is 1) than the ratio of x^1 to x^0 in Fig. 7(b)(1.920). Therefore, the interpretation of the proposed CoxNAM model is also more accurate in nonlinear experiment. In addition, DeepSurv+SHAP spent an additional 218.36 s for the explanation of the result in this experiment again.

Fig. 8 shows the shape functions of the nonlinear experiment. Fig. 8(a) shows the shape functions of the 10 features of the nonlinear experiment in the same coordinate system. We can see from the figure that x^0 and x^1 play an important role in the prediction of the risk function and that the shape functions of x^0 and x^1 are very similar. The shape function value of $x^2 \sim x^9$ is approximately 0, which demonstrates



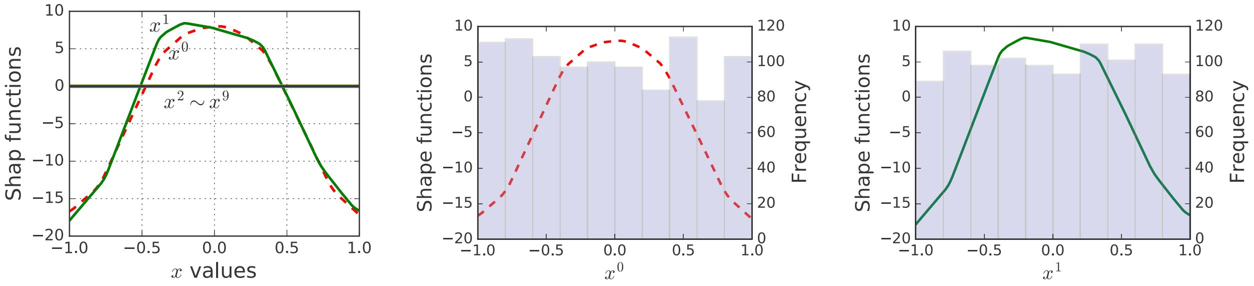
(a) All features (b) Feature x^0 (c) Feature x^1

Fig. 6. Shape functions of linear experiment.



(a) CoxNAM (b) DeepSurv+SHAP

Fig. 7. The importance of features in nonlinear experiment.



(a) All features (b) Feature x^0 (c) Feature x^1

Fig. 8. Shape functions of linear experiment.

Table 5
Average running time of one round for all experiments (s)

Experiment	CPH	RSF	DeepSurv	CoxNAM
Simulated Linear	1.84	735.09	130.31	107.96
Simulated Nonlinear	2.35	918.30	235.55	170.72
Rotterdam & GBSG	0.85	14.69	27.34	44.07
Gastric cancer prognosis	1.82	1077.74	775.10	804.17

that these features are not very important for the risk function. These findings are consistent with the previous assumptions. Figs. 8(b) and 8(c) show the shape functions of x^0 and x^1 and the number of samples supported. The figure shows that the values of the features are uniformly distributed in the [−1,1) interval. The shape functions of x^0 and x^1 are roughly quadratic nonlinear and symmetric about the y-axis. These results are in line with the previous assumptions.

Fig. 9 shows the importance of features related to breast cancer patients' prognosis in the Rotterdam & GBSG data. Based on the calculation results of the feature importance, we can analyze the importance of the feature. It can be seen from the figure that the progesterone receptor, positive lymph nodes, and the estrogen receptor are important factors affecting the prognosis of breast cancer patients. In addition, tumor size, hormone therapy and age are also important factors.

Fig. 10 shows the shape functions of the breast cancer patients' prognosis-related features in the Rotterdam & GBSG dataset. Based on the trend of the shape function, we can analyze the impact on the risk function as the feature value changes. For specific individual features, hormone therapy can reduce the risk of death for breast cancer patients. Breast cancer patients have an increased risk of death as tumor size increases. Breast cancer patients in menopausal states have a higher risk of death. The breast cancer patients' risk of death increases with age. The higher the number of positive lymph nodes, the greater the breast cancer patient's risk of death. An increase in progesterone receptor or estrogen receptor reduces a patient's risk of death.

Figs. 11 and 12 show the importance of related features related to the prognosis of gastric cancer and their influence on the results. Fig. 11 shows that clinical information and age have a greater impact on the

Table 6
Basic information on Patients 01117360 and 05603394.

Features	01117360	05603394
Sex	Male	Female
Age	52	64
Year of diagnosis	2009	2010
Race	White	White
Origin	Spanish-Hispanic-Latino	Non-Spanish-Hispanic-Latino
Primary Site	162	164
Grade	Poorly differentiated, undifferentiated	Poorly differentiated, undifferentiated
Tumor Size	7 millimeters	size not stated
Lymph Node Ratio	0	1
Diagnostic Confirmation	Positive histology	Positive histology
Histology	Cystic, mucinous and serous neoplasms	Cystic, mucinous and serous neoplasms
Stage	Localized	Distant
Surg Prim Site	Gastrectomy, NOS (partial, subtotal, hemi-)	Gastrectomy, NOS (partial, subtotal, hemi-)
Scope Reg LN Sur	4 or more regional lymph nodes removed	1 to 3 regional lymph nodes removed
Surg Oth Reg/Dis	None	None
Survival months	129 months	6 months
Event	0	1
Partial hazard	0.5972	5.2513

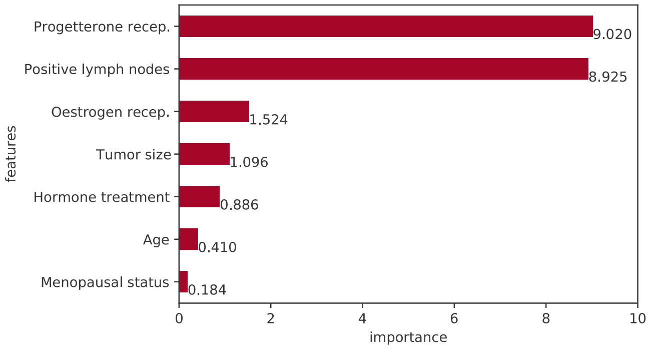


Fig. 9. The importance of features related to breast cancer patient prognosis.

5-year survival of gastric cancer patients. The lymph node ratio, stage and age of gastric cancer patients have a significant influence on their survival, which is consistent with the experience of clinicians regarding factors important for gastric cancer patient survival. The age of gastric cancer patients has always been a key factor affecting the survival rate of gastric cancer patients. In addition, primary surgical site, tumor size, histology, grade, race, and primary site are also important prognostic factors.

Fig. 12 shows the shape functions of gastric cancer prognosis. The impact of changes in the value of the feature on the mortality risk can be understood based on the shape function of a specific feature. For example, the higher the lymph node ratio value is, the greater the mortality risk of gastric cancer patients. As the tumor stage increases, the mortality risk of the patient gradually increases. The older the patient is, the greater the mortality risk. A surgical primary site of 0 indicates that no surgical procedure was performed at the primary site. This type of patient has a higher mortality risk, whereas other types of surgery performed have a lower risk. As the tumor size increased, the mortality risk of patients also gradually increased. Similarly, the higher the tumor grade is, the greater the mortality risk to gastric cancer patients.

3.4. Example of survival analysis for patients with gastric cancer

Based on parametric fitting and nonparametric baseline hazard function $h_0(t)$ estimation, we can predict the hazard functions of gastric cancer patients and then predict the survival functions of gastric cancer patients. We used specific patients with different survival outcomes to analyze their hazard function and survival function. We selected Patients 01117360 and 05603394 as examples, and their basic information is shown in Table 6. Patient 01117360 is a five-year survival patient with a survival time of 129 months. Patient 05603394 is a five-year death patient with a survival time of 6 months. Fig. 13 illustrates the hazard function and survival function of two specific gastric cancer patients with different survival outcomes.

Fig. 13(a) shows the hazard functions of the two specific gastric cancer patients. Patient 05603394 faces a mortality risk that is higher than the baseline risk. Patient 01117360 faces a mortality risk that is lower than the baseline risk. It can also be seen from the figure that gastric cancer patients face a higher risk of death within 3 years (36 months), especially at approximately 1 year. This finding also explains why 3-year survival or 5-year survival is used to evaluate the cancer outcome of patients in clinic. After 3 years or 5 years, patients face a lower risk of death and can be considered cured.

Fig. 13(b) shows the survival functions of the two specific gastric cancer patients. The survival function of Patient 05603394 declines rapidly and is lower than the baseline survival function. The survival function declines rapidly between 0 and 1 year, and the probability of survival is already low. Patient 01117360 has a slow decline in survival function that is higher than the baseline survival function and still has a high probability of survival at 5 years.

4. Discussion

Focusing on the limitations of traditional statistical survival analysis models that cannot capture nonlinearity and the inexplicable limitations of deep survival analysis models, we proposed an interpretable deep survival analysis model. We faced three main challenges in this process: (i) How to balance the performance and interpretability of proposed model. To this end, we designed an interpretable network and fused it with CPH for survival analysis; (ii) How to design the corresponding loss function. Unlike traditional machine learning, survival analysis requires special likelihood functions to solve parameters. We also incorporated loss function design methods in machine learning such as regularization; (iii) How to measure and evaluate the effectiveness of the interpretation obtained by the model. We designed a calculation method about importance of features based on their shape functions, and we addressed the challenge of evaluating the effectiveness of the interpretation with synthetic datasets in which we can set the importance of features in advance. We validated the performance and interpretability of the CoxNAM on two synthetic and two real datasets. It can also be seen from the experimental results in Section 3.2 that CoxNAM has the highest performance on the C-index. Although some experiments have no statistically significant difference

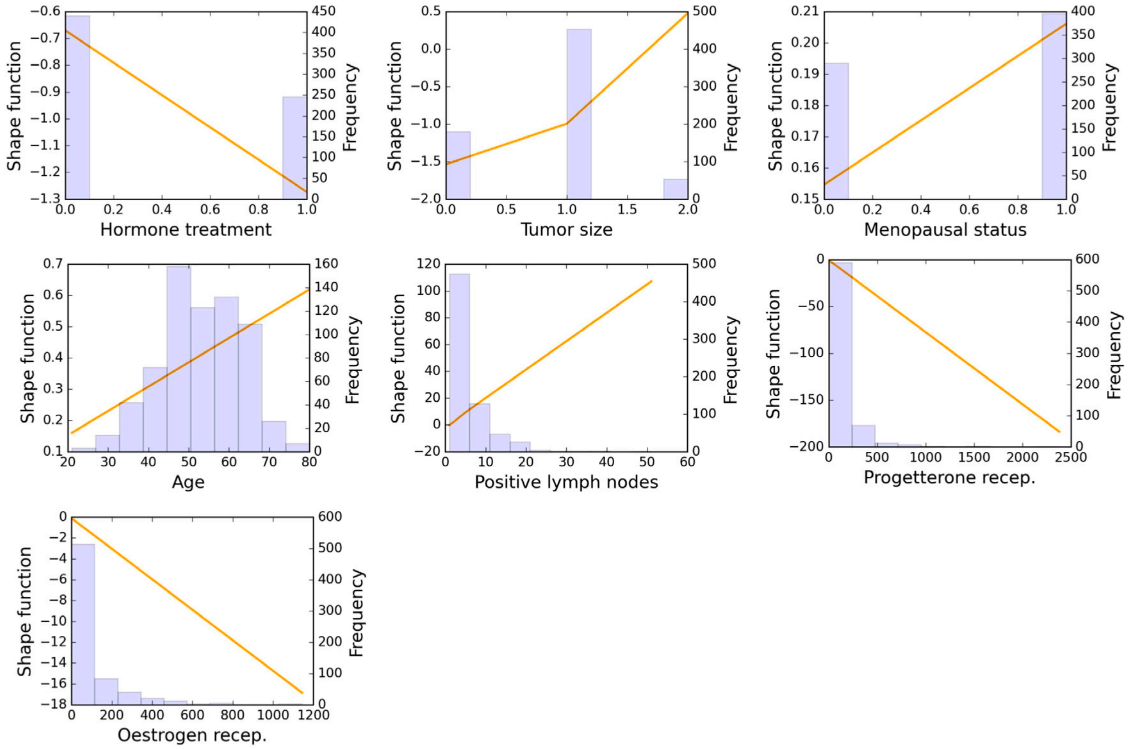


Fig. 10. Shape functions of Rotterdam & German Breast Cancer Study Group.

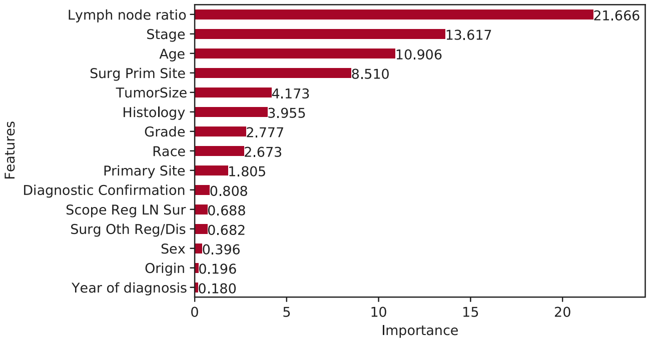


Fig. 11. The importance of features related to gastric cancer patient prognosis.

from DeepSurv, CoxNAM's performance is at least not bad. In addition, CoxNAM can provide additional interpretation of the results.

We also compared the explainability of the CoxNAM with DeepSurv+SHAP. Combined with the feature importance analysis of two synthetic datasets in Section 3.3, in summary, the proposed model has three differences from DeepSurv+SHAP: (i) The two methods belong to different types. DeepSurv+SHAP belongs to an explainable method, which explains the results afterwards with the help of explanation tools. The model we proposed belongs to an interpretable method and is a white-box model, which can directly obtain the basis while obtaining the results. Therefore, the two frameworks belong to different types; (ii) There is a difference in the effect of the two models on the interpretation of the results. According to the analysis results on the feature importance of the two synthetic datasets (as shown in Figs. 5 and 7), we can find that CoxNAM is more in line with the advance assumptions and more accurate in two experiment;(iii) The computational complexity of the two models is different. Exact calculations for the SHAP method need several hours, although approximate calculations also need additional minutes (additional 205.31 s in linear experiment and 218.36 s needed in nonlinear experiment). The method we propose can get the corresponding interpretability at the same time as the results are obtained.

So combining the analysis of performance and predictability of the proposed method and the compared methods, we summarized the main contributions of our research as follows:

(a) An interpretable deep survival analysis model that combines machine learning methodology with the traditional statistical survival analysis model was proposed with the ability to model nonlinearly and output the importance of related features for survival analysis;
(b) We designed a method to calculate the importance of related features. We can also study the influence of the values of related features on the final results according to the value of the shape function. The obtained information can provide a reference for relevant decision-making and facilitates the application of the model in practice;
(c) The proposed model was tested by numerical experiments with two controlled synthetic datasets and two public real datasets to illustrate its ability about capturing the nonlinearity and the interpretability.

We also summarized the advantages and disadvantages of this research. Through the analysis of the results, we think that our model mainly has the following two advantages: (i)The model we proposed can better capture the nonlinear relation between features and targets. The model we proposed combines deep learning technology and can fit more complex functional relationships. Therefore, our model is an

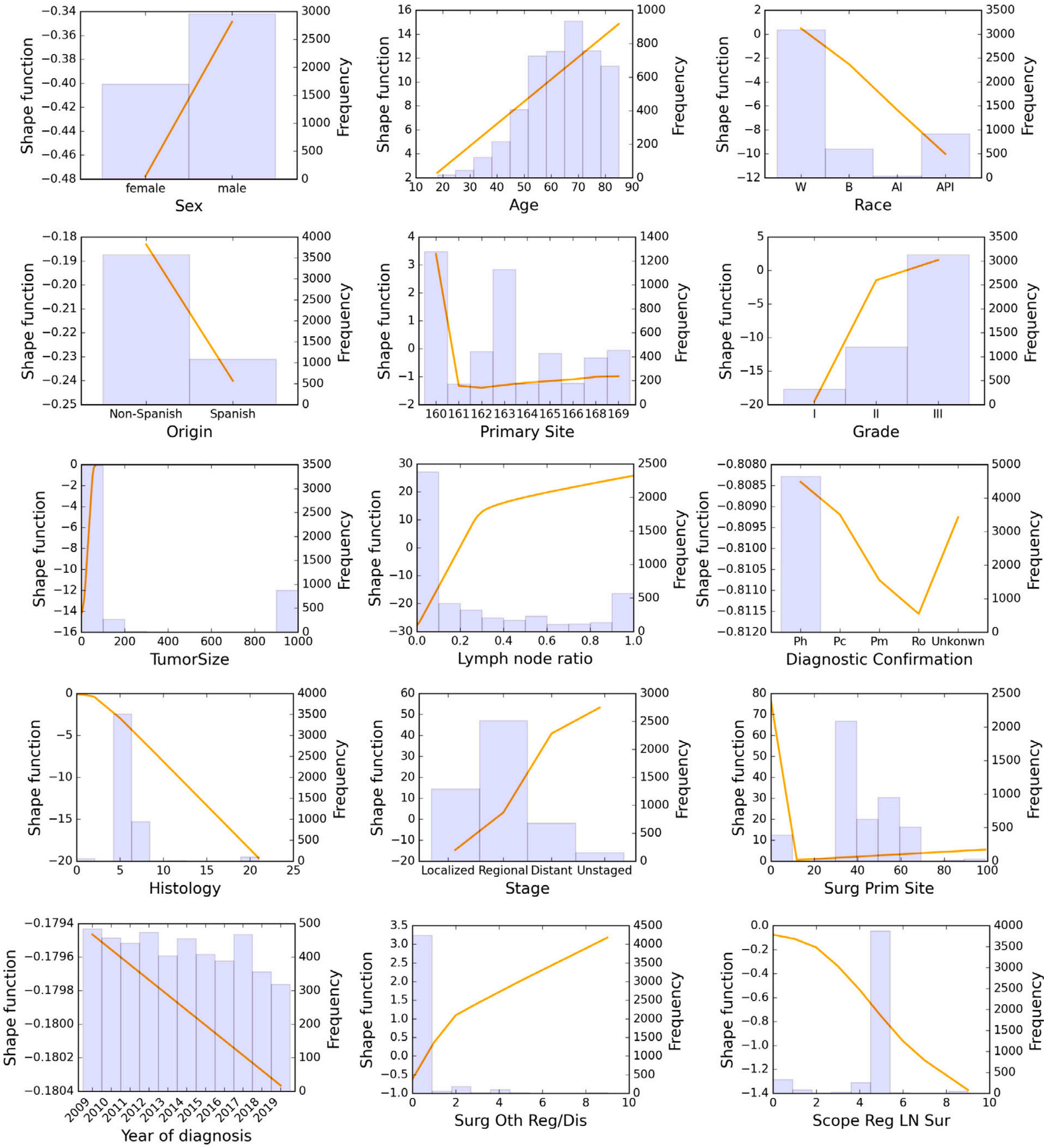
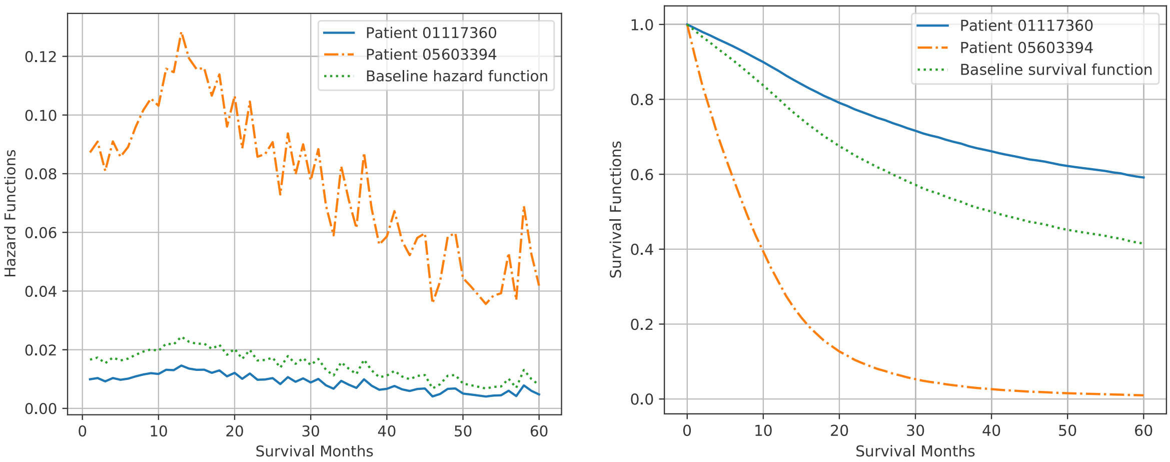


Fig. 12. Shape functions of gastric cancer prognosis.

extension of the traditional survival analysis models, which breaks through the limitations of the linear assumption; (ii) The model we proposed is a white-box model, which can show the mechanism of how the model obtains the predicted results. So that we can see the basis of the results obtained, which will make the results more convincing and easier to adopt. We also need to clearly point out that our model still has some disadvantages. We employed standard NAM to predict risk, so the proposed model currently does not take into account the capture of interaction relation between variables. In addition, we unified the network structure of each variable for simplicity, but it may not be the optimal setting.

This research is a type of study that combines machine learning models with traditional statistical models. The machine learning method is used to improve the performance of traditional statistical models while retaining their advantages, which further enriches the relevant theories of different research paradigms' fusion and can provide reference for the research paradigms' fusion of other similar work.

In addition, we also took two representative gastric cancer patients as examples to conduct survival analysis, obtained their risk functions and survival functions, and demonstrated how the model can be applied in the prognosis management of gastric cancer patients. Cancer poses a huge threat to our health, and research on prognosis management



(a) Hazard function

(b) Survival Function

Fig. 13. Example of risk function and survival function for gastric cancer patients.

and clinical decision-making for cancer patients is of great significance. Survival analysis holds great value for cancer patients, clinicians, researchers, and policy-makers (Mariotto et al., 2014). Survival analysis for cancer patients can help patients understand their life expectancy and protect their mental health. On the other hand, it can also assist clinicians in formulating precise treatment plans to ensure treatment effects. In addition, survival analysis is of great significance for formulating and optimizing cancer prevention and treatment, providing a reference for the implementation of systems, such as palliative care and hospice care, and providing support for the formulation of medical welfare systems and policies and the distribution of medical resources. Therefore, we applied the CoxNAM model to the gastric cancer survival analysis task. In the task of gastric cancer survival analysis, we focused on the time period between diagnosis and death, which serves as the event of interest. The part about gastric cancer prognosis of the description provides a reference for how the model can be used in gastric cancer prognosis and which is an important reference for the processes of other similar problems.

5. Conclusion and future research

This research proposed an interpretable deep survival analysis model for survival analysis that can better model the relation between related features and the time of an event of interest occurring, and the proposed model can also obtain the shape functions to interpret the importance of related features and analyze the effect of feature values. We used numerical experiments on two synthetic datasets and one public real-world breast cancer dataset to verify the performance of the model. The results indicated that the proposed model can effectively model the relation between related features and the target, capture important features, and analyze the impact of feature values. At the same time, we compared the interpretability with the SHAP framework on the two synthetic datasets and the results demonstrated the effectiveness of the proposed model's interpretation. We also applied the model to the study of gastric cancer patient prognosis. Experimental results performed on the real-world gastric cancer dataset from the publicly available SEER database indicate that the proposed model has better performance on the C-index. We can also obtain important features related to gastric cancer patient survival time and analyze the effect of feature values, which can influence the hazard functions of gastric cancer patients. The proposed method shows promising performance and realistic interpretability. The model can potentially be extended to survival analysis problems in multiple domains for relevant decision-making.

In future research, we will focus on the following important issues. Firstly, it would be valuable to verify the proposed model on other larger real-world survival analysis datasets from more data sources. Secondly, it is worth exploring other network structures to improve the performance of the model. Then, the model structure can be further designed to capture the interaction between variables. Finally, it is also valuable to focus on the application of survival analysis results, especially in the management of patient prognosis, which has the potential to achieve greater practical value.

CRediT authorship contribution statement

Liangchen Xu: Conceptualization, Methodology, Software, Writing – original draft. **Chonghui Guo:** Supervision, Writing – review & editing, Funding acquisition.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

The authors do not have permission to share data.

Acknowledgments

This research is supported by the National Natural Science Foundation of China (Grant No. 71771034), the Fundamental Funds for the Central Universities of China (Grant No. DUT21YG108), the Dalian High Level Talents Innovation Support Plan (Grant No. 2021RD01), and the Liaoning Province Applied Basic Research Program Project (Grant No. 2023JH2/101300208).

Appendix

See Table A.1.

Table A.1
The hyperparameters of the experiments.

Experiments	Parameters	Synthetic datasets	Rotterdam & GBSG	Gastric cancer prognosis
RSF	*n*_estimators	50	80	100
	min_samples_split	5	7	10
	min_samples_leaf	10	12	15
	max_features	“sqrt”	“sqrt”	“sqrt”
DeepSurv	hidden_sizes	[16,32,16]	[16,32,16]	[16,32,64,32,16]
	activation	“Relu”	“Relu”	“Relu”
	L2_reg	0.5	0.8	1
	epochs	500	500	500
	learning_rate	3.194e−3	3.194e−3	3.194e−3
	lr_decay_rate	3.173e−4	3.173e−4	3.173e−4
	optimizer	“Adam”	“Adam”	“Adam”
	dropout_rate	0.1	0.1	0.1
CoxNAM	hidden_sizes	[16,32,16]	[16,32,16]	[32,64,32]
	activation	“Relu”	“Relu”	“Relu”
	L2_reg	0.5	0.8	0.5
	epochs	500	500	500
	learning_rate	3.194e−3	3.194e−3	3.194e−3
	lr_decay_rate	3.173e−4	3.173e−4	3.173e−4
	optimizer	“Adam”	“Adam”	“Adam”
	dropout_rate	0.1	0.1	0.1

References

Agarwal, R., Melnick, L., Frosst, N., Zhang, X., Lengerich, B., Caruana, R., et al. (2021). Neural additive models: Interpretable machine learning with neural nets. *Advances in Neural Information Processing Systems*, *34*, 4699–4711.

Andersen, P. K., Borgan, O., Gill, R. D., & Keiding, N. (2012). *Statistical models based on counting processes*. Springer Science & Business Media.

Austin, P. C. (2012). Generating survival times to simulate cox proportional hazards models with time-varying covariates. *Statistics in Medicine*, *31*(29), 3946–3958.

Behrad, F., & Abadeh, M. S. (2022). An overview of deep learning methods for multimodal medical data mining. *Expert Systems with Applications*, *200*, Article 117006.

Béjar, J., Pérez, R., Vilalta, A., Álvarez-Napagao, S., & Garcia-Gasulla, D. (2022). Large scale prediction of sick leave duration with nonlinear survival analysis algorithms. *Expert Systems with Applications*, *198*, Article 116760.

Chen, X., Li, Y., Li, X., Cao, X., Xiang, Y., Xia, W., et al. (2021). An interpretable machine learning prognostic system for locoregionally advanced nasopharyngeal carcinoma based on tumor burden features. *Oral Oncology*, *118*, Article 105335.

Cox, D. R. (1972). Regression models and life-tables. *Journal of the Royal Statistical Society. Series B. Statistical Methodology*, *34*(2), 187–202.

Cutler, S. J., & Ederer, F. (1958). Maximum utilization of the life table method in analyzing survival. *Journal of Chronic Diseases*, *8*(6), 699–712.

de Lima Lemos, R. A., Silva, T. C., & Tabak, B. M. (2022). Propension to customer churn in a financial institution: A machine learning approach. *Neural Computing and Applications*, *34*(14), 11751–11768.

Dolatsara, H. A., Chen, Y.-J., Evans, C., Gupta, A., & Megahed, F. M. (2020). A two-stage machine learning framework to predict heart transplantation survival probabilities over time with a monotonic probability constraint. *Decision Support Systems*, *137*, Article 113363.

Foekens, J. A., Peters, H. A., Look, M. P., Portengen, H., Schmitt, M., Kramer, M. D., et al. (2000). The urokinase system of plasminogen activation and prognosis in 2780 breast cancer patients. *Cancer Research*, *60*(3), 636–643.

Gao, Y., Cai, G.-Y., Fang, W., Li, H.-Y., Wang, S.-Y., Chen, L., et al. (2020). Machine learning based early warning system enables accurate mortality risk prediction for COVID-19. *Nature communications*, *11*(1), 1–10.

Grant, S. W., Hickey, G. L., & Head, S. J. (2019). Statistical primer: multivariable regression considerations and pitfalls. *European Journal of Cardio-Thoracic Surgery*, *55*(2), 179–185.

Hu, Y., Zhao, L., Li, Z., Dong, X., Xu, T., & Zhao, Y. (2022). Classifying the multi-omics data of gastric cancer using a deep feature selection method. *Expert Systems with Applications*, *200*, Article 116813.

Ishwaran, H., Kogalur, U. B., Blackstone, E. H., & Lauer, M. S. (2008). Random survival forests. *The Annals of Applied Statistics*, *2*(3), 841–860.

Janssen, F. M., Aben, K. K., Heesterman, B. L., Voorham, Q. J., Seegers, P. A., & Moncada-Torres, A. (2022). Using explainable machine learning to explore the impact of synoptic reporting on prostate cancer. *Algorithms*, *15*(2), 49.

Kalbfleisch, J. D., & Prentice, R. L. (1973). Marginal likelihoods based on Cox’s regression and life model. *Biometrika*, *60*(2), 267–278.

Kaplan, E. L., & Meier, P. (1958). Nonparametric estimation from incomplete observations. *Journal of the American Statistical Association*, *53*(282), 457–481.

Katzman, J. L., Shaham, U., Cloninger, A., Bates, J., Jiang, T., & Kluger, Y. (2018). DeepSurv: personalized treatment recommender system using a Cox proportional hazards deep neural network. *BMC Medical Research Methodology*, *18*(1), 1–12.

Kovalev, M. S., Utkin, L. V., & Kasimov, E. M. (2020). SurvLIME: A method for explaining machine learning survival models. *Knowledge-Based Systems*, *203*, Article 106164.

Kvamme, H., Borgan, Ø., & Scheel, I. (2019). Time-to-event prediction with neural networks and Cox regression. arXiv:1907.00825.

Lee, C., Yoon, J., & Van Der Schaar, M. (2019). Dynamic-DeepHit: A deep learning approach for dynamic survival analysis with competing risks based on longitudinal data. *IEEE Transactions on Biomedical Engineering*, *67*(1), 122–133.

Lee, C., Zame, W., Yoon, J., & Van Der Schaar, M. (2018). DeepHit: A deep learning approach to survival analysis with competing risks. In *Proceedings of the AAAI Conference on artificial intelligence* (pp. 2314–2321).

Mariotto, A. B., Noone, A.-M., Howlader, N., Cho, H., Keel, G. E., Garshell, J., et al. (2014). Cancer survival: an overview of measures, uses, and interpretation. *Journal of the National Cancer Institute Monographs*, *2014*(49), 145–186.

Moncada-Torres, A., van Maaren, M. C., Hendriks, M. P., Siesling, S., & Geleijnse, G. (2021). Explainable machine learning can outperform Cox regression predictions and provide insights in breast cancer survival. *Scientific Reports*, *11*(1), 1–13.

Naeem, H., Alshammari, B. M., & Ullah, F. (2022). Explainable artificial intelligence-based IoT device malware detection mechanism using image visualization and fine-tuned CNN-based transfer learning model. *Computational Intelligence and Neuroscience*, *2022*.

Nagpal, C., Li, X., & Dubrawski, A. (2021). Deep survival machines: Fully parametric survival regression and representation learning for censored data with competing risks. *IEEE Journal of Biomedical and Health Informatics*, *25*(8), 3163–3175.

Nezhad, M. Z., Sadati, N., Yang, K., & Zhu, D. (2019). A deep active survival analysis approach for precision treatment recommendations: application of prostate cancer. *Expert Systems with Applications*, *115*, 16–26.

Pouyanfar, S., Sadiq, S., Yan, Y., Tian, H., Tao, Y., Reyes, M. P., et al. (2018). A survey on deep learning: Algorithms, techniques, and applications. *ACM Computing Surveys*, *51*(5), 1–36.

Sauerbrei, W., Bastert, G., Bojar, H., Beyerle, C., Neumann, R., Schmoor, C., et al. (2000). Randomized 2× 2 trial evaluating hormonal treatment and the duration of chemotherapy in node-positive breast cancer patients: an update based on 10 years’ follow-up. *Journal of Clinical Oncology*, *18*(1), 94.

Sim, J.-a., Kim, Y., Kim, J. H., Lee, J. M., Kim, M. S., Shim, Y. M., et al. (2020). The major effects of health-related quality of life on 5-year survival prediction among lung cancer survivors: applications of machine learning. *Scientific Reports*, *10*(1), 1–12.

Smith, M., & Alvarez, F. (2021). Identifying mortality factors from Machine Learning using Shapley values–a case of COVID19. *Expert Systems with Applications*, *176*, Article 114832.

Society of Gastric Cancer of China Anti-Cancer Association secretariat (2022). CACA guidelines for holistic integrative management of gastric cancer. *Holistic Integrative Oncology*, *1*(1), 3.

Tibshirani, R. (1997). The lasso method for variable selection in the Cox model. *Statistics in Medicine*, *16*(4), 385–395.

Ullah, F., Moon, J., Naeem, H., & Jabbar, S. (2022). Explainable artificial intelligence approach in combating real-time surveillance of COVID19 pandemic from CT scan and X-ray images using ensemble model. *The Journal of Supercomputing*, *78*(17), 19246–19271.

Vale-Silva, L. A., & Rohr, K. (2021). Long-term cancer survival prediction using multimodal deep learning. *Scientific Reports*, *11*(1), 1–12.

Wang, P., Li, Y., & Reddy, C. K. (2019). Machine learning for survival analysis: A survey. *ACM Computing Surveys*, *51*(6), 1–36.

Wang, Y., Wang, D., Ye, X., Wang, Y., Yin, Y., & Jin, Y. (2019). A tree ensemble-based two-stage model for advanced-stage colorectal cancer survival prediction. *Information Sciences*, *474*, 106–124.

Wei, L.-J. (1992). The accelerated failure time model: a useful alternative to the Cox regression model in survival analysis. *Statistics in Medicine*, *11*(14–15), 1871–1879.

Yang, M., Lim, M. K., Qu, Y., Li, X., & Ni, D. (2023). Deep neural networks with L1 and L2 regularization for high dimensional corporate credit risk prediction. *Expert Systems with Applications*, *213*, Article 118873.