



# SurvNet: A Novel Deep Neural Network for Lung Cancer Survival Analysis With Missing Values

Jiayong Wang<sup>1</sup>, Nan Chen<sup>2</sup>, Jixiang Guo<sup>1</sup>, Xiuyuan Xu<sup>1</sup>, Lunxu Liu<sup>2\*</sup> and Zhang Yi<sup>1\*</sup>

<sup>1</sup> Machine Intelligence Laboratory, College of Computer Science, Sichuan University, Chengdu, China, <sup>2</sup> Department of Thoracic Surgery, West China Hospital and West China School of Medicine, Sichuan University, Chengdu, China

## OPEN ACCESS

### Edited by:

Changqiang Wu,  
North Sichuan Medical College,  
China

### Reviewed by:

Quan Guo,  
Michigan State University,  
United States  
Xi Wang,  
The Chinese University of  
Hong Kong, China

### \*Correspondence:

Lunxu Liu  
lunxu\_liu@aliyun.com  
Zhang Yi  
zhangyi@scu.edu.cn

### Specialty section:

This article was submitted to  
Cancer Imaging and  
Image-directed Interventions,  
a section of the journal  
Frontiers in Oncology

**Received:** 30 July 2020

**Accepted:** 04 December 2020

**Published:** 20 January 2021

### Citation:

Wang J, Chen N, Guo J, Xu X, Liu L  
and Yi Z (2021) SurvNet: A Novel Deep  
Neural Network for Lung Cancer  
Survival Analysis With Missing Values.  
*Front. Oncol.* 10:588990.  
doi: 10.3389/fonc.2020.588990

Survival analysis is important for guiding further treatment and improving lung cancer prognosis. It is a challenging task because of the poor distinguishability of features and the missing values in practice. A novel multi-task based neural network, SurvNet, is proposed in this paper. The proposed SurvNet model is trained in a multi-task learning framework to jointly learn across three related tasks: input reconstruction, survival classification, and Cox regression. It uses an input reconstruction mechanism cooperating with incomplete-aware reconstruction loss for latent feature learning of incomplete data with missing values. Besides, the SurvNet model introduces a context gating mechanism to bridge the gap between survival classification and Cox regression. A new real-world dataset of 1,137 patients with IB-IIA stage non-small cell lung cancer is collected to evaluate the performance of the SurvNet model. The proposed SurvNet achieves a higher concordance index than the traditional Cox model and Cox-Net. The difference between high-risk and low-risk groups obtained by SurvNet is more significant than that of high-risk and low-risk groups obtained by the other models. Moreover, the SurvNet outperforms the other models even though the input data is randomly cropped and it achieves better generalization performance on the Surveillance, Epidemiology, and End Results Program (SEER) dataset.

**Keywords:** survival analysis, prognosis prediction, deep neural networks, multi-task learning, missing value

## INTRODUCTION

In clinical research, the development of effective survival analysis methods for censored data is always required to evaluate the relationship between the risk factors and event of interest (1, 2). It has been widely applied to modeling the prognosis of cancers to help to optimize and improve cancer treatment (3–6).

Lung cancer is one of the most heterogeneous cancers and has distinct prognoses. A great deal of work has been conducted on lung cancer prognostic prediction in recent decades, among which the series of tumor node metastasis classification (TNM classification) for lung cancer is the most famous one (4, 7, 8). It has been the guideline for clinical treatment. In the eighth edition of TNM classification for lung cancer, the five-year survival rate of the IB-IIA stage ranged from 65 to 73%, which is relatively high. However, in the practice, many IB-IIA stage patients present with a recurrence and die within five years after treatment. Distinguishing the IB-IIA stage patients with a

high risk of recurrence and death from low-risk patients is worthwhile for guiding further treatment and may improve the lung cancer prognosis. Additionally, the clinicopathologic variable used in TNM classification is limited for personalized prediction for different patients and it is limited to integrate new variables into the existing prognosis models (9). There is a great need for a new survival analysis method to establish fine-grained prognoses for individual patients with IB-IIA stage lung cancer for more accurate individual prediction by integrating an expanding number of prognostic factors.

Cox (2, 10, 11) proportional hazards regression is one of the most well-known survival analysis methods. It has been implemented in many famous software toolboxes and been widely used in many prognosis prediction tasks (8), such as TNM classification for lung cancer (4, 7, 8). The Cox proportional hazards model is semi-parametric and is subject to a linear model (12). It makes an important assumption about the hazard function, which is that covariances that affect the hazard rate are independent. However, in practice, the relationship between variables and the outcome is complex and unknown and there may be interactions among variables (13). Deep neural networks (DNNs) is apparent to be a promising method to solve these problems.

The DNN is a class of biologically inspired computational models towards artificial intelligence. It has been proven that DNNs can approximate any non-linear function when provided with sufficient neurons. Generally, a DNN can be a very complex non-linear model and learn latent features from data directly (14). It has achieved many impressive results in various applications, such as image classification (15, 16), natural language processing (17–20), and biomedical analysis (14, 21–23) in addition to survival analysis (1, 2, 7, 11, 24–26).

Generally, most of DNN based methods for survival analysis could be divided into two paradigms. The first is to formulate the survival analysis as a classification problem to evaluate survival probability at different fixed time points (9, 27, 28). In (27), neural networks were used to improve the prediction accuracy of the five-year survival of patients with breast cancer. Lundina et al. demonstrated that a neural network model trained on some prognostic factors can accurately predict specific 5-, 10-, and 15-year breast cancer survival (9).

The other paradigm is to extend Cox regression with DNNs, in which DNNs are used to extract the features of the patient and trained using Cox-like cost function with the gradient-based method. In (12), the authors proposed the Cox-Net model for prognosis prediction on high-throughput omics data and implemented it with the Theano math library in Python to achieve an efficient computational time using GPUs. Huang et al. modified the Cox-Net method to use multi-omics survival analysis learning on breast cancer (26). Moreover, (16), DNN was applied to cardiac motion analysis for human survival prediction and outperformed the traditional Cox models. However, Cox-Net inherited the limitation of Cox models, which is that it was not designed to estimate the probability of survival at a fixed time. Therefore, it is necessary to study how to design a unified model that integrates the good

properties of the aforementioned two paradigms to improve the performance.

Moreover, the dataset used in survival analysis commonly contains incomplete data with missing values in practice. In many cases, most of the patients with missing values are excluded (2, 29). Omitting patients with missing values limits the number of patients to train the prognosis model and may introduce substantial biases in the study, whereas using patients with missing values may harm the performance. The learning of the latent feature of incomplete data with missing values in survival analysis by using DNNs should be evaluated further.

To address the aforementioned problems, in this study, we propose a multi-task based neural network model, SurvNet, for survival analysis of real-world datasets of patients with IB-IIA stage lung cancer. The main contributions are as follows:

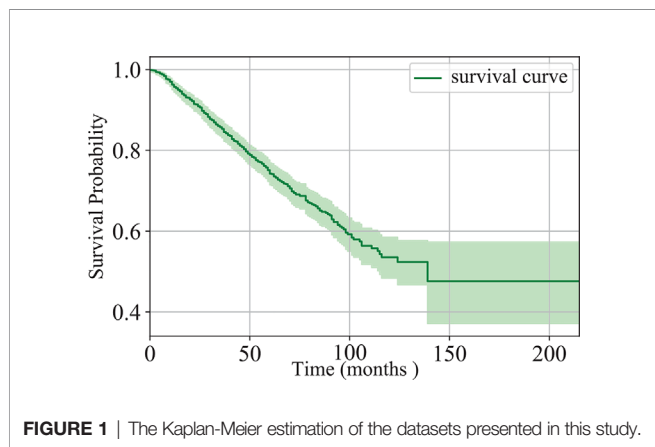
- An input reconstruction mechanism cooperating with incomplete-aware reconstruction loss is proposed in the SurvNet for latent feature learning of incomplete data with missing values.
- A context gating mechanism is proposed in the SurvNet to bridge the gap between survival classification and Cox regression for prognosis prediction.
- The proposed SurvNet model is trained in a multi-task learning framework to jointly learn across three related tasks: input reconstruction, survival classification, and Cox regression.
- A new real-world dataset is collected to evaluate the performance of the prognosis prediction models for IB-IIA stage non-small lung cancer.

The proposed method is compared with the traditional Cox model and Cox-Net in the experiments. The experiment results demonstrate that the proposed SurvNet outperforms the other models with a much higher concordance index ( $C_{index}$ ). The difference between high and low risk groups obtained by SurvNet is more significant than that of high and low risk groups obtained by the other models. Furthermore, it achieves better performance on incomplete data with missing values and better generalization performance on Surveillance, Epidemiology, and End Results Program (SEER) dataset.

## MATERIALS AND METHODS

### Datasets

In this study, we collected the data of 1,280 patients with IB-IIA stage non-small cell lung cancer at West China Hospital, from 2005 to 2018. There are 1,137 patients remaining after the exclusion of patients with unknown survival time. Of the 1,137 patients, 346 died and the others are missing follow-up or still alive. The survival time of the patients is in the range of (1,215) months. **Figure 1** shows the Kaplan-Meier estimation of the dataset. Clearly, the five-year survival probability of the patients at the IB-IIA stage is relatively high. However, 42, 97, 160, 219, and 263 patients died in 1-, 2-, 3-, 4-, and 5-year respectively. Distinguishing the IB-IIA stage patients with a high risk of



recurrence or death from low-risk patients is worthwhile for guiding further treatment and improving the non-small lung cancer prognosis.

In this study, nine clinicopathologic variables are taken into account for prognosis prediction. The distribution of the variables in the final patient series used in the study is shown in **Table 1**. In practice, it is difficult to ensure all of the variables were recorded for each patient. As illustrated in **Table 1**, there are lots of missing values (denoted as “unknown”). Of 1,137 patients, 961 contains missing values (at least one of the clinicopathologic variables is missing). It brings a great challenge for prognosis prediction models.

Of nine clinicopathologic variables, age and tumor size are continuous variables, whereas the others are encoded using

**TABLE 1** | Distribution of clinicopathologic variables in our datasets of patients with IB-IIA stage lung cancer. Missing values are denoted as “unknown”.

Clinicopathologic variables	Value	Num of patients
Age	Range (18–86)	1,137
Tumor Size	Range (0.3–5)	778
	<b>Unknown</b>	359
Sex	Female	670
	Male	467
Tumor Location	Left Upper	287
	Left Lower	186
	Right Upper	367
	Right Middle	87
	Right Lower	210
Differentiation	Lower	501
	Middle or High	542
	<b>Unknown</b>	94
Cancer Type	Adenocarcinoma	843
	Squamous carcinoma	294
Lymph node management	Dissection	1,015
	Biopsy or No Management	97
	<b>Unknown</b>	25
Pleura Immersion	Yes	656
	No	87
	<b>Unknown</b>	394
Operation	Lobectomy	53
	Sub Lobectomy	1,084

discrete values, for example, -1 for females and 1 for males. The missing values are filled with zeros. Thus, the clinicopathologic variables of each patient are represented by a 9 dimension vector. Formally, the proposed dataset can be formulated into a set of triplets  $\{(x_i, s_i, t_i) \mid i = 1, 2, \dots, n\}$ , where  $n$  is the number of the patients,  $x_i \in R^9$  is a vector of 9 clinicopathologic variables that describes the  $i$ -th patient,  $s_i$  is the patient’s end state, that is, 1 for dead or 0 for alive, and  $t_i$  is the patient’s survival time.

### Problem Definition

In this study, the aim is to dichotomize the patients in the dataset into high and low risk groups according to their prognosis index. This can be formulated as:

$$d_x = \begin{cases} 1, & \text{if } p_x > PI \\ 0, & \text{if } p_x \leq PI \end{cases} \quad (1)$$

where  $PI$  is a constant and the prognosis index  $p_x$  is calculated by

$$p_x = F(x) \quad (2)$$

where  $F$  is generally a complex non-linear function. Thus, the core task of the prognosis prediction task is to determine a suitable function  $F$ .

However, it is difficult to estimate the prognosis index function  $F$  for fine-grained prognosis prediction of IB-IIA stage lung cancer. There are three challenges:

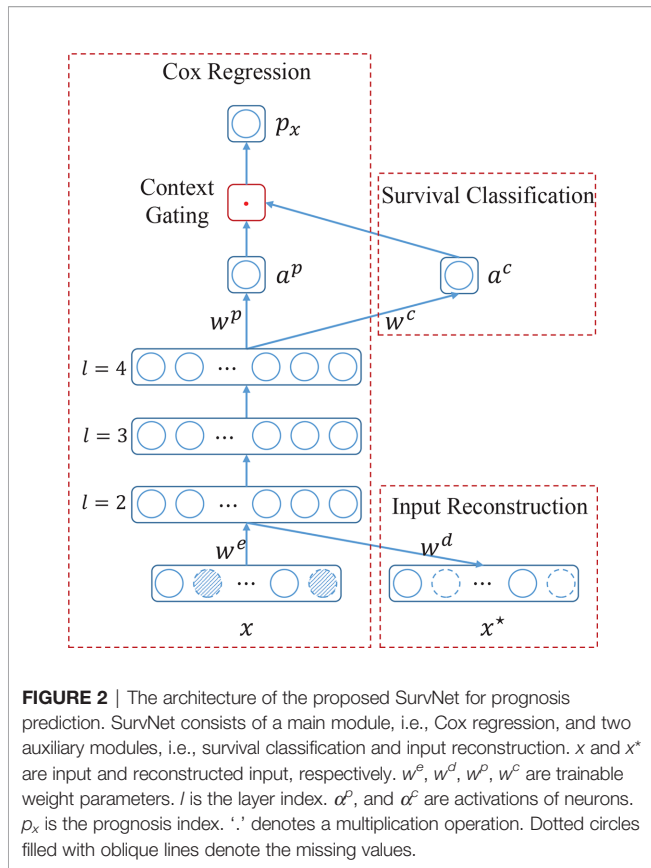
- For most patients, the end event (death) has not yet happened. It is known as censoring. In other words, we could not get actual survival times for these patients.
- Most of the patients contain at least one missing value. Omitting patients with missing values may bias the result, whereas using patients with missing values may harm the performance. It is a great obstacle for machine learning methods (30, 31).
- The distinguishing feature is difficult to learn for patients with IB-IIA stage lung cancer.

### Multi-Task Based SurvNet for Prognosis Prediction

To overcome the aforementioned difficulties, in this study, a novel multi-task based neural network namely SurvNet is proposed for the prognosis prediction of IB-IIA stage lung cancer. As **Figure 2** illustrated, the proposed SurvNet consists of three modules: Cox regression module, survival classification module, and input reconstruction module. Cox regression module is the main backbone of SurvNet and is used to represent the Function  $F$  for prognosis prediction. Survival classification module and input reconstruction module are auxiliary modules that aim to improve the performance of SurvNet for fine-grained prognosis prediction on incomplete data with missing values.

#### Input Reconstruction Module

The missing value is a common phenomenon in survival analysis. The suitable method to deal with the missing values is always desired to improve the performance of prognosis prediction models. In the proposed SurvNet, we use zeros to



**FIGURE 2** | The architecture of the proposed SurvNet for prognosis prediction. SurvNet consists of a main module, i.e., Cox regression, and two auxiliary modules, i.e., survival classification and input reconstruction.  $x$  and  $x^*$  are input and reconstructed input, respectively.  $w^e, w^d, w^p, w^c$  are trainable weight parameters.  $l$  is the layer index.  $a^p$ , and  $a^c$  are activations of neurons.  $p_x$  is the prognosis index. ‘.’ denotes a multiplication operation. Dotted circles filled with oblique lines denote the missing values.

fill the missing values and then learn the latent feature of incomplete data by input reconstruction.

Given an input  $x$ , it is first encoded into latent feature vector  $a^2$  (the output of layer 2). Then  $a^2$  would be feed into input reconstruction module to be decoded into  $x^*$ , a reconstruction of input  $x$ . Formally, it can be formulated as:

$$\begin{cases} a^2 = f(w^e x) \\ x^* = w^d a^2, \end{cases} \quad (3)$$

where  $w^e$  and  $w^d$  are encoder and decoder weights,  $x$  and  $x^*$  are input and reconstruction respectively.  $f$  denotes the non-linear activation function. Actually, the input layer, 2<sup>nd</sup> layer, and the input reconstruction layer composes an autoencoder network (See **Figure 2**).

Generally, mean square error (MSE) is used to make  $x^*$  approximate  $x$  as accurately as possible. However, it is not suitable for input with missing values since we do not know the true values at such locations and do not want to reconstruct a new vector with missing values too. To address this problem, an incomplete-aware cost function is proposed to learn the latent feature of incomplete data with missing values.

Let the binary vector  $r$  denote the locations of missing values in input  $x$ :

$$r(j) = \begin{cases} 0, & \text{the } j\text{-th element of } x \text{ is missing} \\ 1, & \text{otherwise} \end{cases} \quad (4)$$

The proposed incomplete-aware cost function is formulated as:

$$J_{re} = \frac{1}{2n} \sum_{i=1}^n \|r_i \cdot (x_i^* - x_i)\|_2 \quad (5)$$

where  $i$  is the sample index and ‘.’ denotes the elementwise product.

### Survival Classification Module

To fully  $x_p$ , utilize the relationship among the input variables  $x_p$ , end state  $s_i$  and survival time  $t_i$ , we introduce an auxiliary survival classification module to the SurvNet. The output of this module  $a^c$  denotes the probability of the patient living over  $T$  years or not. In other words, it learns a patient’s survival probability at some fixed time point  $T$ . Formally,  $a^4$  could be calculated as:

$$a^c = \sigma(w^c a^4) \quad (6)$$

where  $w^c$  and  $a^4$  are weight connection and output of layer 4, and  $\sigma(z) = 1/(1 + \exp(-z))$ . In this study, layer 3 is a batch norm layer.  $a^4$  could be calculated as:

$$\begin{cases} a^4 = \tanh(w^3 a^3) \\ a^3 = \frac{a^2 - E[a^2]}{\sqrt{\text{Var}[a^2] + \epsilon}}, \end{cases} \quad (7)$$

where  $w^3$  is weight connection and  $\tanh(z) = (\exp(-z) - \exp(z))/(\exp(-z) + \exp(z))$ .  $E$  and  $\text{Var}$  denote the expectation and variance, respectively.  $\epsilon$  is a small constant.

In this manuscript, the learning of survival classification module could be formulated as a binary classification task, which aims to minimize the following cost function:

$$J_c = \sum_i^n \delta_i (d_i \cdot \log(a^c) + (1 - d_i) \cdot (\log(1 - a^c))). \quad (8)$$

where  $d_i$  is the survival state of the patient  $i$ , which is defined as

$$d_i = \begin{cases} 1, & \text{if } t_i > T \\ 0, & \text{if } t_i < T \text{ and } s_i == 1, \end{cases} \quad (9)$$

and  $\delta_i$  denotes whether a patient is a valid sample classification task and it is defined as

$$\delta_i = \begin{cases} 1, & \text{if } t_i < T \text{ and } s_i == 0 \\ 0, & \text{otherwise.} \end{cases} \quad (10)$$

It means that patients censored before  $T$  are ignored.

### Cox Regression Module

This is the main backbone of the proposed SurvNet. As **Figure 2** illustrated, it consists of several successive feedforward layers, such as fully connected layer, batch normalization (32), and dropout layer, and a new context gating submodule that does not exist in the traditional Cox-Net (12).

Given an input  $x$ , the high-level representation  $a^L$  could be calculated layer by layer. Specifically, the activation  $a^p$  is computed as

$$a^p = w^p a^L. \quad (11)$$

In the traditional Cox-Net (12),  $\alpha^p$  would be expressed as log hazard ratio in Cox regression. However, in the proposed SurvNet, the distribution of log hazard ratio  $\alpha^p$  is adjusted by survival probability  $\alpha^c$  by using context gating mechanism:

$$p_x = a^p \cdot a^c. \tag{12}$$

Then, we take  $p_x$  as log hazard ratio in Cox regression and use the following log partial likelihood for Cox regression:

$$J_{cox} = \sum_{i: s_i=1}^n \left( p_{x_i} - \log \left( \sum_{j: t_j \leq t_i} p_{x_j} \right) \right), \tag{13}$$

where  $i$  and  $j$  are sample indexes.

The context gating mechanism is inspired by the attention mechanism where the input is adjusted by the attention coefficient. It is notable that the proposed context gating mechanism bridges the gap between Cox regression and survival classification to improve the performance. On the one hand, in the Cox regression, it is supposed that the larger the survival time, the larger the prognosis index. It reveals that the prognosis indexes of patients that are alive at some fixed time point  $T$  should be larger than the patients that died at that time point. On the other hand, the survival classification aims to predict the survival state of a given patient at a fixed time point  $T$ . As Equation (12) illustrates, the survival prediction  $\alpha^c$  serves as a context coefficient that adjusts the hazard ratio  $\alpha^p$  automatically thus to produce a better prognosis index that has good distribution at time point  $T$ .

### Multi-Task Learning

To train the proposed SurvNet, three learning tasks are optimized synchronously. The final cost function could be formulated as:

$$J = \alpha J_{cox} + \beta J_C + \gamma J_{re}, \tag{14}$$

where  $\alpha$ ,  $\beta$ , and  $\gamma$  are the coefficients that balance the Cox regression, survival classification, and input reconstruction tasks. By using gradient-based algorithms, the (local) minimal of cost function  $J$  could be found iteratively.

## EXPERIMENTS

### Evaluation Metrics

To evaluate the performance of the proposed model, two metrics were used. One is the Harrell’s concordance index ( $C_{index}$ ), which is valued from 0 to 1. It is an extension of the area under the receiver operating characteristic curve to censored time-to-event data (16, 26).

Generally, it is defined as

$$C_{index} = \frac{\sum_{i,j} s_i \cdot I(p_i, p_j) \cdot I(t_i, t_j)}{\sum_{i,j} s_i \cdot I(t_i, t_j)} \tag{15}$$

Where  $i$  and  $j$  are sample indexes.  $s$ ,  $t$ , and  $p$  are end state, survival time and hazard ratio of a given sample, respectively.  $I(z_1, z_2)$  is defined as:

$$I(z_1, z_2) = \begin{cases} 1, & \text{if } z_1 < z_2 \\ 0, & \text{otherwise.} \end{cases} \tag{16}$$

The other metric is the survival analysis with the log-rank test. Kaplan-Meier survival curves are generated by dichotomizing all patients in the testing dataset into low-risk and high-risk groups *via* the median hazard ratio. The corresponding log-rank p-value indicates the ability of the model to differentiate two risk groups. The lower the p-values, the better the model performance.

## Running Configuration

### Datasets

To train the prognosis models, the presented dataset was randomly split into train set (682 patients), validation set (227 patients), and test set (228 patients). Furthermore, we also obtained a SEER dataset (9,534 patients) by selecting the IB-IIA stage lung cancer patients from SEER to test the generalization performance of the models.

### Models

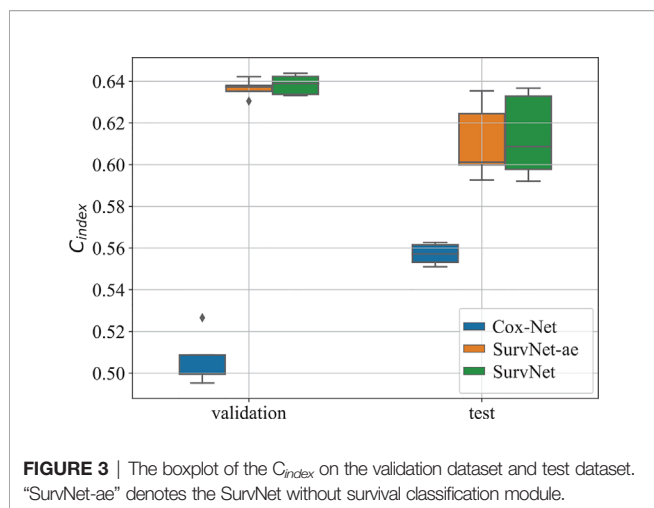
The proposed SurvNet was compared with the traditional Cox proportional hazards model and neural network extended Cox model (Cox-Net). For a fair comparison, the Cox-Net shared the same architecture with the Cox regression module in SurvNet except for the context gating module. The network settings is presented in **Table 2**. Besides, we set  $T = 36$  for SurvNet and the coefficients  $\alpha$ ,  $\beta$ , and  $\gamma$  were set to 0.2, 1, 3, respectively. The RMSProp (33) with default learning parameters in Pytorch was used as the optimizer and the weight decay was set to 0.00001. All of the networks run 100 epochs with batch size 64. For each run, the weight parameters that achieved the best  $C_{index}$  on the validation dataset were used to evaluate the performance of the model on the test dataset.

### Performance on Our Dataset

To eliminate the influence of initial values of neural networks, we run Cox-Net, SurvNet, and SurvNet-ae (SurvNet without survival classification module) five times. For each running, the model with the highest  $C_{index}$  on the validation dataset is selected to evaluate the performance on the test dataset. The boxplot of the  $C_{index}$  is presented in **Figure 3**. It demonstrates that the proposed SurvNet with and without survival classification module outperformed the Cox-Net significantly by the using input reconstruction module to learning the latent feature of incomplete data with missing values. And the proposed survival classification module further improves the network’s performance. Besides, the best  $C_{index}$  of traditional Cox model, Cox-Net, and the proposed SurvNet are 0.5612, 0.5627, 0.6367,

**TABLE 2** | Network settings for Cox-Net and SurvNet.

Layer	Neurons
1	9 (input)
2	64 (tanh)
3	Batch normalization
4	32 (tanh)



**FIGURE 3** | The boxplot of the  $C_{index}$  on the validation dataset and test dataset. “SurvNet-ae” denotes the SurvNet without survival classification module.

respectively. The proposed SurvNet outperforms the other models significantly.

Furthermore, by interpreting the outputs of the models as the log hazard ratio, two groups (high risk and low risk) are obtained by using Eq. (1) where  $PI$  was set to the median of log hazard ratios. The Kaplan-Meier estimation of the Cox model, Cox-Net and the proposed SurvNet on the test dataset are presented in **Figure 4**. The log-rank  $p$ -values (the lower the better) of the three methods are 0.293, 0.072, 0.002. It is obvious that the difference between high and low risk groups obtained by neural network based models is more significant than that of high and low risk groups obtained by the Cox model. Moreover, the difference between high and low risk groups obtained by SurvNet is most significant. It demonstrates that the proposed SurvNet achieves the best performance.

In addition, the distribution of survival times of patients in each group is presented in **Figure 5**. The proposed SurvNet achieves the largest median survival time for low risk group and the lowest median survival time for high risk group. It demonstrates that the proposed method improves the performance of fine-grained prognosis prediction for IB-IIA stage non-small cell lung cancer.

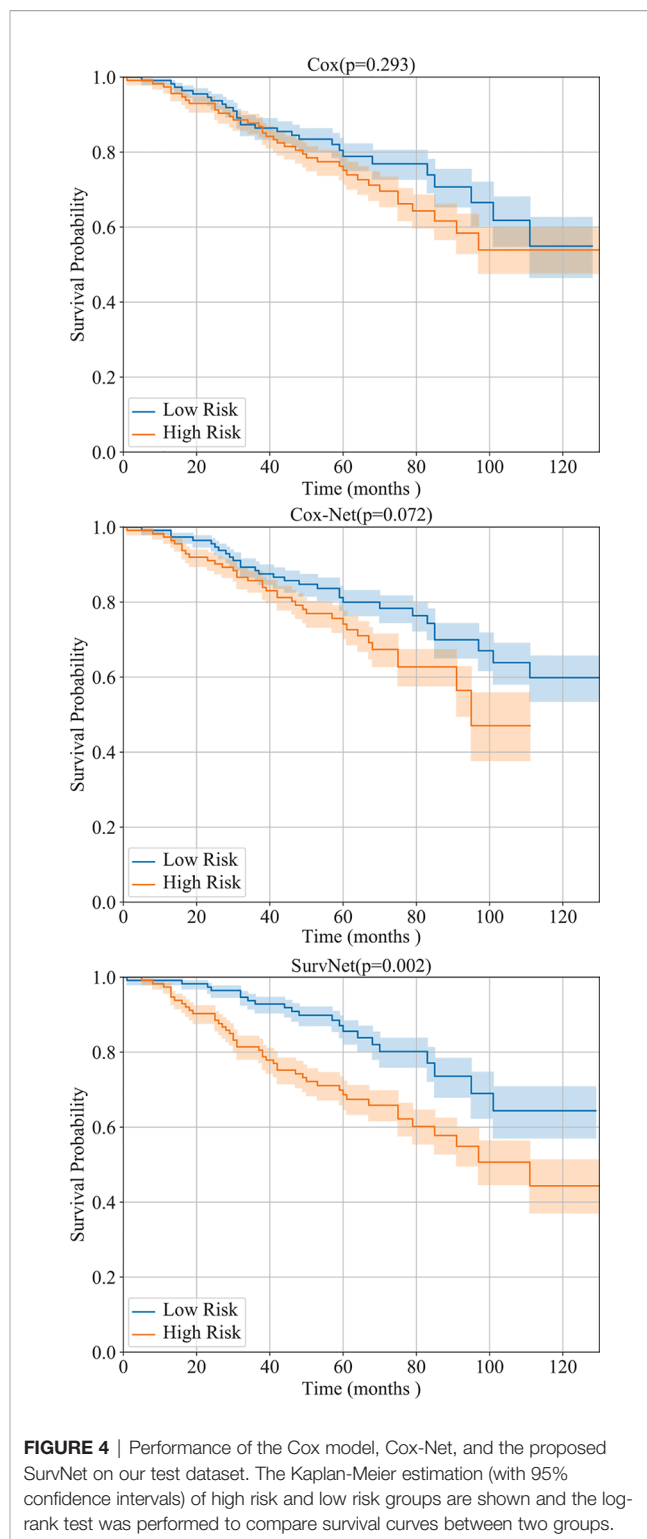
## Robustness on Missing Values

To further evaluate the robustness of prognosis models on incomplete data with missing values, we randomly zeroed the values of the input vector in the test dataset with drop probability  $dp$  and then evaluated the performance of the trained models. For each drop probability  $dp$ , we run each model 100 times.

The boxplot of  $C_{index}$  is illustrated in **Figure 6**. As drop probability gets large, the performance of the three models gets worse. Notably, the proposed SurvNet performed more stable and the  $C_{index}$  of the SurvNet is always larger than that of the Cox model and Cox-Net significantly.

## Generalization Performance on SEER Dataset

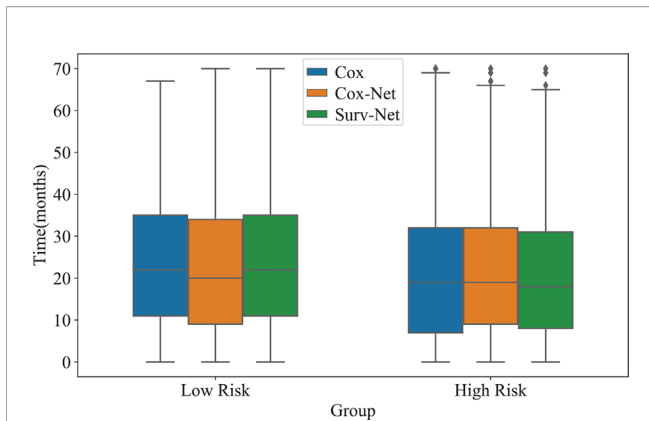
The generalization performance is an important measurement of prognosis models. SEER dataset has been widely used in the



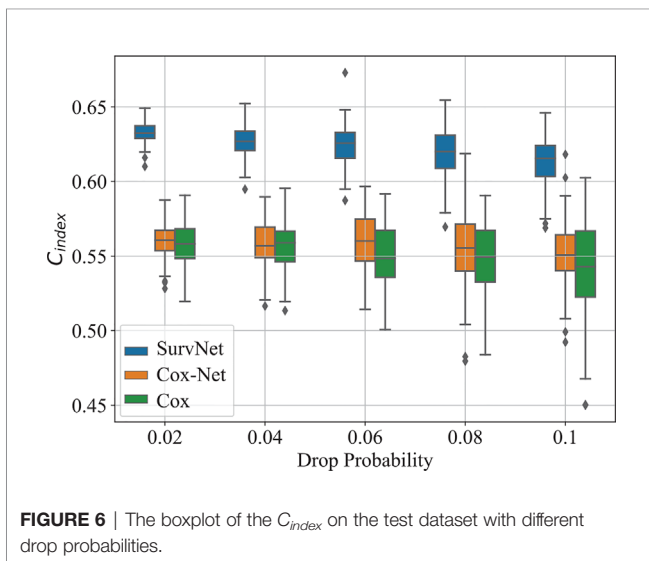
**FIGURE 4** | Performance of the Cox model, Cox-Net, and the proposed SurvNet on our test dataset. The Kaplan-Meier estimation (with 95% confidence intervals) of high risk and low risk groups are shown and the log-rank test was performed to compare survival curves between two groups.

literature. In this study, we focused on the IB-IIA stage non-small lung cancer and obtained a dataset of 9,534 patients.

We evaluated the models, which have been trained using our dataset, on the obtained SEER dataset. The  $C_{index}$  of Cox model, Cox-Net, and SurvNet are 0.5955, 0.5617, and 0.6003,



**FIGURE 5** | The distribution of survival time of high-risk group and low risk group obtained by three models.

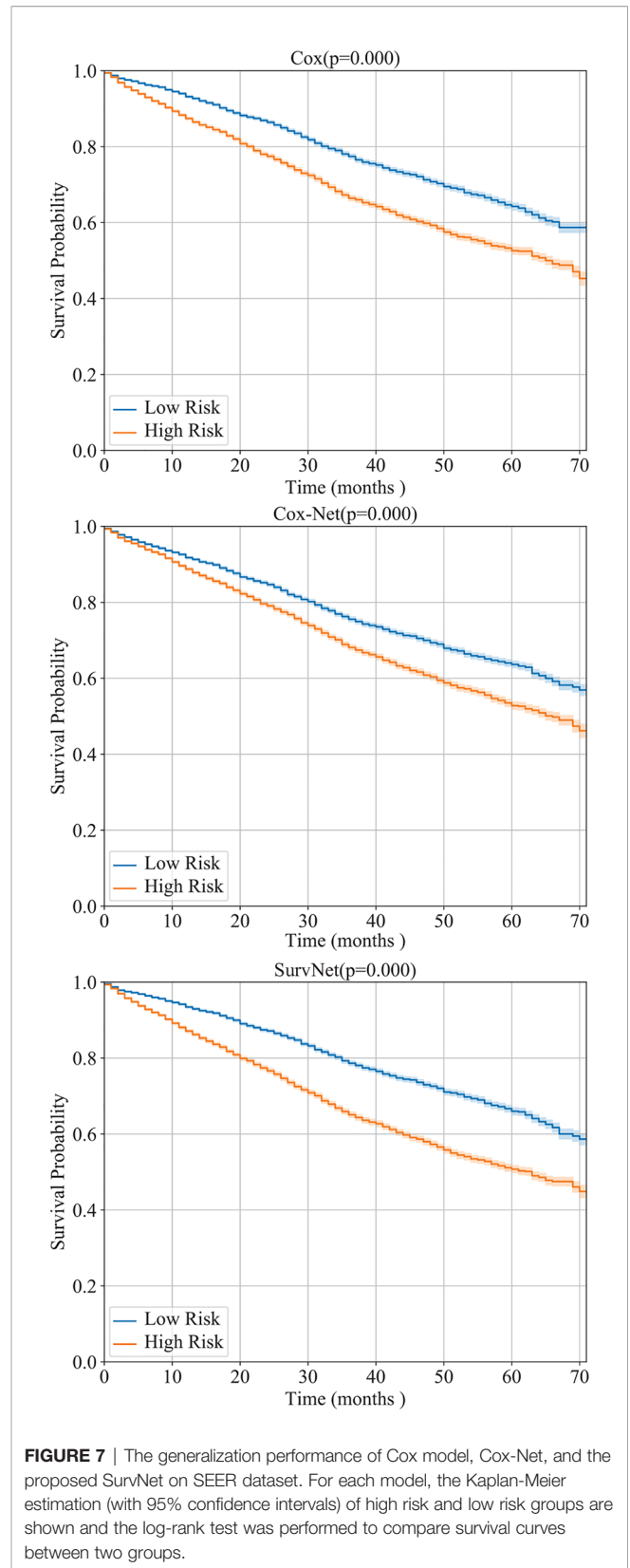


**FIGURE 6** | The boxplot of the  $C_{index}$  on the test dataset with different drop probabilities.

respectively. Besides, as the Kaplan-Meier estimation presented in **Figure 7** shows, the difference between high and low risk groups obtained SurvNet is more significant than that of high and low risk groups obtained by other two models. The proposed SurvNet achieves better generalization performance than the Cox model and Cox-Net.

## CONCLUSION

Prognosis prediction for IB-IIA stage lung cancer is important for improving the accuracy of the management of lung cancer. In this study, a new real-world dataset is collected and a novel multi-task based neural network, SurvNet, is proposed to further improve the prognosis prediction for IB-IIA stage lung cancer. In the proposed SurvNet, the input reconstruction module overcomes the problems by missing values and the proposed context gating mechanism



**FIGURE 7** | The generalization performance of Cox model, Cox-Net, and the proposed SurvNet on SEER dataset. For each model, the Kaplan-Meier estimation (with 95% confidence intervals) of high risk and low risk groups are shown and the log-rank test was performed to compare survival curves between two groups.

could bridge the gap between Cox regression and survival classification. By training in a multi-task framework, the proposed SurvNet outperforms the traditional Cox model and Cox-Net significantly. It achieved higher  $C_{index}$ s and lower  $p$ -values on the proposed dataset and better generalization performance on the SEER dataset. It is apparent to be a promising method for survival analysis tasks. A limitation of the proposed SurvNet may lie on the survival classification module which just considers survivals on some fixed time point rather than a set of non-overlap time intervals. Future work will be focused on how to integrate survival classification module that classifies the survivals into a set of time intervals with the Cox regression module to further improve the performance on prognosis prediction.

## DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

## REFERENCES

- Pölsterl S, Conjeti S, Navab N, Katouzian A. Survival analysis for high-dimensional, heterogeneous medical data: Exploring feature extraction as an alternative to feature selection. *Artif Intell Med* (2016) 72:1–11. doi: 10.1016/j.artmed.2016.07.004
- Jing B, Zhang T, Wang Z, Jin Y, Liu K, Qiu W, et al. A deep survival analysis method based on ranking. *Artif Intell Med* (2019) 98:1–9. doi: 10.1016/j.artmed.2019.06.001
- Wang Y, Wang D, Ye X, Wang Y, Yin Y, Jin Y. A tree ensemble-based two-stage model for advanced-stage colorectal cancer survival prediction. *Inf Sci* (2019) 474:106–24. doi: 10.1016/j.ins.2018.09.046
- Rami-Porta R, Bolejack V, Crowley J, Ball D, Kim J, Lyons G, et al. The iaslc lung cancer staging project: proposals for the revisions of the t descriptors in the forthcoming eighth edition of the tnm classification for lung cancer. *J Thoracic Oncol* (2015) 10:990–1003. doi: 10.1097/JTO.0000000000000559
- Kim S, Park T, Kon M. Cancer survival classification using integrated data sets and intermediate information. *Artif Intell Med* (2014) 62:23–31. doi: 10.1016/j.artmed.2014.06.003
- Zupan B, Demšar J, Kattan MW, Beck J, Bratko I. Machine learning for survival analysis: a case study on recurrence of prostate cancer. *Artif Intell Med* (2000) 20:59–75. doi: 10.1016/S0933-3657(00)00053-1
- Goldstraw P, Chansky K, Crowley J, Rami-Porta R, Asamura H, Eberhardt WE, et al. The iaslc lung cancer staging project: proposals for revision of the tnm stage groupings in the forthcoming (eighth) edition of the tnm classification for lung cancer. *J Thoracic Oncol* (2016) 11:39–51. doi: 10.1016/j.jtho.2015.09.009
- Detterbeck FC, Boffa DJ, Kim AW, Tanoue LT. The eighth edition lung cancer stage classification. *Chest* (2017) 151:193–203. doi: 10.1016/j.chest.2016.10.010
- Lundin M, Lundin J, Burke HB, Toikkanen S, Pylkkanen L, Joensuu H. Artificial neural networks applied to survival prediction in breast cancer. *Oncology* (1999) 57:281–6. doi: 10.1159/000012061
- Cox DR. Regression models and life-tables. *J R Stat Society: Ser B (Methodological)* (1972) 34:187–202. doi: 10.1111/j.2517-6161.1972.tb00899.x
- Lin H, Zelterman D. *Modeling survival data: extending the Cox model*. Taylor & Francis (2002). doi: 10.1198/tech.2002.s656
- Ching T, Zhu X, Garmire LX. Cox-nnet: An artificial neural network method for prognosis prediction of high-throughput omics data. *PLoS Comput Biol* (2018) 14:e1006076. doi: 10.1371/journal.pcbi.1006076
- Chi C-L, Street WN, Wolberg WH. Application of artificial neural network-based survival analysis on two breast cancer datasets. In: *AMIA Annual*

## AUTHOR CONTRIBUTIONS

ZY and LL designed the research. ZY, JW, JG, and XX implemented the proposed method and analyzed the data. LL and NC were responsible for the dataset. JW wrote the manuscript. All authors discussed the results and commented on the manuscript. All authors contributed to the article and approved the submitted version.

## FUNDING

This work was supported by the National Natural Science Foundation of China (grant no. 61906127), the National Major Science and Technology Projects (grant no. 2018AAA0100201), and the Major Scientific and Technological Projects of the New Generation of Artificial Intelligence in Sichuan Province in 2018 (grant no. 2018GZDZX0035).

*Symposium Proceedings*, vol. 2007. American Medical Informatics Association (2007). p. 130.

- Wang J, Ju R, Chen Y, Zhang L, Hu J, Wu Y, et al. Automated retinopathy of prematurity screening using deep neural networks. *EBioMedicine* (2018a) 35:361–8. doi: 10.1016/j.ebiom.2018.08.033
- Silver D, Schrittwieser J, Simonyan K, Antonoglou I, Huang A, Guez A, et al. Mastering the game of go without human knowledge. *Nature* (2017) 550:354. doi: 10.1038/nature24270
- Raghavendra U, Fujita H, Bhandary SV, Gudigar A, Tan JH, Acharya UR. Deep convolution neural network for accurate diagnosis of glaucoma using digital fundus images. *Inf Sci* (2018) 441:41–9. doi: 10.1016/j.ins.2018.01.051
- Cho K, van Merriënboer B, Gulcehre C, Bahdanau D, Bougares F, Schwenk H, et al. Learning Phrase Representations using RNN Encoder Decoder for Statistical Machine Translation. In: *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing*. In: *EMNLP*. Stroudsburg, PA, USA: Association for Computational Linguistics (2014). p. 1724–34. doi: 10.3115/v1/D14-1179
- Vinyals O, Toshev A, Bengio S, Erhan D. Show and tell: A neural image caption generator. In: *Proceedings of the IEEE conference on computer vision and pattern recognition* (2015). p. 3156–64. doi: 10.1109/CVPR.2015.7298935
- Wang J, Zhang L, Guo Q, Yi Z. Recurrent Neural Networks With Auxiliary Memory Units. *IEEE Trans Neural Networks Learn Syst* (2018c) 29:1652–61. doi: 10.1109/TNNLS.2017.2677968
- Wang J, Zhang L, Chen Y, Yi Z. A new delay connection for long short-term memory networks. *Int J Neural Syst* (2018b) 28:1750061. doi: 10.1142/S0129065717500617
- Acharya UR, Fujita H, Oh SL, Hagiwara Y, Tan JH, Adam M. Application of deep convolutional neural network for automated detection of myocardial infarction using eeg signals. *Inf Sci* (2017) 415:190–8. doi: 10.1016/j.ins.2017.06.027
- Ng E, Acharya UR, Keith LG, Lockwood S. Detection and differentiation of breast cancer using neural classifiers with first warning thermal sensors. *Inf Sci* (2007) 177:4526–38. doi: 10.1016/j.ins.2007.03.027
- Daoud H, Bayoumi M. Efficient Epileptic Seizure Prediction based on Deep Learning. *IEEE Trans Biomed Circuits Syst* (2019) 13(5):804–13. doi: 10.1109/TBCAS.2019.2929053
- Bello GA, Dawes TJ, Duan J, Biffi C, de Marvao A, Howard LS, et al. Deep-learning cardiac motion analysis for human survival prediction. *Nat Mach Intell* (2019) 1:95. doi: 10.1038/s42256-019-0019-2
- Ching T, Zhu X, Garmire L. Cox-nnet: an artificial neural network cox regression for prognosis prediction. *BioRxiv* (2016) 093021. doi: 10.1101/093021



26. Huang Z, Zhan X, Xiang S, Johnson TS, Helm B, Yu CY, et al. Salmon: Survival analysis learning with multi-omics neural networks on breast cancer. *Front Genet* (2019) 10:166. doi: 10.3389/fgene.2019.00166
27. Burke HB, D. B. R, Goodman PH. Artificial neural networks improve the accuracy of cancer survival prediction. *Cancer* (1997) 79:857–62. doi: 10.1002/(SICI)1097-0142(19970215)79:4<857::AID-CNCR24>3.0.CO;2-Y
28. Park K, Ali A, Kim D, An Y, Kim M, Shin H. Robust predictive model for evaluating breast cancer survivability. *Eng Appl Artif Intell* (2013) 26:2194–205. doi: 10.1016/j.engappai.2013.06.013
29. Kalderstam J, Edén P, Bendahl P-O, Strand C, Fernö M, Ohlsson M. Training artificial neural networks directly on the concordance index for censored data using genetic algorithms. *Artif Intell Med* (2013) 58:125–32. doi: 10.1016/j.artmed.2013.03.001
30. Deb R, Liew AW-C. Missing value imputation for the analysis of incomplete traffic accident data. *Inf Sci* (2016) 339:274–89. doi: 10.1016/j.ins.2016.01.018
31. Aydilek IB, Arslan A. A hybrid method for imputation of missing values using optimized fuzzy c-means with support vector regression and a genetic algorithm. *Inf Sci* (2013) 233:25–35. doi: 10.1016/j.ins.2013.01.021
32. Ioffe S, Szegedy C. Batch normalization: Accelerating deep network training by reducing internal covariate shift. *In Int Conf Mach Learning* (2015) 448–56.
33. Zou F, Shen L, Jie Z, Zhang W, Liu W. A sufficient condition for convergences of adam and rmsprop. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (2019). p. 11127–35. doi: 10.1109/CVPR.2019.01138

**Conflict of Interest:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2021 Wang, Chen, Guo, Xu, Liu and Yi. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.