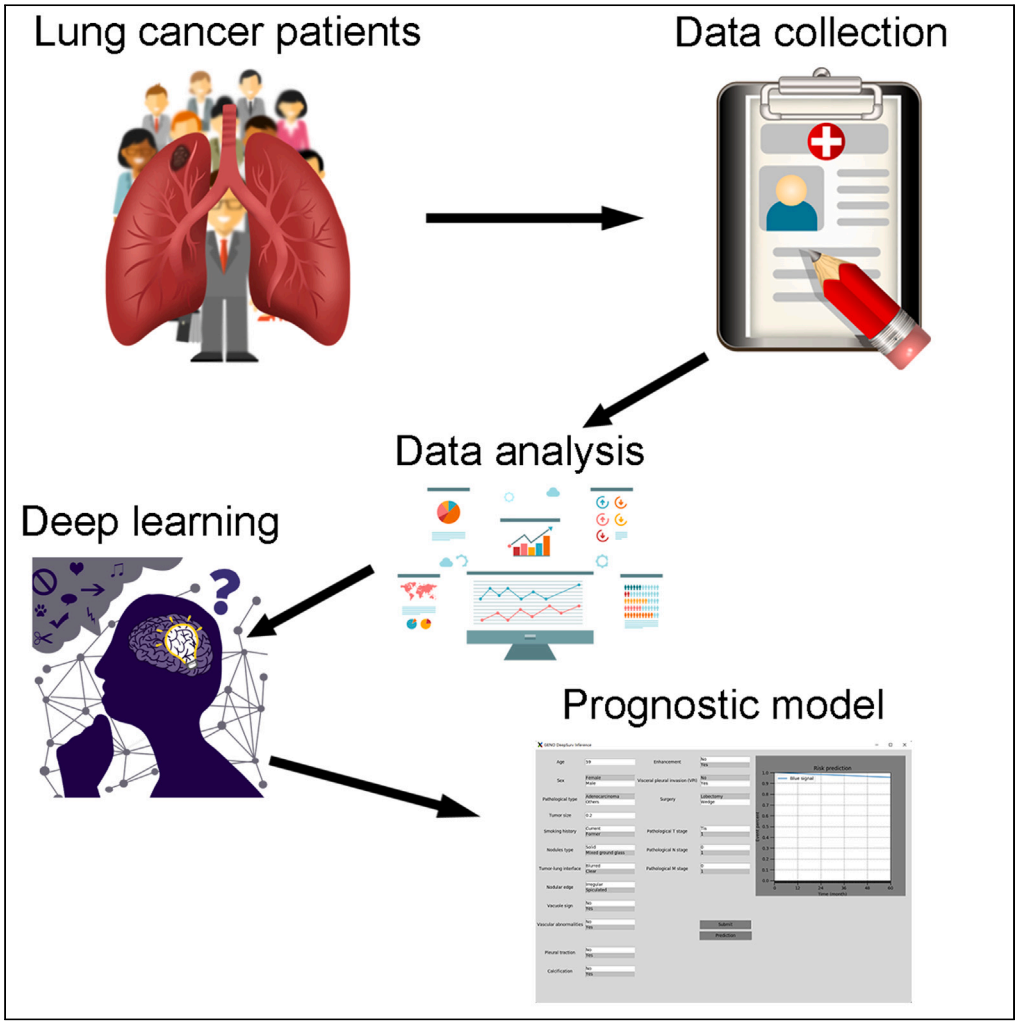
iScience

CellPress
OPEN ACCESS

Article

The predictive value of modified-DeepSurv in overall survivals of patients with lung cancer



Jie Lei, Xin Xu, Junrui Xu, ..., Renquan Zhang, Zhemin Zhang, Tao Jiang

zhangrenquan@live.cn (R.Z.)
zhemindoc@163.com (Z.Z.)
jiangtaochest@163.com (T.J.)

Highlights

Modified-DeepSurv is more consistent with the patients in the real medical environment

The performance of Modified-DeepSurv is superior to that of the conventional CPH model

The Modified-DeepSurv can effectively predict the survival of lung cancer patients

The Modified-DeepSurv visualization was realized by a user-friendly graphic interface

Lei et al., iScience 26, 108200
November 17, 2023
https://doi.org/10.1016/j.isci.2023.108200

iScience

CellPress
OPEN ACCESS

Article

The predictive value of modified-DeepSurv in overall survivals of patients with lung cancer

Jie Lei,[1,8] Xin Xu,[2,8] Junrui Xu,[3,8] Jia Liu,[4] Yi Wang,[5] Chao Wu,[6,7] Renquan Zhang,[3,*] Zhemin Zhang,[2,*] and Tao Jiang[1,9,*]

SUMMARY

The traditional prognostic model may induce the possibility of incorrect assessment of mortality risk under the assumption of linearity. It is urgent to develop a non-linearity precise prognostic model for achieving personalized medicine in lung cancer. In our study, we develop and validate a prognostic model "Modified-DeepSurv" for patients with lung carcinoma based on deep learning and evaluate its value for prognosis, while Cox proportional hazard regression was used to develop another model "CPH." The C-index of the Modified-DeepSurv and CPH was 0.956 (95% confidence interval [CI]: 0.946–0.974) and 0.836 (95% CI: 0.774–0.896), respectively, in the training cohort, while the C-index of the Modified-DeepSurv and CPH was 0.932 (95%CI: 0.908–0.964) and 0.777 (95%CI: 0.633–0.919), respectively, in the test dataset. The Modified-DeepSurv model visualization was realized by a user-friendly graphic interface. Modified-DeepSurv can effectively predict the survival of lung cancer patients and is superior to the conventional CPH model.

INTRODUCTION

Lung cancer is the second most diagnosed cancer worldwide.[1] In China, the morbidity and mortality of lung cancer rank first among all types of cancer.[2] The prognosis of lung cancer patients with different clinical stages is different, and even the survival rate could vary within the same stage.[3,4] The tumor-node-metastasis (TNM) staging system in the 8th Edition of the American Joint Committee on Cancer (AJCC) is the routine method to predict prognosis.[5] Recently, it has been found that age, sex, histopathology, and treatment choices could be independent prognostic factors that significantly contribute to the individualized prediction of survival.[6] Therefore, it is urgent to construct a precise prognostic model for achieving personalized medicine and further improving the survival rates of lung cancer patients.[7]

Cox proportional hazard (CPH) is a semi-parametric model that calculates the effects of observed covariates on the risk of an occurring event.[8] The model was popular as a way of predicting outcomes.[9] In practice, most CPH models lack a fixed hazard ratio (HR) and ignore interactions between risk factors. These deficiencies may increase the possibility of incorrect mortality risk assessment under linearity assumptions. The nomogram based on CPH is a reliable tool that has demonstrated the ability to quantify risk factors by combining and clarifying significant clinical characteristics of oncology.[6,10] However, these models have several limitations for the precise evaluation of overall survival (OS) and progression-free survival (PFS).[11] Because the clinical characteristics in the real world are mostly nonlinear, these nomogram models were based on linearity assumptions rather than nonlinear analyses.[12] Therefore, nonlinear functions are required to fit survival data in the real world accurately and aim to improve survival models' performance.[11,13]

Artificial intelligence (AI), consisting of machine learning (ML) and deep learning (DL), has been applied to a lot of fields of medicine.[14–16] ML, a branch of AI that enables the detection of relationships from complex datasets, has recently been employed for the survival prediction of lung oncological outcomes. DeepSurv is an extension of DL-based survival analysis that combines a CPH model with a modern DL algorithm[11] and has been used to estimate the survival risks with a recommender system in multiple cancers.[14,17,18] She et al. conducted a study applying DeepSurv to the prognosis prediction of 17,322 patients with lung cancer,[19] which used a static framework-TensorFlow operation mode, to construct a calculation diagram of TensorFlow with a fixed state operational process. In contrast, considering the data characteristics of our study, we choose the

[1]Department of Thoracic Surgery, The Second Affiliated Hospital, Air Force Medical University, Xi'an 710038, China
[2]Department of Respiratory Medicine, Shanghai Pulmonary Hospital, School of Medicine, Tongji University, Shanghai 200433, China
[3]Department of Thoracic Surgery, First Affiliated Hospital of Anhui Medical University, Hefei 230022, China
[4]Operations Management Department, People's Hospital of Xinjiang Uygur Autonomous Region, No. 91 Tianchi Road, Tianshan District, Urumqi 830001, China
[5]Department of Epidemiology and Biostatistics, School of Public Health and Management, Wenzhou Medical University, Wenzhou 325035, China
[6]Department of Respiratory and Critical Care Medicine, People's Hospital of Xinjiang Uygur Autonomous Region, No. 91 Tianchi Road, Tianshan District, Urumqi 830001, China
[7]Xinjiang Clinical Research Center for Interstitial Lung Diseases, People's Hospital of Xinjiang Uygur Autonomous Region, No. 91 Tianchi Road, Tianshan District, Urumqi 830001, China
[8]These authors contributed equally
[9]Lead contact
*Correspondence: zhangrenquan@live.cn (R.Z.), zheminoc@163.com (Z.Z.), jiangtaochest@163.com (T.J.)
https://doi.org/10.1016/j.isci.2023.108200

Table 1. Baseline characteristics between training and test dataset (original, missing data reported)

	Overall (n = 1907)	Training (n = 1430)	Test (n = 477)	p value
Age (year, median [Q1-Q3])	59.0 [52.0–65.0]	59.0 [51.0–65.0]	60.0 [52.0–65.0]	0.355
Sex (n, %)				0.357
Female	1095 (57.4%)	812 (56.8%)	283 (59.3%)	
Male	812 (42.6%)	618 (43.2%)	194 (40.7%)	
Pathological type (n, %)				0.457
Adenocarcinoma	1779 (93.3%)	1330 (93.0%)	449 (94.1%)	
Others	128 (6.7%)	100 (7.0%)	28 (5.9%)	
Tumor size				0.596
(Median [Q1-Q3])	1.30 [1.00–2.00]	1.30 [1.00–2.00]	1.20 [0.800–2.00]	
Missing	469 (24.6%)	360 (25.2%)	109 (22.9%)	
Smoking history (n, %)				0.415
Never smoked	248 (13.0%)	193 (13.5%)	55 (11.5%)	
Former smoker	308 (16.2%)	234 (16.4%)	74 (15.5%)	
Current smoker	1309 (68.6%)	970 (67.8%)	339 (71.1%)	
Missing	42 (2.2%)	33 (2.3%)	9 (1.9%)	
Nodules type (n, %)				0.742
Mixed ground glass	304 (15.9%)	222 (15.5%)	82 (17.2%)	
Pure ground glass	438 (23.0%)	327 (22.9%)	111 (23.3%)	
Solid	981 (51.4%)	738 (51.6%)	243 (50.9%)	
Missing	184 (9.6%)	143 (10.0%)	41 (8.6%)	
Tumor-lung interface (n, %)				1
Blurred	1032 (54.1%)	772 (54.0%)	260 (54.5%)	
Clear	676 (35.4%)	506 (35.4%)	170 (35.6%)	
Missing	199 (10.4%)	152 (10.6%)	47 (9.9%)	
Nodular edge (n, %)				0.562
Irregular	425 (22.3%)	311 (21.7%)	114 (23.9%)	
Lobulated	126 (6.6%)	100 (7.0%)	26 (5.5%)	
Smooth	883 (46.3%)	660 (46.2%)	223 (46.8%)	
Spiculated	269 (14.1%)	203 (14.2%)	66 (13.8%)	
Missing	204 (10.7%)	156 (10.9%)	48 (10.1%)	
Vacuole sign (n, %)				0.485
No	1621 (85.0%)	1211 (84.7%)	410 (86.0%)	
Yes	152 (8.0%)	118 (8.3%)	34 (7.1%)	
Missing	134 (7.0%)	101 (7.1%)	33 (6.9%)	
Vascular abnormalities (n, %)				0.869
No	1619 (84.9%)	1214 (84.9%)	405 (84.9%)	
Yes	154 (8.1%)	114 (8.0%)	40 (8.4%)	
Missing	134 (7.0%)	102 (7.1%)	32 (6.7%)	
Pleural traction (n, %)				0.098
No	1121 (58.8%)	855 (59.8%)	266 (55.8%)	
Yes	650 (34.1%)	472 (33.0%)	178 (37.3%)	
Missing	136 (7.1%)	103 (7.2%)	33 (6.9%)	
Calcification (n, %)				0.351
No	1700 (89.1%)	1269 (88.7%)	431 (90.4%)	
Yes	71 (3.7%)	57 (4.0%)	14 (2.9%)	

(*Continued on next page*)

Table 1. *Continued*

	Overall (n = 1907)	Training (n = 1430)	Test (n = 477)	p value
Missing	136 (7.1%)	104 (7.3%)	32 (6.7%)	
Enhancement (n, %)				0.313
No	1511 (79.2%)	1136 (79.4%)	375 (78.6%)	
Yes	197 (10.3%)	141 (9.9%)	56 (11.7%)	
Missing	199 (10.4%)	153 (10.7%)	46 (9.6%)	
Visceral pleural invasion (VPI) (n, %)				0.849
No	1284 (67.3%)	978 (68.4%)	306 (64.2%)	
Yes	95 (5.0%)	71 (5.0%)	24 (5.0%)	
Missing	528 (27.7%)	381 (26.6%)	147 (30.8%)	
Surgery (n, %)				0.591
Lobectomy	1151 (60.4%)	854 (59.7%)	297 (62.3%)	
Segmentectomy	331 (17.4%)	252 (17.6%)	79 (16.6%)	
Wedge	344 (18.0%)	255 (17.8%)	89 (18.7%)	
Missing	4 (0.2%)	4 (0.3%)	0 (0%)	
T (n, %)				0.156
Tis	1211 (63.5%)	911 (63.7%)	300 (62.9%)	
1	205 (10.7%)	159 (11.1%)	46 (9.6%)	
2	16 (0.8%)	10 (0.7%)	6 (1.3%)	
3	23 (1.2%)	19 (1.3%)	4 (0.8%)	
4	299 (15.7%)	209 (14.6%)	90 (18.9%)	
Missing	153 (8.0%)	122 (8.5%)	31 (6.5%)	
N (n, %)				0.55
0	1565 (82.1%)	1175 (82.2%)	390 (81.8%)	
1	68 (3.6%)	47 (3.3%)	21 (4.4%)	
2	115 (6.0%)	84 (5.9%)	31 (6.5%)	
3	3 (0.2%)	2 (0.1%)	1 (0.2%)	
Missing	156 (8.2%)	122 (8.5%)	34 (7.1%)	
M (n, %)				0.89
0	1719 (90.1%)	1283 (89.7%)	436 (91.4%)	
1	39 (2.0%)	30 (2.1%)	9 (1.9%)	
Missing	149 (7.8%)	117 (8.2%)	32 (6.7%)	
Status (n, %)				0.447
0	1834 (96.2%)	1372 (95.9%)	462 (96.9%)	
1	73 (3.8%)	58 (4.1%)	15 (3.1%)	
Survival duration (month, median[Q1-Q3]	15.3 [9.20–24.3]	15.3 [9.20–24.3]	15.3 [9.20–24.4]	0.62

VPI: Visceral pleural invasion.

Pytorch operation mode, which is based on a dynamic framework. The operation process is optimally arranged for different values and can build protocols and results in a short time. Furthermore, only some studies compare the performance of traditional statistics (such as the well-known CPH model) with DL techniques in terms of the ability to predict survival for malignant lung tumors. Thus, we developed a Modified-DeepSurv model based on the Pytorch operation mode, which contains 18 variables to predict the survival of lung cancer patients, and compared it with the CPH model in terms of discrimination and calibration. Further, the weights of the variables in the model were also explored and ranked.

RESULTS

Baseline characteristics

A total of 2,521 lung cancer patients were included in this study. One-thousand nine-hundred and seven patients were included in the final analysis (1,430 in the training set, 477 in the test set), while 188 and 426 patients were excluded for missing survival information and ambiguous

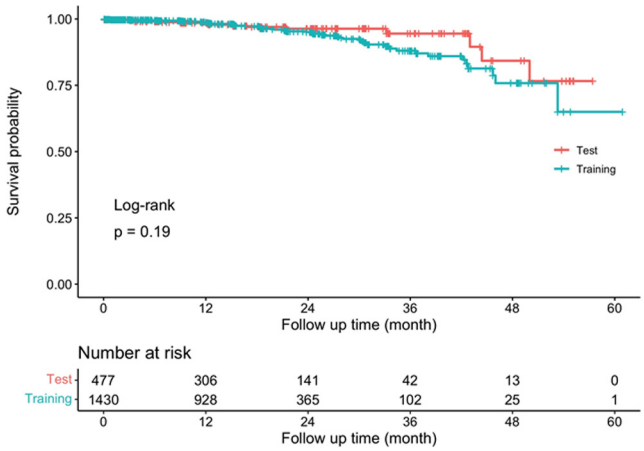


Figure 1. Kaplan-Meier curves of overall survival (OS) distribution of the training and test datasets

pathological diagnosis, respectively. In the training dataset, the median age of the subjects was 59.0 years, 43.2% of the subjects were men, and 93.0% had adenocarcinoma. In the test dataset, the median age of the subjects was 60.0 years, 40.7% were men, and 94.1% had adenocarcinoma. Solid nodules account for 51.6% of the training dataset and 50.9% of the test dataset. There were no significant differences in smoking history, maximum tumor diameter, tumor-lung interface, nodular edge, vacuole sign, vascular abnormalities, pleural traction, calcification, enhancement, visceral pleural invasion (VPI), surgery, pathological T stage, pathological N stage, pathological M stage, and survival duration between the two datasets (all $p > 0.05$). There was no significant difference between the two datasets after multiple imputations. The clinical and pathological characteristics of the training and test datasets were shown in Tables 1 and S1.

OS analysis and independent factors that affect OS

During the observation, 73 patients (3.8%) succumbed to lung cancer. In the training set, 58 cases (4.1%) died during follow-up, and the median survival duration was 15.3 months (95% confidence interval [CI]: 9.20–24.3 months) (Table 1). In the test dataset, 15 patients (3.1%) died during follow-up, and the median survival duration was 15.3 months (95% CI: 9.20–24.4 months) (Table 1). The Kaplan-Meier OS distributions of the training and test sets were similar to each other ($p = 0.19$) (Figure 1). Those with higher ages (HR, 95% CI: 1.04, 1.00–1.07), blurred tumor-lung interface (HR, 95% CI: 3.08, 1.27–7.46), calcification (HR, 95% CI: 3.33, 95% CI, 1.54–7.19), N stage (N2 HR, 95% CI: 4.45 (2.46–8.06), N3 HR, 95% CI: 10.41 (2.27–47.63)), and M stage (HR, 95% CI: 4.22, 95% CI, 1.77–10.07) were associated with shorter OS. The longer OS was correlated with pure ground glass (HR, 95% CI: 0.18, 0.04–0.73). Detailed results of the univariate and multivariate CPH prediction results for OS in the training dataset were presented in Table 2.

Comparison of survival model performance

In the training dataset, the C-index of Modified-DeepSurv was higher than that of CPH, 0.956 (0.946–0.974) vs. 0.836 (0.774–0.896), and in the test dataset, the C-index of Modified-DeepSurv was higher than that of CPH, 0.932 (0.908–0.964) vs. 0.777 (0.633–0.919) (Table 3). The calibration curve also indicated that Modified-DeepSurv also showed good calibration in both the training and the test cohort (Figure 2). The importance of variables in the training set of the Modified-DeepSurv was analyzed (Table S2).

Model visualization

In the prediction window, the system invokes a prediction model (Figure 3; Video S1), and the Modified-DeepSurv model is used to predict patients' survival probability. The analysis results are visualized in a graphic view as a survival curve, indicating the patient input's survival probability over time.

DISCUSSION

"Medical +AI" is the trend of current science and technology development. It is also an important approach to achieving widely beneficial medical care, assisting doctors in carrying out clinical diagnosis and treatment conveniently, reasonably, and scientifically. In this study, the Modified-DeepSurv model was constructed to predict lung patients' OS based on the large real-world database and the DL algorithm. The model was compared with the model constructed by Cox regression. The performance of the Modified-DeepSurv model is superior to the Cox regression model. This study also demonstrated that DL algorithms could provide a novel solution to assist clinicians in treatment decisions or clinical trial design through personalized, predictive models.

In this study, the CPH model combined gender, age, the history of the tumor, histological grade, tumor shape, tumor size, surgery, and TNM stage, and its C-index was lower than the Modified-DeepSurv model both in the training cohort and the test cohort. This may be caused by some

Table 2. Cox regression analysis in the training dataset

	Univariate analysis			Multivariate analysis		
Characteristics	HR	95% CI	p	HR	95% CI	p
Age	1.04	1.01–1.07	0.012	1.04	1–1.07	0.024
Sex						
Female	–	–	–	–	–	–
Male	1.92	1.13–3.26	0.016	–	–	–
Pathological type						
Others	–	–	–	–	–	–
Adenocarcinoma	0.39	0.2–0.78	0.007	–	–	–
Tumor size	1.17	1.03–1.33	0.015	–	–	–
Smoking history						
Never smoked	–	–	–	–	–	–
Former smoker	1.51	0.78–2.94	0.223	–	–	–
Current smoker	1.75	0.92–3.34	0.089	–	–	–
Nodules type						
Solid	–	–	–	–	–	–
Pure ground glass	0.12	0.03–0.48	0.003	0.18	0.04–0.73	0.017
Mixed ground glass	0.14	0.02–1.05	0.056	0.24	0.03–1.74	0.157
Tumor-lung interface						
Clear	–	–	–	–	–	–
Blurred	3.52	1.5–8.23	0.004	3.08	1.27–7.46	0.013
Nodular edge						
Smooth	–	–	–	–	–	–
Spiculated	2.79	1.34–5.81	0.006	–	–	–
Lobulated	1.54	0.5–4.72	0.452	–	–	–
Irregular	2.96	1.51–5.8	0.002	–	–	–
Vacuole sign						
No	–	–	–	–	–	–
Yes	1.4	0.55–3.51	0.48	–	–	–
Vascular abnormalities						
No	–	–	–	–	–	–
Yes	0.68	0.16–2.81	0.595	–	–	–
Pleural traction						
No	–	–	–	–	–	–
Yes	1.63	0.97–2.73	0.066	–	–	–
Calcification						
No	–	–	–	–	–	–
Yes	2.96	1.4–6.27	0.005	3.33	1.54–7.19	0.002
Enhancement						
No	–	–	–	–	–	–
Yes	2.04	1.15–3.62	0.015	–	–	–
Visceral pleural invasion (VPI)						
No	–	–	–	–	–	–
Yes	1.78	0.87–3.65	0.113	–	–	–
Surgery						
Wedge	–	–	–	–	–	–

(*Continued on next page*)

Table 2. *Continued*

Characteristics	Univariate analysis			Multivariate analysis		
	HR	95% CI	p	HR	95% CI	p
Segmentectomy	0.16	0.04–0.7	0.015	–	–	–
Lobectomy	0.76	0.42–1.4	0.382	–	–	–
T						
Tis	–	–	–	–	–	–
1	5	0.68–36.61	0.113	–	–	–
2	12.71	1.67–96.49	0.014	–	–	–
3	39.81	3.59–441.91	0.003	–	–	–
4	13.89	1.61–119.94	0.017	–	–	–
N						
0	–	–	–	–	–	–
1	2.82	1.17–6.79	0.021	2.17	0.89–5.27	0.088
2	5.88	3.28–10.53	<0.001	4.45	2.46–8.06	<0.001
3	14.69	3.48–61.99	<0.001	10.41	2.27–47.63	0.003
M						
0	–	–	–	–	–	–
1	5.09	2.29–11.33	<0.001	4.22	1.77–10.07	0.001

VPI: Visceral pleural invasion.

shortcomings of the model itself. The CPH model is a semi-parametric model that can be used to calculate the effect of observed covariates on the risk of event occurrence, such as death or cancer recurrence.[20] The risk model assumes that a patient's risk of death is a linear combination of covariates, an assumption known as the "Assumption of equal proportions". However, in real-world datasets, the assumption of equal proportions is often not satisfied.[21] Therefore, more survival models are needed to fit survival data to nonlinear risk functions better. In the large amount of patient health data, DL or ML can be integrated into electronic health records and provide clinicians with valuable prognostic information.[14,19]

DeepSurv algorithm is a risk network that applies DL technology to Cox regression, and it is a deep feedforward neural network.[11] The influence of patients' covariates on their risk is predicted by the network learning weights.[22] DL techniques may be a more verifiable prediction method due to the ability to handle large datasets with complex, nonlinear, heterogeneous distributions.[23,24] DL is unique, and it can construct models by applying Boolean logic, absolute condition, conditional probability, and other unconventional logarithmic strategies. In the previous study, the DeepSurv algorithm used TensorFlow operation mode, a static framework, to construct a calculation diagram of TensorFlow, and then different data can be input. The operation process is a fixed state.[19] This kind of inflexible operation method will inevitably lead to low efficiency.[11,14] In our study, we choose the Pytorch operation mode, which is more suitable for small sample projects. Pytorch is based on a dynamic framework. In the operation process, it will be arranged optimally according to different values, and it can establish protocols and produce results in a short time. However, it also has some limitations: the analysis lacks transparency. Although DL involves multilayer analysis that may make meaningful predictions, these layers often cannot be interpreted meaningfully.

In our result, the importance of variables showed that surgery is the first important variable for OS. The choice of surgical method is very important for the survival of lung cancer patients. However, in a previous study, the type of surgery was not a prognostic factor for either lung cancer-related or non-lung cancer-related OS in geriatric lung cancer patients.[25] However, in a nationwide propensity-matched study, there were significant differences in 1-year or 5-year survival rates among those who underwent different surgical methods for patients with lung cancer.[26] The effect of surgical methods on survival remains controversial. Therefore, prospective studies with larger samples are needed to confirm the importance of types of surgery on survival. Nodule type was the second most important variable for survival in the Modified-DeepSurv model. In a recent study, nodule type was associated with OS. Moreover, pure ground glass opacity (GGO) was positively correlated with OS in

Table 3. C-index of CPH and Modified-DeepSurv models

Model	C-index, 95% CI	
	Training Dataset	Test Dataset
CPH	0.836 (0.774–0.896)	0.777 (0.633–0.919)
Modified-DeepSurv	0.956 (0.946–0.974)	0.932 (0.908–0.964)

C-index: concordance index, CPH: Cox proportional hazard.

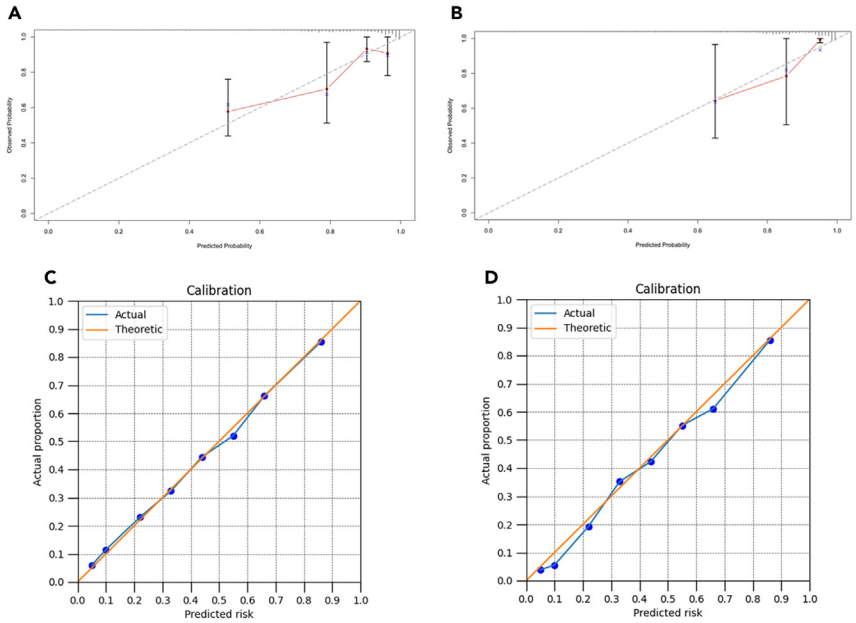


Figure 2. Calibration curves of the predicting models

Calibration curves estimated the calibration for predicting the survival rate in training (A), and test datasets (B) for the CPH model, and in training (C), and test datasets (D) for the Modified-DeepSurv model, respectively.

pathological stage I invasive lung adenocarcinoma.[27] In a large sample size study, nodule type was associated with OS, and it is an independent risk factor for survival in patients with resected stage I non-small cell lung cancer (NSCLC).[28] These results suggest that the surgery method and the type of nodules are the key factors affecting the survival of patients. These results are similar to those obtained by traditional CPH methods, which need to be confirmed by clinical practice. Smoking history was a well-known prognostic factor for lung cancer. In our Modified-Deepsurv model, smoking history was the ninth most important variable; however, it was not included in the multivariate Cox model. It is possible that Modified-DeepSurv model retains more variables than traditional CPH model, avoiding useful variables that are eliminated from the variable screening process before building the model. Converting methodology into informative clinical tools is always a great concern of researchers and clinicians. A graphic interface is a tool that visualizes the complex operation process to help doctors assess the survival status of patients and give appropriate care recommendations.[29,30] In our study, we developed an easy-to-use survival prediction tool based on the Modified-DeepSurv model. When a patient's information is entered, the survival probability of that patient can be predicted intuitively.

The Modified-DeepSurv model was constructed by a DL algorithm in this study. The predictive variables in this model were age, sex, pathological type, smoking history, maximum tumor diameter, nodules type, tumor-lung interface, nodular edge, vacuole sign, vascular abnormalities, pleural traction, calcification, enhancement, VPI, surgery, pathological T stage, pathological N stage, and pathological M stage, which could predict the survival of lung cancer patients better than the CHP model. In the era of big data, DL and AI play an important role in assessing prognosis by improving the quantitative ability of patient risk estimation and providing a new direction for developing more accurate prognostic prediction methods.

Limitations of the study

Firstly, as a retrospective study, the information bias caused by follow-up cannot be avoided; limited follow-up time resulted in limited number of deaths, and that may influence the application scenarios of the prediction model. Secondly, the external validation is necessary. Although the prediction model was well constructed and tested, for studies that require big data to complete model construction, the sample size of our study is still small; thus, a larger amount of data from different clinical centers is needed for validation in the future. Thirdly, DL algorithms were still the most difficult method to interpret,[31] despite the fact that the trained DL model can be directly integrated into the electronic medical record system, and clinicians can put patient data into Web-based forms to predict OS for lung cancer patients easily and effectively.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead contact

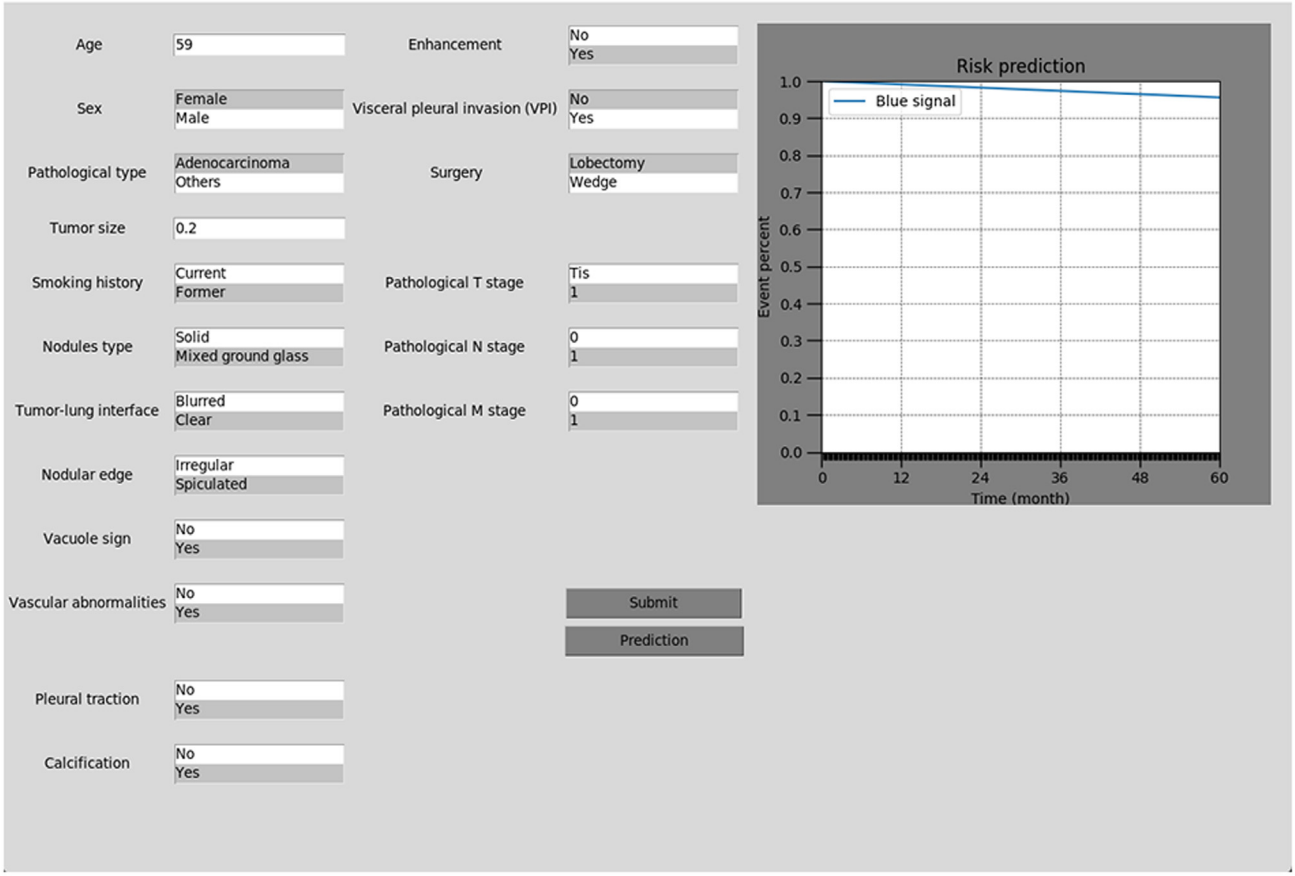


Figure 3. User-friendly interface of Modified-DeepSurv model which facilitates survival prediction

- Materials availability
- Data and code availability

- EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS
- METHOD DETAILS
  - Data collection and outcome definition
  - Deep learning algorithm
  - The data processing
  - Model visualization
- QUANTIFICATION AND STATISTICAL ANALYSIS

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.isci.2023.108200.

ACKNOWLEDGMENTS

The work was funded by the National Natural Science Foundation of China (52172285).

AUTHOR CONTRIBUTIONS

Conception and design of the research: R.Z., Z.Z., and T.J.; Acquisition of data: J. Lei and X.X.; Provision of study materials or patients: J. Lei and J.X.; Analysis and interpretation of data: Y.W., X.X., and J. Liu; Statistical analysis: J.X. and C.W.; Drafting the manuscript: All authors; Final approval of manuscript: All authors.

DECLARATION OF INTERESTS

The authors declare no competing interests.

INCLUSION AND DIVERSITY

We support inclusive, diverse, and equitable conduct of research.

Received: June 5, 2023
Revised: July 10, 2023
Accepted: October 11, 2023
Published: October 18, 2023

REFERENCES

1. Sung, H., Ferlay, J., Siegel, R.L., Laversanne, M., Soerjomataram, I., Jemal, A., and Bray, F. (2021). Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. CA. Cancer J. Clin. *71*, 209–249.
2. Xia, C., Dong, X., Li, H., Cao, M., Sun, D., He, S., Yang, F., Yan, X., Zhang, S., Li, N., and Chen, W. (2022). Cancer statistics in China and United States, 2022: profiles, trends, and determinants. Chin. Med. J. *135*, 584–590.
3. Zhu, L., Chen, S., Ma, S., and Zhang, S. (2016). Glasgow prognostic score predicts prognosis of non-small cell lung cancer: a meta-analysis. SpringerPlus *5*, 439.
4. Kawaguchi, T., Takada, M., Kubo, A., Matsumura, A., Fukai, S., Tamura, A., Saito, R., Maruyama, Y., Kawahara, M., and Ignatius Ou, S.H. (2010). Performance status and smoking status are independent favorable prognostic factors for survival in non-small cell lung cancer: a comprehensive analysis of 26,957 patients with NSCLC. J. Thorac. Oncol. *5*, 620–630.
5. Amin, M.B., Greene, F.L., Edge, S.B., Compton, C.C., Gershenwald, J.E., Brookland, R.K., Meyer, L., Gress, D.M., Byrd, D.R., and Winchester, D.P. (2017). The Eighth Edition AJCC Cancer Staging Manual: Continuing to build a bridge from a population-based to a more "personalized" approach to cancer staging. CA. Cancer J. Clin. *67*, 93–99.
6. Liang, W., Zhang, L., Jiang, G., Wang, Q., Liu, L., Liu, D., Wang, Z., Zhu, Z., Deng, Q., Xiong, X., et al. (2015). Development and validation of a nomogram for predicting survival in patients with resected non-small-cell lung cancer. J. Clin. Oncol. *33*, 861–869.
7. Weissman, S.M. (2018). Personalized medicine: a new horizon for medical therapy. Precis. Clin. Med. *1*, 1–2.
8. Prentice, R.L., and Zhao, S. (2021). Regression Models and Multivariate Life Tables. J. Am. Stat. Assoc. *116*, 1330–1345.
9. Xie, D., Marks, R., Zhang, M., Jiang, G., Jatoi, A., Garces, Y.I., Mansfield, A., Molina, J., and Yang, P. (2015). Nomograms Predict Overall Survival for Patients with Small-Cell Lung Cancer Incorporating Pretreatment Peripheral Blood Markers. J. Thorac. Oncol. *10*, 1213–1220.
10. Wang, Q., Wang, Y., Wang, X., Nakamura, Y., Hydbring, P., Yamauchi, Y., Zhao, X., and Cao, M. (2022). Development and validation of a dynamic survival nomogram for metastatic non-small cell lung cancer based on the SEER database and an external validation cohort. Transl. Lung Cancer Res. *11*, 1678–1691.
11. Katzman, J.L., Shaham, U., Cloninger, A., Bates, J., Jiang, T., and Kluger, Y. (2018). DeepSurv: personalized treatment recommender system using a Cox proportional hazards deep neural network. BMC Med. Res. Methodol. *18*, 24.
12. Li, B., Cui, Y., Diehn, M., and Li, R. (2017). Development and Validation of an Individualized Immune Prognostic Signature in Early-Stage Nonsquamous Non-Small Cell Lung Cancer. JAMA Oncol. *3*, 1529–1537.
13. Kvamme, H., Borgan, Ø., and Scheel, I. (2019). Time-to-Event Prediction with Neural Networks and Cox Regression.
14. Byun, S.S., Heo, T.S., Choi, J.M., Jeong, Y.S., Kim, Y.S., Lee, W.K., and Kim, C. (2021). Deep learning based prediction of prognosis in nonmetastatic clear cell renal cell carcinoma. Sci. Rep. *11*, 1242.
15. Adeoye, J., Koohi-Moghadam, M., Lo, A.W.I., Tsang, R.K.Y., Chow, V.L.Y., Zheng, L.W., Choi, S.W., Thomson, P., and Su, Y.X. (2021). Deep Learning Predicts the Malignant-Transformation-Free Survival of Oral Potentially Malignant Disorders. Cancers *13*, 6054.
16. Huang, L., Lu, X., Huang, X., Zou, X., Wu, L., Zhou, Z., Wu, D., Tang, D., Chen, D., Wan, X., et al. (2021). Intelligent difficulty scoring and assistance system for endoscopic extraction of common bile duct stones based on deep learning: multicenter study. Endoscopy *53*, 491–498.
17. Yin, M., Lin, J., Liu, L., Gao, J., Xu, W., Yu, C., Qu, S., Liu, X., Qian, L., Xu, C., and Zhu, J. (2022). Development of a Deep Learning Model for Malignant Small Bowel Tumors Survival: A SEER-Based Study. Diagnostics *12*, 1247.
18. Yang, J., Guo, F., Lyu, T., Yan, L.N., Wen, T.F., Yang, J.Y., Wu, H., Wang, W.T., Song, J.L., Xu, H., and Zhang, Q.H. (2020). [Research of artificial intelligence-based clinical decision support system for primary hepatocellular carcinoma]. Zhonghua Yixue Zazhi *100*, 3870–3873.
19. She, Y., Jin, Z., Wu, J., Deng, J., Zhang, L., Su, H., Jiang, G., Liu, H., Xie, D., Cao, N., et al. (2020). Development and Validation of a Deep Learning Model for Non-Small Cell Lung Cancer Survival. JAMA Netw. Open *3*, e205842.
20. Zucker, D.M., Zhou, X., Liao, X., Li, Y., and Spiegelman, D. (2019). A modified partial likelihood score method for Cox regression with covariate error under the internal validation design. Biometrics *75*, 414–427.
21. Benítez-Parejo, N., Rodríguez del Águila, M.M., and Pérez-Vicente, S. (2011). Survival analysis and Cox regression. Allergol. Immunopathol. *39*, 362–373.
22. Jared, L., Katzman, U.S., Alexander, C., Bates, J., Jiang, T., and Kluger, Y. (2016). Deep Survival: A Deep Cox Proportional Hazards Network.
23. Wang, G., Lam, K.M., Deng, Z., and Choi, K.S. (2015). Prediction of mortality after radical cystectomy for bladder cancer by machine learning techniques. Comput. Biol. Med. *63*, 124–132.
24. McCarthy, J.F., Marx, K.A., Hoffman, P.E., Gee, A.G., O'Neil, P., Ujwal, M.L., and Hotchkiss, J. (2004). Applications of machine learning and high-dimensional visualization in cancer detection, diagnosis, and management. Ann. N. Y. Acad. Sci. *1020*, 239–262.
25. Lin, Y.J., Chiang, X.H., Lu, T.P., Hsieh, M.S., Lin, M.W., Hsu, H.H., and Chen, J.S. (2021). Thoracoscopic Lobectomy Versus Sublobar Resection for pStage I Geriatric Non-Small Cell Lung Cancer. Front. Oncol. *11*, 777590.
26. Valo, J.K., Kytö, V., Sipilä, J., Rautava, P., Sihvo, E., and Gunn, J. (2020). Thoracoscopic surgery for lung cancer is associated with improved survival and shortened admission length: a nationwide propensity-matched study. Eur. J. Cardio. Thorac. Surg. *57*, 100–106.
27. Zhai, W.Y., Wong, W.S., Duan, F.F., Liang, D.C., Gong, L., Dai, S.Q., and Wang, J.Y. (2022). Distinct Prognostic Factors of Ground Glass Opacity and Pure-Solid Lesion in Pathological Stage I Invasive Lung Adenocarcinoma. World J. Oncol. *13*, 259–271.
28. Wang, C., Wu, Y., Li, J., Ren, P., Gou, Y., Shao, J., Zhou, Y., Xiao, X., Tuersun, P., Liu, D., et al. (2020). Distinct clinicopathologic factors and prognosis based on the presence of ground-glass opacity components in patients with resected stage I non-small cell lung cancer. Ann. Transl. Med. *8*, 1133.
29. Wang, F., Casalino, L.P., and Khullar, D. (2019). Deep Learning in Medicine-Promise, Progress, and Challenges. JAMA Intern. Med. *179*, 293–294.
30. Simon, G., DiNardo, C.D., Takahashi, K., Cascone, T., Powers, C., Stevens, R., Allen, J., Antonoff, M.B., Gomez, D., Keane, P., et al. (2019). Applying Artificial Intelligence to Address the Knowledge Gaps in Cancer Care. Oncol. *24*, 772–782.
31. Bur, A.M., Holcomb, A., Goodwin, S., Woodroof, J., Karadaghy, O., Shnayder, Y., Kakarala, K., Brant, J., and Shew, M. (2019). Machine learning to predict occult nodal metastasis in early oral squamous cell carcinoma. Oral Oncol. *92*, 20–25.
32. Akaike, H. (1974). A new look at the statistical model identification. IEEE Trans. Automat. Contr. *19*, 716–723.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Software and algorithms		
Python (version 3.6.9)	Python Software Foundation	https://www.python.org
R (version 4.2.1)	The R Foundation	https://www.r-project.org/
pyTorch (version 1.4.0)	pyTorch software	https://pytorch.org/

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources should be directed to and will be fulfilled by the Lead Contact, Tao Jiang (jiangtaochest@163.com).

Materials availability

This study did not generate new unique reagents.

Data and code availability

- Deidentified final results supporting this study are available for research purposes upon reasonable written request to the corresponding author. Access to such data is available from the date of publication and requires a Data Access Agreement, which is examined and approved by the ethics committees who approved this research.
- The code used for reproducing our analysis result are no available.
- Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS

A retrospective, multicenter cohort study was conducted. Patients with lung cancer diagnosed by pathology in the Tangdu Hospital, Xinjiang Peoples' Hospital and The First Affiliated Hospital of Anhui Medical University from April 2014 to May 2022 were included in constructing the model and followed up until July 2022. The patients diagnosed without determinate pathology or survival data were excluded. This study was conducted in accordance with the World Medical Association Declaration of Helsinki and approved by the Ethics Committee of all participating hospitals. Because this study was retrospective and data were analyzed anonymously, informed consent from all patients was not required. All the patients were Han Chinese from mainland China and provided written informed consent for the use of clinical information and tissue specimens.

METHOD DETAILS

Data collection and outcome definition

Data were extracted from the electronic medical record system of the above hospitals. Sex, age, histological type, TNM stage, tumor size, metastasis status, operation, survival time and survival status were collected. OS was defined as the time from surgery to death or the end.

Deep learning algorithm

DeepSurv algorithm is a risk network that applies deep learning technology to nonlinear Cox regression. It is a deep feedforward neural network. DeepSurv uses a multilayer perceptron to self-learn the effects of a covariate. Priori selection and interaction of the covariates should be considered in designing the CPH model, but DeepSurv has the advantage of not considering this. The network parameter setting is the key to building the Modified-DeepSurv model. Modified-DeepSurv comprises 1 input layer, multiple hidden layers, and 1 node with tanh activation and output. In our study, the number of input vector is 18, the total number of hidden layer nodes is 450, and the activation function is SeLU. We used the Adam optimizer with a learning rate of 0.07 and a learning rate decay of 0.003. The input vector X needs normalization before training. The number of hidden layer nodes is obtained by a large number of experiments during training. The random forest approach is used to rank the importance of the input vectors.

The data processing

Multiple imputations (MIs)

Each variable was observed to determine the missing proportion, and variables with excessive missing proportions were deleted. To ensure the regularity of data distribution to the greatest extent, MIs were performed on the missing variables. The tool used the R software

(https://www.r-project.org/) mice package, and the MIs were performed five times, with maxit = 10 and seed = 1234. After comprehensive consideration, the fifth imputation dataset was used for the final analysis.

Data transformation

The two models are independent in the way they do data transformation. For the development of Modified-DeepSurv model, firstly, the text data was digitized. In detail, for binary categorical data, sex, it was 0 for female and 1 for male; Pathological type, it was 0 for others and 1 for adenocarcinoma; Tumor interface, it was 0 for blurred and 10 for clear; Vacuole sign, vascular abnormalities, pleural traction, calcification, enhancement, and visceral pleural invasion (VPI), it was 0 for no and 10 for yes, respectively; For multi-categorical data, smoking history, it was 0 for never, 10 for current, and 20 for former; Nodules type, it was 0 for solid, 10 for mixed ground glass, and 20 for pure ground glass nodule (pGGN); The nodular edge, it was 0 for irregular, 10 for spiculated, 20 for smooth, and 30 for lobulated; Surgery, it was 0 for lobectomy, 10 for wedge, 20 for segmentectomy. Second, the digitized data were normalized: 0 and 1 do not need processing, with the floating-point values reserved for 0.0 and 1.0. For normal data: floating-point value normalized between -1.0 and 1.0 (normal distribution). For categorical data, linear scaling to floating-point values is between -1.0 and 1.0. Time to event data, normalized to a floating point value between 0.0 and 4.0. Digitized and normalized were processed by python3.6.9. For CPH model, before data analysis, data were assigned a value, with an event as 1 for dependent variables and no event as 0. For continuous independent variables, no need to assign it. For binary categorical independent variables, sex, it was 0 for female and 1 for male; Pathological type, it was 0 for others and 1 for adenocarcinoma; Tumor interface, it was 0 for blurred and 1 for clear; Vacuole sign, vascular abnormalities, pleural traction, calcification, enhancement, and visceral pleural invasion (VPI), it was 0 for no and 1 for yes, respectively. For multi-categorical independent variables, smoking history, it was 0 for never, 1 for current, 2 for former; Nodules type, it was 0 for solid, 1 for mixed ground glass, and 2 for pGGN; nodule edge: 0 for irregular, 1 for spiculated, 2 for smooth, 3 for lobulated; Surgery, it was 0 for lobectomy, 1 for wedge, 2 for segmentectomy. In either of the above two models, the numbers are used as a numerals, not for the values they represent, just as symbols.

Model visualization

We developed a user-friendly risk prediction interface programming by python3.6.9 to facilitate survival prediction of the Modified-DeepSurv model. This interface consists of 2 windows: the user input window and the survival prediction window. The user input window is designed to help users input all entries regarding patient characteristics into Modified-DeepSurv model by manual. The user input window allows users to predict the survival probability based on specific patient information by clicking the predict buttons.

QUANTIFICATION AND STATISTICAL ANALYSIS

The data of included patients were randomly assigned into the training and test dataset in the ratio of 3:1. Measurement data were expressed as median, interquartile range (IQR). Categorical data were represented as numbers (%), and comparison between groups was analyzed using chi-square test. Kaplan-Meier was used to calculate the survival rate, and the log-rank test was used to compare survival rates between groups. Cox proportional hazards regression was used to analyze the factors affecting OS in the training cohort, while the hazard ratio (HR) and its corresponding 95% confidence interval (CI) were calculated. Analyzed variables including patient age, sex, pathological type, smoking history, maximum tumor diameter, nodules type, tumor-lung interface, nodular edge, vacuole sign, vascular abnormalities, pleural traction, calcification, enhancement, VPI, surgery, pathological T stage, pathological N stage, and pathological M stage. The final multivariate COX model was developed by stepwise regression to obtain the best result with the smallest Akaike information criterion (AIC).[32] Performances of the CPH and deep learning model were compared using the fivefold cross-validated Harrel's C-index. The fivefold cross-validation technique was used in all CPH and Modified-DeepSurv training processes. The C-index and calibration curve were used to evaluate the discrimination and calibration of the model. The Bootstrap method was used for calculation, and the resampling times were 1000.

The survival package of R 4.2.1 software (https://www.r-project.org/) was used to construct the Cox proportional hazard regression model, and the PyTorch DeepSurv repo in Python 3.6.9 was used to construct the Modified-DeepSurv model (https://github.com/jaredleekatzman/DeepSurv). All other data were analyzed by R 4.2.1 software. All tests were two-sided, and $p<0.05$ was considered statistically significant.