Appendix 1

Data collection

The training cohort of patients comprised only patients with microscopically confirmed lung cancer. Patients without follow-up information were excluded, as were patients who received chemotherapy before surgery or underwent resection for a recurrent lung cancer (r-stage cases).

The testing cohort of lung cancer patients all underwent surgical treatment alone or combined with chemotherapy alone or with radiotherapy.

Study data preparation

The SEER database was downloaded with SEER*Stat version 8.3.6.1 and only 18 survival-related variables were selected: sex, age, race, marital status at diagnosis, CS-tumor size, CS extension (2004–15), CS lymph nodes (2004–15), CS metastasis at diagnosis (2004–), regional nodes exam (1988–), regional nodes positive (1988–), chemotherapy recode, grade, laterality, radiation sequence with surgery, radiation recode, CS site-specific factor 1 (2004–), histologic type ICD-O-3, and sequence number for the patient. The definition and recode rules of these variables referred to the SEER database (https://seer.cancer.gov/analysis), and those for the GYFY database were also recoded according to these rules.

The training data were randomly split patient-wise into a training set (80%) and a validation set (20%). Because some patients had multiple lines/multiple results in SEER database, the patient-wise separation ensured that lines belonging to same patient would only be split to the same subset (training or validation set).

Data encoding

The dataset included categorical features (e.g., sex, marital status at diagnosis), ordinal features (e.g., grade), and numerical features (e.g., CS-tumor size, age). To standardize the input features, numerical features were normalized to (0,1), and categorical features were transformed using a one-hot scheme, and ordinal features were converted into dummy variables.

Deep learning algorithm description

In this work, we used DeepSurv (1) for survival prediction. DeepSurv is an extension to the standard survival Cox proportional hazards model (CPH). It is a non-linear model that predicts a patient's risk of death.

The standard CPH is a linear proportional hazards model that estimates the risk function $h(x)$ by a linear function $\hat{h}_\beta(x) = \beta^T x$. Cox regression was performed by optimizing the Cox partial likelihood, which is defined as:

$$L_c(\beta) = \prod_{i:E_i=1} \frac{\exp\left(\hat{h}_\beta(x_i)\right)}{\sum_{j \in R(T_i)} \exp\left(\hat{h}_\beta(x_i)\right)} \quad [1]$$

where the values X_i, T_i, and E_i are the data, event time, and event indicators for the i-th observation, respectively. The risk set $R(t) = \{i: T_i \geq t\}$ is the set of patients still at risk of failure at time t. The product is defined over the set of patients with an observable event $E_i = 1$.

As an extension to the linear CPH model, DeepSurv estimates the risk by a non-linear function $\hat{h}_\theta(x)$ parameterized by the weights of the network. The network of DeepSurv is constructed using several fully connected layers. The loss function of DeepSurv is defined as negative log partial likelihood of Eq. [1] as above:

$$l(\theta) := -\sum_{i:E_i=1} \left(\hat{h}_\theta(x_i) - \log \sum_{j \in R(T_i)} e^{\hat{h}_\theta(x_j)} \right) \quad [2]$$

AI certainty

Gal *et al.* (2) shows that model uncertainty can be estimated using dropout neural networks (NNs) by extracting information

https://dx.doi.org/10.21037/tlcr-23-84

from existing models. In order to compute AI certainty, we first activate the dropout layer during inference time, and then infer the test set for N time to get N results. The variance of these N results was computed to get the model uncertainty U. The final AI certainty C = 1-U.

Basic concept of transfer learning

Transfer learning is a machine learning technique in which the model developed for one task is used as the starting point of a different but related task. In the context of deep NNs, transfer learning is achieved by first pretraining the weights on a large related dataset and using them as the initialization for the target task. With transfer learning, knowledge gained while learning on the pretraining task is applied when learning the target task. Transfer learning is widely applied and shown to be effective, especially for tasks where the training data are limited.

Model evaluation and statistical analysis

The outcome was measured based on overall survival, the period (in months) between diagnosis and death or loss of follow-up from any cause, as reported in the SEER and GYFY databases. Features were expressed as counts and percentages for categorical variables and as the mean [standard deviation (SD)] or median (range) for continuous variables. Qualitative and quantitative differences between subgroups were analyzed using the chi-squared test or Fisher's exact test for categorical parameters and Student's t-test or the Mann-Whitney U test for continuous parameters, as appropriate. Univariable and multivariable Cox proportional hazards regression models were used to estimate the effects of various variables on the hazard of lung cancer occurrence and develop the lung cancer prediction model. The cumulative incidence of death was estimated by the Kaplan-Meier (K-M) method and compared using the log-rank test. The concordance index (C-index) was used to assess the discriminatory powers of the models, and the survival calibration curve was calculated to evaluate the calibration of the probability of survival as predicted by the model versus the observed probability. Statistical analysis was performed with R (Version 4.0.0). $P<0.05$ was statistically significant, and all tests were two-sided.

References

1. Katzman JL, Shaham U, Cloninger A, et al. DeepSurv: personalized treatment recommender system using a Cox proportional hazards deep neural network. BMC Med Res Methodol 2018;18:24.
2. Gal Y, Ghahramani Z. Dropout as a bayesian approximation: Representing model uncertainty in deep learning. International Conference on Machine Learning, 2016:1050-9.

https://dx.doi.org/10.21037/tlcr-23-84

Table S1 Univariable cox regression analysis of 18 variables in 2 cohorts

Variables	Training cohort		Testing cohort	
	HR (95% CI)	P value	HR (95% CI)	P value
Age	1.0181 (1.0179–1.0184)	<0.0001	1.0187 (1.0101–1.0274)	<0.0001
CS-tumor size	1.0007 (1.0007–1.0007)	<0.0001	1.0003 (0.9999–1.0008)	0.1647
CS extension (2004-2015)	1.0011 (1.0011–1.0012)	<0.0001	1.0012 (1.0009–1.0016)	<0.0001
CS lymph nodes (2004-2015)	1.0009 (1.0009–1.0009)	<0.0001	1.0010 (1.0008–1.0013)	<0.0001
CS Mets at dx (2004+)	1.0127 (1.0126–1.0128)	<0.0001	1.0013 (0.9986–1.0040)	0.3433
Regional nodes exam (1988+)	0.9977 (0.9976–0.9978)	<0.0001	1.0018 (0.9987–1.0049)	0.2501
Regional nodes positive (1988+)	1.0134 (1.0133–1.0135)	<0.0001	1.0085 (1.0064–1.0106)	<0.0001
Sex				
Female-male	0.8051 (0.8005–0.8097)	<0.0001	0.4865 (0.4005–0.5911)	<0.0001
Marital status at diagnosis				
Married-others	0.8372 (0.8325–0.8420)	<0.0001	1.5850 (0.9767–2.5720)	0.0623
Chemotherapy recode				
Yes-no	0.8813 (0.8762–0.8864)	<0.0001	1.1414 (0.9589–1.3586)	0.1368
Grade				
Grade II—well differentiated	1.5963 (1.5665–1.6265)	<0.0001	4.8958 (2.2989–10.4265)	<0.0001
Grade III/Grade IV—well differentiated	2.7209 (2.6731–2.7696)	<0.0001	8.1926 (3.8545–17.4132)	<0.0001
Unknown—well differentiated	3.7915 (3.7264–3.8577)	<0.0001	9.162 (4.2935–19.5512)	<0.0001
Laterality				
Left-others	0.5186 (0.5128–0.5243)	<0.0001	0.7523 (0.3991–1.4182)	0.3789
Right-others	0.5190 (0.5134–0.5246)	<0.0001	0.7094 (0.3778–1.3321)	0.2855
Radiation sequence with surgery				
No radiation and/or cancer-directed surgery-others	1.3259 (1.3119–1.3402)	<0.0001	0.6551 (0.4895–0.8766)	0.0044
Radiation recode				
None/unknown-others	0.9499 (0.9444–0.9555)	<0.0001	0.7388 (0.5353–1.0196)	0.0655
Lung-surgery to primary site (1988-2015)				
Lobectomy/bilobectomy-others	0.2264 (0.2240–0.2287)	<0.0001	0.7315 (0.6020–0.8887)	0.0017
CS site-specific factor 1 (2004+)				
No separate tumor nodules noted-others	0.7231 (0.7185–0.7278)	<0.0001	0.7492 (0.6307–0.8899)	0.001
Histologic Type ICD-O-3				
Adenocarcinoma-others	0.6740 (0.6700–0.6781)	<0.0001	0.7032 (0.5879–0.8410)	0.0001
Sequence number				
One primary only-others	1.4068 (1.3978–1.4159)	<0.0001	0.8542 (0.5758–1.2671)	0.4334

CS, collaborative staging.

https://dx.doi.org/10.21037/tlcr-23-84

Table S2 Multivariable cox regression analysis of 18 variables in 2 cohorts

Variables	Training cohort		Testing cohort	
	HR(95%CI)	P value	HR (95%CI)	P value
Age	1.0122 (1.0119–1.0124)	<0.0001	1.0134 (1.0048–1.0220)	0.0022
CS-tumor size	1.0001 (1.0001–1.0001)	<0.0001		
CS extension (2004-2015)	1.0004 (1.0004–1.0004)	<0.0001	1.0009 (1.0005–1.0013)	<0.0001
CS lymph nodes (2004-2015)	1 (0.9999–1)	<0.0001	1.0006 (1.0002–1.0010)	0.0009
CS Mets at dx (2004+)	1.0049 (1.0048–1.0050)	<0.0001		
Regional nodes exam (1988+)	0.9974 (0.9973–0.9975)	<0.0001		
Regional nodes positive (1988+)	1.0055 (1.0054–1.0056)	<0.0001	1.0029 (0.9997–1.0061)	0.0747
Sex				
Female-male	0.8043 (0.7996–0.8091)	<0.0001	0.5106 (0.4166–0.6259)	<0.0001
Marital status at diagnosis				
Married-others	0.9181 (0.9127–0.9236)	<0.0001		
Chemotherapy recode				
Yes-no	0.7429 (0.7381–0.7477)	<0.0001		
Grade				
Grade II—well differentiated	1.5474 (1.5184–1.5769)	<0.0001	3.854 (1.8028–8.2392)	0.0005
Grade III/Grade IV—well differentiated	2.0813 (2.0439–2.1194)	<0.0001	5.716 (2.6738–12.2193)	<0.0001
Unknown—well differentiated	1.9918 (1.9565–2.0277)	<0.0001	6.5816 (3.062–14.1468)	<0.0001
Laterality				
Left-others	1.1138 (1.1–1.1277)	<0.0001		
Right-others	1.1327 (1.119–1.1466)	<0.0001		
Radiation sequence with surgery				
No radiation and/or cancer-directed surgery-others	0.853 (0.8427–0.8634)	<0.0001	0.7488 (0.5579–1.0049)	0.0539
Radiation recode				
None/unknown-others	1.1537 (1.1459–1.1616)	<0.0001		
Lung-surgery to primary site (1988-2015)				
Lobectomy/bilobectomy-others	0.4566 (0.4503–0.463)	<0.0001	0.9267 (0.75–1.145)	0.4804
CS site-specific factor 1 (2004+)				
No separate tumor nodules noted-others	0.9326 (0.9264–0.9389)	<0.0001	0.7888 (0.6632–0.9382)	0.0074
Histologic Type ICD-O-3				
Adenocarcinoma-others	0.9187 (0.913–0.9244)	<0.0001	1.0579 (0.8755–1.2781)	0.5601
Sequence number				
One primary only-others	1.3031 (1.2946–1.3117)	<0.0001		

CS, collaborative staging.

https://dx.doi.org/10.21037/tlcr-23-84