

## Journal Pre-proofs

Artificial intelligence predictive system of individual survival rate for lung adenocarcinoma

Tingshan He, Jing Li, Peng Wang, Zhiqiao Zhang

PII: S2001-0370(22)00162-3  
DOI: <https://doi.org/10.1016/j.csbj.2022.05.005>  
Reference: CSBJ 1515

To appear in: *Computational and Structural Biotechnology Journal*

Received Date: 12 February 2022  
Revised Date: 5 May 2022  
Accepted Date: 5 May 2022

Please cite this article as: T. He, J. Li, P. Wang, Z. Zhang, Artificial intelligence predictive system of individual survival rate for lung adenocarcinoma, *Computational and Structural Biotechnology Journal* (2022), doi: <https://doi.org/10.1016/j.csbj.2022.05.005>

This is a PDF file of an article that has undergone enhancements after acceptance, such as the addition of a cover page and metadata, and formatting for readability, but it is not yet the definitive version of record. This version will undergo additional copyediting, typesetting and review before it is published in its final form, but we are providing this version to give early visibility of the article. Please note that, during the production process, errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

© 2022 The Author(s). Published by Elsevier B.V. on behalf of Research Network of Computational and Structural Biotechnology.



Artificial intelligence predictive system of individual survival rate for  
lung adenocarcinoma

Tingshan He<sup>#</sup>, Jing Li<sup>#</sup>, Peng Wang, Zhiqiao Zhang<sup>\*</sup>

Department of Infectious Diseases, Shunde Hospital, Southern Medical  
University, Shunde, Guangdong, China

<sup>#</sup> The co-first authors of the current study.

<sup>\*</sup> Corresponding author: Zhiqiao Zhang.

Address: Shunde Hospital, Southern Medical University, Shunde 528303,  
Guangdong, China; E-mail: sdgrxjbk@163.com.

**Abstract**

**Background:** The current research aimed to develop an artificial intelligence predictive system for individual survival rate of lung adenocarcinoma (LUAD).

**Methods:** Independent risk variables were identified by multivariate Cox regression. Artificial intelligence predictive system was constructed using three different data mining algorithms.

**Results:** Stage, PM, chemotherapy, PN, age, PT, sex, and radiation\_surgery were determined as risk factors for LUAD patients. For 12-month survival rate in model cohort, concordance indexes of RFS, MTLR, and Cox models were 0.852, 0.821, and 0.835, respectively. For 36-month survival rate in model cohort, concordance indexes of RFS, MTLR, and Cox models were 0.901, 0.864, and 0.862, respectively. For 60-month survival rate in model cohort, concordance indexes of RFS, MTLR, and Cox models were 0.899, 0.874, and 0.866, respectively. The concordance indexes in validation dataset were similar to those in model dataset.

**Conclusions:** The current study designed an individualized survival predictive system, which could provide individual survival curves using three different artificial intelligence algorithms. This artificial intelligence predictive system could directly convey treatment benefits by comparing individual mortality risk curves under different treatments. This artificial

intelligence predictive tool is available at

[https://zhangzhiqiao11.shinyapps.io/Artificial\\_Intelligence\\_Survival\\_Prediction\\_System\\_AI\\_E1001/](https://zhangzhiqiao11.shinyapps.io/Artificial_Intelligence_Survival_Prediction_System_AI_E1001/).

**Keywords:** Lung adenocarcinoma; artificial intelligence; prognostic model; overall survival.

Journal Pre-proofs

## 1. Background

Lung adenocarcinoma (LUAD) is one of the most common malignant tumours, accounting for 1.8 million cancer related deaths [1]. The prognosis of LUAD patients is still unsatisfactory until today [2]. At present, there were many predictive models in predicting survival rate for LUAD patients at the group level [3, 4]. However, the prognosis of LUAD patients with different clinical characteristics is complicated till now. Therefore, the prognostic prediction of one special group is far from meeting the need of individualized treatment decisions for a special individual patient.

In recent years, artificial intelligence has made great progress in cancer research, diagnosis, prognostic prediction, and treatment. Various algorithms have been used to find the lncRNAs closely related to different diseases, so as to provide valuable biomarkers for clinical diagnosis [5-8]. Artificial intelligence predictive models based on gene expression data could predict prognosis for different tumors [9, 10]. Artificial intelligence algorithms based on gene expression data could also be used to predict the efficacy of tumor treatments [11, 12]. Tumor imaging recognition based on deep learning technology is helpful for early diagnosis and accurate classification for tumor [13]. The above researches suggested that artificial intelligence has broad application prospects in cancer research, diagnosis, prognostic prediction, and

treatment.

From 2012 to 2013, Professor Gary S Collins developed several on-line prognostic predictive tools to predict mortality for different tumours [14-18]. In recent years, several studies have predicted individual survival curves for cancer patients based on different algorithms [19-21]. Our research team developed several precise medicine predictive systems to predict individual survival curves for different cancers before clinical treatment based on genetic data [22-29]. Several experts proposed valuable suggestions for improving our precision medicine tools presented in our previous articles. Can precision medicine predictive tools provide individualized mortality curves for patients receiving radiotherapy or chemotherapy? Can the current precision medicine predictive tools convey treatment benefits by comparing the individual mortality curves of patients under different treatments, which might be more valuable for optimizing individual treatment decisions?

Therefore, according to previous suggestions, our team planned to develop an artificial intelligence predictive tool in predicting and compare individual survival curves of LUAD patients under different treatments.

## **2. Methods**

## 2.1 Study cohorts

All datasets were obtained from the Surveillance Epidemiology and End Results (SEER) database (2010-2015). All included patients were diagnosed with lung adenocarcinoma (ICD-O-3 code: 8140). To eliminate the confounding effects of other causes, living subjects with survival time  $< 12$  months were removed from present study ( $n = 852$ ).

## 2.2 Research methods

Induction and deduction are common research methods in scientific research [30]. The data type in the current study belongs to cohort study data. The current research used induction method to summarize cohort research information of LUAD patients, so as to obtain general rules of prognosis of LUAD patients. Then, the current research used the deductive method to study the prognosis of individual patients from the general rules of the overall cohort.

## 2.3 Artificial intelligence algorithms

The random survival forest algorithm was performed in accordance with the original studies [31-34]. The multitask logistic regression (MTLR) algorithm performed in line with the suggestions of previous articles [35, 36]. The Cox proportional hazards algorithm was performed based on advices in original articles [37, 38].

## 2.4 Statistical analyses

R software 3.5.2 was used to run statistical analysis and relevant algorithms [22-29]. The research methods and statistical analysis steps were as follows: Continuous data with non normal distribution were presented as median (first and third quantiles). The continuity data were compared using nonparametric test. The counting variables were compared using chi square test. The random survival forest method was used to evaluate variable importance. Multivariate Cox regression was carried out for determining risk factors of LUAD. The prognostic score was constructed according to the coefficient of above risk factors. Kaplan-Meier curve was carried out for presenting prognosis of different cohorts. The area under the time-dependent receiver operating characteristic curve and Brier score were used to assess accuracy of different prognostic models. The flow chart of methodology was presented in Figure 1.



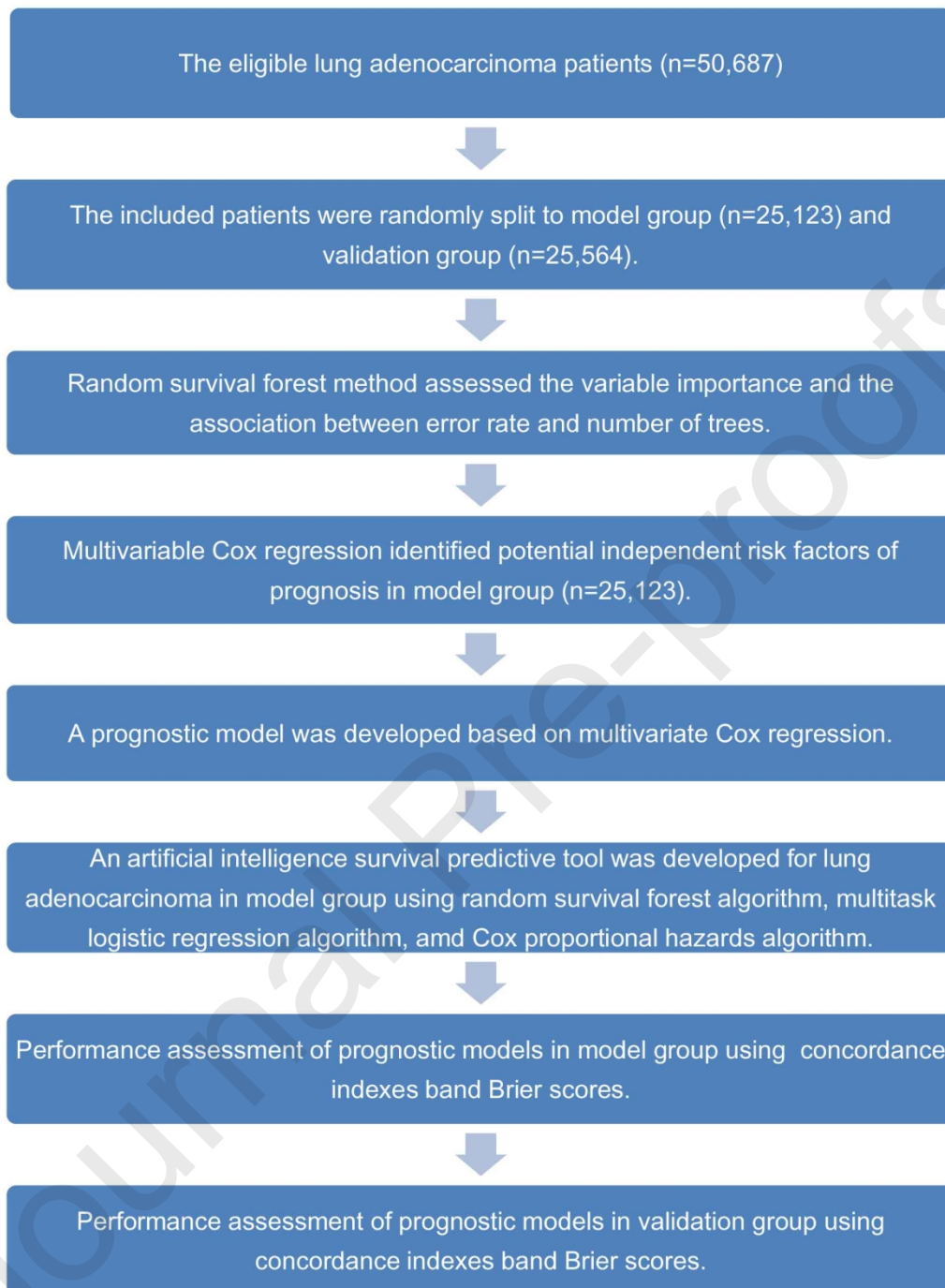


Figure 1. Flow chart of study methodology

### 3. Results

#### 3.1 Study datasets

The included patients (n=50,687) were randomly split to model group and validation group. Baseline features for patients in model group and validation group were presented in Table 1. After random grouping, there existed significant differences in mortality, PN, and laterality between model group and validation group, whereas there was no significant difference for other variables between model cohort and validation cohort.

### **3.2 Variable importance assessment and selection**

Random survival forest method was carried out to assess the variable importance and the association between error rate and number of trees. As shown in Supplementary document 1, the variable importance from high to low was as follows: stage, PM, chemotherapy, PN, age, PT, gender, and radiation\_surgery.

Through multivariable Cox regression, stage, PM, chemotherapy, PN, age, PT, sex, and radiation\_surgery were identified as independent risk factors of prognosis in model group (Table 2). In validation group, stage, PM, chemotherapy, PN, age, PT, sex, and radiation\_surgery were determined as risk factors of LUAD.

### **3.3 Prognostic models**

Through multivariate Cox regression, a prognostic model was developed

with following formula: Prognostic score =  
 $(0.786 * \text{Stage}) + (-0.692 * \text{Chemotherapy}) + (0.156 * \text{PN}) + (0.0140 * \text{Age}) + (0.2684 * \text{Gender}) + (0.070 * \text{PT}) + (-0.135 * \text{Radiation\_Surgery}) + (0.091 * \text{PM})$ . A RFS model and MTLR model were built to predict the prognosis using eight above prognostic variables identified in the Cox regression.

### 3.4 Artificial intelligence predictive system

The current study developed an artificial intelligence survival predictive tool for LUAD (Fig.2). This tool is available at

[https://zhangzhiqiao11.shinyapps.io/Artificial\\_Intelligence\\_Survival\\_Prediction\\_System\\_AI\\_E1001/](https://zhangzhiqiao11.shinyapps.io/Artificial_Intelligence_Survival_Prediction_System_AI_E1001/).

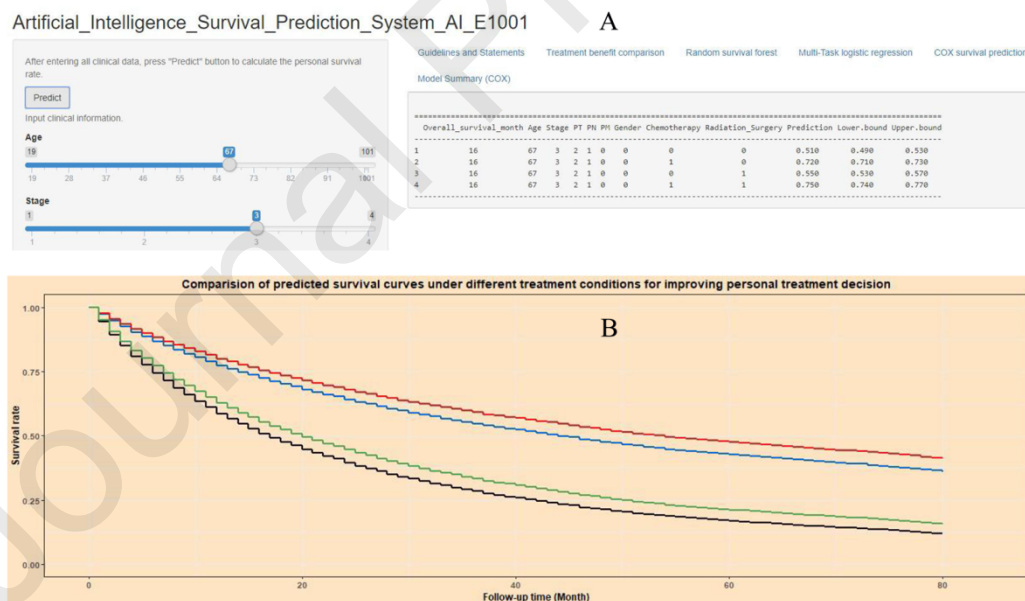


Figure 2. Home page of Artificial intelligence survival prediction system: (A). Data input page and digital result display page; (B). Comparison of predicted survival curves under different treatment conditions.

The generated survival rate and 95% confidence interval were shown

in Fig. 2A. Moreover, Fig. 2B provided comparisons of four predicted survival curves (red, blue, green, and black line in Fig. 2B) under different treatments.

Three individualized mortality risk curves were presented for RFS (Fig. 3A), MTLR (Fig. 3B), and Cox models (Fig. 3C). The abscissa of each point in the survival curve (Fig. 3) corresponded to a specific time point, and the ordinate represented the predicted survival probability of a special individual patient at that specific time point.

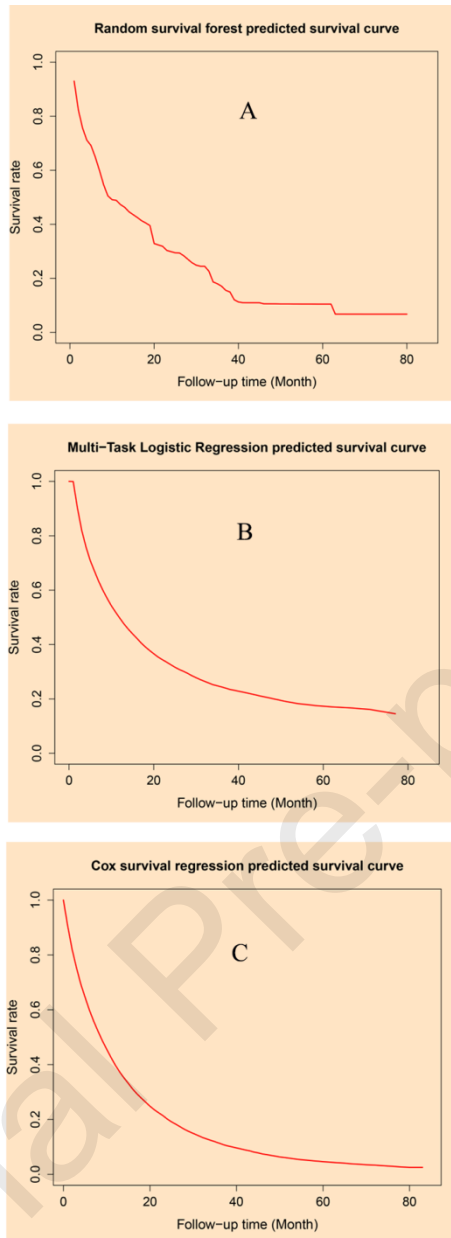


Figure 3. Home page of Artificial intelligence survival prediction system: (A). Predictive personal survival curve by random survival forest; (B). Predictive personal survival curve by Multi-task logistic regression; (C). Predictive personal survival curve by Cox survival regression.

### 3.5 Performance of prognostic models

The survival curve chart (Fig. 4) indicated that three artificial intelligence prognostic models could discriminate high mortality risk patients from low mortality risk patients in the model cohort.

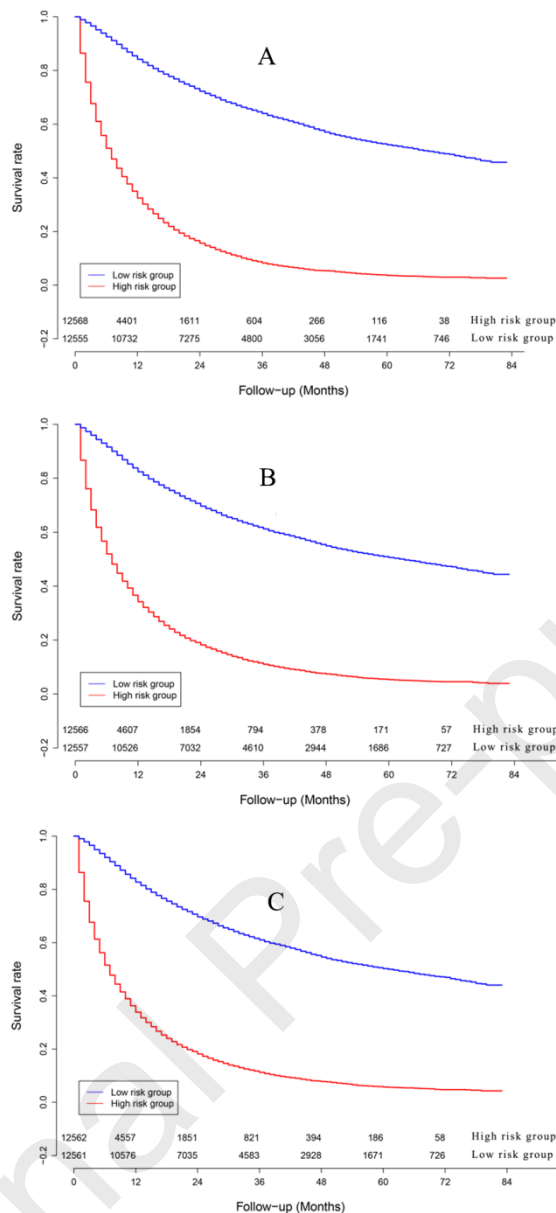


Figure 4. Survival curves of high risk patients and low risk patients in model cohort: (A). Random survival forest; (B). Multi-task logistic regression; (C). Cox survival regression.

For 12-month survival rate (Fig. 5A), the concordance indexes of RFS, MTLR, and Cox models were 0.852, 0.821, and 0.835, respectively. For 36-month survival rate (Fig. 5B), the concordance indexes of RFS, MTLR, and Cox models were 0.901, 0.864, and 0.862, respectively. For 60-month survival rate (Fig. 5C), the concordance indexes of RFS, MTLR, and Cox models were 0.899, 0.874, and 0.866, respectively.

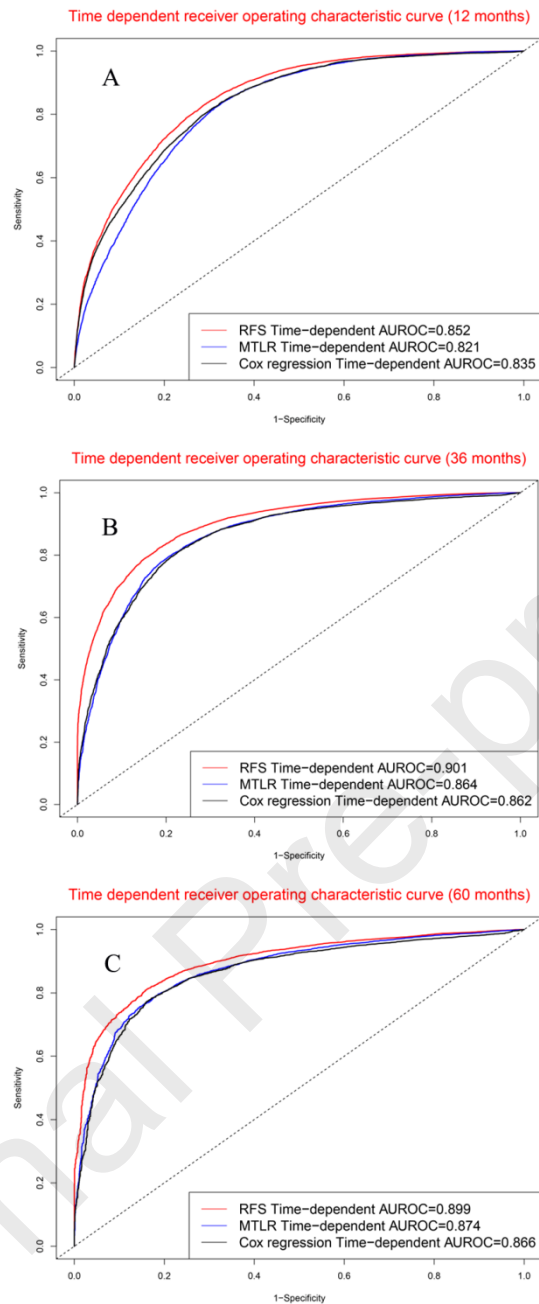


Figure 5. Time-dependent receiver operating characteristic curves in model cohort: (A). 12 months; (B). 36 months; (C). 60 months.

The survival curve chart (Supplementary document 2) indicated that three artificial intelligence prognostic models could discriminate high mortality risk patients from low mortality risk patients in the validation cohort. For 12-month survival rate (Fig. 6A), the concordance indexes of RFS, MTLR, and Cox model were 0.824, 0.834, and 0.834, respectively.

For 36-month survival rate (Fig. 6B), the concordance indexes of RFS, MTLR, and Cox models were 0.851, 0.857, and 0.853, respectively. For 60-month survival rate (Fig. 6C), the concordance indexes of RFS, MTLR, and Cox models were 0.870, 0.876, and 0.871, respectively.

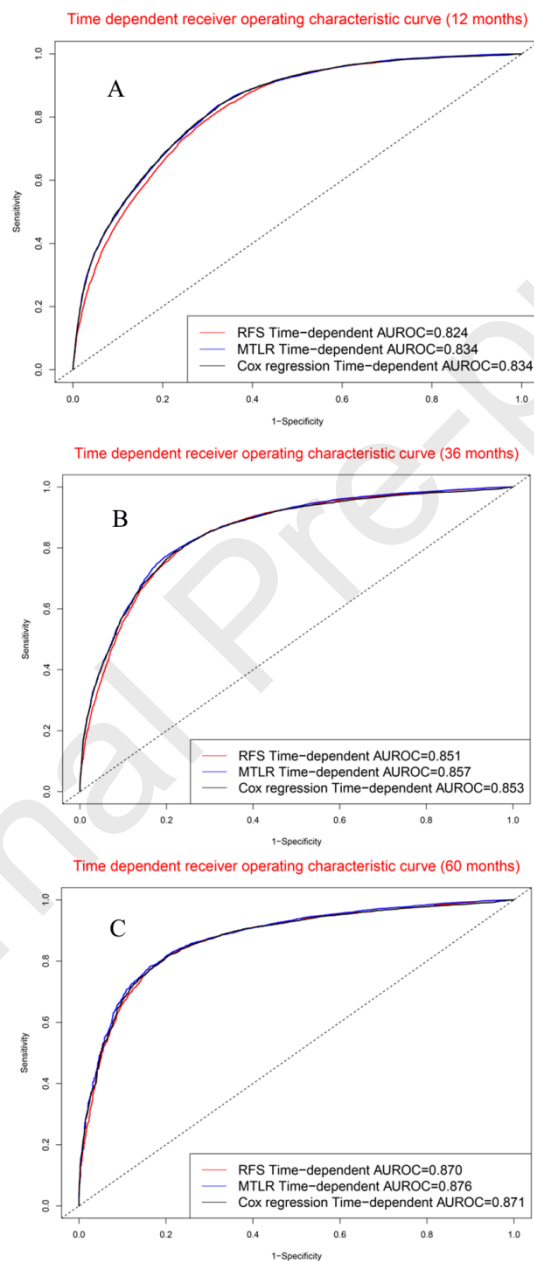


Figure 6. Time-dependent receiver operating characteristic curves in validation cohort: (A). 12 months; (B). 36 months; (C). 60 months.

Brier scores of RFS, MTLR, and Cox models were 0.124, 0.152, and 0.143, respectively, indicating that accuracy of RFS Model was better



than that of MTLR model and Cox model.

Supplementary document 3, 4, and 5 showed calibration plots of RFS, MTLR, and Cox models in the model cohort. Supplementary document 6, 7, and 8 showed calibration plots of RFS, MTLR, and Cox models in the validation cohort.

#### **4. Discussion**

The current study established an interesting artificial intelligence survival predictive system for LUAD patients. Three different artificial intelligence algorithms could provide individual survival curves that supported and corroborated one another. More importantly, this artificial intelligence survival predictive system could successfully predict and compare individual mortality risk curves under four treatments, providing clinical benefit comparisons at the individual level to optimize individualized treatment decisions.

The current study provides a convenient individual mortality risk predictive tool for lung adenocarcinoma patients. For example, the 16-month(user selected time-point) survival rate were 0.51 for non-treatment status (black line in Figure 2), 0.72 for chemotherapy status (blue line in Figure 2), 0.55 for radiation\_surgery status (green line in Figure 2), and 0.75 for combination therapy status (red line in Figure 2) for a special patients with the following parameters: age 67 years, stage 3,

PT 2, PN 1, PM 0, and gender female. Through the predicted survival rate at a specific time-point in the upper part of Figure 2 and the individual survival curve in the lower part of Figure 2, patients can easily get their own individual survival curve, so as to optimize individualized treatment decisions.

Several predictive models were constructed for predicting overall survival for lung cancer at the group level [3, 4, 22]. However, these prognostic models could only forecast mortality risk at group level with unique clinical features. Our predictive system could provide individualized survival curves at individual level, which is importance for individualized treatment decisions. Additionally, our survival predictive system predicted and compared the individual survival curves under different treatments, which is valuable for patients to make optimal medical decisions before treatment.

Considering the opaque nature of the operation process of artificial intelligence algorithms, the current research provided three individual survival curves predicted using different artificial intelligence algorithms for clinical application. The concordances of prognostic models based on MTLR, RFS, and Cox algorithms were 0.703, 0.650, and 0.698, respectively, for glioblastoma multiforme patients (the higher the concordance, the higher the accuracy of the prognostic model), whereas the Brier scores were 0.039, 0.059, and 0.040, respectively, (the smaller

the Brier score, the higher the accuracy of the prognostic model) [39], indicating that MTLR algorithm was superior to RFS and Cox algorithms for prognostic prediction. The MTLR model had an AUROC of 0.92 and a brier score of 0.08, suggesting good clinical application value for prognostic prediction [36]. The RSF algorithm performed better than Cox algorithm for predicting the prognosis of major adverse cardiac and cerebrovascular event patients [40]. Harvard University artificial intelligence research team developed an artificial intelligence predictive tool for predicting prognosis of glioblastoma and provided individual predictive information of predicted survival time, one-year survival rate, and overall survival curves [41]. The concordances of prognostic models based on RFS and Cox algorithms were 0.680 and 0.690, respectively, for glioblastoma patients in another prognostic study [41]. Combined with the concordance indexes of different algorithms in the current research and the conclusions of previous studies, we first recommended the survival curve predicted by MTLR algorithm, and the survival curves predicted by RFS algorithm and Cox algorithm might be used as the second and third recommendation survival predicted curves.

Random survival forest algorithm has the following abilities: dealing with multicollinearity effects, selecting the most important parameters in accordance with defined tree threshold, and assessing the variable relative importance [42, 43]. The RFS algorithm has been recommended for

prognostic models and was reported to be superior to the Cox model in terms of predictive accuracy [44-46]. It was reported that multitask learning algorithm was superior to Cox algorithm in cancer survival analysis [47]. The concordance indexes and calibration plots of the RFS and MTLR models indicated good predictive performance, which was similar to that of the Cox model. RFS model scored higher than Cox and MTLR models on Brier score. Our results indicated that the RFS and MTLR models have good clinical application values, which were not inferior to the Cox model in survival analysis.

Our artificial intelligence predictive tool could predict the survival curve of lung cancer patients under different treatments and its 95% confidence interval. An individualized survival predictive function is very important for identifying patients at high mortality risk. Our artificial intelligence predictive tool is helpful for providing valuable predictive information for individual patient survival rate in optimizing the comprehensive management and individualized treatment for LUAD patients. In clinical work, it is necessary for lung cancer patients who were predicted as high mortality risk by artificial intelligence predictive tool to consider receiving more active and timely antitumor treatment to improve the prognosis.

It was true that artificial intelligence methods have made great progress in the fields of tumor diagnosis, treatment, prognostic prediction, and

research. However, in the clinical field, artificial intelligence method can never become a substitute for professional medical personnel, but exists as assistants and tools of medical personnel. Medical treatment is a comprehensive prevention and treatment system covering physical, psychological, and social relations, rather than a simple superposition of high-end equipment, cold technology, and complex algorithms. For clinicians, artificial intelligence technology helps to optimize the diagnosis and treatment of tumors.

Limitations: First, because lung cancer has high clinical heterogeneity, the treatment of lung cancer was too complex to form a unified treatment. Meanwhile, the great progress of radiotherapy, chemotherapy, and surgery were not conducive to forming the clinical subgroups. Although the SEER database provided limited treatment information (including radiotherapy, chemotherapy, and surgery), the treatment information was not sufficient to divide patients into stable subgroups. Second, all study patients were included from 2010 to 2015, resulting in a relatively short follow-up time (minimum follow-up time: 36 months; maximum follow-up time: 83 months). A long follow-up time is valuable for ascertaining the clinical application value of survival predictive system for long time points of greater than 83 months. Third, as nonparametric algorithms, the RFS and MTLR algorithms could't be directly expressed by conventional mathematical formulas, weakening the interpretability

and clinical application of these predictive models to a certain extent.

Fourth, external research datasets provide more convincing evidences for the conclusions of prognostic studies. However, we failed to identify a long-term tumour research dataset similar to the dataset provided by the SEER database. Independent external follow-up datasets are of great value for the construction and verification of tumour prognostic research.

## **5. Conclusions**

The current study designed an individualized survival predictive system, which could provide individual survival curves using three different artificial intelligence algorithms. This artificial intelligence predictive system could directly convey treatment benefits by comparing individual mortality risk curves under different treatments. This artificial intelligence predictive tool is available at

[https://zhangzhiqiao11.shinyapps.io/Artificial\\_Intelligence\\_Survival\\_Prediction\\_System\\_AI\\_E1001/](https://zhangzhiqiao11.shinyapps.io/Artificial_Intelligence_Survival_Prediction_System_AI_E1001/).

## **Acknowledgements**

We are thankful to Mrs Qingmei Liu for assistance to current research.

## **REFERENCES**

1. Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A:

- Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries.** *CA Cancer J Clin* 2018, **68**(6):394-424.
2. Lin JJ, Cardarella S, Lydon CA, Dahlberg SE, Jackman DM, Janne PA, Johnson BE: **Five-Year Survival in EGFR-Mutant Metastatic Lung Adenocarcinoma Treated with EGFR-TKIs.** *Journal of thoracic oncology : official publication of the International Association for the Study of Lung Cancer* 2016, **11**(4):556-565.
  3. Dong X, Zhang R, He J, Lai L, Alolga RN, Shen S, Zhu Y, You D, Lin L, Chen C *et al*: **Trans-omics biomarker model improves prognostic prediction accuracy for early-stage lung adenocarcinoma.** *Aging* 2019, **11**(16):6312-6335.
  4. Luo C, Lei M, Zhang Y, Zhang Q, Li L, Lian J, Liu S, Wang L, Pi G, Zhang Y: **Systematic construction and validation of an immune prognostic model for lung adenocarcinoma.** *Journal of cellular and molecular medicine* 2019.
  5. Wei H, Liao Q, Liu B: **iLncRNAdis-FB: Identify lncRNA-Disease Associations by Fusing Biological Feature Blocks Through Deep Neural Network.** *IEEE/ACM transactions on computational biology and bioinformatics* 2021, **18**(5):1946-1957.
  6. Tan H, Sun Q, Li G, Xiao Q, Ding P, Luo J, Liang C: **Multiview**

- Consensus Graph Learning for lncRNA-Disease Association Prediction. *Frontiers in genetics* 2020, 11:89.
7. Gong Y, Zhu W, Sun M, Shi L: Bioinformatics Analysis of Long Non-coding RNA and Related Diseases: An Overview. *Frontiers in genetics* 2021, 12:813873.
  8. Fan W, Shang J, Li F, Sun Y, Yuan S, Liu JX: IDSSIM: an lncRNA functional similarity calculation model based on an improved disease semantic similarity method. *BMC bioinformatics* 2020, 21(1):339.
  9. Yang L, Wu Y, Xu H, Zhang J, Zheng X, Zhang L, Wang Y, Chen W, Wang K: Identification and Validation of a Novel Six-lncRNA-Based Prognostic Model for Lung Adenocarcinoma. *Frontiers in oncology* 2021, 11:775583.
  10. Wu Y, Yang L, Zhang L, Zheng X, Xu H, Wang K, Weng X: Identification of a Four-Gene Signature Associated with the Prognosis Prediction of Lung Adenocarcinoma Based on Integrated Bioinformatics Analysis. *Genes* 2022, 13(2).
  11. Yu X, Wang Z, Chen Y, Yin G, Liu J, Chen W, Zhu L, Xu W, Li X: The Predictive Role of Immune Related Subgroup Classification in Immune Checkpoint Blockade Therapy for Lung Adenocarcinoma. *Frontiers in genetics* 2021, 12:771830.
  12. Auslander N, Zhang G, Lee JS, Frederick DT, Miao B, Moll T,



- Tian T, Wei Z, Madan S, Sullivan RJ et al: Robust prediction of response to immune checkpoint blockade therapy in metastatic melanoma. *Nature medicine* 2018, 24(10):1545-1549.
13. Coccia M. 2020. **Deep learning technology for improving cancer care in society: New directions in cancer imaging driven by artificial intelligence.** *Technology in Society*, vol. 60, February, pp. 1-11, art. n. 101198, <https://doi.org/10.1016/j.techsoc.2019.101198>
  14. Collins GS, Altman DG: **Identifying patients with undetected colorectal cancer: an independent validation of QCancer (Colorectal).** *British journal of cancer* 2012, 107(2):260-265.
  15. Collins GS, Altman DG: **Identifying patients with undetected renal tract cancer in primary care: an independent and external validation of QCancer® (Renal) prediction model.** *Cancer epidemiology* 2013, 37(2):115-120.
  16. Collins GS, Altman DG: **Identifying patients with undetected gastro-oesophageal cancer in primary care: External validation of QCancer® (Gastro-Oesophageal).** *European journal of cancer (Oxford, England : 1990)* 2013, 49(5):1040-1048.
  17. Collins GS, Altman DG: **Identifying women with undetected ovarian cancer: independent and external validation of QCancer(®) (Ovarian) prediction model.** *European journal of*

- cancer care* 2013, **22**(4):423-429.
18. Collins GS, Altman DG: **Identifying patients with undetected pancreatic cancer in primary care: an independent and external validation of QCancer(®) (Pancreas).** *The British journal of general practice : the journal of the Royal College of General Practitioners* 2013, **63**(614):e636-642.
  19. Senders JT, Staples P, Mehrtash A, Cote DJ, Taphoorn MJB, Reardon DA, Gormley WB, Smith TR, Broekman ML, Arnaout O: **An On-line Calculator for the Prediction of Survival in Glioblastoma Patients Using Classical Statistics and Machine Learning.** *Neurosurgery* 2019.
  20. Chang MC: **Development of individual survival estimating program for cancer patients' management.** *Healthcare informatics research* 2015, **21**(2):134-137.
  21. Rupji M, Zhang X, Kowalski J: **CASAS: Cancer Survival Analysis Suite, a web based application.** *F1000Research* 2017, **6**:919.
  22. Cheng C, Wang Q, Zhu M, Liu K, Zhang Z: **Integrated analysis reveals potential long non-coding RNA biomarkers and their potential biological functions for disease free survival in gastric cancer patients.** *Cancer cell international* 2019, **19**:123.
  23. Zhang Z, He T, Huang L, Ouyang Y, Li J, Huang Y, Wang P, Ding

- J: Two precision medicine predictive tools for six malignant solid tumors: from gene-based research to clinical application.**  
*Journal of translational medicine* 2019, **17**(1):405.
24. Zhang Z, Li J, He T, Ding J: **Bioinformatics Identified 17 Immune Genes as Prognostic Biomarkers for Breast Cancer: Application Study Based on Artificial Intelligence Algorithms.**  
*Frontiers in oncology* 2020, **10**:330.
25. Zhang Z, Li J, He T, Ouyang Y, Huang Y, Liu Q, Wang P, Ding J: **The competitive endogenous RNA regulatory network reveals potential prognostic biomarkers for overall survival in hepatocellular carcinoma.** *Cancer science* 2019, **110**(9):2905-2923.
26. Zhang Z, Li J, He T, Ouyang Y, Huang Y, Liu Q, Wang P, Ding J: **Two predictive precision medicine tools for hepatocellular carcinoma.** *Cancer cell international* 2019, **19**:290.
27. Zhang Z, Liu Q, Wang P, Li J, He T, Ouyang Y, Huang Y, Wang W: **Development and internal validation of a nine-lncRNA prognostic signature for prediction of overall survival in colorectal cancer patients.** *PeerJ* 2018, **6**:e6061.
28. Zhang Z, Ouyang Y, Huang Y, Wang P, Li J, He T, Liu Q: **Comprehensive bioinformatics analysis reveals potential lncRNA biomarkers for overall survival in patients with**

- hepatocellular carcinoma: an on-line individual risk calculator based on TCGA cohort.** *Cancer cell international* 2019, **19**:174.
29. Zhu M, Wang Q, Luo Z, Liu K, Zhang Z: **Development and validation of a prognostic signature for preoperative prediction of overall survival in gastric cancer patients.** *OncoTargets and therapy* 2018, **11**:8711-8722.
30. Coccia M. 2018. **An introduction to the methods of inquiry in social sciences**, J. Adm. Soc. Sci. - JSAS - vol. 5, n. 2, pp. 116-126, <http://dx.doi.org/10.1453/jsas.v5i2.1651>.
31. Hsich E, Gorodeski EZ, Blackstone EH, Ishwaran H, Lauer MS: **Identifying important risk factors for survival in patient with systolic heart failure using random survival forests.** *Circulation Cardiovascular quality and outcomes* 2011, **4**(1):39-45.
32. Ruysinck J, van der Hertten J, Houthoof R, Ongenaes F, Couckuyt I, Gadeyne B, Colpaert K, Decruyenaere J, De Turck F, Dhaene T: **Random Survival Forests for Predicting the Bed Occupancy in the Intensive Care Unit.** *Computational and mathematical methods in medicine* 2016, **2016**:7087053.
33. Nasejje JB, Mwambi H: **Application of random survival forests in understanding the determinants of under-five child mortality in Uganda in the presence of covariates that satisfy the proportional and non-proportional hazards assumption.**

- BMC research notes* 2017, **10**(1):459.
34. Hamidi O, Poorolajal J, Farhadian M, Tapak L: **Identifying Important Risk Factors for Survival in Kidney Graft Failure Patients Using Random Survival Forests.** *Iranian journal of public health* 2016, **45**(1):27-33.
35. Alaeddini A, Hong SH: **A Multi-way Multi-task Learning Approach for Multinomial Logistic Regression\*. An Application in Joint Prediction of Appointment Miss-opportunities across Multiple Clinics.** *Methods of information in medicine* 2017, **56**(4):294-307.
36. Bisaso KR, Karungi SA, Kiragga A, Mukonzo JK, Castelnuovo B: **A comparative study of logistic regression based machine learning techniques for prediction of early virological suppression in antiretroviral initiating HIV patients.** *BMC medical informatics and decision making* 2018, **18**(1):77.
37. LD F, DY L: **Time-dependent covariates in the Cox proportional-hazards regression model.** *Annual review of public health* 1999, **20**(undefined):145-157.
38. JL K, U S, A C, J B, T J, Y K: **DeepSurv: personalized treatment recommender system using a Cox proportional hazards deep neural network.** *BMC medical research methodology* 2018, **18**(1):24.

39. Haider H, Hoehn B, Davis S, Greiner R: **Effective Ways to Build and Evaluate Individual Survival Distributions.** *Journal of Machine Learning Research* 2020, **21**:1-63.
40. Farhadian M, Dehdar Karsidani S, Mozayanimonfared A, Mahjub H: **Risk factors associated with major adverse cardiac and cerebrovascular events following percutaneous coronary intervention: a 10-year follow-up comparing random survival forest and Cox proportional-hazards model.** *BMC cardiovascular disorders* 2021, **21**(1):38.
41. Senders JT, Staples P, Mehrtash A, Cote DJ, Taphoorn MJB, Reardon DA, Gormley WB, Smith TR, Broekman ML, Arnaout O: **An On-line Calculator for the Prediction of Survival in Glioblastoma Patients Using Classical Statistics and Machine Learning.** *Neurosurgery* 2020, **86**(2):E184-e192.
42. Shi M, Xu G: **Development and validation of GMI signature based random survival forest prognosis model to predict clinical outcome in acute myeloid leukemia.** *BMC medical genomics* 2019, **12**(1):90.
43. Zhang L, Chen S, Wang B, Su Y, Li S, Liu G, Zhang X: **An eight-long noncoding RNA expression signature for colorectal cancer patients' prognosis.** *Journal of cellular biochemistry* 2019, **120**(4):5636-5643.

44. Wang H, Li G: **A Selective Review on Random Survival Forests for High Dimensional Data.** *Quantitative bio-science* 2017, **36(2):85-96.**
45. Wang H, Shen L, Geng J, Wu Y, Xiao H, Zhang F, Si H: **Prognostic value of cancer antigen -125 for lung adenocarcinoma patients with brain metastasis: A random survival forest prognostic model.** *Scientific reports* 2018, **8(1):5670.**
46. Adham D, Abbasgholizadeh N, Abazari M: **Prognostic Factors for Survival in Patients with Gastric Cancer using a Random Survival Forest.** *Asian Pacific journal of cancer prevention : APJCP* 2017, **18(1):129-134.**
47. Gu W, Zhang Z, Xie X, He Y: **An Improved Multi-task Learning Algorithm for Analyzing Cancer Survival Data.** *IEEE/ACM transactions on computational biology and bioinformatics* 2019.

## **Declarations**

## **Ethics approval**

The current study was approved by ethics committee of Shunde Hospital, Southern Medical University and exempted from informed consent.

### **Consent for publication**

All authors reviewed the current paper and consented for publication.

### **Availability of data**

Study dataset is available from SEER database  
(<https://seer.cancer.gov/>).

### **Competing interests**

The authors have no conflicts of interest.

### **Funding**

Foshan Science and Technology Bureau (2020001004584).



Table 1. Clinical features of included patients

Parameters	Model group	Validation group	Test value	<i>P</i> value
Number [n]	25123	25564		
Death [n(%)]	16054(63.9)	16568(64.8)	4.518	0.034
Total survival time (month)	16(5,32)	15(5,32)	323167295	0.107
Age (year)	67(59,75)	67(59,75)	319958070	0.760
Male [(n)%]	13220(52.6)	13478(52.7)	0.048	0.826
Stage 1 [n(%)]	5838(23.2)	5826(22.8)	2.920	0.404
Stage 2 [n(%)]	1902(7.6)	2019(7.9)		
Stage 3 [n(%)]	4494(17.9)	4565(17.9)		
Stage 4 [n(%)]	12889(51.3)	13154(51.5)		
AJCC PT (T0) [n(%)]#	140(0.6)	180(0.7)	4.931	0.294
AJCC PT (T1) [n(%)]#	6468(25.7)	6563(25.7)		
AJCC PT (T2) [n(%)]#	7730(30.8)	7795(30.5)		
AJCC PT (T3) [n(%)]#	5097(20.3)	5183(20.3)		
AJCC PT (T4) [n(%)]#	5688(22.6)	5843(22.9)		
AJCC PN (N0) [n(%)]#	10592(42.2)	10556(41.3)	12.226	0.007
AJCC PN (N1) [n(%)]#	2193(8.7)	2342(9.2)		
AJCC PN (N2) [n(%)]#	8576(34.1)	8995(35.2)		
AJCC PN (N3) [n(%)]#	3762(15.0)	3671(14.4)		
AJCC PM (M0) [n(%)]#	12234(48.7)	12410(48.5)	0.111	0.740
AJCC PM (M1) [n(%)]#	12889(51.3)	13154(51.5)		
Chemotherapy[n(%)]	11834(47.1)	12056(47.2)	0.014	0.907
Radiation_Surgery[n(%)]	21869(87.0)	22287(87.2)	0.189	0.663
Grade 1[n(%)]	2035(13.2)	2092(13.3)	0.941	0.815
Grade 2[n(%)]	5919(38.5)	6082(38.7)		
Grade 3[n(%)]	7257(47.2)	7396(47.1)		
Grade 4[n(%)]	151(1.0)	139(0.9)		
Laterality(Right)[n(%)]	14647(58.3)	15084(59.0)	6.401	0.041
Laterality(Left)[n(%)]	9926(39.5)	9865(38.6)		
Laterality(Other)[n(%)]	550(2.2)	615(2.4)		
White [n(%)]	19538(77.9)	19773(77.5)	3.542	0.617
Black[n(%)]	3126(12.5)	3227(12.7)		
American Indian, Aleutian, Alaskan Native [n(%)]	108(0.4)	133(0.5)		

Chinese[n(%)]	591(2.4)	583(2.3)		
Japanese[n(%)]	186(0.7)	193(0.8)		
Other[n(%)]	1534(6.1)	1598(6.3)		
Regional_Nodes_Positive(n)	98(1,98)	98(1,98)	319907291	0.4
Regional_Nodes_Examined(n)	0(0,7)	0(0,7)	321556157.5	0.765
Tumor_Size(mm)	34(21,52)	34(22,52)	269170130.5	0.332

Note: Continuous variables were presented as median (the first quantile, the third quantile); #AJCC: American Joint Committee on Cancer.

Table 2. Results of Cox regression analyses

Variable	Univariate analysis			Multivariate analysis			
	HR <sup>#</sup>	95% CI*	P-value	Coefficient	HR <sup>#</sup>	95% CI*	P-value
Model cohort (n=25123)							
Gender(Male/Female)	1.365	1.323-1.407	<0.001	0.28	1.323	1.283-1.365	<0.001
Age(High/Low)	1.108	1.074-1.142	<0.001	0.244	1.276	1.236-1.317	<0.001
Stage(3-4/1-2)	6.057	5.779-6.349	<0.001	1.286	3.618	3.378-3.875	<0.001
PT(3-4/0-2)	2.209	2.141-2.279	<0.001	0.149	1.16	1.122-1.199	<0.001
PN(1-3/0)	2.825	2.729-2.925	<0.001	0.431	1.539	1.477-1.603	<0.001
PM(1/0)	4.67	4.511-4.835	<0.001	0.937	2.551	2.445-2.662	<0.001
Chemotherapy (Yes/No)	1.362	1.319-1.406	<0.001	-0.726	0.484	0.467-0.501	<0.001
Radiation_Surgery (Yes/No)	0.987	0.944-1.032	0.565	-0.146	0.864	0.824-0.906	<0.001
Validation cohort (n=25564)							
Gender(Male/Female)	1.349	1.308-1.391	<0.001	0.247	1.28	1.242-1.320	<0.001
Age(High/Low)	1.113	1.080-1.147	<0.001	0.233	1.262	1.223-1.302	<0.001
Stage(3-4/1-2)	5.688	5.435-5.953	<0.001	1.176	3.24	3.030-3.465	<0.001
PT(3-4/0-2)	2.224	2.156-2.294	<0.001	0.204	1.226	1.186-1.267	<0.001
PN(1-3/0)	2.804	2.709-2.901	<0.001	0.471	1.602	1.537-1.669	<0.001
PM(1/0)	4.502	4.351-4.657	<0.001	0.927	2.528	2.424-2.636	<0.001
Chemotherapy (Yes/No)	1.311	1.271-1.353	<0.001	-0.752	0.471	0.455-0.488	<0.001
Radiation_Surgery (Yes/No)	0.984	0.940-1.029	0.466	-0.123	0.884	0.845-0.926	<0.001

Note: <sup>#</sup>HR, hazard ratio; \*CI, confidence interval.

## Declarations

## Consent for publication

All authors have reviewed the manuscript and consented for publication.

### **Availability of data and materials**

The study data are available at the SEER database (<https://seer.cancer.gov/>).

### **Competing interests**

The authors have no conflicts of interest.

### **Funding**

Foshan Science and Technology Bureau (2020001004584).

### **Author contributions**

Conceptualization, methodology and resources: ZZ, PW, LJ, and TH. investigation, data curation, formal analysis, validation, software, project administration, and supervision: ZZ, PW, LJ, and TH. writing and visualization: ZZ and PW; funding acquisition: ZZ.

### **1. Statement of conflicts of interest**

There were no conflicts of interest in the present study.

2. Declaration of funding interests:

This study was supported by Foshan Science and Technology Bureau (2020001004584). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Operation instructions and result interpretation of artificial intelligence survival prediction system**

**For a patients with the following parameters: age 70 years, stage 4, PT 3, PN**

**1.PM 1, gender male**

After entering all clinical data, press "Predict" button to calculate the personal survival rate.

Input clinical information.

**Age**  
 19 28 37 46 55 64 70 73 82 91 100 101

**Stage**  
 1 2 3 4

**PT**  
 0 1 2 3 4

**PN**  
 0 1 2 3

**PM**  
 1

**Gender**  
 1

**Chemotherapy**  
 0

**Radiation\_Surgery**  
 0

**inputed individual parameters**

Overall_survival_month	Age	Stage	PT	PN	PM	Gender	Chemotherapy
1	16	70	4	3	1	1	0
2	16	70	4	3	1	1	1
3	16	70	4	3	1	1	0
4	16	70	4	3	1	1	1

**Result interpretation**

For this patient, the predicted survival rates are: 0.83 for non-treatment (black line), 0.304 for chemotherapy (blue line), 0.114 for radiation (red line), and 0.114 for chemotherapy plus radiation (green line).

